Fundamentos Matemáticos. Diagonalización.

R. Alberich, A. Mir

22 de mayo de 2007

Parte I

Fundamentos Matemáticos. Matrices

La Materia de trabajo

Matriz (tabla) de datos.

Nuestra materia de trabajo serán tablas de datos similares a las siguientes

$\mathbf{X}$	x_1	x_2	$\ldots$	x_j	$\ldots$	x_p
1	x_{11}	x_{12}	$\ldots$	x_{1j}	$\ldots$	x_{1p}
2	x_{21}	x_{22}	$\ldots$	x_{2j}	$\ldots$	x_{2p}
3	x_{31}	x_{32}	$\ldots$	x_{3j}	$\ldots$	x_{3p}
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	
i	x_{i1}	x_{i2}	$\ldots$	x_{ij}	$\ldots$	x_{ip}
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	
n	x_{n1}	x_{n2}	$\ldots$	x_{nj}	$\ldots$	x_{np}

- Las columnas son las observaciones de las variable $x_1, \ldots, x_j, \ldots, x_p$.
- Las filas son los valores de las observaciones en los individuos.

La manera de expresar de la forma más simple un tabla genérica de este tipo es en forma de matriz:

$$\mathbf{X} = (x_{ij})_{i=1,\ldots,n;j=1,\ldots,p}$$

Más concretamente una matriz es:

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}$$

- El orden de una matriz es su número de filas y columnas.

- Así la matriz A es de orden $m \times n$
- Si $m = n$ se dice que es una matriz cuadrada de orden n.
- Si es de orden $1 \times n$ se le dice vector fila $v = (v_1, \ldots, v_n)$.
- Si es de orden $n \times 1$ se le dice vector columna
$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

Suma de matrices

Sólo se pueden sumar matrices de las mismas dimensiones
Sean $A = (a_{ij})_{i=1,\ldots,m;j=1,\ldots,n}$ y $B = (b_{ij})_{i=1,\ldots,m;j=1,\ldots,n}$ dos matrices de dimensiones $m \times n$ se define la suma de A y B entonces

$$A + B = (a_{ij} + b_{ij})_{i=1,\ldots,n;j=1,\ldots,p}$$

Ejemplo

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}; \qquad B = \begin{pmatrix} -1 & -2 & 1 \\ -4 & 1 & 1 \end{pmatrix}$$

$$A + B = \begin{pmatrix} 0 & 0 & 4 \\ 0 & 6 & 7 \end{pmatrix}$$

Producto de matrices

Sólo se pueden multiplicar dos matrices A y B si la primera tiene el mismo número de filas que columnas la segunda. Sean $A = (a_{ij})_{i=1,\ldots,m;j=1,\ldots,n}$; $B = (b_{jk})_{j=1,\ldots,n;k=1,\ldots,r}$ dos matrices de dimensiones $m \times n$ y $n \times r$ respectivamente. Así el número de columnas de A es n y es igual al número filas de B como

$$A \cdot B = C = (c_{ij})_{i=1,\ldots,n;j=1,\ldots,r}$$

Donde $c_{ij} = a_{i1}b_{1j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}$

Ejemplos: $v = (v_1, \ldots, v_n)$; $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$

$$v \cdot u = v = (v_1, \ldots, v_n) \cdot \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = v_1 u_1 + v_2 u_2 + \cdots v_n u_n$$

$$u \cdot v = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \cdot (v_1, \ldots, v_n) = \begin{pmatrix} u_1 v_1 & u_1 v_2 & \cdots & u_1 v_n \\ u_2 v_1 & u_2 v_2 & \cdots & u_2 v_n \\ \vdots & \vdots & \ldots & \vdots \\ u_n v_1 & u_n v_2 & \cdots & u_n v_n \end{pmatrix}$$

$$A=\begin{pmatrix}3&0\\2&-1\\5&4\\-2&3\end{pmatrix};\qquad B=\begin{pmatrix}-2&4&6\\-1&3&0\end{pmatrix}$$

El producto de A por B es

$$A\cdot B=\begin{pmatrix}3(-2)+0(-1)&3\cdot 4+0\cdot 3&3\cdot 6+0\cdot 0\\2(-2)+(-1)(-1)&2\cdot 4+(-1)\cdot 3&2\cdot 6+(-1)0\\5(-2)+4(-1)&5\cdot 4+4\cdot 3&5\cdot 6+4\cdot 0\\(-2)(-2)+3(-1)&(-2)4+3\cdot 3&(-2)6+3\cdot 0\end{pmatrix}=$$

$$\begin{pmatrix}-6&12&18\\-3&5&12\\-14&32&30\\1&1&-12\end{pmatrix}$$

Producto de un escalar por una matriz

Dado un escalar $\alpha \in \mathbb{R}$ y una matriz $A = (a_{ij})_{i=1,\ldots,n;j=1,\ldots m}$
Se define el producto de α por A como

$$\alpha \cdot A = (\alpha a_{ij})_{i=1,\ldots,n;j=1,\ldots m}.$$

Ejemplos

$$2 \cdot \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 2 & 6 \end{pmatrix}$$

$$2 \cdot (1, 2, 5) = (2, 4, 10)$$

$$2 \cdot \begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 10 \end{pmatrix}$$

Transposición de Matrices

Dada una matriz A llamaremos transpuesta de A a aquella matriz que tiene por filas (columnas) las columnas (filas) de A; la denotaremos por A^t.

Ejemplo

Sea $A = \begin{pmatrix} 6 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$ entonces la matriz transpuesta de A es $A^t = \begin{pmatrix} 6 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}$

Sea $v = (1, 2, 3)$; entonces $v^t = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$

Propiedades

1) $A+B=B+A$; $(A+B)+C=A+(B+C)$
2) $(A\cdot B)\cdot C=A\cdot(B\cdot C)$
3) $A\cdot(B+C)=A\cdot B+A\cdot C$
4) $A\cdot(\alpha\cdot B)=\alpha\cdot(A\cdot B)$
5) $(A+B)^t=A^t+B^t$; $(A\cdot B)^t=B^t\cdot A^t$

Ejemplo Sean $A = \begin{pmatrix} 3 & 0 \\ 2 & -1 \\ 5 & 4 \\ -2 & 3 \end{pmatrix}$; $B = \begin{pmatrix} -2 & 4 & 6 \\ -1 & 3 & 0 \end{pmatrix}$

Entonces $A \cdot B = \begin{pmatrix} -6 & 12 & 18 \\ -3 & 5 & 12 \\ -14 & 32 & 30 \\ 1 & 1 & -12 \end{pmatrix}$

mientras que

$$B^t \cdot A^t = \begin{pmatrix} -2 & -1 \\ 4 & 3 \\ 6 & 0 \end{pmatrix} \cdot \begin{pmatrix} 3 & 2 & 5 & -2 \\ 0 & -1 & 4 & 3 \end{pmatrix} =$$

$$= \begin{pmatrix} -6 & -3 & -14 & 1 \\ 12 & 5 & 32 & 1 \\ 18 & 12 & 30 & -12 \end{pmatrix} = (A \cdot B)^t$$

Matriz Identidad

Llamaremos matriz identidad de orden n a la matriz de orden $n \times n$

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

Propiedad

- Si A es una matriz de orden n entonces $A \cdot I_n = I_n \cdot A = A$.
- Si A es una matriz de orden $n \times m$ entonces $I_n \cdot A = A \cdot I_m = A$.

Matriz Simétrica. Matriz Diagonal

Matriz Simétrica

Una matriz cuadrada A es simétrica si $A^t = A$

Ejemplo

Sean $A = \begin{pmatrix} 3 & 6 & 11 \\ 6 & 5 & 7 \\ 11 & 7 & 1 \end{pmatrix}$; $B = \begin{pmatrix} 10 & 1 \\ 2 & 10 \end{pmatrix}$

A es simétrica y B no lo es.

Matriz Diagonal

Una matriz cuadrada $A = (a_{ij})_{i,j=1,\ldots n}$ es diagonal sii $a_{ij} = 0$ para $i, j = 1, \ldots n$ con $i \neq j$.

Es decir una matriz cuadrada es diagonal si todos los elementos que no estén en su diagonal principal (la que va de la parte superior izquierda a la inferior derecha) son ceros.

Traza de una matriz

Dada una matriz cuadrada $A = (a_{ij})_{i,j=1,\ldots n}$ se define la traza de A como $tr(A) = \sum_{i=1}^{n} a_{ii}$
Es decir la traza de una matriz es la suma de los elementos de su diagonal principal.

Ejemplo

$tr\begin{pmatrix} 2 & 5 \\ 4 & 6 \end{pmatrix} = 8$; $tr\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = 3$

$tr\left(\begin{pmatrix} 2 & 5 \\ 4 & 6 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\right) = tr\begin{pmatrix} 3 & 5 \\ 4 & 8 \end{pmatrix} = 11 = 8 + 3 =$
$tr\begin{pmatrix} 2 & 5 \\ 4 & 6 \end{pmatrix} + tr\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$

Propiedad

$$tr(A+B) = tr(A) + tr(B)$$

La traza de la suma de dos matrices es igual a la suma de sus trazas.

Determinantes.

Supondremos que se tiene familiaridad con el concepto de determinante. Lo definiremos brevemente:

- Llamamos determinante de una matriz cuadrada a un número real que es el resultado de sumas de productos de elementos de cada fila y columna poderadas por ± 1.

Sea $A_n = (a_{i,j})_{i,j=1,\ldots,n}$ una matriz cuadrada de orden n

- Para orden $n=1$ se define como $\det(A) = \left|a_{11}\right| = a_{11}$
- Para orden $n=2$ se define como $\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - (a_{21} + a_{12})$

- Para órdenes superiores se define recursivamente basándose en determinantes de submatrices de orden inferior. Se elige una fila (o columna) y para cada elemento a_{1j} se calcula

$$(-1)^{i+j} a_{ij} \det \left(\begin{array}{c} \text{Matriz resultante de suprimir} \\ \text{la fila i y la columna j} \end{array} \right)$$

- Para orden 3 se define como

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = (-1)^{1+1} a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} +$$
$$(-1)^{2+1} a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + (-1)^{3+1} a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

- Para determinantes de matrices de orden superior se generaliza la fórmula anterior.
- Es útil el siguiente mapa de signos de los coeficientes de una matriz, por ejemplo para orden 4:

$$\begin{pmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{pmatrix}$$

Inversa de una matriz

Dada una matriz A de orden n llamaremos inversa de la matriz A, y cuando exista la denotaremos por A^{-1}, a otra matriz del mismo orden tal que

$$A^{-1} \cdot A = A \cdot A^{-1} = I_n.$$

Propiedades

- Si existe una matriz cuadra tal que $B \cdot A = I_n$ entonces $A \cdot B = I_n$ y $B = A^{-1}$. El mismo resultado es cierto si multiplicamos por la derecha.
- La inversa, si existe, es única.
- Un matriz cuadrada A tiene inversa sii $det(A) \neq 0$. Las matrices cuadradas con determinante nulo no tienen inversa. Si una matriz tiene determinante cero se dice que es singular.

Submatrices. Menores. Matriz Adjunta. Adjunto

- Dada una matriz A llamaremos submatriz a la matriz resultante de eliminar una cierta cantidad de filas y/o columnas.
- Llamaremos menor de orden r de una matriz A al determinante de una submatriz cuadrada de orden r.
- Llamaremos matriz adjunta de i, j de la matriz A a la submatriz resultante de eliminar la fila i y la columna j. Denotaremos esta matriz por A^*_{ij}
- Llamaremos adjunto del elemento a_{ij} de la matriz A al escalar

$$(-1)^{i+j} \cdot det(A^*_{ij})$$

Cálculo de la inversa de una matriz.

Dada una matriz cuadrada A de orden n se puede calcular la inversa mediante varios algoritmos exactos o aproximados. Un algoritmo para calcular de forma exacta la inversa de una matriz es el siguiente:

- Se calcula $det(A)$ si es cero (matriz singular) entonces no tiene inversa.
- Se sustituye cada elemento de la matriz A por el de su correspondiente adjunto. A la matriz resultante se la denomina *Matriz Adjunta*.
- La matriz inversa es $A^{-1} = \frac{(Matriz \quad Adjunta)^t}{det(A)}$ o bien de una forma más mnemotécnica; "la inversa de la matriz A es la traspuesta de la matriz adjunta de A partida por el determinante de A"

Ejemplo

Sea $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}$

$det(A) = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{vmatrix} =$

$1 \cdot 2 \cdot 3 + 1 \cdot 1 \cdot 2 + 1 \cdot 2 \cdot 1 - 1 \cdot 2 \cdot 1 - 1 \cdot 1 \cdot 3 - 1 \cdot 2 \cdot 2 = 1 \neq 0$

Por lo tanto A es no singular y tiene inversa.

$$MatrizAdjunta = \begin{pmatrix} +\begin{vmatrix} 2 & 2 \\ 2 & 3 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 1 & 3 \end{vmatrix} & +\begin{vmatrix} 1 & 2 \\ 1 & 2 \end{vmatrix} \\ -\begin{vmatrix} 1 & 1 \\ 2 & 3 \end{vmatrix} & +\begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} & -\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} \\ +\begin{vmatrix} 1 & 1 \\ 2 & 2 \end{vmatrix} & -\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} & +\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} \end{pmatrix}$$

Entonces $A^{-1} = \frac{(MatrizAdjunta)^t}{det(A)} = \frac{\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}}{1}$

Propiedad

$$det(A^{-1}) = \frac{1}{det(A)}$$

Combinación lineal. Independencia lineal

- Una combinación lineal de los vectores $v_1, \ldots, v_n$ es cualquier expresión del tipo

$$\alpha_1 \cdot v_1 + \cdots + \alpha_n \cdot v_n$$

donde $\alpha_i \in \mathbb{R}$ para $i = 1, \ldots, n$.

- Diremos que el vector v se puede poner como combinación lineal de los vectores $v_1, \ldots, v_n$ si existen escalares $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ tales que

$$v = \alpha_1 \cdot v_1 + \cdots + \alpha_n \cdot v_n.$$

- Diremos que los vectores $v_1, \ldots, v_n$ son linealmente independientes si la única combinación lineal de éstos que es igual al vector 0 es la trivial (es decir $\alpha_i = 0$ para $i = 1, \ldots, n$)

Ejemplo
Veamos si los vectores $(1,2)$ y $(2,0)$ son l.i.
$\alpha_1(1,2)+\alpha_2(2,0)=(0,0)$; $(\alpha_1+2\alpha_2, 2\alpha_1+0)=(0,0)$
Se obtiene el sistema de ecuaciones lineales:

$$\left.\begin{array}{rcrl} \alpha_1 & + & 2\alpha_2 = & 0 \\ 2\alpha_1 & + & 0 = & 0 \end{array}\right\}$$

Resolviendo el sistema obtenemos que la única solución es la trivial $\alpha_1=\alpha_2=0$ lo que nos dice que estos dos vectores son linealmente independientes.

Consideremos ahora los vectores $(1,2)$ y $(2,4)$. Estos vectores no son l.i.

$$\alpha_1(1,2)+\alpha_2(2,4)=(0,0).$$

El sistema que se obtiene es

$$\left.\begin{array}{rcrcr} \alpha_1 & + & 2\alpha_2 = & 0 \\ 2\alpha_1 & + & 4\alpha_2 = & 0 \end{array}\right\}$$

obtenemos el sistema equivalente

$$\left.\begin{array}{rcrcr} \alpha_1 & + & 2\alpha_2 = & 0 \end{array}\right\}$$

de donde α_2 toma cualquier valor real y $\alpha_1=-2\alpha_2$. Por ejemplo una solución particular no trivial es $\alpha_2=1$ y $\alpha_1=-2$. Efectivamente $-2(1,2)+(2,4)=(0,0)$ por lo tanto los dos vectores no son linealmente independientes.

Producto escalar

Dados dos vectores

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \text{ y } u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

se define el producto escalar de v y u como

$$< v, u >= v^t \cdot u = \sum_{i=1}^{n} v_i \cdot u_i$$

Norma de un vector

Dado un vector

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

se define la norma ($\|-\|$) del vector como

$\|v\| = \sqrt{< v, v >} = \sqrt{v^t \cdot v} = \sqrt{\sum_{i=1}^{n} v_i^2}$

Definición Llamaremos ángulo de dos vectores al que se forma uniendo cada vector como si fuera un punto con el origen. El ángulo de los vectores u y v lo denotemos por $\angle u, v$.

Propiedad

$$< v, u >= \|v\| \cdot \|u\| \cdot \cos(\angle u, v).$$

Definición

- Diremos que un vector es unitario si su norma es uno. Diremos que dos vectores son ortogonales si son no nulos y su producto escalar es cero.
- Diremos que los vectores $v_1, \ldots v_k$ son ortogonales si son ortogonales dos a dos.
- Diremos que los vectores $v_1, \ldots v_k$ son ortonormales si son ortogonales y todos son unitarios.

Nota:
Efectivamente si $< v, u >= 0$ y ambos vectores son no nulos, se tiene que

$$\|v\| \cdot \|u\| \cos(\angle u, v) = 0$$

por lo tanto $\cos(\angle u, v) = 0$ y $\angle u, v = \pi/2$ o 90 grados y por lo tanto son perpendiculares (ortogonales) en el sentido popular.
Todo vector u con $\|u\| \neq 0$ tiene un vector unitario asociado $\frac{1}{\|u\|} \cdot u$.

Propiedad Si los vectores $v_1, v_2, \ldots v_k$ son ortogonales dos a dos entonces son linealmente independientes. (Ejercicio demostralo para $k = 3$).

Definición

- Sea P una matriz cuadrada de orden k. Diremos que P es una matriz ortogonal si $P^t \cdot P = I_k$, es decir si $P^{-1} = P^t$.
- Dicho de otra forma: una matriz cuadrada es ortonormal si su inversa es igual a su transpuesta.
- Evidentemente una matriz es ortogonal si el conjunto de sus vectores fila (columna) son ortonormales.

Rango de una matriz.

Definición

- Sea A una matriz de orden $m \times n$ llamaremos rango de A ($rg(A)$) al numero máximo de vectores fila (o columna) que son l.i.
- Así tenemos que $rg(A) \leq \min\{m, n\}$
- En particular una matriz de orden n tendrá a lo más rango n.

Definición

- Llamaremos menor de una matriz al determinante de una submatriz cuadrada.
- Llamaremos orden del menor a la dimensión de esa matriz.

Propiedad El rango de una matriz es igual al orden del mayor menor no nulo

Ejemplo

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

Menores de orden tres

$\begin{vmatrix}1&1&1\\1&1&1\\0&0&1\end{vmatrix}=0$ luego $rg(A)<3$.

Los menores de orden 2 son de la forma

$\begin{vmatrix}1&1\\1&1\end{vmatrix}=0$; $\begin{vmatrix}1&1\\0&0\end{vmatrix}=0$; $\begin{vmatrix}1&1\\0&1\end{vmatrix}=1$ luego $rg(A)=2$. Esta matriz sólo tiene dos vectores fila o columna linealmente independientes.

Ejemplo ¿Son los vectores $(1,2,3)$, $(0,1,1)$ y $(0,1,1)$ l.i?

Como $\begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} = 1 \neq 0$ luego los tres vectores son l.i. La matriz formada por los tres vectores fila tiene rango 3.

