**Álgebra Matricial** (ADM-Peña)

1. **Introducción**

   - La media de los datos es proporcional a la proyección del vector de datos sobre la dirección del vector constante, (el que tiene todas sus coordenadas iguales).
   - La desviación típica es la distancia promedio entre el vector de datos y el vector constante.
   - La dependencia lineal entre 2 variables se mide por la covarianza, y el concepto análogo vectorial es el producto escalar, que sirve para estudiar la posición en el espacio de 2 vectores.
   - Si las variables están tipificadas, la covarianza se reduce al coeficiente de correlación, que equivale al producto escalar de vectores de norma unidad.
   - La dependencia lineal entre variables, establece cuántas variables distintas tenemos realmente, decimos que $p$ variables son linealmente dependientes, si podemos obtener los valores de una cualquiera, como combinación lineal del resto, por ejemplo, las 3 variables, número de hombres, de mujeres y de personas, son dependientes, ya que podemos calcular el valor de una de ellas, conocidas las otras dos.
   - Un conjunto de $n$ nº reales $x$ se puede representar como un punto en el espacio de las 2 variables, y se define el vector $x$ como el segmento orientado que une el origen de coordenadas con el punto $x$, (vector de posición), con esta correspondencia a cada punto del espacio le asociamos un vector. A los valores de una variable con $n$ elementos le podemos asociar un vector en $R^n$.

2. **Operaciones con vectores y propiedades**:

   - Suma de 2 vectores: es otro vector con componentes iguales a la suma de las componentes de los sumandos, propiedades: conmutativa y asociativa. La suma de 2 variables genera una nueva variable como suma de las dos iniciales.
   - Producto de una constante por un vector, es otro vector de componentes las del vector multiplicado por la constante. Multiplicar una variable por una constante equivale a un cambio en las unidades de medida.
   - Vector traspuesto $x'$ de otro $x$, es un vector con las mismas componentes pero escritas en fila.
   - Producto escalar o interno de 2 vectores $x$ e $y$, que se escribe $x'y$ o $y'x$, es el escalar obtenido al sumar los productos de sus componentes.
   - Norma o longitud de un vector es la raíz cuadrada del producto escalar $x'x$.
   - El producto escalar se puede calcular como el producto de las normas de los vectores por el coseno del ángulo que forman, o también como el producto de la norma de un vector por la proyección del otro sobre él, si uno de los vectores tiene norma 1, el producto escalar es la proyección del otro vector sobre él.
   - Se define el ángulo entre 2 vectores a través del coseno del ángulo como el cociente entre el producto escalar y las normas de los vectores. Si 2 variables tienen media cero, el coseno del ángulo es su coeficiente de correlación.
   - La desigualdad de Cauchy-Schwarz, se deduce de la expresión del ángulo entre vectores, el valor absoluto del producto escalar es menor o igual que el producto de los módulos: $|x'y| \leq \|x\|\|y\|$.
   - La media es el escalar que define el vector obtenido al proyectar el vector de datos sobre la dirección constante que en $R^n$ es $\frac{1}{\sqrt{n}}$, o bien es la norma tipificada del vector obtenido al proyectar los datos en la dirección del vector constante.
   - La desviación típica mide la variabilidad de los datos y es la distancia tipificada entre el vector de datos y el vector constante. La proyección del vector de datos sobre la dirección del vector constante, produce el vector $\bar{x}\mathbf{1}$ y la norma del vector diferencia $x-\bar{x}\mathbf{1}$, mide la distancia entre el vector de datos y el vector constante: $\dfrac{1}{\sqrt{n}}\|x-\bar{x}\mathbf{1}\| = \sqrt{\dfrac{\sum(x_i-\bar{x})^2}{n}}$.
   - La covarianza mide la dependencia lineal entre 2 variables $x$ e $y$, y es el producto escalar tipificado de los 2 vectores medidos en desviaciones a la media: $\dfrac{1}{n}(x-\bar{x}\mathbf{1})'(y-\bar{y}\mathbf{1}) = \dfrac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n}$.
   - Para variables con media cero, la covarianza es el producto escalar de los vectores normalizados.
   - Para variables de media cero y desviación típica 1, la covarianza es el coeficiente de correlación.
   - La ortogonalidad de 2 variables implica incorrelación entre ellas, ya que $r = 0$.

3. **Dependencia lineal de vectores**: un conjunto de $p$ vectores son l.d. si existen $p$ escalares no todos nulos tales que: $c_1x_1 + \ldots + c_px_p = 0$.

- Si $p$ vectores son l.d. se puede expresar uno de ellos como combinación lineal del resto.
- En $R^p$ el nº máximo de vectores linealmente independientes es $p$.
- Un conjunto de vectores l.i. corresponde a un conjunto de variables que no están relacionadas linealmente de forma exacta.
- La dimensión de un espacio se define como el nº de vectores l.i que lo generan, que a su vez constituye una base de dicho espacio.
- Vector ortogonal a un subespacio es aquél que es ortogonal a cualquier vector de dicho subespacio.
  (a) Complemento ortogonal de un subespacio $E_p$, es el espacio de dimensión $n-p$ que contiene todos los vectores ortogonales a $E_p$.
  (b) Espacio nulo de un vector, es el complemento ortogonal del espacio generado por el vector.

4. **Matriz** es un conjunto de nº dispuestos en filas y columnas, y se puede considerar como un conjunto de vectores fila o columna.

- Dimensión $np$, significa que tiene $n$ filas y $p$ columnas.
- Traspuesta de una matriz es la que se obtiene al intercambiar las filas por las columnas.
- El producto matricial entre 2 matrices es otra matriz que contiene todos los productos escalares entre los vectores fila de la 1ª y los vectores columna de la 2ª. Para que se pueda realizar, el nº de columnas de la 1ª matriz tiene que coincidir con el nº de filas de la 2ª. El producto no es commutativo.
- Rango de una matriz es el nº máximo de vectores fila o columna l.i. El rango de una matriz es igual al de su traspuesta.
- Matriz cuadrada es la que tiene el mismo nº de filas que de columnas, se puede definir la matriz inversa, y existen varias formas de obtener una medida escalar de una matriz cuadrada: la traza, el determinante y construir una forma cuadrática a partir de la matriz.

5. **Operaciones con matrices**:

- La suma de 2 