Fundamentos Matemáticos. Diagonalización.

R. Alberich, A. Mir

22 de mayo de 2007

Parte II

Fundamentos Matemáticos. Diagonalización

Valores y vectores propios de una matriz

(*eigenvalues, eigenvectors*)
Sea A una matriz real cuadrada de orden p y sea $x \in R^p$ una vector columna $x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ de dimensión $p \times 1$ y distinto del vector nulo. Se dice que x es un vector propio de la matriz A si existe $\lambda \in R$ tal que

$$A \cdot x = \lambda x$$

.
El escalar λ recibe el nombre de valor propio de A asociado al vector propio x.

ejemplo

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

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 0 \\ 2 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 9 \\ 3 \\ 7 \end{pmatrix} = 4 \cdot \begin{pmatrix} 9 \\ 3 \\ 7 \end{pmatrix}$$

Luego el vector $x = \begin{pmatrix} 9 \\ 3 \\ 7 \end{pmatrix}$ es un vector propio de la matriz A asociado al valor propio $\lambda = 4$

Nota: Los vectores de la forma $\alpha \cdot \begin{pmatrix} 9 \\ 3 \\ 7 \end{pmatrix}$ también serán vectores propios de A de valor propio λ (ejercicio).

¿Cómo hemos encontrado el valor y el vector propio? pues planteando la ecuación

$$A \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \lambda \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 0 \\ 2 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 + 3x_3 \\ x_1 + x_2 \\ 2x_1 + x_2 \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda x_3 \end{pmatrix}$$

de donde obtenemos el siguiente sistema de ecuaciones lineales homogéneo con un parámetro λ.

$$\left.\begin{array}{rcrcrl} (1-\lambda)x_1 & + & 2x_2 & + & 3x_3 & = 0 \\ x_1 & + & (1-\lambda)x_2 & & & = 0 \\ 2x_1 & + & x_2 & - & (1-\lambda)x_3 & = 0 \end{array}\right\}$$

- Recordar que en un sistema homogéneo es determinado si el rango de la matriz de coeficientes coincide con el numero de incógnitas y en este caso tiene por solución la trivial (todas las incógnitas nulas).
- En este caso esto sería equivalente a que el determinante de la matriz de coeficientes sea nulo. Luego los valores propios serán los valores de λ que hagan que existan soluciones distintas de la trivial (hagan el sistema lineal compatible indeterminado) y por lo tanto son las soluciones de la ecuación:

$$\begin{vmatrix} 1-\lambda & 2 & 3 \\ 1 & 1-\lambda & 0 \\ 2 & 1 & 1-\lambda \end{vmatrix} = 0$$

Calculando el determinante obtenemos que

$$\begin{vmatrix} 1-\lambda & 2 & 3 \\ 1 & 1-\lambda & 0 \\ 2 & 1 & 1-\lambda \end{vmatrix} =$$

$$(1-\lambda)(1-\lambda)(1-\lambda) + 2\cdot 0\cdot 2 + 3\cdot 1\cdot 1$$
$$-(3\cdot(1-\lambda)\cdot 2 + 2\cdot 1\cdot(1-\lambda) + (1-\lambda)\cdot 1\cdot 0) =$$

$$(1-\lambda)^2(-\lambda) + 3 - (6\cdot(1-\lambda) - 2\cdot\lambda)$$

Luego hemos obtenido un polinomio de indeterminada λ
$p(\lambda) = -\lambda^3 + 3\lambda^2 + 5\lambda - 4$
Las raíces de este polinomio serán los candidatos a valor propio: Una de ellas es $\lambda = 4$.

Sustituyendo $\lambda = 4$ en el sistema de ecuaciones homogéneo obtenemos

$$\left.\begin{array}{rcrcrl} -3x_1 & + & 2x_2 & + & 3x_3 & = 0 \\ x_1 & - & 3x_2 & & & = 0 \\ 2x_1 & + & x_2 & - & 3x_3 & = 0 \end{array}\right\}$$

Que da lugar a un sistema homogéneo indeterminado cuyas soluciones son de la forma

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{9}{7}x_3 \\ \frac{3}{7}x_2 \\ x_3 \end{pmatrix}$$

para cualquier valor de $x_3 \in R$; cualquier solución distinta de la trivial ($x_3 \neq 0$) es un vector propio asociado al valor propio 4. En particular para $x_3 = 7$ obtenemos la solución particular anterior.

Polinomio característico.

Establezcamos un método para calcular los valores y vectores propios

Definición

- Dada una matriz A cuadrada de orden p llamaremos polinomio característico de A a :

$$p_A(\lambda) = \left|A - \lambda I_p\right|$$

- Llamaremos ecuación característica de A a

$$p_A(\lambda) = 0.$$

Definición

Llamaremos multiplicidad de un valor propio a su multiplicidad como raíz del polinomio característico, o lo que es lo mismo *el número de veces* que es solución del polinomio característico.

Propiedades

- Los valores propios son las soluciones de la ecuación característica:

$$p_A(\lambda) = \left|A - \lambda I_p\right| = 0$$

.

- Son vectores propios del valor propio $\lambda = \lambda_0$ las soluciones de la ecuación característica:

$$p_A(\lambda) = (A - \lambda_0 I_p) \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

.

- Dado un valor propio real el número de vectores linealmente independientes asociados a él es inferior o igual a su multiplicidad y es siempre mayor o igual que 1.
- Vectores propios de valores propios distintos son linealmente independientes

Ejemplo

Sea $A = \begin{pmatrix} 10 & -3 & -1 \\ 8 & 0 & -2 \\ 8 & -1 & -1 \end{pmatrix}$

El polinomio característico asociado a A es

$P_A(\lambda) = det(A - \lambda I_3) =$

$$\left| \begin{pmatrix} 10 & -3 & -1 \\ 8 & 0 & -2 \\ 8 & -1 & -1 \end{pmatrix} - \lambda \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right|$$

$$P_A(\lambda) = \begin{vmatrix} 10-\lambda & -3 & -1 \\ 8 & -\lambda & -2 \\ 8 & -1 & -1-\lambda \end{vmatrix} = -\lambda^3 + 9\lambda^2 - 20\lambda + 12$$

Busquemos alguna solución entera, recordemos que las raíces enteras de un polinomio con coeficientes enteros son los divisores, tanto positivos como negativos, de su término independiente.
Probemos con $\lambda = 1$

$$\begin{array}{c|cccc} & 1 & -9 & 20 & -12 \\ 1 & & 1 & -8 & 12 \\ \hline & 1 & -8 & 12 & 0 \end{array}$$

Por lo tanto $P_A(\lambda) = (\lambda - 1)(\lambda^2 - 8\lambda + 12) = 0$
Las soluciones son las de las ecuaciones:
$\lambda - 1 = 0$ y $\lambda^2 - 8\lambda + 12 = 0$ que resueltas de forma convencional son :
$\lambda_1 = 6, \lambda_2 = 2, \lambda_3 = 1$. Se suele tomar la convención de ordenarlas de mayor a menor.

Calculemos los vectores propios:
Asociados a $\lambda_1 = 1$; son soluciones del sistema

$$\begin{pmatrix} 10-6 & -3 & -1 \\ 8 & -6 & -2 \\ 8 & -1 & -1-6 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Resolviendo el sistema obtenemos que las soluciones son de la forma

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_3 \\ x_3 \\ x_3 \end{pmatrix}$$

para cualquier valor $x_3 \in R$. Una solución particular se obtiene para $x_3 = 1$ y es $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

De forma similar podemos calcular los vectores propios asociados a $\lambda_2 = 2$

$$\begin{pmatrix} 8 & -3 & -1 \\ 8 & -2 & -2 \\ 8 & -1 & -3 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Las soluciones son de la forma

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{x_3}{2} \\ x_3 \\ x_3 \end{pmatrix}$$

para cualquier valor $x_3 \in R$. Una solución particular se obtiene para $x_3 = 2$ y es $\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$

Por último calculemos los de $\lambda_3 = 1$

$$\begin{pmatrix} 9 & -3 & -1 \\ 8 & -1 & -2 \\ 8 & -1 & -2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Las soluciones son de la forma

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{x_3}{2} \\ \frac{2}{3}x_3 \\ x_3 \end{pmatrix}$$

para cualquier valor $x_3 \in R$. Una solución particular se obtiene para $x_3 = 3$ y es $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$

Diagonalización de matrices

- ¿Por qué hemos estudiado álgebra de matrices en un curso de Análisis de Datos?
- La idea principal, como veremos, es que las matrices acumulan información de los datos. Una vez recogida esta información en una matriz es muy interesante reducirla a una expresión más sencilla o explicable (lo que se suelen denominar propiedades espectrales de las matrices).
- Una de las invenciones matemáticas más útiles para esto son los valores y vectores propios.
- Mencionemos dos de la más utilizadas: la diagonalización de matrices y la descomposición en valores singulares (*Single Value Decomposition* abreviado SVD).

- Existen muchas otras aplicaciones que se utilizan con matrices especiales como las de similaridades o disimilaridades, grafos matemáticos aplicados a proteínas, grafos metabólicos comparación de árboles filogenéticos, etc.
- Por ejemplo algunas de éstas son las técnicas de escalado multidimensional (*multidimensional scaling*, abreviado MDS) o las propias de grafos aplicadas a las llamadas *Social Networks*, redes biológicas etc...
- Además de aplicaciones más recientes a clasificación por ejemplo al estudio de datos provenientes de experiencias en *microarrays*.

Definición Dada una matriz cuadrada A de orden p diremos que es diagonalizable (en $\mathbb{R}$)si existe una matriz D y otra P de orden p tales que

- La matriz D es una matriz diagonal.
- La matriz P es no singular, es decir $\det(P) \neq 0$.
- Se verifica la identidad $D = P^{-1} \cdot A \cdot P$ o lo que es lo mismo $A = P \cdot D \cdot P^{-1}$.

Propiedad

- Sea A una matriz cuadrada de orden p. Si esta matriz tiene todos sus valores propios reales y la multiplicidad de cada valor propio es igual al número de vectores propios linealmente independientes asociados a este valor entonces es diagonalizable (en $\mathbb{R}$)).
- Además su matriz diagonal D asociada corresponde a la que tiene en la diagonal los valores propios de A y como matriz P una formada por P vectores propios l.i. en el mismo orden que los valores propios de la diagonal de D

Ejemplo Sea $A = \begin{pmatrix} 10 & -3 & -1 \\ 8 & 0 & -2 \\ 8 & -1 & -1 \end{pmatrix}$ la matriz del ejemplo de cálculo de valores y vectores propios.

Entonces tomamos $D = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ la matriz diagonal formada por los valores propios de A y, en el mismo orden

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}$$

y se tiene que (ejercicio)

$$P^{-1} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

Ahora si calculamos

$$P^{-1} \cdot A \cdot P = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 10 & -3 & -1 \\ 8 & 0 & -2 \\ 8 & -1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = D$$

Así con esta equivalencia se consigue una matriz semejante a la matriz A que mantiene algunas de sus propiedades, pero con una expresión sumamente más sencilla.

- Efectivamente muchas propiedades de una matriz A y de una de sus matrices diagonales D son iguales.
- De hecho estas matrices son lo que se llama semejantes. Entre otras propiedades cumple que su traza y su determinante son iguales (comprobadlo como ejercicio en el ejemplo anterior).

Propiedad

- $tr(A) = tr(D) = \lambda_1 + \lambda_2 + \cdots + \lambda_p$
- $\det(A) = \det(D) = \lambda_1 \cdot \lambda_2 \cdot \cdots \cdot \lambda_p$

Diagonalización ortogonal.

- Algunas matrices admiten una diagonalización con más propiedades, es la llamada diagonalización ortogonal.
- La matriz diagonal es la misma mientras que los vectores propios de la matriz P se pueden elegir de forma que sean ortogonales.

Propiedad Si P es un matriz simétrica entonces vectores propios correspondientes a valores propios diferentes son ortogonales.

Ejemplo Consideremos la matriz

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

Entonces $\det(A) = 18$ el polinomio característico es

$$P_A(\lambda) = \det(A - \lambda I_3) = \begin{vmatrix} 3-\lambda & 1 & 0 \\ 1 & 3-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{vmatrix}$$

Simplificando y operando obtenemos que
$P_A(\lambda)(3-\lambda)^2 \cdot (2-\lambda) + 0 + 0 - 0 - (2-\lambda) - 0 =$
$(3-\lambda)^2 \cdot (2-\lambda) - (2-\lambda) = (2-\lambda) \cdot \left((3-\lambda)^2 - 1\right) = 0$
Los valores propios son las soluciones de las ecuaciones

- $2-\lambda = 0$
- $(3-\lambda)^2 - 1.$

Por lo tanto los valores propios son
$\lambda_1 = 4$ y $\lambda_2 = \lambda_3 = 2$; así que el valor propio 2 tiene multiplicidad 2
Calculemos los vectores propios asociados a $\lambda_1 = 4$, estos serán las soluciones no triviales de

$$\begin{pmatrix} 3-4 & 1 & 0 \\ 1 & 3-4 & 0 \\ 0 & 0 & 2-4 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

por lo tanto las soluciones son de la forma
$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_2 \\ 0 \end{pmatrix}$ para $x_2 \in R$.

Una solución particular se obtiene tomando $x_2 = 1$ y es

$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Para hacer este vector propio unitario basta dividirlo por su norma que es $\sqrt{2}$ y obtenemos el vector propio unitario asociado a este valor propio:

$$v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$$

En el caso de los vectores propios de $\lambda_2 = \lambda_3 = 2$ procederemos, por el momento de forma similar

$$\begin{pmatrix} 3-2 & 1 & 0 \\ 1 & 3-2 & 0 \\ 0 & 0 & 2-2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

por lo tanto las soluciones son de la forma
$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -x_2 \\ x_2 \\ x_3 \end{pmatrix}$ para $x_2 \in R$.
Una solución particular la obtenemos para $x_2 = 1$ y $x_3 = 0$
es $\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$.

Para hacer este vector propio unitario basta dividirlo por su norma que es $\sqrt{2}$ y obtenemos el vector propio unitario asociado a este valor propio:

$$v_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$$

Ahora tenemos que buscar otro vector propio que sea ortogonal a v_2. Planteamos la ecuación

$$\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}^t \cdot \begin{pmatrix} -x_2 \\ x_2 \\ x_3 \end{pmatrix} = 0$$ de donde se obtiene que

$$\frac{x_2}{\sqrt{2}} + \frac{x_2}{\sqrt{2}} = 0$$

por lo tanto $x_2 = 0$ así los vectores propios asociados al valor propio 2 y que al mismo tiempo son ortogonales a v_2 son de la forma $\begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix}$ para cualquier valor $x_3 \in R$. Una solución particular la obtenemos para $x_3 = 1$ y resulta ser unitaria.

$$v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

Ejemplo

Consideremos ahora la matriz

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

En este caso (se deja como ejercicio) se tiene que $\lambda_1 = 4$, $\lambda_2 = 2$ y $\lambda_3 = 1$ y tres vectores propios son respectivamente $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ y $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

Notemos que al ser la matriz simétrica y todos los valores propios distintos los vectores propios obtenidos son ortogonales. Para hacerlos ortonormales es suficiente con dividir cada vector por su norma (ejercicio)

de forma que $D = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $P = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Así pues $P^{-1} = P^t$ (comprobarlo)
y $P^t \cdot D \cdot P = A$ (comprobarlo)

Definición

Sea A una matriz cuadrada simétrica y de orden p. Diremos que A es una matriz semidefinida positiva si para todo vector

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}.$$

Se verifica que

$$x^t \cdot A \cdot x = (x_1, \cdots, x_p) \cdot A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \geq 0.$$

Si la desigualdad anterior es estricta, salvo para el vector nulo, se dice que la matriz es definida positiva.

Ejemplo La matriz $A = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$.

$(x_1, x_2, x_3) \cdot \begin{pmatrix} 3 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 3x^3 + 2xy + 3y^2 + 2z^2 =$
$2(x^2 + 2xy + y^2) + (x^2 - 2xy + y^2) + 2z^2 \geq 0$
Por lo tanto A es semidefinida positiva, además es definida positiva.

Propiedad

- Una matriz cuadrada simétrica de orden p es semidefinida positiva si y solo sí los determinantes de las k primeras filas o columnas para $k = 1, \ldots, p$ son mayores o iguales que cero.
- Una matriz cuadrada simétrica de orden p es semidefinida positiva si y solo sí tiene todos sus valores propios no negativos ($\lambda_i \geq 0$).
- Si las desigualdades anteriores son estrictas la matriz es definida positiva.

Ejemplo Comprobemos las propiedades con la matriz

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

Los determinantes de las k primeras filas y columnas son:

$$\begin{vmatrix} 3 \end{vmatrix};\begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix};\begin{vmatrix} 3 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{vmatrix}$$

Los valores de estos tres determinantes son $3, 8$ y 16 respectivamente por lo tanto son todos estrictamente positivos y la matriz será definida positiva.
Los valores propios de A, como ya vimos son, $\lambda_1 = 4$, $\lambda_2 = 2$ y $\lambda_3 = 1$. Por lo tanto la matriz es definida positiva.

Nota:

- Las matrices semidefinidas positivas aparecen recurrentemente en el tratamiento de datos, por ejemplos son matrices semidefinidas positivas las matrices de covarianzas y las de correlaciones.
- Además las matrices definidas positivas son las que determinan distintos tipos de producto escalar y sus normas y distancias asociadas.
- Estas distancias son las llamadas distancias euclídeas; intuitivamente las distancias euclídeas son las que tienen las propiedades que consideramos habituales.
- Sin embargo, como veremos, no todas las distancias entre objetos se pueden representar en un espacio euclídeo de una dimensión determinada.

Fundamentos Estadísticos.

R. Alberich, A. Mir

22 de mayo de 2007

Parte III

Fundamentos Estadísticos.

- En esta parte veremos las definiciones y fundamentos básicos de las variables aleatorias vectoriales.
- El objetivo es conocer someramente estas construcciones y distinguirlas de las de las muestras de datos multivariantes.
- Recordar que son los modelos probabilísticos los que nos permiten contrastar hipótesis. Los datos son importantes pero los modelos más.

Variable aleatoria multidimensional.

- Una v.a. multidimensional o vector aleatorio es un vector compuesto por p v.a.

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

- Cada v.a. X_i puede ser discreta, continua o de otros tipos. Esto da lugar a una gran diversidad de vectores aleatorios.
- Diferenciemos entre un vector aleatorio multidimensional y una muestra de una vector aleatorio multidimensional.
- El vector aleatorio corresponde a un modelo teórico mientras que la muestra de ese vector corresponde a una recolección de datos del mismo medidos sobre distinto individuos.

Estas mediciones pueden corresponder a dos grandes grupos:

- Provienen de un experimento diseñado y reproducible para estudiar su comportamiento. Este experimento debe tener en cuenta la representatividad de la muestra, su tamaño y los métodos estadísticos que se utilizarán par inferir las conclusiones deseadas.
- Proviene de datos recopilados (si se quiere de forma estadística pero en su sentido etimológico): bases de datos de proteínas, datos estadísticos de Institutos Oficiales (INE, Eurostat, IBAE). O bien de diferentes procedencias.

La diferencia es clara.

- En un estudio inferencial, del que se desee extraer conclusiones estadísticas digamos *profesionales o científicos*, debemos contar con muestras aleatorias que respalden los resultados del experimento.
- Por ejemplo estamos hablando de resultados pubicados en revistan científicas, o de la industria farmacéutica , encuestas profesionales de cualquier índole (políticas/opinión, sociológicas, sanitarias, ecológicas, de estudio de mercados etc.).
- Además el experimento debe ser reproducible.

Vector de valores esperados. Matriz de Covarianza y de Correlaciones.

- Al igual que en el caso unidimensional los vectores aleatorios tienen función de probabilidad o de densidad, función de distribución, medias, varianzas y otros momentos asociados...
- En el caso de los vectores aleatorios estas cantidades se convierten en vectores y en matrices que también medirán, por ejemplo, sus valores esperados y la variación conjunta de las variables.

Vector de medias.

De la misma forma que en el caso unidimensional un vector aleatorio $\vec{X}$ posee un valor esperado que en este caso es el vector de valores esperados o medias formado por los valores esperados de cada una de las componentes. Se representa como

$$E(\vec{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \mu.$$

Covarianzas.

- Dadas dos variables aleatorias, de las que se conoce su distribución conjunta se define su covarianza como

$$Cov(X_1, X_2) = E((X_1 - \mu_1)(X_2 - \mu_2))$$

- Que también se puede calcular con la siguiente identidad

$$Cov(X_1, X_2) = E(X_1 X_2) - \mu_1 \mu_2.$$

- La covarianza puede tomar cualquier valor real y mide el grado de dependencia lineal entre las variables (es decir si existen α y β reales tales que $X_i = \alpha X_j + \beta$). La unidades de la covarianza son difíciles de comparar, sobre todo si los pares de variables se miden con magnitudes diferentes.
- La covarianza de una variable consigo misma es su varianza

Así como para simplificar la notación se suele llamar μ a valor esperado de una variable, se suele utilizar σ para las covarianzas. Así pondremos

$$Cov(X_i, X_j) = \sigma_{ij}; \qquad Cov(X_i, X_i) = \sigma_{ii} = var(X_i) = \sigma_i^2.$$

Notemos que las unidades de la varianza son unidades cuadradas. La raíz cuadrada de la varianza es σ_i, recibe el nombre de desviación típica o estándar de X_i y recupera las unidades de la variable.

Variable tipificada.

Propiedad Sean a y b dos números reales y X_i una variable aleatoria.

- Entonces podemos construir $X_i + b$ que es otra variable aleatoria. Se suele decir que esta variable es un cambio de origen de la variable X_i pues desplaza todos los valores una cantidad b.
- También podemos construir la v.a. aX_i. Se suele decir que esta variable es un cambio de escala de la variable X_i pues agranda si $a > 1$ o encoge si $0 < a < 1$; mientras que para valores negativos sucede un efecto parecido pero además cambia su signo.

Propiedad

- $E(aX_i + b) = aE(X_i) + b$. La esperanza es un operador lineal.
- $var(aX_i + b) = a^2 var(X_i)$. La varianza no varía con cambios de origen y queda multiplicada por el cuadrado del factor de cambio de escala.

Variable tipificada

- La trasformación de una variable del tipo $Z_i = \frac{X_i - \mu_i}{\sigma_i}$ recibe en nombre de variable típica, tipificada o estándar de X_i.
- Se suele llamar tipificar la variable al proceso de calcular su variable típica.
- Utilizando las propiedades de los valores esperados y las varianzas se obtiene que (ejercicio) $E(Z_i) = 0$ y $var(X_i) = 1$.
- Cuando tipificamos una variable obtenemos otra variable que siempre tiene media cero y varianza uno.
- Este proceso es útil a la hora de comparar distribuciones de variables y en el caso multidimensional se utiliza par disminuir el efecto del tamaño y de las unidades en que estén medidas las variables.

Matriz de covarianzas.

De la misma forma que en el caso unidimensional un vector aleatorio $\vec{X}$ posee un una medida de su dispersión respecto al valor medio; es la llamada matriz de covarianzas.

$$Cov(\vec{X}) = \begin{pmatrix} \sigma_1 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix} =$$

$E((\vec{X} - \vec{\mu}).(\vec{X} - \vec{\mu})^t) = \Sigma.$

O lo que es lo mismo

$$Cov(\vec{X}) = E(\vec{X}\vec{X}^t) - \vec{\mu}\vec{\mu}^t.$$

La matriz de covarianzas se suele representar por Σ.

Matriz de Correlaciones.

Como las covarianzas son difíciles de comparar se utiliza el llamado coeficiente de correlación lineal de Pearson que es una medida adimensional de la variación lineal de dos variables.

Definimos la correlación de las variables X_i y X_j como

$$Cor(X_i, X_j) = \rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}.$$

Nota: Es evidente que la correlación de una variable con si misma es uno; $\rho_{ii} = 1$ (ejercicio).

Y podemos construir la matriz de correlaciones

$$Cor(\vec{X}) = \begin{pmatrix} 1 & \rho_{12} & \ldots & \rho_{1p} \\ \rho_{21} & 1 & \ldots & \rho_{2p} \\ \vdots & \vdots & & \vdots \\ \rho_{p1} & \rho_{p2} & \ldots & 1 \end{pmatrix}.$$

Propiedades

- $-1 \leq \rho_{ij} \leq 1$
- $\rho_{ij} = \rho_{ji}$; $\rho_{ii} = 1$.
- Salvo en el signo es invariante a cambios de origen y escala.
- Si $\rho_{ij} = \pm 1$ las variables tienen una relación lineal perfecta. Es decir $X_i = \alpha X_j + \beta$. La pendiente α tiene el mismo signo que la correlación.
- Si la correlación es cero se dice que las variables son incorreladas (notemos que la correlación es cero sii la covarianza es cero).

Propiedad La matriz de correlaciones de un vector aleatorio es igual a la matriz de correlaciones del vector aleatorio cuyas componentes son sus variables tipificadas. Es decir

Sea $\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$ y sea $\vec{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{pmatrix}$

donde

$$Z_i = \frac{X_i - \mu_i}{\sigma_i} \text{ para } i = 1, \ldots, p.$$

Entonces $Cor(\vec{X}) = Cov(\vec{Z})$

Nota: Evidentemente (ejercicio) $Cov(\vec{Z}) = Cor(\vec{Z})$.

Expresión matricial de la tipificación de un vector aleatorio.

La operación consistente en dado un vector aleatorio obtener el vector formado por sus componentes tipificadas admite una forma matricial que recuerda a la tipificación de una variable.

Sea $\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$, sea $\vec{\mu}$ su vector de medias y

$$A = \begin{pmatrix} \sigma_1^{-1} & 0 \ldots & 0 & \\ 0 & \sigma_2^{-1} \ldots & 0 & \\ \vdots & \vdots & \vdots & \\ 0 & 0 & \ldots & \sigma_p^{-1} \end{pmatrix}$$

la matriz que tiene en su diagonal el inverso de las desviaciones típicas y el resto de valores iguales a cero.

Entonces

$$\vec{Z} = A \cdot (\vec{X} - \vec{\mu}).$$

Comprobarlo como ejercicio el efecto de esta transformación para $p = 3$.

- La expresión anterior es un caso particular de tranformación lineal multivariante.
- Una transformación lineal general se obtiene tomando cualquier otra matriz A y sustituyendo por un vector constante cualqueira el vector de medias. Incluso podemos variar las dimensiones de estas matrices.

- Como hemos visto la matriz de correlaciones es una matriz de covarianzas de un cierto vector aleatorio.
- De este hecho se sigue que las propiedades que verifican las matrices de covarianzas las deben de cumplir las de correlaciones. En particular la siguiente propiedad.

Propiedad Las matrices de covarianzas son semidefinidas positivas.

- De esta propiedad se desprende que son simétricas (ejercicio) y que tienen todos sus valores propios no negativos.
- De hecho estas matrices cumple alguna otra propiedad pero para nuestros propósitos esta es suficiente.

¿Por qué hemos tenido que aprender las nociones de la sección anterior?

- Pues el motivo es que tenemos que distinguir entre muestras y variables. Este concepto es tan importante que es el fundamento de la estadística y forzosamente del método científico.
- Cualquier verdad científica pasa por crear un modelo una teoría un paradigma, que no sea contradictorio en si mismo.
- Para que una teoría sea científica debe de existir un experimento con el que podamos demostrar que es falsa (falsacionismo Popper).

Ejercicio Dar ejemplos de teorías, conjeturas populares etc.. que no sean falsables.

- La estadística se mueve en este terreno de una forma elegante. Uno plantea una teoría la modela estadísticamente mediante por ejemplo vectores aleatorios.
- Luego se diseña un experimento en condiciones adecuadas y los datos no contradirán o falsearán la teoría.
- Claro que no será todo tan sencillo, la estadística siempre nos dirá cuan probables son los resultados obtenidos suponiendo que la teoría es cierta.
- Así que la estadística sirve muy bien para demostrar qué datos son falsos. Demostrar que algo es cierto mediante muestras y métodos estadísticos es algo más complicado.

Datos Multivariantes.

Supondremos que hemos observado, recopilado, obtenido etc... p variables (si no se dice lo contrario numéricas) en un conjunto de n individuos u objetos. Es decir lo que tenemos son n observaciones consistentes a su vez en la observación conjunta sobre un individuo de p variables.
Claramente se pueden expresar esta observaciones de forma matricial.

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^t \\ \mathbf{x}_2^t \\ \vdots \\ \mathbf{x}_n^t \end{pmatrix}$$

Donde utilizamos las siguientes notaciones

- Denotamos por

$$\mathbf{x}_i^t = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \ldots \\ x_{ip} \end{pmatrix}^t = (x_{i1}, x_{i2}, \ldots, x_{ip})$$

  Es decir $\mathbf{x}_i^t$ son los vectores filas compuestos por las observaciones de las P variables sobre el i-ésimo individuo.

- Denotamos por

$$\mathbf{x}_{(j)} = \begin{pmatrix} x_{ij} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}$$

- Es decir $\mathbf{x}_{(j)}$ son los vectores columna compuestos por las n observaciones de la j-ésima variable.
- Así podemos expresar la matriz de datos como $\mathbf{X} = (\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \ldots, \mathbf{x}_{(p)})$
- Por último denotamos por $\vec{X} = (X_1, X_2, \ldots, X_p)$ el vector aleatorio cuyas componentes son las variables aleatorias observadas.

Vector de medias, varianzas.

- Con estas notaciones podemos recuperar algunas definiciones ya conocidas de los estadísticos más usuales de una muestra.
- La media aritmética que es un estimador del valor esperado de cada variables.
- La varianza y la desviación típica muestral que estiman los parámetros poblacionales del mismo nombre.

Media de una variable

$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ es la media aritmética de la variable j-ésima en esta muestra. Así el vector de medias aritméticas al que denotaremos por

$$\overline{\mathbf{x}} = \begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \vdots \\ \overline{x}_p \end{pmatrix}$$

es un estimador de $E(\vec{X}) = \vec{\mu}$.

Desviación de una observación respecto a la media

$$d_{ij} = (x_{ij} - \overline{x}_j)^2$$

Varianza muestral

La varianza muestral es el estadístico que estima la varianza σ_j de la variable X_j.
Tiene dos versiones

- La varianza muestral MLE, acrónimo del inglés *maximum likelihood estimator*, estimador máximo verosímil que es aquel valor de σ_j^2 que hace más probable la muestra obtenida.

$$s_j^2 = \frac{1}{n} \sum_{i=1}^{n} d_{ij} = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \overline{x}_j)^2 = \frac{1}{n} \sum_{i=1}^{n} x_{ij}^2 - \overline{x}_j^2.$$

- El estimador insesgado de la varianza. Los estimadores insesgados son aquellos cuyo valor esperado es el verdadero valor del parámetro.

$$\begin{aligned}\tilde{s}_j^2 = & \frac{1}{n-1}\sum_{i=1}^{n} d_{ij} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\overline{x}_j)^2 \\ = & \frac{1}{n-1}\sum_{i=1}^{n} x_{ij}^2 - \frac{n}{n-1}\overline{x}_j^2.\end{aligned}$$

- La relación que existen entre ambas formulas de la varianza muestral es la siguiente:

$$\frac{n}{n-1}s_j^2 = \tilde{s}_j^2$$

Nota: Algunos autores utilizan el nombre de cuasivarianza para denominar a $\tilde{s}_j^2$. En general no se da esta distinción y se utiliza la notación s_j^2 en ambos caso lo que da lugar a confusiones. De todas maneras para valores muestrales grandes las diferencias entre ambos estimadores de la varianza se hacen pequeñas. Los paquetes estadísticos suelen calcular por defecto la cuasivarianza.

Desviación típica muestral
Es la raíz cuadrada positiva de la varianza (de la que estemos utilizando)

Coeficiente de variación
Es una medida de la variación muestral estandarizada (es mejor utilizar la solamente para variables positivas)

$$cv_j = \frac{s_j}{\overline{x}_j}$$

Coeficiente de asimetría, coeficiente de apuntamiento
Otros coeficientes que veremos en prácticas.

Centralización de una matriz de datos.

- Al igual que para el caso univariante centrar una variable es hacer que tenga media 0. Para ello bastará restar a sus observaciones la media.
- En el caso de observaciones multivariantes tendremos que realizar esta operación en todas las columnas de datos

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix} = \frac{1}{n}\begin{pmatrix} \sum_{i=1}^{n} x_{i1} \\ \sum_{i=1}^{n} x_{i2} \\ \vdots \\ \sum_{i=1}^{n} x_{ip} \end{pmatrix} = \begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \vdots \\ \overline{x}_p \end{pmatrix}$$

Cálculo matricial del vector de medias.

- Esta igualdad admite calculo matricial . Sea $\mathbf{1}_n$ un vector columna de n filas todas iguales a 1

Ejemplo El efecto que produce la multiplicación de una matriz 3×4 por el vector $\mathbf{1}_4$ queda muy claro con el siguiente ejemplo

$$\begin{pmatrix} 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 12 \\ 10 \end{pmatrix}.$$

Entonces

$$\frac{1}{n}\mathbf{X}^t\mathbf{1}_n = \frac{1}{n}\begin{pmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ \vdots & \vdots & & \vdots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} =$$

$$\frac{1}{n}\begin{pmatrix} x_{11} + x_{21} + \dots x_{n1} \\ \vdots \\ x_{1p} + x_{2p} + \dots x_{np} \end{pmatrix} =$$

$$\begin{pmatrix} \overline{x}_1 \\ \vdots \\ \overline{x}_p \end{pmatrix} = \overline{\mathbf{x}}$$

Propiedad

$$\overline{\mathbf{x}} = \frac{1}{n}\mathbf{X}^t\mathbf{1}_n$$

Matriz de datos centrados

Dada una matriz de datos $\mathbf{X}$ llamaremos matriz de datos centrados y la denotaremos por $\tilde{\mathbf{X}}$ a la matriz de datos resultante de restar a cada columna de $\mathbf{X}$ su media aritmética.

$$\tilde{\mathbf{X}} = \begin{pmatrix} x_{11} - \overline{\mathbf{x}}_1 & x_{12} - \overline{\mathbf{x}}_2 & \ldots & x_{1p} - \overline{\mathbf{x}}_p \\ x_{21} - \overline{\mathbf{x}}_1 & x_{22} - \overline{\mathbf{x}}_2 & \ldots & x_{2p} - \overline{\mathbf{x}}_p \\ \vdots & \vdots & & \vdots \\ x_{n1} - \overline{\mathbf{x}}_1 & x_{n2} - \overline{\mathbf{x}}_2 & \ldots & x_{np} - \overline{\mathbf{x}}_p \end{pmatrix}$$

El resultado es una matriz de datos donde todas las variables tienen media aritmética cero.

Veamos que la operación de centrado tiene una expresión matricial:

Notación: Llamamos matriz centralizadora de orden n a:
$H_n = I_n - \frac{1}{n}\mathbf{1}_n \cdot \mathbf{1}_n^t$

Es decir $H_n = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix} - \frac{1}{n} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \cdot (1, 1, \ldots, 1) =$

$$\begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix} - \frac{1}{n} \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \ldots & 1 \end{pmatrix} =$$

$$\begin{pmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \ldots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & \ldots & -\frac{1}{n} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \ldots & 1 - \frac{1}{n} \end{pmatrix}$$

Propiedad

- $\tilde{\mathbf{X}} = H_n \cdot \mathbf{X}$, que nos da una expresión matricial del centrado.
- La matriz H cumple que $H \cdot H = H$, es lo que se llama una matriz indempotente (comprobadlo como ejercicio).
- Además H es simétrica tiene rango $n-1$ y $H \cdot \mathbf{1_n} = 0$ (ejercicio).

Tipificación Tabla de datos

- Dado un conjunto de datos llamaremos datos tipificados a los datos resultantes de restar a cada columna de datos u media y dividir el resultado por su desviación típica.
- De esta forma obtenemos datos tipificados que tienen media artimética 0 y varianza 1.
- La tipificación se puede relizar de forma tradicional, o bien matricialmente. La operación de centrado ya nos resta las medias. La división por las desviaciones típicas se puede realizar también matricialmente

Propiedad Sea $\mathbf{Z}$ la matriz de datos resultante de tipificar la matriz de datos $\mathbf{X}$. Sea $\mathbf{D}^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{s_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{s_2} & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \frac{1}{s_p} \end{pmatrix}$

Es decir es la matriz que contiene en la diagonal la inversa de las desviaciones típicas. Entonces:

$$\mathbf{Z} = H_n \cdot \mathbf{X} \cdot \mathbf{D}^{-1/2} = \tilde{\mathbf{X}} \cdot \mathbf{D}^{-1/2}.$$

Ejemplo Sea $X = \begin{pmatrix} 1 & -1 & 3 \\ 1 & 0 & 3 \\ 2 & 3 & 0 \\ 3 & 0 & 1 \end{pmatrix}$
una matriz de datos con $p = 3$ variables y $n = 4$ observaciones por individuo.
Calculemos la matriz de datos tipificado $\mathbf{Z}$.
La matriz de la inversa de las desviaciones típicas es (ejercicio):

$$\mathbf{D}^{-1/2} = \begin{pmatrix} \frac{1}{\sqrt{\frac{11}{16}}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{\frac{9}{4}}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{\frac{27}{16}}} \end{pmatrix}.$$

Entonces

$$\mathbf{Z} = H_4 \cdot \mathbf{X} \cdot \mathbf{D}^{-1/2} = H_4 = \begin{pmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{pmatrix} \cdot$$

$$\begin{pmatrix} 1 & -1 & 3 \\ 1 & 0 & 3 \\ 2 & 3 & 0 \\ 3 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{\frac{11}{16}}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{\frac{9}{4}}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{\frac{27}{16}}} \end{pmatrix} =$$

$$\begin{pmatrix} -\frac{3}{\sqrt{11}} & -1 & \frac{5}{3\sqrt{3}} \\ -\frac{3}{\sqrt{11}} & -\frac{1}{3} & \frac{5}{3\sqrt{3}} \\ \frac{1}{\sqrt{11}} & \frac{5}{3} & -\frac{7}{3\sqrt{3}} \\ \frac{5}{\sqrt{11}} & -\frac{1}{3} & -\frac{1}{\sqrt{3}} \end{pmatrix}$$

Covarianza.

- Se define la covarianza muestral de las variables x_i y x_j como

$$S_{ij} = \frac{1}{n}\sum_{k=1}^{n}(x_{ki} - \overline{x}_i)(x_{kj} - \overline{x}_j) = \frac{1}{n}\sum_{k=1}^{n} x_{ki}x_{kj} - \overline{x}_i\overline{x}_j.$$

- La expresión anterior es un estimador máximo verosímil de σ_{ij}. También se tiene un estimador insesgado que consiste en dividir por $n-1$ en lugar de n.
- La covarianza muestral estima la relación lineal entre las variables.
- Puede tomar cualquier valor. Si es cero se dice que las muestras son incorreladas.
- $s_{ij} = s_{ji}$.
- $s_{ii} = s_i^2$.

Matriz de covarianzas.

Dada una tabla de datos llamaremos matriz de covarianzas (o de varianzas-covarianzas) a la matriz

$$\mathbf{S} = (s_{ij})_{i,j=1,\ldots,p} = \begin{pmatrix} s_{11} & s_{12} & \ldots & s_{1p} \\ s_{21} & s_{22} & \ldots & s_{2p} \\ \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & \ldots & s_{pp} \end{pmatrix}$$

- La matriz de covarianzas muestral representa la variabilidad conjunta de los datos multidimensionales.
- La matriz de covarianzas es simétrica.
- $tr(\mathbf{S}) = \sum_{i=1}^{p} s_i^2 \geq 0$.
- Las matrices de covarianzas son definidas positivas.
- Tienen todos los valores propios no negativos y por lo tanto $\det(\mathbf{S}) \geq 0$.

Expresión matricial de S.

$$\mathbf{S} = \frac{1}{n}\tilde{\mathbf{X}}^t \cdot \tilde{\mathbf{X}} = \frac{1}{n}\mathbf{X}^t \cdot H \cdot \mathbf{X}.$$

$$\mathbf{S} = \frac{1}{n}\begin{pmatrix} x_{11}-\overline{\mathbf{x}}_1 & x_{12}-\overline{\mathbf{x}}_2 & \dots & x_{1p}-\overline{\mathbf{x}}_p \\ x_{21}-\overline{\mathbf{x}}_1 & x_{22}-\overline{\mathbf{x}}_2 & \dots & x_{2p}-\overline{\mathbf{x}}_p \\ \vdots & \vdots & & \vdots \\ x_{n1}-\overline{\mathbf{x}}_1 & x_{n2}-\overline{\mathbf{x}}_2 & \dots & x_{np}-\overline{\mathbf{x}}_p \end{pmatrix}^t \cdot$$

$$\begin{pmatrix} x_{11}-\overline{\mathbf{x}}_1 & x_{12}-\overline{\mathbf{x}}_2 & \dots & x_{1p}-\overline{\mathbf{x}}_p \\ x_{21}-\overline{\mathbf{x}}_1 & x_{22}-\overline{\mathbf{x}}_2 & \dots & x_{2p}-\overline{\mathbf{x}}_p \\ \vdots & \vdots & & \vdots \\ x_{n1}-\overline{\mathbf{x}}_1 & x_{n2}-\overline{\mathbf{x}}_2 & \dots & x_{np}-\overline{\mathbf{x}}_p \end{pmatrix}$$

$$\mathbf{S} = \frac{1}{n}\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}^t \cdot$$

$$\begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \dots & -\frac{1}{n} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \dots & 1-\frac{1}{n} \end{pmatrix} \cdot$$

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} .$$

Ejemplo. Sea $X = \begin{pmatrix} 1 & -1 & 3 \\ 1 & 0 & 3 \\ 2 & 3 & 0 \\ 3 & 0 & 1 \end{pmatrix}$ una matriz de datos con $p = 3$ variables y $n = 4$ observaciones por individuo. De la forma tradicional se ponen los datos en una tabla de la siguiente forma:

i	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2	x_1x_2	x_1x_3	x_2x_3
1	1	-1	3	1	1	9	-1	3	-3
2	1	0	3	1	0	9	0	3	0
3	2	3	0	4	9	0	6	0	0
4	3	0	1	9	0	1	0	3	0
	7	2	7	15	10	19	5	9	-3

Así tenemos que

- $\overline{x}_1 = \frac{7}{4}$, $\overline{x}_2 = \frac{2}{4}$ y $\overline{x}_3 = \frac{7}{4}$
- $s_1^2 = \frac{1}{4}\sum_{i=1}^{4} x_{i1}^2 - \overline{x}_1^2 = \frac{6}{3} - \left(\frac{7}{4}\right)^2 = \frac{11}{16}$
- $s_2^2 = \frac{1}{4}\sum_{i=1}^{4} x_{i2}^2 - \overline{x}_2^2 = \frac{10}{4} - \left(\frac{2}{4}\right)^2 = \frac{9}{4}$
- $s_3^2 = \frac{1}{4}\sum_{i=1}^{4} x_{i3}^2 - \overline{x}_3^2 = \frac{19}{4} - \left(\frac{7}{4}\right)^2 = \frac{27}{16}$
- $s_{12} = \frac{1}{4}\sum_{i=1}^{n} x_{i1}x_{i2} - \overline{x}_1\overline{x}_2 = \frac{5}{4} - \frac{7}{4}\frac{2}{4} = \frac{3}{8}$
- $s_{13} = \frac{1}{4}\sum_{i=1}^{n} x_{i1}x_{i3} - \overline{x}_1\overline{x}_3 = \frac{9}{4} - \frac{7}{4}\frac{7}{4} = -\frac{13}{16}$
- $s_{23} = \frac{1}{4}\sum_{i=1}^{n} x_{i2}x_{i3} - \overline{x}_2\overline{x}_3 = \frac{-3}{4} - \frac{2}{4}\frac{7}{4} = -\frac{13}{8}$

Luego la matriz de covarianzas es

$$\mathbf{S} = \begin{pmatrix} \frac{11}{16} & \frac{3}{8} & -\frac{13}{16} \\ \frac{3}{8} & \frac{9}{4} & -\frac{13}{8} \\ -\frac{13}{16} & -\frac{13}{8} & \frac{27}{16} \end{pmatrix}$$

Hagamos los cálculos de forma matricial

En este caso la matriz $H_4 = \begin{pmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{pmatrix}$

Entonces

$$\mathbf{S} = \frac{1}{4}X^t \cdot H.X = \frac{1}{4} \begin{pmatrix} 1 & 1 & 2 & 3 \\ -1 & 0 & 3 & 0 \\ 3 & 3 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 & 3 \\ 1 & 0 & 3 \\ 2 & 3 & 0 \\ 3 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 1 & 2 & 3 \\ -1 & 0 & 3 & 0 \\ 3 & 3 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} -\frac{3}{4} & -\frac{3}{2} & \frac{5}{4} \\ -\frac{3}{4} & -\frac{1}{2} & \frac{5}{4} \\ \frac{1}{4} & \frac{5}{2} & -\frac{7}{4} \\ \frac{5}{4} & -\frac{1}{2} & -\frac{3}{4} \end{pmatrix} =$$

$$= \frac{1}{4} \begin{pmatrix} \frac{11}{4} & \frac{3}{2} & -\frac{13}{4} \\ \frac{3}{2} & 9 & -\frac{13}{2} \\ -\frac{13}{4} & -\frac{13}{2} & \frac{27}{4} \end{pmatrix} = \begin{pmatrix} \frac{11}{16} & \frac{3}{8} & -\frac{13}{16} \\ \frac{3}{8} & \frac{9}{4} & -\frac{13}{8} \\ -\frac{13}{16} & -\frac{13}{8} & \frac{27}{16} \end{pmatrix} = \mathbf{S}$$

También podemos calcular de forma matricial $\overline{\mathbf{x}}$:

$$\overline{\mathbf{x}} = \frac{1}{4}\mathbf{X}^t\mathbf{1}_4 = \frac{1}{4}\left(\begin{array}{cccc} 1 & 1 & 2 & 3 \\ -1 & 0 & 3 & 0 \\ 3 & 3 & 0 & 1 \end{array}\right) \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 7 \\ 2 \\ 7 \end{pmatrix} = \begin{pmatrix} \frac{7}{4} \\ \frac{2}{4} \\ \frac{7}{4} \end{pmatrix}$$

Ejercicio Se deja como ejercicio el cáculo de la matriz centrada $\tilde{\mathbf{X}}$.

Variables redundantes

Decimos que en una tabla de datos hay variables redundantes cuando una o más variables aportan la misma información que otra.
Una de las maneras en que sucede esto es si por ejemplo la variable x_i cumple que

$$x_i = a_1 x_{i_1} + \cdots + a_k x_{i_k} + b$$

es decir tienen relación lineal.
La matriz de covarianzas es muy útil para descubrir las variables redundantes de tipo lineal. Lo veremos en la siguiente propiedad.

Propiedad

Sea $\mathbf{S}$ un matriz de covarianzas de dimensión p.

- El número de variables redundantes es igual al número de valores propios iguales a cero.
- Si $\det(S) = 0$ entonces existe una variable redundante.
- Si $rg(\mathbf{S}) = k$ entonces existen $p - k$ variables redundantes.

Ejemplo

i	x_1	x_2	x_3
1	1	0	-1
2	1	2	1
3	1	1	0
4	0	3	0

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

Calculemos, en esta ocasión, $\overline{\mathbf{x}}$ en primer lugar.

$$\overline{\mathbf{x}} = \frac{1}{4} \cdot \mathbf{X}^t \cdot \mathbf{1}_4 = \frac{1}{4} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 2 & 1 & 3 \\ -1 & 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{4} = \begin{pmatrix} \frac{3}{4} \\ \frac{3}{2} \\ 0 \end{pmatrix}$$

Ahora podemos restar manualmente las medias a $\mathbf{X}$ para obtener $\tilde{\mathbf{X}}$ (ejercicio)
O podemos calcular directamente

$$\tilde{\mathbf{X}} = H_4 \cdot \mathbf{X} = \begin{pmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & -1 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \\ 0 & 3 & 0 \end{pmatrix} =$$

$$\begin{pmatrix} \frac{1}{4} & -\frac{3}{2} & -1 \\ \frac{1}{4} & \frac{1}{2} & 1 \\ \frac{1}{4} & -\frac{1}{2} & 0 \\ -\frac{3}{4} & \frac{3}{2} & 0 \end{pmatrix}$$

Ahora $\mathbf{S} = \frac{1}{4}\tilde{\mathbf{X}}^t \cdot \mathbf{X} = \frac{1}{4}\begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{3}{4} \\ -\frac{3}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{3}{2} \\ -1 & 1 & 0 & 0 \end{pmatrix} \cdot$
$\begin{pmatrix} \frac{1}{4} & -\frac{3}{2} & -1 \\ \frac{1}{4} & \frac{1}{2} & 1 \\ \frac{1}{4} & -\frac{1}{2} & 0 \\ -\frac{3}{4} & \frac{3}{2} & 0 \end{pmatrix} = \begin{pmatrix} \frac{3}{16} & -\frac{3}{8} & 0 \\ -\frac{3}{8} & \frac{5}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$

Si calculamos $\det(\mathbf{S}) = 0$ (ejercicio). Luego existe al menos una variable redundante.

Veamos si hay más. Calculemos los valores propios.
El polinomio característico de $\mathbf{S}$ es

$$p_{\mathbf{S}} = \begin{vmatrix} \frac{3}{16}-\lambda & -\frac{3}{8} & 0 \\ -\frac{3}{8} & \frac{5}{4}-\lambda & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2}-\lambda \end{vmatrix} = -\lambda^3 + \frac{31\lambda^2}{16} - \frac{9\lambda}{16}$$

Que sólo tiene una solución nula. Por lo tanto sólo hay una variable redundante
¿A alguien se le ocurre una relación lineal entres las variables?

- La dificultad de la interpretación de la matriz de covarianzas como medida de variabilidad radica en que son muchas cantidades.
- Desafortunadamente no hay una sola cantidad que mida la variabilidad multivariante de forma sobresaliente. Veamos dos
- **Varianza total** $T = tr(\mathbf{S}) = \sum_{i=1}^{p} s_i^2 = \sum_{i=1}^{n} \lambda_i$. La varianza media será $\frac{T}{p}$.
- **Varianza Generalizada**

$$\det(\mathbf{S}) = \lambda_1 \cdot \cdots \cdot \lambda_p.$$

La desviación típica generalizada será $\sqrt{\det(\mathbf{S})}$ que cuando el conjunto de datos se representa en R^p es el área, volumen o hipervolumen del conjunto de datos

Correlación lineal de Pearson.

Se define la correlación lineal de Pearson de las muestras de las variables x_i y x_j como

$$r_{ij} = \frac{s_{ij}}{s_i s_j}$$

La correlación r_{ij} estima el parámetro poblacional $\rho_{ij} = Cor(X_i, X_j)$.

Propiedades

- $-1 \leq r_{ij} \leq 1$.
- $r_{ii} = 1$.
- La correlación tiene el mismo signo que la covarianza
- $r_{ij} = \pm 1$ si y sólo sí $x_j = a x_i + b$, para alguna par de valores a y b. Además la pendiente de la recta tiene el mismo signo que la correlación.

Matriz de correlaciones

Llamaremos matriz de correlaciones de la tabla de datos $\mathbf{X}$ a

$$\mathbf{R} = (r_{ij})_{i,j=1,\ldots,p} = \begin{pmatrix} 1 & r_{12} & \ldots & r_{1p} \\ r_{21} & 1 & \ldots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \ldots & 1 \end{pmatrix}$$

Propiedades

- La matriz $\mathbf{R}$ es semidefinida positiva.
- Si todas las variables son incorreladas entonces $\mathbf{R} = I_p$ y $\det(\mathbf{R}) = 1$
- Respecto a las variables redundantes cumple las mismas propiedades que la matriz de covarianzas. Por ejemplo si $\det(R) = 0$ hay al menos una variable redundante.
- $\det(\mathbf{R}) \leq 1$.

Expresión matricial de la matriz de correlaciones.

Sea $\mathbf{D}^{\frac{1}{2}} = \begin{pmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & s_p \end{pmatrix}$

Entonces su inversa es $\mathbf{D}^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{s_1} & 0 & \dots & 0 \\ 0 & \frac{1}{s_2} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \frac{1}{s_p} \end{pmatrix}$

Propiedades

- Expresión matricial de la matriz de correlaciones $\mathbf{R} = \mathbf{D}^{-\frac{1}{2}} \cdot \mathbf{S} \cdot \mathbf{D}^{-\frac{1}{2}}$.
- También tenemos esta otra expresión matricial de la matrzi de covarianzas $\mathbf{S} = \mathbf{D}^{\frac{1}{2}} \cdot \mathbf{R} \cdot \mathbf{D}^{\frac{1}{2}}$ La matriz de covarianzas de los datos tipificados es la matriz de correlaciones de $\mathbf{X}$. Escrito de otra manera si $\mathbf{Z}$ es la matriz de datos tipificados de $\mathbf{X}$ entonces

$$S_{\mathbf{Z}} = \mathbf{R}_{\mathbf{X}}$$

- Esta última propiedad confirma que la matriz de correlaciones es también de covarianzas y cumple sus propiedades.

Ejercicio

Consideremos la siguiente matriz de datos

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

a) Calcular la matriz de correlaciones.

b) Obtener la tabla de datos tipificados $\mathbf{Z}$.

c) Calcular la covarianza de los datos tipificados y comprobar que es igual a la matriz de correlaciones de $\mathbf{X}$.

d) Sin hacer cálculos determinar la matriz de correlaciones de $\mathbf{Z}$.

Otros tipos de correlaciones.

- **Correlaciones parciales.** Hasta ahora hemos visto la relación lineal enter cada par de variables. Pero en un estudio conjunto nos podría interesar la relación lineal de dos variables eliminando el efecto de las demás, este concepto recibe el nombre de correlación parcial.
- **Correlaciones ordinales.** Otros tipos de correlaciones son las llamadas correlaciones ordinales. Lo que se busca es la relación entre los rangos que aproximadamente es el número de orden en la obervación de la variable. Las más conocidas con la correlación ordinal de Spearman y la de Kendall.

Regresion Lineal Múltiple

R. Alberich, A. Mir

22 de mayo de 2007

Parte IV

Regresión lineal simple

Introducción

Introducción

Consideremos los pares de observaciones de dos variables:

$$\{(x_i, y_i) | i = 1, 2, \ldots, n\}$$

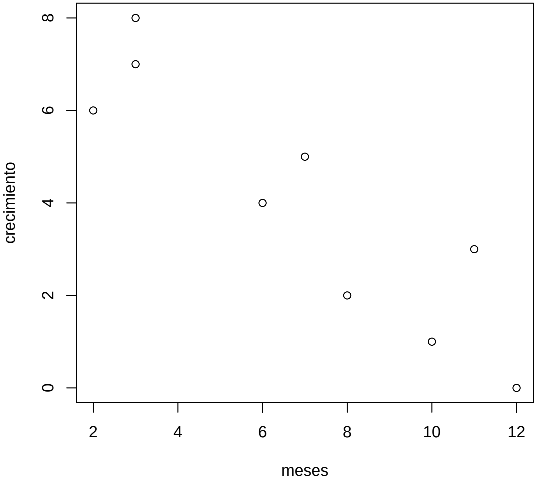
- La variable y es la variable dependiente o de respuesta
- La variable x es la variable de control o independiente o de regresión.
- El problema que se intenta resolver es encontrar la mejor relación funcional que explique la variable y conocido el valor de la variable x:Y/x. En nuestro caso esa función será una recta.

Ejemplo

Ejemplo: Consideremos los datos siguientes donde x representa los meses e y representa el crecimiento de un determinado tipo de planta en mm.

Meses	Crecimiento
12	0
10	1
8	2
11	3
6	4
7	5
2	6
3	7
3	8

Ejemplo



El modelo de Regresión lineal simple

En realidad, en un análisis más riguroso, el modelo de regresión lineal:

$$\mu_{Y/x} = \beta_0 + \beta_1 x,$$

donde $\mu_{Y/x}$ es el valor esperado que toma la variable y cuando la variable de control vale x, mientras que β_0 (término independiente) y β_1 (pendiente) son dos parámetros a determinar.

Dada una muestra calcularemos las estimaciones b_0 y b_1 de β_0 y de β_1 respectivamente. Notemos que para muestras diferentes las estimaciones serán diferentes.
Una vez obtenidas las estimaciones podemos calcular la recta de regresión estimada, que es:

$$\hat{y} = b_0 + b_1 x$$

Regresión lineal simple por Mínimos cuadrados

- Existen diversas maneras de calcular las estimaciones de los coeficientes de una regresión lineal: Regresión ortogonal, métodos robustos, regresión mínimo cuadrática o de mínimos cuadrados,... Nosotros optaremos por el método más habitual que es el de mínimos cuadrados (m.c.).
- Modelo

$$Y_i = \beta_0 + \beta_1 x_i + E_i,$$

  Donde E_i es una nueva variable llamada error o residuo.
- Una vez planteado el modelo y dada una muestra el modelo se debe ajustar a los datos de ésta:

$$y_i = \beta_0 + \beta_1 x_i + E_i, \text{ para } i = 1, 2, \ldots n.$$

Regresión lineal simple por Mínimos cuadrados

- Cuando ajustamos por las estimaciones b_0 y b_1 obtenemos la recta de regresión ajustada

$$\hat{y} = b_0 + b_1 x.$$

- Podemos calcular para cada par de observaciones:

$$y_i = b_0 + b_1 x_i + e_i, \qquad \hat{y}_i = b_0 + b_1 x_i \text{ para } i = 1, 2, \ldots n.$$

- Entonces el error o residuo de la i-ésima observación, $i = 1, 2, \ldots, n$, es

$$e_i = y_i - \hat{y}_i.$$

Cálculo de b_0 y b_1 por M.C.

- Los valores de b_0 y b_1 buscados son los que minimizan el error cuadrático:

$$SSE = \sum_{i=1}^{n} e_i^2.$$

- Estos valores serán los estimadores de β_0 y β_1 por el método de mínimos cuadrados.

Regresión lineal simple por Mínimos cuadrados

Cáculo de los coeficientes

- En primer lugar tenemos que:

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - y_i)^2 = \sum_{i=1} (y_i - b_0 - b_1 x_i)^2.$$

- Calculando las derivadas parciales respecto a b_0 y a b_1, e igualando a cero:

$$\frac{\partial SSE}{\partial b_0} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) = 0,$$

$$\frac{\partial SSE}{\partial b_1} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) x_i = 0.$$

Regresión lineal simple por Mínimos cuadrados
Cálculo de los coeficientes. Ecuaciones normales

- Las ecuaciones anteriores reciben el nombre de ecuaciones normales:

$$\left.\begin{aligned} nb_0 + \sum_{i=1}^n x_i b_1 &= \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i b_0 + \sum_{i=1}^n x_i^2 b_1 &= \sum_{i=1}^n x_i y_i \end{aligned}\right\}$$

- Las soluciones de estas ecuaciones son:

$$b_1 = \frac{n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}, \quad b_0 = \frac{\sum_{i=1}^n y_i - b_1\sum_{i=1}^n x_i}{n}.$$

Ejemplo

En el ejemplo anterior, tenemos:

$$b_1 = \frac{n\sum\limits_{i=1}^n x_i y_i - \sum\limits_{i=1}^n x_i \sum\limits_{i=1}^n y_i}{n\sum\limits_{i=1}^n x_i^2 - (\sum\limits_{i=1}^n x_i)^2} = \frac{9\cdot 175 - 62\cdot 36}{9\cdot 536 - (175)^2} = -0{,}6704,$$

$$b_0 = \frac{\sum\limits_{i=1}^n y_i - b_1\sum\limits_{i=1}^n x_i}{n} = \frac{36 - (-0{,}6704)\cdot 62}{9} = 8{,}6184.$$

Regresión lineal simple por Mínimos cuadrados

Cálculo de los coeficientes. Definición de los momentos de primer y segundo orden.

- Definimos las medias y varianzas de las variabes x e y como:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n},$$

$$s_x^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2,$$

$$s_y^2 = \frac{\sum_{i=1}^{n} (y_i - \overline{y})^2}{n} = \frac{\sum_{i=1}^{n} y_i^2}{n} - \overline{y}^2,$$

Regresión lineal simple por Mínimos cuadrados

Cálculo de los coeficientes. Definición de los momentos de primer y segundo orden.

- Definimos la covarianza entre las variabes x e y como:

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n} = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \overline{x}\overline{y}.$$

Ejemplo anterior

- Los momentos de primer orden son para el ejemplo anterior:

$$\overline{x} = \frac{62}{9} = 6{,}89, \quad \overline{y} = \frac{36}{9} = 4.$$

- Los momentos de segundo orden son:

$$s_x^2 = \frac{536}{9} - 6{,}89^2 = 12{,}09877, s_y^2 = \frac{204}{9} - 4^2 = 6{,}67,$$

$$s_{xy} = \frac{175}{9} - 6{,}89 \cdot 4 = -8{,}111111.$$

Regresión lineal simple por Mínimos cuadrados

Cálculo de los coeficientes en función de los momentos de primer y segundo orden

- Los coeficientes de la recta de regresión son en función de las medias, varianzas y covarianza entre las variables x e y:

$$b_1 = \frac{s_{xy}}{s_x^2}, \quad b_0 = \overline{y} - b_1\overline{x}.$$

- Ejemplo anterior:

$$b_1 = \frac{-8{,}11}{12{,}09877} = -0{,}6704,$$

$$b_0 = 4 - (-0{,}6704)\cdot 6{,}89 = 8{,}6184.$$

Propiedades de los estimadores

- La recta de regresión pasa por el vector de medias $(\overline{x}, \overline{y})$, es decir:

$$b_0 + b_1\overline{x} = \overline{y}$$

- La media de los valores estimados es igual a la media de los observados

$$\overline{\hat{y}} = \frac{\sum_{i=1}^{n} \hat{y}_i}{n} = \overline{y}$$

Ejemplo anterior

- Comprobemos que la recta de regresión pasa por el vector de medias que será $(6{,}89, 4)$:

$$\hat{y} = b_0 + b_1 x, \text{ si } x = 6{,}89, \text{ queda } 8{,}6184 - 0{,}6704 \cdot 6{,}89 \approx 4.$$

- Veamos que la media de los valores estimados es la misma que los valores observados; o sea, 4. Los valores estimados son:

$$0{,}57347,\ 1{,}91429,\ 3{,}25510,\ 1{,}24388,\ 4{,}59591,\ 3{,}92551,\ 7{,}27755,\ 6{,}60714,\ 6{,}60714$$

Si hallamos la media, vale efectivamente 4.

Consideraciones sobre el modelo de regresión lineal

- Se supone que los errores del modelo E_i tienen una distribución normal de media 0 y desviación típica σ.
- Los errores de la estimación por mínimos cuadrados tienen media 0.
  Efectivamente,

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0.$$

  En conclusión

$$\overline{e} = \frac{\sum_{i=1}^{n} e_i}{n} = 0,$$

  y

$$s_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n} = \frac{SSE}{n}$$

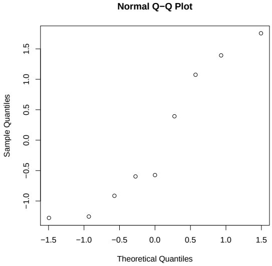
  .

Ejemplo anterior

- Para comprobar la normalidad se realiza un QQ test. Los errores son los siguientes:

  $$-0{,}57347,\ -0{,}91429,\ -1{,}25510,\ 1{,}75612,\ -0{,}59591,\ 1{,}07449,\ -1{,}27755,\ 0{,}39286,\ 1{,}3928571,$$

  El qq-test de los errores aparece en el gráfico siguiente:

  Para que la normalidad se cumpla los puntos deben estar lo más alineados posible.

Definición de las sumas de cuadrados

- Llamaremos suma de cuadrados de los residuales o del error a
$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$
- Llamaremos suma de cuadrados de totales a
$$SST = \sum_{i=1}^{n} (y_i - \overline{y})^2.$$
- Llamaremos suma de cuadrados de la regresión a
$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2.$$

Relación entre las sumas de cuadrados

- En una regresión lineal por el método de mínimos cuadrados se tiene que:

$$SST = SSR + SSE.$$

- La expresión anterior es equivalente a

$$S_y^2 = S_{\hat{y}}^2 + S_e^2.$$

Ejemplo anterior

- Las sumas de cuadrados son:

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 11{,}06020,$$
$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 = 48{,}9398,$$
$$SST = \sum_{i=1}^{n}(y_i - \overline{y})^2 = 60.$$

Se puede comprobar que se cumple la igualdad $SST = SSE + SSR$.

El coeficiente de determinación R^2 y la estimación de la varianza.

- Se define como
$$R^2 = \frac{SSR}{SST}.$$
- En el caso de regresión lineal m.c. se cumple que:
$R^2 = \frac{S_{\hat{y}}^2}{S_y^2}$, $R^2 = 1 - \frac{SSE}{SST}$, $R^2 = 1 - \frac{S_e^2}{S_y^2}$.
- $R^2 = r_{xy}^2$, donde $r_{xy} = \frac{s_{xy}}{s_x s_y}$.
- Por lo tanto R^2 es la proporción de varianza de la variable y que queda explicada por la regresión lineal.
- Una estimación insesgada de σ^2 (la varianza del error E) en m.c. es
$$S^2 = \frac{SSE}{n-2}.$$

Ejemplo anterior

- El coeficiente R^2 valdrá: $R^2 = \frac{48{,}9398}{60} = 0{,}8157$. Por tanto, se explica el 81,57 % de la varianza del crecimiento de la planta.
- Estimación insesgada de la varianza:

$$S^2 = \frac{SSE}{n-2} = \frac{11{,}06020}{7} = 1{,}5800.$$

Intervalos de confianza

- Suponemos de que los residuos siguen una ley normal.
- Intervalo de confianza al nivel $(1-\alpha)100\,\%$ para el parámetro β_1: $(\mu_{Y/x} = \beta_0 + \beta_1 x)$

$$b_1 - \frac{t_{n-2,\alpha/2} S}{\sqrt{n S_x^2}} < \beta_1 < b_1 + \frac{t_{n-2,\alpha/2} S}{\sqrt{n S_x^2}}$$

- Intervalo de confianza al nivel $(1-\alpha)100\,\%$ para el parámetro β_0:

$$b_0 - \frac{t_{n-2,\alpha/2} S \sqrt{\sum_{i=1}^n x_i^2}}{n S_x} < \beta_1 < b_0 + \frac{t_{n-2,\alpha/2} S \sqrt{\sum_{i=1}^n x_i^2}}{n S_x}$$

Intervalos de confianza

- Intervalo de confianza al nivel $(1-\alpha)100\,\%$ para la respuesta media μ_{Y/x_0}:

$$\hat{y}_0 - t_{n-2,\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x_0-\overline{x})^2}{nS_x^2}} < \mu_{Y/x_0} <$$

$$\hat{y}_0 + t_{n-2,\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x_0-\overline{x})^2}{nS_x^2}}$$

- Intervalo de confianza al nivel $(1-\alpha)100\,\%$ para el valor de y_0 cuando $x = x_0$:

$$\hat{y}_0 - t_{n-2,\alpha/2} S \sqrt{1 + \frac{1}{n} + \frac{(x_0-\overline{x})^2}{ns_x^2}} < \mu_{Y/x_0} <$$

$$\hat{y}_0 + t_{n-2,\alpha/2} S \sqrt{1 + \frac{1}{n} + \frac{(x_0-\overline{x})^2}{ns_x^2}}$$

Ejemplo anterior

- Intervalo de confianza al nivel 95 % para la respuesta media μ_{Y/x_0}:

$$\hat{y}_0 - t_{n-2,\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{n s_x^2}} < \mu_{Y/x_0} <$$
$$\hat{y}_0 + t_{n-2,\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{n s_x^2}}$$
$$\hat{y}_0 - t_{7,0{,}025} \sqrt{\frac{1{,}58}{9} + \frac{(x_0 - 6{,}89)^2}{9 \cdot 12{,}09877}} < \mu_{Y/x_0} <$$
$$\hat{y}_0 + t_{7,0{,}025} \sqrt{\frac{1{,}58}{9} + \frac{(x_0 - 6{,}89)^2}{9 \cdot 12{,}09877}}$$
$$\hat{y}_0 - 2{,}36 \cdot \sqrt{0{,}1756 + \frac{(x_0 - 6{,}89)^2}{108{,}889}} < \mu_{Y/x_0} <$$
$$\hat{y}_0 + 2{,}36 \cdot \sqrt{0{,}1756 + \frac{(x_0 - 6{,}89)^2}{108{,}889}}$$

Si cogemos $x_0 = 11$ meses, el intervalo anterior vale: $-0{,}1134 < \mu_{Y/x_0} < 2{,}6012$.

ANOVA en la recta de regresión lineal

Muy brevemente el Análisis de la Varianza (ANalisys Of VAriance) consiste en contrastar si la media de una variable en k poblaciones independientes, con distribución normal de igual varianza, son iguales contra que al menos dos son distintas.

$$\left\{ \begin{array}{ll} H_0: & \mu_1 = \mu_2 = \ldots = \mu_k \\ H_1: & \text{no todas las medias son iguales} \end{array} \right.$$

En el caso de la regresión lineal, usaremos la técnica anterior para contrastar si las medias de los grupos que conforman las variables son iguales o no (el grupo k está formado por los valores cuya media vale μ_{Y/x_k}). En caso afirmativo, decir que las medias son iguales es equivalente a afirmar que $\beta_1 = 0$ y, por lo tanto, el modelo de regresión lineal no es bueno. Por tanto, para que el modelo sea bueno, hemos de rechazar la hipótesis nula en el contraste ANOVA

ANOVA en la recta de regresión lineal

- Test a realizar:

$$\begin{cases} H_0: & \beta_1 = 0 \\ H_1: & \beta_1 \neq 0 \end{cases}$$

- Tabla a calcular:

Fuente de variación	Suma de cuadrados	g. l.	Cuadrados medios	F
Regresión	SSR	1	SSR	SSR/S^2
Error	SSE	$n-2$	$S^2 = \frac{SSE}{n-2}$	
Total	SST	$n-1$		

ANOVA en la recta de regresión lineal

- Ahora rechazamos la hipótesis nula al nivel de significación α si $f > f_{\alpha,1,n-2}$ donde $f_{\alpha,1,n-2}$ es el valor de una distribución F de con grados de libertad 1 y $n-2$.
- Esta prueba, en el caso de regresión lineal simple tiene un efecto a otra parecida en la que se contrasta con una t de student.

Ejemplo anterior

- Tabla ANOVA:

Fuente de variación	Suma de cuadrados	g. l.	Cuadrados medios	F
Regresión	48,9398	1	48,9398	30,974
Error	11,0602	7	$S^2 = 1{,}5800$	
Total	60	8		

- Cogemos $\alpha = 0{,}05$. El valor $f_{0{,}05,1,7}$ vale 5,59. Como $f = 30{,}974 > 5{,}59$, rechazamos la hipótesis nula y concluimos que $\beta_1 \neq 0$. Por tanto, nuestro modelo es adecuado según este análisis.

Parte V

Regresión lineal múltiple

Introducción

- Tenemos k variables independientes $x_1, \ldots, x_k$ y una variable dependiente y.
- Postulamos el modelo de regresión lineal como:

$$\mu_{Y,x_1,\ldots,x_k} = \beta_0 + \beta_1 x_1 + \ldots \beta_k x_k.$$

Los parámetros β_i son desconocidos y se pueden estimar a partir de una muestra:

$$\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) | i = 1, 2, \ldots, n\}$$

de la que se exige que $n > k$, es decir el número de observaciones sea mayor que el número de variables.

Introducción

- El modelo es el siguiente: Consideramos una conjunto de k variables aleatorias $X_1, X_2, \ldots, X_k$. Suponemos que existen variables aleatorias respuestas $Y_1, \ldots, Y_k$ cuya relación con las anteriores es:

$$Y_i = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + E_i,$$

  donde E_i son variables aleatorias que representan el error aleatorio del modelo asociado a la respuesta Y_i.
- El problema es estimar los parámetros β_i a partir de una muestra de datos que representan una muestra aleatoria simple de las variables X_i y de la variable Y de tamaño n:

$$\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) | i = 1, 2, \ldots, n\}.$$

Introducción

- Llamaremos y_i al valor estimado de la variable Y_i usando las estimaciones b_i de los parámetros β_i:

  $$y_i = b_0 + b_1 x_{i1} + \cdots + b_k x_{ik} + e_i \text{ para } i = 1, 2, \ldots, n$$

  donde e_i será la estimación de la variable error residual E_i asociado a la respuesta Y_i.
- Llamaremos

  $$\hat{y}_i = b_0 + b_1 x_{i1} + \cdots + b_k x_{ik}.$$

  Entonces $e_i = y_i - \hat{y}_i$.

Introducción

- Definimos los vectores siguientes:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}, \hat{\mathbf{y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix}, \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}.$$

- Definimos la matriz siguiente:

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$$

Introducción

- Podemos escribir el modelo de regresión múltiple matricialmente como:

$$\begin{aligned} \mathbf{y} = &\quad \mathbf{X}\mathbf{b} + \mathbf{e}, \\ \hat{\mathbf{y}} = &\quad \mathbf{X}\mathbf{b}, \\ \mathbf{e} = &\quad \mathbf{y} - \hat{\mathbf{y}}. \end{aligned}$$

Cálculo de los coeficientes b_i usando el método de mínimos cuadrados

- Definimos el error cuadrático SSE como:

$$\begin{aligned} SSE = & \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ = & \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - \cdots - b_{ik} x_k)^2. \end{aligned}$$

- Los estimadores por el método de mínimos cuadrados serán los valores $b_0, b_1, \ldots, b_k$ que minimicen SSE.
- Para resolver este problema calculamos las derivadas parciales de SSE respecto a cada b_i para $i = 1, 2, \ldots, n$ y se obtiene el un sistema de ecuaciones que recibe el nombre de ecuaciones normales.

Cálculo de los coeficientes b_i usando el método de mínimos cuadrados

$$\left.\begin{array}{rl}
nb_0 + b_1 \sum_{i=1}^n x_{i1} + \ldots + b_k \sum_{i=1}^n x_{ik} & = \sum_{i=1}^n y_i \\
b_0 \sum_{i=1}^n x_{i1} + b_1 \sum_{i=1}^n x_{i1}^2 + b_2 \sum_{i=1}^n x_{i1}x_{i2} + \ldots + & \\
b_k \sum_{i=1}^n x_{i1}x_{ik} & = \sum_{i=1}^n x_{i1}y_i \\
\ldots & \ldots \\
b_0 \sum_{i=1}^n x_{ik} + b_1 \sum_{i=1}^n x_{ik}x_{i1} + b_2 \sum_{i=1}^n x_{ik}x_{i2} + \ldots + & \\
b_k \sum_{i=1}^n x_{ik}^2 & = \sum_{i=1}^n x_{ik}y_i
\end{array}\right\}$$

Cálculo de los coeficientes b_i usando el método de mínimos cuadrados

- El sistema anterior se puede expresar en forma matricial de la forma siguiente:

$$\left(\mathbf{X}^\top \mathbf{X}\right) \cdot \mathbf{b} = \mathbf{X}^\top \cdot \mathbf{y}.$$

- La solución buscada del sistema anterior será:

$$\mathbf{b} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \cdot \left(\mathbf{X}^\top \mathbf{y}\right).$$

Ejemplo

- Se postula que la estatura de un niño recién nacido (y) tiene una relación con su edad en días x_1, su estatura al nacer en cm. (x_2), su peso en Kg. al nacer (x_3) y su estatura en cm. al nacer (x_4). Se pudo obtener una pequeña muestra con $n = 9$ niños cuyos resultados fueron:
-

y	x_1	x_2	x_3	x_4
57,5	78	48,2	2,75	29,5
52,8	69	45,5	2,15	26,3
61,3	77	46,3	4,41	32,2
67	88	49	5,52	36,5
53,5	67	43	3,21	27,2
62,7	80	48	4,32	27,7
56,2	74	48	2,31	28,3
68,5	94	53	4,3	30,3
69,2	102	58	37,1	28,7

Ejemplo

- La matriz **X** es:

$$\mathbf{X} = \begin{pmatrix} 1 & 78 & 48,2 & 2,75 & 29,5 \\ 1 & 69 & 45,5 & 2,15 & 26,3 \\ 1 & 77 & 46,3 & 4,41 & 32,2 \\ 1 & 88 & 49 & 5,52 & 36,5 \\ 1 & 67 & 43 & 3,21 & 27,2 \\ 1 & 80 & 48 & 4,32 & 27,7 \\ 1 & 74 & 48 & 2,31 & 28,3 \\ 1 & 94 & 53 & 4,3 & 30,3 \\ 1 & 102 & 58 & 37,1 & 28,7 \end{pmatrix}$$

Ejemplo

- El vector **y** es:

$$\mathbf{y} = \begin{pmatrix} 57,5 \\ 52,8 \\ 61,3 \\ 67 \\ 53,5 \\ 62,7 \\ 56,2 \\ 68,5 \\ 69,2 \end{pmatrix}$$

Ejemplo

- El producto $\mathbf{X}^\top \mathbf{X}$ es el siguiente:

$$\begin{pmatrix} 9 & 729 & 439 & 66{,}07 & 266{,}7 \\ 729 & 60123 & 35947{,}2 & 6108{,}19 & 21715{,}3 \\ 439 & 35947{,}2 & 21568{,}18 & 3541{,}008 & 13026{,}01 \\ 66{,}07 & 6108{,}19 & 3541{,}008 & 1491{,}306 & 1948{,}561 \\ 266{,}7 & 21715{,}3 & 13026{,}01 & 1948{,}561 & 7980{,}83 \end{pmatrix}$$

- El producto $\mathbf{X}^\top \mathbf{y}$ es el siguiente:

$$\mathbf{X}^\top \mathbf{y} = \begin{pmatrix} 548{,}7 \\ 45001 \\ 26946{,}89 \\ 4346{,}109 \\ 16348{,}29 \end{pmatrix}$$

Ejemplo

- Resolviendo el sistema $(\mathbf{X}^\top \mathbf{X}) \cdot \mathbf{b} = \mathbf{X}^\top \mathbf{y}$ obtenemos como solución:

$$\mathbf{b} = \begin{pmatrix} 40{,}7752 \\ 0{,}9911 \\ -1{,}19901 \\ -0{,}0502 \\ -0{,}0416 \end{pmatrix}$$

- La recta de regresión estimada es :

$$\hat{y} = 40{,}7752 + 0{,}9911 x_1 - 1{,}19901 x_2 - 0{,}0502 x_3 - 0{,}0416 x_4.$$

Propiedades de la recta de regresión

- La recta de regresión ajustada pasa por el vector de medias. O sea, si llamamos $\overline{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}$, para $i = 1, \ldots, k$, $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$, se verifica que:

$$\overline{y} = b_0 + b_1\overline{x}_1 + \cdots + b_k\overline{x}_k.$$

- La suma de los errores e_i es 0: $\sum_{i=1}^{n} e_i = 0$ y por lo tanto su media también: $\overline{e} = 0$.
- La media de los valores estimados coincide con la media de los valores de la muestra: $\overline{\hat{y}} = \overline{y}$.

Ejemplo

- Veamos que la recta de regresión pasa por el vector de medias. Éste vale:

  $$\overline{x}_0 = 1,\ \overline{x}_1 = 81,\ \overline{x}_2 = 48{,}778,\ \overline{x}_3 = 7{,}341,\ \overline{x}_4 = 29{,}633.$$

  La media del vector $\mathbf{y}$ vale: $\overline{y} = 60{,}967$.
  Se cumple:

  $$\begin{aligned} 60{,}967 \approx\ & 40{,}7752 + 0{,}9911 \cdot 81 - 1{,}19901 \cdot 48{,}778 \\ & -0{,}0502 \cdot 7{,}341 - 0{,}0416 \cdot 29{,}633. \end{aligned}$$

Ejemplo

- Veamos que la media de los errores es nula. Los valores de los vectores $\hat{\mathbf{y}}$ y $\mathbf{e}$ son:

$$\hat{\mathbf{y}} = \mathbf{X} \cdot b = \begin{pmatrix} 58{,}922 \\ 53{,}403 \\ 60{,}014 \\ 67{,}444 \\ 54{,}328 \\ 61{,}140 \\ 55{,}269 \\ 68{,}913 \\ 69{,}267 \end{pmatrix}, \quad \mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \begin{pmatrix} -1{,}422 \\ -0{,}603 \\ 1{,}286 \\ -0{,}444 \\ -0{,}828 \\ 1{,}560 \\ 0{,}930 \\ -0{,}413 \\ -0{,}067 \end{pmatrix}$$

- Puede comprobarse que la media del vector $\mathbf{e}$ es 0 y que $\overline{\hat{y}} = \overline{y} = 60{,}967$.

Sumas de cuadrados en la regresión

- Llamaremos suma de cuadrados de los residuales o del error a
$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$
- Llamaremos suma de cuadrados de totales a
$$SST = \sum_{i=1}^{n} (y_i - \overline{y})^2.$$
- Llamaremos suma de cuadrados de la regresión a
$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2.$$
- Se verifica:
$$SST = SSR + SSE,$$
o equivalentemente:
$$S_y^2 = S_{\hat{y}}^2 + S_e^2.$$

Ejemplo

- La suma de cuadrados del error vale en el ejemplo anterior:

$$SSE = (57{,}5-58{,}922)^2+\cdots+(69{,}2-69{,}267)^2 = 8{,}397.$$

- La suma de cuadrados totales vale:

$$SST = (57{,}5-60{,}967)+\cdots+(69{,}2-60{,}967)^2 = 321{,}24.$$

- La suma de cuadrados de la regresión es:

$$\begin{aligned} SSR = & \ (58{,}922-60{,}967)^2+\cdots+(69{,}267-60{,}967)^2 \\ = & \ 312{,}8434. \end{aligned}$$

- Puede observarse que se cumple:

$$SST = SSE + SSR, \quad 321{,}24 = 8{,}397 + 312{,}8434.$$

Definición del coeficiente de determinación

- Definimos el coeficiente de determinación como
$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST},$$
o también
$$R^2 = \frac{S^2_{\hat{y}}}{S^2_y} = 1 - \frac{S^2_e}{S^2_y}.$$
R^2 se interpreta como la proporción de varianza de la variable y que es explicada por el modelo de regresión múltiple.
- Definimos el coeficiente de determinación ajustado como:
$$R^2a = R^2 - \frac{k(1-R^2)}{n-k-1}.$$
- Definimos el coeficiente de correlación múltiple r de la variable y respecto de las variables $x_1, \ldots, x_k$ como $r = \sqrt{R^2}$.

Ejemplo

- El coeficiente de determinación vale en nuestro ejemplo:

$$R^2 = \frac{312{,}8434}{321{,}24} \approx 0{,}974.$$

- El coeficiente de determinación ajustado vale:

$$R^2 a = 0{,}974 - \frac{4 \cdot (1 - 0{,}974)}{9 - 4 - 1} = 0{,}948.$$

- El coeficiente de correlación múltiple r de la variable y respecto de las variables $x_1,\ x_2,\ x_3,\ x_4$ vale: $r = \sqrt{0{,}974} = 0{,}987.$

Consideraciones sobre el modelos de regresión múltiple

- Suponemos que las variables aleatorias error E_i son independientes e idénticamente distribuidas según una normal de media 0 y varianza σ^2.
- Bajo el supuesto anterior, los estimadores $b_0, \ldots, b_k$ de $\beta_0, \ldots, \beta_k$ son insesgados. O sea, $E(b_i) = \beta_i,\ i = 0, \ldots, k$.
- La matriz $(X^\top X)^{-1}\sigma^2$ es la matriz de covarianzas de $\beta_0, \ldots, \beta_k$.
- Un estimador insesgado de σ^2 es

$$S^2 = \frac{SSE}{n-k-1}.$$

Ejemplo

- Una estimación de la varianza σ^2 será:

$$S^2 = \frac{8{,}397}{9-4-1} = 2{,}099.$$

- Una estimación de la matriz de covarianzas de $\beta_0, \ldots, \beta_4$: $(X^\top X)^{-1} S^2 =$

$$\begin{pmatrix} 18{,}9 & 1530{,}3 & 921{,}5 & 138{,}7 & 559{,}8 \\ 1530{,}3 & 126207{,}1 & 75458{,}5 & 12822{,}0 & 45583{,}6 \\ 921{,}5 & 75458{,}5 & 45274{,}8 & 7433{,}1 & 27343{,}5 \\ 138{,}7 & 12822{,}0 & 7433{,}1 & 3130{,}5 & 4090{,}3 \\ 559{,}8 & 45583{,}6 & 27343{,}5 & 4090{,}3 & 16753{,}0 \end{pmatrix}$$

ANOVA en regresión lineal múltiple

- El contraste ANOVA en la regresión lineal múltiple nos permite contrastar la adecuación del modelo. Se trata de contrastar:

$$\begin{cases} H_0: & \beta_1 = \beta_2 = \cdots = \beta_k = 0 \\ H_1: & \text{hay alguna } \beta_i \neq 0 \end{cases}$$

- Si aceptamos H_0 estamos diciendo que la estimación dada por la regresión es constante. Por tanto el modelo no sería adecuado.
- En la tabla siguiente aparece los pasos necesarios para realizar el contraste. Como puede observarse, se usa como estadístico de contraste el cociente $\frac{MSR}{MSE}$ que, suponiendo normalidad, sigue una distribución F de Snédecor de $k, n-k-1$ grados de libertad:

ANOVA en regresión lineal múltiple

F.V.	S.C.	g.l.	C.M.	f
Regresión	SSR	k	$MSR = \frac{SSR}{k}$	$f = \frac{MSR}{MSE}$
Error	SSE	$n-k-1$	$MSE = \frac{SSE}{n-k-1}$	
Total	SST	$n-1$		

F.V.: Fuente de variación.

S.C.: Suma de cuadrados.

g.l.: grados de libertad.

C.M.: Cuadrados medios.

Rechazaremos la hipótesis nula $H_0 : \beta_1 = \ldots = \beta_k = 0$ al nivel de significación α si $f > f_{\alpha,k,n-k-1}$ donde $f_{\alpha,k,n-k-1}$ es el valor crítico de una distribución F de con grados de libertad k y $n-k-1$.

Ejemplo

- La tabla ANOVA es en nuestro ejemplo:

F.V.	S.C.	g.l.	C.M.	f
Regresión	312,84	4	$MSR = 78{,}21$	$f = 37{,}258$
Error	8,397	4	$MSE = 2{,}099$	
Total	321,24	8		

- Cogiendo $\alpha = 0{,}05$, el valor crítico $f_{0{,}05,4,4}$ vale 6,388. Como $f = 37{,}258 > f_{0{,}05,4,4} = 6{,}388$, rechazamos la hipótesis nula y concluimos que el modelo es adecuado según este análisis.

ANOVA en regresión lineal múltiple

- El modelo de regresión lineal con este conjunto de x's puede no es el único que se puede utilizar. Es posible que con algunas transformaciones de las x's mejore el valor de f.
- El modelo podría ser más eficaz si se incluyen otras variables o podría continuar siendo casi igual de eficaz si se eliminan algunas (principio de parsimonia).

Intervalos de confianza

- Un intervalo de confianza al nivel $(1-\alpha)100\,\%$ para la respuesta media $\mu_{Y/x_{10},x_{20},\ldots,x_{k0}}$ es

$$\hat{y}_0 - t_{\alpha/2,n-k-1} S\sqrt{x_0^\top (X^\top X)^{-1} x_0} <$$

$$\mu_{Y/x_{10},x_{20},\ldots,x_{k0}} < \hat{y}_0 + t_{\alpha/2,n-k-1} S\sqrt{x_0^\top (X^\top X)^{-1} x_0}$$

  donde $t_{\alpha/2,n-k-1}$ es el valor crítico de una t de student con $n-k-1$ grados de libertad,
  $x_0 = (1, x_{10}, x_{20}, \ldots, x_{k0})^\top$ y
  $\hat{y}_0 = b_0 + b_1 x_{10} + \cdots + b_k x_{k0}$.
- La cantidad $S\sqrt{x_0^\top (X^\top X)^{-1} x_0}$ recibe el nombre de error estándar de predicción.

Ejemplo

- Para $\alpha = 0{,}05$, hallemos un intervalo de confianza para la respuesta media $\mu_{Y/x_{10},x_{20},x_{30},x_{40}}$, para $x_{10} = 69,\ x_{20} = 45{,}5,\ x_{30} = 2{,}15,\ x_{40} = 26{,}3$.
- El valor crítico $t_{0{,}025,4}$ vale 2,776. El valor $\hat{y}_0$ valdrá:

$$\begin{aligned} \hat{y}_0 = & \quad \beta_0 + \sum_{i=1}^{4} b_i x_{i0} \\ = & \quad 40{,}775 + 0{,}991 \cdot 69 - 1{,}199 \cdot 45{,}5 - 0{,}05 \cdot 2{,}15 \\ & -0{,}041 \cdot 26{,}3 = 53{,}403. \end{aligned}$$

- El valor de $x_0^\top (X^\top X)^{-1} x_0$ es:

$$(1, 69, 45{,}5, 2{,}15, 26{,}3) \cdot (X^\top X)^{-1} \cdot \begin{pmatrix} 1 \\ 69 \\ 45{,}5 \\ 2{,}15 \\ 26{,}3 \end{pmatrix} = 1{,}722.$$

Ejemplo

El intervalo de confianza será:

$$53{,}403 - 2{,}776 \cdot \sqrt{2{,}099 \cdot 1{,}722} < \mu_{Y/x_{10},x_{20},x_{30},x_{40}}$$
$$< 53{,}403 + 2{,}776 \cdot \sqrt{2{,}099 \cdot 1{,}722} =$$
$$48{,}123 < \mu_{Y/x_{10},x_{20},x_{30},x_{40}} < 58{,}682.$$

Intervalos de confianza

- Un intervalo de confianza al nivel $(1-\alpha)100\,\%$ para una predicción individual y_0 para los valores de la variables dependientes $x_{10}, x_{20}, \ldots, x_{k0}$ es

$$\hat{y}_0 - t_{\alpha/2,n-k-1} S\sqrt{1 + x_0^\top (X^\top X)^{-1} x_0} < y_0 <$$

$$\hat{y}_0 + t_{\alpha/2,n-k-1} S\sqrt{1 + x_0^\top (X^\top X)^{-1} x_0},$$

donde $t_{\alpha/2,n-k-1}$ es el valor crítico de una t de student con $n-k-1$ grados de libertad, y $x_0 = (1, x_{10}, x_{20}, \ldots, x_{k0})^\top$.

El problema de la selección del modelo. Colinealidad

- Dado un problema regresión lineal múltiple podemos ajustar todos los submodelos lineales posibles

$$
\begin{aligned}
Y = & \ \beta_0 + \beta_1 x_1, \\
Y = & \ \beta_0 + \beta_1 x_1 + \beta_2 x_2, \\
& \ \ldots \\
Y = & \ \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k,
\end{aligned}
$$

  ya que es posible que entre variables x_i exista una fuerte relación lineal y entonces *sobren del modelo*. Este problema recibe el nombre de colinealidad.
- Existen también otro tipo de problemas que se pueden resolver usando la técnica de la regresión lineal múltiple. Véase como ejemplo el problema de la regresión polinomial:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k.$$

El problema de la selección del modelo. Colinealidad

Una solución para ver qué modelo lineal es el más simple y adecuado es recurrir a los llamados métodos secuenciales de selección del modelo lineal, como son los siguientes:

- Regresión paso paso (Stepwise).
- Selección hacia adelante(Forward).
- Selección hacia atrás (Backward).

Parte VI

Análisis de Componentes Principales.

Análisis de Componentes Principales.

R. Alberich, A. Mir

22 de mayo de 2007

Matriz (tabla) de datos.

Ind.	x_1	x_2	$\ldots$	x_k	v_1	v_2
1						
2						
3						
$\vdots$						
n						
s_1						
s_2						

- Donde las variables $x_1, \ldots, x_n$ describen una realidad común de los n individuos observados.
- Las variables v_1, v_2 son de perfil (o explicativas) y los individuos s_1, s_2 son individuos suplementarios o ilustrativos.
- Tanto los individuos como las variables suplementarias ayudan a interpretar la variabilidad de los datos.

Objetivos del análisis

- Reducción de la dimensionalidad (factorialidad).
- Lo que se busca es un espacio de variables más reducido y fácil de interpretar.
- El problema es que si reducimos el número de variables es posible que no representemos toda la variabilidad de los datos originales.
- El idea básica es:
  Consentir una pérdida de información para lograr una ganancia en la significación

Análsis Factorial

- Algunos autores consideran el ACP como una parte del Anàlisis Factorial
- En las técnicas de AF se postula que la variabilidad total se puede explicar mediante distintos tipos de factores:
  - factores comunes subyacentes (F_i).
  - factores específicos de las variables (E_i).
  - fluctuaciones aleatorias (A_i).

$$X_1 = \alpha_{11} F_1 + \alpha_{12} F_2 + \cdots + \alpha_{1k} F_k + E_1 + A_1$$

$$X_2 = \alpha_{21} F_1 + \alpha_{22} F_2 + \cdots + \alpha_{2k} F_k + E_2 + A_2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$X_p = \alpha_{p1} F_1 + \alpha_{p2} F_2 + \cdots + \alpha_{pk} F_k + E_p + A_p$$

- Podríamos decir que en un Análisis Factorial se fija a priori la cantidad de varianza de cada variable que debe quedar interpretada por los factores comunes.
- Este valor recibe el nombre de comunalidad y se suele representar como h_i^2.

Así tenemos

- h_i^2 comunalidad de la variable X_i, es la varianza explicada por $F_1, F_2, \ldots F_k$.
- $s_i^2 - h_i^2$ es la varianza de la variable X_i que se queda en los factores específicos y aleatorios.

Var. observada = Var. común + Var. específica y aleat..

El problema de los Componentes Principales.

"Todos los factores son comunes"

$$X_1 = \alpha_{11} CP_1 + \alpha_{12} CP_2 + \cdots + \alpha_{1p} CP_p$$

$$X_2 = \alpha_{21} CP_1 + \alpha_{22} CP_2 + \cdots + \alpha_{2p} CP_p$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$X_p = \alpha_{p1} CP_1 + \alpha_{p2} CP_2 + \cdots + \alpha_{pp} CP_p$$

Se trata de encontrar unas nuevas variables $CP_1, \ldots CP_p$, a las que llamaremos componentes principales, de forma que:

- Se cumplan las condiciones anteriores.
- El origen de las variables esté situado en el vector de medias o centro de gravedad de las observaciones.
- Sean incorreladas entre si $Cor(CP_i, CP_j) = 0$ para $i \neq j = 1, \ldots, p$.
- $Var(CP_1) \geq Var(CP_2) \geq \cdots \geq Var(CP_p)$ y hagan máximas estas varianzas.
- Se conserva la varianza total (inercia) de la nube de puntos.

Tipos de A.C.P:

- Sobre los datos centrados: a cada variable se le resta su media $x_i - \overline{x}_i$.
- Sobre los datos tipificados $\frac{x_i - \overline{x}_i}{s_i}$.
- En el primer caso las variables centradas tienen media cero y la misma varianza que las variables originales: se le suele llamar ACP de covarianzas.
- En el segundo caso las variables tipificadas tienen media cero y varianza 1: se le suele llamar ACP de correlaciones o normado.

Recordemos que dada una matriz de datos X ($n \times p$ es decir de n individuos y p variables) representábamos por $\tilde{X}$ la matriz de datos centrada. Entonces:

- La matriz de covarianzas de X viene dada por $\mathbf{S} = 1/n\tilde{X}^t\tilde{X}$
- Si llamamos Z a la tabla de los datos tipificados la matriz de correlaciones viene dada por $\mathbf{R} = 1/n\tilde{Z}^t Z$

Teorema

- Los componentes principales vienen determinados por los vectores propios ortonormales (ordenados de mayor a menor valor propio) de la matriz de covarianzas (para datos centrados) y de la matriz de correlaciones (para los datos tipificados).
- Así en el ACP de covarianzas cada variable interviene con su propia varianza mientras que el ACP de correlaciones todas las variables tienen varianza 1.

ACP covarianzas:

- Sea **S** la matriz de covarianzas de orden p. Calculamos sus valores propios

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$$

y los correspondientes vectores propios ortonormales (perpendiculares y de norma 1)

$$u_1, u_2, \ldots, u_p$$

- Las direcciones de los componentes principales quedan determinadas por su respectivo vector.
- **Cálculo de las coordenadas de los componentes :**

$$CP_i = \tilde{X} u_i, \text{ para } i = 1, \ldots, p.$$

ACP correlaciones.

Sea **R** la matriz de correlaciones de orden p. Calcularemos sus valores propios

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$$

y los correspondientes vectores propios ortonormales.

$$u_1, u_2, \ldots, u_p$$

Las direcciones de los componentes principales quedan determinadas por el vector propio correspondiente.
Cálculo de las coordenadas de las componentes :
$CP_i = Zu_i,$ para $i = 1, \ldots, p$
donde Z es la matriz de datos tipificados.

Propiedades ACP covarianzas.

Sea X una matriz de datos $n \times p$ y sea

$$\mathbf{S} = \begin{pmatrix} s_1^2 & s_{12} & \dots & s_{1p} \\ s_{21} & s_2^2 & \dots & s_{2p} \\ \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & \dots & s_p^2 \end{pmatrix}$$

Recordemos que s_i^2 es la varianza de la variable x_i y que s_{ij} son las covarianzas de la variables i y j.
Además la Varianza Total $= tr(\mathbf{S}) = \sum_{i=1}^{p} s_i^2$

Propiedades ACP covarianzas.

- $Var(CP_i) = \lambda_i$. La varianza de cada componente principal es su valor propio.
- $\sum_{i=1}^{n} CP_i = \sum_{i=1}^{n} \lambda_i = tr(\mathbf{S}) = \sum_{i=1}^{n} s_i^2$. Por lo tanto los componentes principales reproducen la varianza total
- Los componentes principales tienen correlación cero entre sí (son incorrelados) por lo tanto su matriz de covarianzas es

$$\mathbf{S}_{CP} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_p \end{pmatrix}$$

Propiedades ACP covarianzas.

- $\det(\mathbf{S}_{CP}) = \prod_{i=1}^{n} \lambda_i = \det(\mathbf{S})$. Luego los componentes principales conservan la varianza generalizada.
- La proporción de varianza explicada por la componente j-ésima es
$$\frac{\lambda_j}{\sum_{i=1}^{n} \lambda_i}.$$
Además al ser incorrelados la proporción de varianza explicada por los k primeros componentes es
$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i}.$$
- $Cov(X_i, CP_i) = \lambda_i u_{ji}$; $Corr(X_j, CP_i) = \frac{\sqrt{\lambda_i} u_{ji}}{s_j}$ donde u_{ji} es la j-ésima componente del vector propio u_i.

Propiedades ACP covarianzas.

- La primera componente principal es la recta que conserva mayor inercia de la nube de puntos.
- Los dos primeros componentes principales forma el plano que conserva mayor inercia de la nube de puntos.
- Lo mismo sucede con los espacios formados por los k primeros componentes

Propiedades ACP correlaciones.

Sea X una matriz de datos $n \times p$ y sea

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21}^2 & 1 & \dots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & s_{p2} & \dots & 1 \end{pmatrix}$$

Su matriz de correlaciones. Se verifican las siguientes propiedades:

- Recordemos que la diagonal es 1 pues es la varianza de los datos tipificados y que r_{ij} son las correlaciones lineales de la variables i y j.
- Además la Varianza Total $= tr(\mathbf{R}) = p$

Propiedades ACP correlaciones.

- $Var(CP_i) = \lambda_i$. El valor propio del componente es igual a su varianza
- $\sum_{i=1}^{n} CP_i = \sum_{i=1}^{n} \lambda_i = tr(\mathbf{R}) = p$. Por lo tanto los componentes principales reproducen la varianza total y ésta es igual al numero de variables p.
- Los componentes principales tienen correlación cero entre sí (son incorrelados) por lo tanto su matriz de covarianzas (que este caso es igual a la de correlaciones es

$$\mathbf{S}_{CP} = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \lambda_p \end{pmatrix}$$

Propiedades ACP correlaciones.

- $\det(\mathbf{S}_{CP}) = \prod_{i=1}^{n} \lambda_i = \det(\mathbf{S})$. Luego los componentes principales conservan la varianza generalizada.
- La proporción de varianza explicada por cada componente es

$$\frac{\lambda_i}{p}.$$

  Además al ser incorreladas la proporción de varianza explicada por los k primeros componentes es

$$\frac{\sum_{i=1}^{k} \lambda_i}{p}.$$

- $Corr(Z_j, CP_i) = \sqrt{\lambda_i} u_{ji}$ donde u_{ji} es la j-ésima componente del vector propio u_i.

Propiedades ACP correlaciones.

- La primera componente principal es la recta que conserva mayor inercia de la nube de puntos.
- Los dos primeros componentes principales forma el plano que conserva mayor inercia de la nube de puntos.
- Lo mismo sucede con los espacios formados por los k primeros componentes

Etapas de un ACP

- Determinar las variables e individuos que intervienen en el análisis, las variables de perfil y los individuos ilustrativos.
- Decidir si se realiza el análisis sobre los datos brutos (matriz de covarianzas) o sobre los datos tipificados (matriz de correlaciones.
- Reducción de la dimensionalidad; tenemos que decidir cuántas componente retenemos. La cantidad de varianza retenida será:

Comp.	Valor propio	Cantidad retenida
Cp_1	λ_1	$\lambda_1/\sum_{i=1}^{p}\lambda_i$
Cp_2	λ_2	$(\lambda_1+\lambda_2)/\sum_{i=1}^{p}\lambda_i$
Cp_3	λ_3	$(\lambda_1+\lambda_2+\lambda_3)/\sum_{i=1}^{p}\lambda_i$
$\ldots$	$\ldots$	$\ldots$
Cp_p	λ_p	$(\lambda_1+\ldots+\lambda_p)/\sum_{i=1}^{p}=1$

Retención de componentes

Una vez realizado el ACP tengo que decidir que número de componentes se retienen. Existen diversos métodos:

- Fijar una cantidad fija de varianza un 80 % un 90 %...
- Método de la Media aritmética. Se retienen todas las componentes hasta que $\lambda_h \geq \overline{\lambda} = \frac{\sum_{i=1}^{p} \lambda_i}{p}$
  En el caso normado esto equivale a retener las componentes con valores propios mayores que 1. (es el método que aplica por defecto del SPSS)
- Pruebas Hipótesis valores propios nulos:

$$\left\{ \begin{array}{l} H_0 : \lambda_m = \ldots \lambda_p = 0 \\ H_1 : \text{no todos nulos} \end{array} \right.$$

- Pruebas Hipótesis de Anderson:

$$\left\{\begin{array}{l} H_0 : \lambda_m = \ldots \lambda_p \\ H_1 : \text{no todos iguales} \end{array}\right.$$

- Gráfico de sedimentación (screenplot) Se dibuja la poligonal de los valores propios y se retienen hasta el primero que "sedimenta".
- Otras pruebas Lebart y Fenelon.
- Baston roto de Frontier.
- Prueba E de Ibañez.

Adecuación de los datos al ACP

- Coeficiente de adecuación muestral (Kaiser Meyer y Olkin):

$$KMO = \frac{\sum_j \sum_{h\neq j} r_{ij}^2}{\sum_j \sum_{h\neq j} r_{jh}^2 + \sum_j \sum_{h\neq j} a_{jh}}$$

  donde r_{ij} son los coef. de correlación entre las variables i y j, mientras que los a_{ij} son los coef. de correlación parcial entre las variables i y j (equivalentes a las correlaciones entre los residuos de la regresiones de estas dos variables con las restantes.
- Niveles de KMO $\geq 0{,}5$ son considerados aceptables.

Test de esfericiadad de Barlett

- $$\begin{cases} H_0 : \text{La matriz de correlaciones es la identidad} \\ H_1 : \text{es distinta de la identidad} \end{cases}$$

  Interesa rechazar la hipótesis nula, pues si $\mathbf{R} = I$ las componentes pricipales son las propias variables.

Este test, al igual que casi todas las propiedades de los estimadores en ACP, requiere multinormalidad en la distribución de las variables.

Comunalidades

En un ACP la comunalidad de la variable X_j retenida por las k primeras componentes es la proporción de varianza de la variable que queda explicada por esas componentes. Por ejemplo.

- Si retenemos sólo el componente CP_1 la comunalidad de la variable X_j es:

$$h_j = r_{j1}^2 = \left(u_{1j}\sqrt{\lambda_1}\right)^2$$

- Si retenemos los componentes CP_1 y CP_2 la comunalidad de la variable X_j es:

$$h_j = r_{j1}^2 + r_{2j}^2 = \left(u_{1j}\sqrt{\lambda_1}\right)^2 + \left(u_{2j}\sqrt{\lambda_2}\right)^2$$

Interpretación de las variables y los individuos.

- Las variables también pueden representar de forma simultanea con los individuos en los componentes principales.
- Esta representación se hace mediante las coordenadas que la matriz de componentes (cargas factoriales) que nos explican las correlaciones de cada factor con cada variable.

En esta representación gráfica cada variable aparece como un punto de forma que :

- Cada variable está representada por el vector que une el origen de coordenadas cono el punto.
- Todos están en círculo unidad (círculo de correlación).
- A medida que cada variable se acerca a la circunferencia unidad está mejor representado por las componentes retenidas y viceversa.
- El ángulo entre variables y componentes nos da una idea de su correlación, al nivel de retención de varianza total que tengamos.
- Así variable perpendiculares tenderán a ser incorreladas.
- Los valores de una variable crecen en la dirección de ésta.

Para acabar dos cosas más...

Análisis Factorial Confirmatorio y Exploratorio

- El Análisis factorial confirmatorio se realiza sobre modelos establecidos de factores y se hacen inferencias sobre sus propiedades.
- El análisis factorial descriptivo ayuda a la descripción de los datos y a la búsqueda de factores.

Relación del ACP con otras técnicas de análisis de datos

- Regresión Lineal Múltiple
- Clasificación.
- Análisis de correspondencias simples y múltiples.
- .. y muchas otras más

Parte VII

Métodos de clasificación automática.

Descripción del problema

- Tenemos un conjunto de individuos con unas ciertas medidas multidimensionales.
- Queremos ver si existe una "forma natural" de clasificar a dichos individuos en grupos, suponiendo que tenemos una matriz de distancias que nos da la distancia entre cada par de individuos.
- Los grupos que formemos tienen que ser lo más homogéneos posible y las diferencias entre los grupos, lo más acentuadas posible.

Pasos a realizar en la clasificación

- Escoger una medida de proximidad entre los individuos. O sea, hallar la matriz de distancias.
- Escoger el algoritmo de construcción de grupos. Usando la matriz de distancias definida anteriormente, indicar cómo se formarán los distintos grupos.

Cómo calcular la matriz de distancias

- Tenemos n individuos de los que se han tomado p mediciones. O sea, tenemos una matriz $n \times p$ de n individuos con p variables: $\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$
- A partir de la matrix $\mathbf{X}$ hemos de construir una matriz de similaridad-disimilaridad $\mathbf{D}$ $n \times n$ entre los objetos. O sea, d_{ij} será la similitud o disimilaridad entre el individuo i y el individuo j: $\mathbf{D} = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{pmatrix}$
- En el caso en que $\mathbf{D}$ sea una matriz de similaridad, cuánto menor sean los valores d_{ij}, más alejados estarán los individuos i y j y en el segundo caso, más cercanos estarán.

Introducción

- Supongamos que la matriz $\mathbf{X}$ sólo coge los valores 0 o 1. En este caso, diremos que la estructura de los datos es binaria.
- De cara a definir la matriz de similaridad, consideremos los datos referentes al individuo i (columna i de la matriz $\mathbf{X}$: $(x_{i1}, \ldots, x_{ip})^{\top}$) y al individuo j (columna j de la matriz $\mathbf{X}$: $(x_{j1}, \ldots, x_{jp})^{\top}$)
- Definimos las cantidades siguientes:

$$
\begin{aligned}
a_1 &= \sum_{k=1}^{p} \mathcal{I}(x_{ik} = x_{jk} = 1), \\
a_2 &= \sum_{k=1}^{p} \mathcal{I}(x_{ik} = 0, x_{jk} = 1), \\
a_3 &= \sum_{k=1}^{p} \mathcal{I}(x_{ik} = 1, x_{jk} = 0), \\
a_4 &= \sum_{k=1}^{p} \mathcal{I}(x_{ik} = x_{jk} = 0),
\end{aligned}
$$

donde $\mathcal{I}$(condición) cuenta cuantos elementos de la matriz $\mathbf{X}$ cumplen la condición.

Definición de la matriz de distancias

- Definimos la similaridad entre los individuos i y j como:

$$d_{ij} = \frac{a_1 + \delta a_4}{a_1 + \delta a_4 + \lambda(a_2 + a_3)}.$$

- Los parámetros δ y λ son factores de peso. La tabla siguiente muestra ejemplos de similaridades más comunes:

Nombre	δ	λ	Definición
Jaccard	0	1	$\frac{a_1}{a_1+a_2+a_3}$
Tanimoto	1	2	$\frac{a_1+a_4}{a_1+2(a_2+a_3)+a_4}$
Simple	1	1	$\frac{a_1+a_4}{p}$
Rusell y Rao	-	-	$\frac{a_1}{p}$
Dice	0	0,5	$\frac{2a_1}{2a_1+a_2+a_3}$
Kulczynski	-	-	$\frac{a_1}{a_2+a_3}$

¿Por qué no se usa la distancia euclídea?

- La distancia euclídea no sería adecuada en este caso ya que trata los valores 0 y 1 de la misma forma y en la mayoría de aplicaciones donde se tratan datos binarios, deben ser tratados de formas distintas.
- Por ejemplo, si $x_{ik} = 1$ significa que el individuo i tiene un cierto conocimiento sobre el lenguaje k y si $x_{ik} = 0$ significa que no lo tiene, el valor $x_{ik} = 0$ tiene que tratarse de forma distinta del valor $x_{ik} = 1$.

Ejemplo

- Consideremos 3 tipos de marca de automóviles, Renault, Rover y Toyota. Se ha realizado una encuesta en 40 personas para que opinen de las variables siguientes: Economía (EC), servicio (SER), valor no depreciado (V), precio (P)(valor 1 para los coches más baratos), diseño (DI), deportividad (DE), seguridad (SEG) y facilidad de Manejo (FM). Los resultados van de 1 (muy bueno) a 6 (muy malo).
- Las medias de las valoraciones de los 40 encuestados son:

Modelo	EC	SER	V	P	DI	DE	SEG	FM
Renault	2.7	3.3	3.4	3.0	3.1	3.4	3.0	2.7
Rover	3.9	2.8	2.6	4.0	2.6	3.0	3.2	3.0
Toyota	2.5	2.9	3.4	3.0	3.2	3.1	3.2	2.8

Ejemplo

- Sea $\mathbf{X}$ la matriz de datos anterior:

$$\mathbf{X} = \begin{pmatrix} 2{,}7 & 3{,}3 & 3{,}4 & 3{,}0 & 3{,}1 & 3{,}4 & 3{,}0 & 2{,}7 \\ 3{,}9 & 2{,}8 & 2{,}6 & 4{,}0 & 2{,}6 & 3{,}0 & 3{,}2 & 3{,}0 \\ 2{,}5 & 2{,}9 & 3{,}4 & 3{,}0 & 3{,}2 & 3{,}1 & 3{,}2 & 2{,}8 \end{pmatrix}$$

  Se trata de una matriz 3×8 de 3 individuos con 8 variables.

- A partir de la matriz anterior, construimos la matriz binaria siguiente: $y_{ik} = \begin{cases} 1, & \text{si } x_{ik} > \overline{x}_k, \\ 0, & \text{en caso contrario.} \end{cases}$

- La matriz $\mathbf{Y}$ es:

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

Ejemplo

- Si aplicamos la medida de Jaccard, obtenemos la matriz siguiente de similaridad:

$$\mathbf{D} = \begin{pmatrix} 1{,}00 & 0{,}00 & 0{,}33 \\ & 1{,}00 & 0{,}25 \\ & & 1{,}00 \end{pmatrix}$$

- La medida de Tanimoto da el siguiente resultado:

$$\mathbf{D} = \begin{pmatrix} 1{,}00 & 0{,}23 & 0{,}60 \\ & 1{,}00 & 0{,}45 \\ & & 1{,}00 \end{pmatrix}$$

- La medida simple da como matriz de similaridad:

$$\mathbf{D} = \begin{pmatrix} 1{,}00 & 0{,}38 & 0{,}75 \\ & 1{,}00 & 0{,}62 \\ & & 1{,}00 \end{pmatrix}$$

Introducción

- En el caso general de datos continuos, se usan las distancias generadas por las normas $L_r, \ r \geq 1$. Esto es, la distancia entre el individuo i (columna i de la matriz $\mathbf{X}$: $(x_{i1}, \ldots, x_{ip})^\top$) y al individuo j (columna j de la matriz X: $(x_{j1}, \ldots, x_{jp})^\top$) viene dada por:

$$d_{ij} = ||\mathbf{x_i} - \mathbf{x_j}||_r = \left(\sum_{k=1}^{p} |x_{ik} - x_{jk}|^r\right)^{\frac{1}{r}}.$$

- Es claro que en este caso $d_{ii} = 0$. Por tanto, usamos una matriz de distancias o disimilaridad. Si $r = 2$, tenemos la distancia euclídea usual.

Escalado de los datos

- Una suposición usual es que los datos estén en la misma escala. Si no fuese el caso, la definición de la matriz de distancias se realiza mediante una métrica **A**. En el caso de la distancia euclídea, la modificación sería:

$$d_{ij} = ||\mathbf{x_i} - \mathbf{x_j}||_{\mathbf{A}} = (\mathbf{x_i} - \mathbf{x_j})^\top \mathbf{A} (\mathbf{x_i} - \mathbf{x_j}).$$

- La métrica **A** más usada es la siguiente:

$$\mathbf{A} = \mathrm{diag}\left(s_{\mathbf{x_1 x_1}}^{-1}, \ldots, s_{\mathbf{x_p x_p}}^{-1} \right),$$

  donde $s_{\mathbf{x_k x_k}}$ es la varianza de la variable k-ésima.

- La distancia euclídea quedará de la forma siguiente:

$$d_{ij} = \sum_{k=1}^{p} \frac{(x_{ik} - x_{jk})^2}{s_{\mathbf{x_k x_k}}}.$$

Ejemplo

- La siguiente tabla muestra el gasto medio en varios tipos de comida para distintos tipos de familias en Francia: trabajadores manuales (TM), empleados (EM) y directivos (DIR) con distintos número de hijos: 2, 3, 4 o 5 hijos.
  Significado de las siglas:

TF	Tipo familia
P	Pan
VE	Vegetales
F	Fruta
C	Carne
A	Aves
L	Leche
VI	Vino

Ejemplo

TF	P	VE	F	C	A	L	VI
TM2	332	428	354	1437	526	247	427
EM2	293	559	388	1527	567	239	258
DIR2	372	767	562	1948	927	235	433
TM3	406	563	341	1507	544	324	407
EM3	386	608	396	1501	558	319	363
DIR3	438	843	689	2345	1148	243	341
TM4	534	660	367	1620	638	414	407
EM4	460	699	484	1856	762	400	416
DIR4	385	789	621	2366	1149	304	282
TM5	655	776	423	1848	759	495	486
EM5	584	995	548	2056	893	518	319
DIR5	515	1097	887	2630	1167	561	284

Ejemplo

- La matriz de distancia euclidea para los datos anteriores es:

$$10^4 \cdot \begin{pmatrix}
0.00 & 5.82 & 58.19 & 3.54 & 5.15 & 151.44 & 16.91 & 36.15 & 147.99 & 51.84 & 102.56 & 271.83 \\
 & 0.00 & 41.73 & 4.53 & 2.93 & 120.59 & 13.52 & 25.39 & 116.31 & 43.68 & 76.81 & 226.87 \\
 & & 0.00 & 44.14 & 40.10 & 24.12 & 29.95 & 8.17 & 25.57 & 20.81 & 20.30 & 88.62 \\
 & & & 0.00 & 0.76 & 127.85 & 5.62 & 21.70 & 124.98 & 31.21 & 72.97 & 231.57 \\
 & & & & 0.00 & 121.05 & 5.70 & 19.85 & 118.77 & 30.82 & 67.39 & 220.72 \\
 & & & & & 0.00 & 96.57 & 48.16 & 1.80 & 60.52 & 28.90 & 29.56 \\
 & & & & & & 0.00 & 9.20 & 94.87 & 11.07 & 42.12 & 179.84 \\
 & & & & & & & 0.00 & 46.95 & 6.17 & 18.76 & 113.03 \\
 & & & & & & & & 0.00 & 61.08 & 29.62 & 31.86 \\
 & & & & & & & & & 0.00 & 15.83 & 116.11 \\
 & & & & & & & & & & 0.00 & 53.77 \\
 & & & & & & & & & & & 0.00
\end{pmatrix}$$

Ejemplo

- Vamos a escalar los datos. Calculemos primero $\mathbf{A} = \mathrm{diag}\left(s_{\mathbf{x_1 x_1}}^{-1}, \ldots, s_{\mathbf{x_7 x_7}}^{-1}\right)$:

$$\begin{aligned}\mathbf{A} = \mathrm{diag}(\quad & 9{,}50 \cdot 10^{-5}, 3{,}05 \cdot 10^{-5}, 4{,}00 \cdot 10^{-5}, \\ & 6{,}97 \cdot 10^{-6}, 1{,}75 \cdot 10^{-5}, 7{,}95 \cdot 10^{-5}, \\ & 2{,}12 \cdot 10^{-4})\end{aligned}$$

- La distancia euclídea escalada será:

$$\mathbf{D} = \begin{pmatrix} 0.00 & 6.85 & 10.04 & 1.66 & 2.66 & 24.90 & 8.28 & 8.56 & 24.61 & 21.55 & 30.68 & 57.48 \\ & 0.00 & 13.11 & 6.59 & 3.75 & 20.12 & 13.13 & 12.38 & 15.88 & 31.52 & 25.65 & 46.64 \\ & & 0.00 & 8.03 & 7.27 & 4.99 & 9.27 & 3.88 & 7.46 & 14.92 & 15.08 & 26.89 \\ & & & 0.00 & 0.64 & 20.06 & 2.76 & 3.82 & 19.63 & 12.81 & 19.28 & 45.01 \\ & & & & 0.00 & 17.00 & 3.54 & 3.81 & 15.76 & 14.98 & 16.89 & 39.87 \\ & & & & & 0.00 & 17.51 & 9.79 & 1.58 & 21.32 & 11.36 & 13.40 \\ & & & & & & 0.00 & 1.80 & 17.92 & 4.39 & 9.93 & 33.61 \\ & & & & & & & 0.00 & 10.50 & 5.70 & 7.97 & 24.41 \\ & & & & & & & & 0.00 & 24.75 & 11.02 & 13.07 \\ & & & & & & & & & 0.00 & 9.13 & 29.78 \\ & & & & & & & & & & 0.00 & 9.39 \\ & & & & & & & & & & & 0.00 \end{pmatrix}$$

Caso en que la matriz $\mathbf{X}$ es una tabla de contingencia

- $\mathbf{X}$ es una tabla de contingencia cuando representa una matriz de frecuencias.
- Definimos $x_{i\bullet} = \sum_{j=1}^{p} x_{ij}$ y $\frac{x_{ij}}{x_{i\bullet}}$ la distribución condicional de la fila i-ésima. De la misma forma, definimos $x_{\bullet j} = \sum_{i=1}^{n} x_{ij}$ y $\frac{x_{ij}}{x_{\bullet j}}$ la distribución condicional de la columna j-ésima. Sea $x_{\bullet\bullet} = \sum_{i=1}^{n} = x_{i\bullet} = \sum_{j=1}^{p} x_{\bullet j}$.
- En este caso la distancia entre las filas i_1 e i_2 es la siguiente:

$$d^2(i_1, i_2) = \sum_{j=1}^{p} \frac{1}{\left(\frac{x_{\bullet j}}{x_{\bullet\bullet}}\right)} \left(\frac{x_{i_1 j}}{x_{i_1\bullet}} - \frac{x_{i_2 j}}{x_{i_2\bullet}}\right)^2.$$

Introducción

- Existen dos métodos de çlustering": algoritmos jerárquicos y algoritmos de partición.
  - Los algoritmos jerárquicos pueden dividirse entre procedimientos aglomerativos y procedimientos de división.
  - Los algoritmos jerárquicos aglomerativos empiezan con la partición más fina posible (cada individuo constituye un clúster) y los va agrupando.
  - Los algoritmos jerárquicos de división empiezan con la partición más burda posible (todos los individuos constituyen el clúster") y va dividiendo los clústers en clúster más pequeños.

  - Los algoritmos de partición empiezan con un estado inicial de clústers dado y van cambiando dicho estado hasta que un determinada función es optimizada.

Introducción

La diferencia esencial entre los dos métodos de “clustering” es que en los métodos jerárquicos las asignaciones de los elementos a los grupos, una vez hallados éstos, no se pueden cambiar; en cambio, en los métodos de partición, sí.

Pasos a realizar

- Construir la partición más fina posible.
- Calcular la matriz de distancias D.
- Realizar los pasos siguientes hasta que todos los individuos estén en un mismo clúster:
  - Encontrar los dos clústers con la distancia más próxima.
  - Definir un nuevo clúster compuesto por los dos clústers anteriores encontrados.
  - Calcular la nueva matriz de distancias reducida D entre los nuevos clústers.

Pasos a realizar

- Si dos grupos, P y Q tienen que agregarse en un nuevo clúster, el cálculo de la distancia entre el nuevo clúster $P+Q$ y otro clúster cualquiera R se realiza de la forma siguiente:

$$\begin{aligned} d(R, P+Q) = \; & \delta_1 d(R,P) + \delta_2 d(R,Q) + \delta_3 d(P,Q) + \\ & \delta_4 |d(R,P) - d(R,Q)|, \end{aligned}$$

donde los δ_j son parámetros a escoger. Cada elección de dichos parámetros da lugar a un algoritmo aglomerativo distinto.

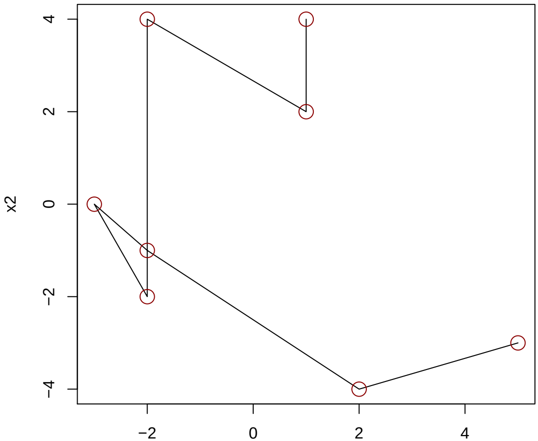
Tabla de algoritmos aglomerativos

Definimos n_P el número de objetos en el clúster P.

Nombre	δ_1	δ_2	δ_3	δ_4
Enlace simple	$1/2$	$1/2$	0	$-1/2$
Enlace completo	$1/2$	$1/2$	0	$1/2$
Enlace promedio	$1/2$	$1/2$	0	0
Enlace promedio con pesos	$\frac{n_P}{n_P+n_Q}$	$\frac{n_Q}{n_P+n_Q}$	0	0
Centroide	$\frac{n_P}{n_P+n_Q}$	$\frac{n_Q}{n_P+n_Q}$	$-\frac{n_P n_Q}{(n_P+n_Q)^2}$	0
Mediana	$1/2$	$1/2$	$-1/4$	0
Ward	$\frac{n_R+n_P}{n_R+n_P+n_Q}$	$\frac{n_R+n_Q}{n_R+n_P+n_Q}$	$-\frac{n_R}{n_R+n_P+n_Q}$	0

Ejemplo

- Consideremos los puntos siguientes en $\mathbb{R}^2: \begin{array}{l}(5,-3),\ (2,-4),\ (-2,1),\ (-3,0),\\ (-2,-2),\ (-2,4),\ (1,2),\ (1,4).\end{array}$



Ejemplo

- La matriz de distancias (distancia euclídea) entre los 8 puntos es:

$$D=\begin{pmatrix}
0{,}00 & 3{,}16 & 7{,}28 & 8{,}54 & 7{,}07 & 9{,}90 & 6{,}40 & 8{,}06\\
 & 0{,}00 & 5{,}00 & 6{,}40 & 4{,}47 & 8{,}94 & 6{,}08 & 8{,}06\\
 & & 0{,}00 & 1{,}41 & 1{,}00 & 5{,}00 & 4{,}24 & 5{,}83\\
 & & & 0{,}00 & 2{,}24 & 4{,}12 & 4{,}47 & 5{,}66\\
 & & & & 0{,}00 & 6{,}00 & 5{,}00 & 6{,}71\\
 & & & & & 0{,}00 & 3{,}61 & 3{,}00\\
 & & & & & & 0{,}00 & 2{,}00\\
 & & & & & & & 0{,}00
\end{pmatrix}$$

Ejemplo

- Aplicación del algoritmo. Como vemos, los puntos más cercanos son el 3 y el 5. Por tanto, los nuevos clústers serán:

$$\{1\},\ \{2\},\ \{3,5\},\ \{4\},\ \{6\},\ \{7\},\ \{8\}.$$

- Calculemos la nueva matriz de distancias usando el enlace simple entre los 7 clústers:

$$D_2 = \begin{pmatrix} 0{,}00 & 3{,}16 & 7{,}07 & 8{,}54 & 9{,}90 & 6{,}40 & 8{,}06 \\ & 0{,}00 & 4{,}47 & 6{,}40 & 8{,}94 & 6{,}08 & 8{,}06 \\ & & 0{,}00 & 1{,}41 & 5{,}00 & 4{,}24 & 5{,}83 \\ & & & 0{,}00 & 4{,}12 & 4{,}47 & 5{,}66 \\ & & & & 0{,}00 & 3{,}61 & 3{,}00 \\ & & & & & 0{,}00 & 2{,}00 \\ & & & & & & 0{,}00 \end{pmatrix}$$

Ejemplo

- La distancia más cercana está ahora entre los clústers $\{3,5\}$ y el clúster $\{4\}$. Los nuevos clúster son:

$$\{1\},\ \{2\},\ \{3,4,5\},\ \{6\},\ \{7\},\ \{8\}.$$

  La nueva matriz de distancias será:

$$D_3 = \begin{pmatrix} 0{,}00 & 3{,}16 & 7{,}07 & 9{,}90 & 6{,}40 & 8{,}06 \\ & 0{,}00 & 4{,}47 & 8{,}94 & 6{,}08 & 8{,}06 \\ & & 0{,}00 & 4{,}12 & 4{,}24 & 5{,}66 \\ & & & 0{,}00 & 3{,}61 & 3{,}00 \\ & & & & 0{,}00 & 2{,}00 \\ & & & & & 0{,}00 \end{pmatrix}$$

Ejemplo

- La distancia más cercana se encuentra entre los clústers $\{7\}$ y $\{8\}$. Los nuevos clústers serán:

$$\{1\},\ \{2\},\ \{3,4,5\},\ \{6\},\ \{7,8\}.$$

La nueva matriz de distancias será:

$$D_4 = \begin{pmatrix} 0{,}00 & 3{,}16 & 7{,}07 & 9{,}90 & 6{,}40 \\ & 0{,}00 & 4{,}47 & 8{,}94 & 6{,}08 \\ & & 0{,}00 & 4{,}12 & 4{,}24 \\ & & & 0{,}00 & 3{,}00 \\ & & & & 0{,}00 \end{pmatrix}$$

Ejemplo

- La distancia más cercana se encuentra entre los clústers $\{6\}$ y $\{7,8\}$. Los nuevos clústers serán:

$$\{1\},\ \{2\},\ \{3,4,5\},\ \{6,7,8\}.$$

La nueva matriz de distancias será:

$$D_5 = \begin{pmatrix} 0{,}00 & 3{,}16 & 7{,}07 & 6{,}40 \\ & 0{,}00 & 4{,}47 & 6{,}08 \\ & & 0{,}00 & 4{,}12 \\ & & & 0{,}00 \end{pmatrix}$$

Ejemplo

- La distancia más cercana se encuentra entre los clústers $\{1\}$ y $\{2\}$. Los nuevos clústers serán:

$$\{1,2\},\ \{3,4,5\},\ \{6,7,8\}.$$

La nueva matriz de distancias será:

$$D_6 = \begin{pmatrix} 0{,}00 & 4{,}47 & 6{,}08 \\ & 0{,}00 & 4{,}12 \\ & & 0{,}00 \end{pmatrix}$$

Ejemplo

- La distancia más cercana se encuentra entre los clústers $\{3,4,5\}$ y $\{6,7,8\}$. Los nuevos clústers serán:

$$\{1,2\},\ \{3,4,5,6,7,8\}.$$

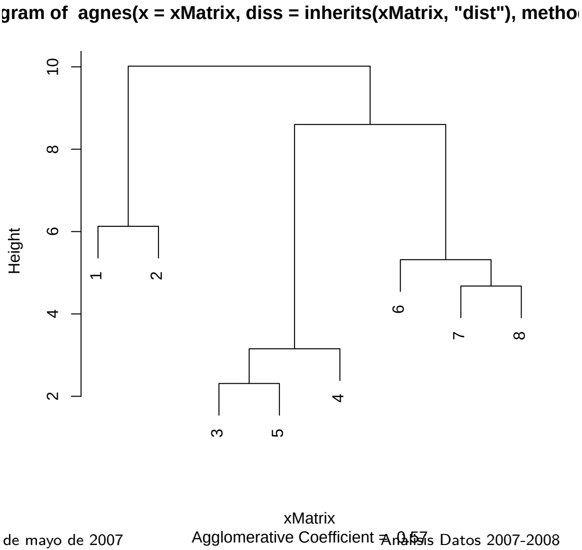
La nueva matriz de distancias será:

$$D_7 = \begin{pmatrix} 0{,}00 & 4{,}47 \\ & 0{,}00 \end{pmatrix}$$

- El último paso del algoritmo será juntar los dos últimos clusters obteniendo un clúster único compuesto por los 8 puntos.

Ejemplo

Todo el ejemplo queda resumido en lo que se denomina un dendograma:



Propiedades de los algoritmos aglomerativos.

- En el procedimiento del enlace simple, la distancia entre el clúster R y el clúster unión de P y Q, $P+Q$ se puede hallar como:

  $$d(R, P+Q) = \text{mín}(d(R,P), d(R,Q)).$$

  Por dicho motivo, se denomina algoritmo del vecino más próximo. Dicho algoritmo tiende a construir clústers grandes. Clústers que difieren pero que no están bien separados pueden unirse en un sólo clúster si tienen dos individuos próximos.

Propiedades de los algoritmos aglomerativos.

- En el procedimiento del enlace completo, se trata de corregir este tipo de agrupamiento considerando distancias grandes. De hecho, la distancia entre el clúster R y el clúster unión de P y Q, $P+Q$ se puede hallar como:

  $$d(R,P+Q)=\max(d(R,P),d(R,Q)).$$

  Por dicho motivo, se denomina el algoritmo del vecino más alejado. Dicho algoritmo agrupa clústers donde todos los puntos están próximos.
- En el procedimiento del enlace promedio, se propone una solución intermedia entre los dos procedimientos anteriores. En este caso, se calcula la distancia promedio:

  $$d(R,P+Q)=\frac{n_P}{n_P+n_Q}d(R,P)+\frac{n_Q}{n_P+n_Q}d(R,Q).$$

Propiedades de los algoritmos aglomerativos.

- El procedimiento del centroide es similar al procemidiento promedio pero usa la distancia geométrica entre el clúster R y el centro de gravedad promediado entre los clústers P y Q.
- El algoritmo de Ward se diferencia de los otros procedimientos respecto de la unificación. Dicho algoritmo no junta grupos con distancias pequeñas sino que junta grupos en los que no se incremente la heterogeneidad de los mismos "demasiado". El objetivo de dicho procedimiento es que la variación dentro de los clústers no aumente de forma drástica. El resultado es que los grupos son lo más homogéneos posible.

Propiedades de los algoritmos aglomerativos.

- Se define la heterogeneidad de un clúster R como:

$$I_R = \frac{1}{n_R} \sum_{i=1}^{n_R} d^2(x_i, \overline{x}_R),$$

  donde $\overline{x}_R$ es el centro de gravedad o la media del clúster R. Si d es la distancia euclídea, I_R representa la varianza del clúster R.
- Cuando dos clústers P y Q se unen, el nuevo clúster $P+Q$ tiene una heterogeneidad I_{P+Q}. Puede demostrarse que el aumento de heterogeneidad viene dada por:

$$\Delta(P,Q) = \frac{n_P n_Q}{n_P + n_Q} d^2(P,Q).$$

- El algoritmo de Ward, por tanto, puede definirse como el algorimo que minimiza $\Delta(P,Q)$.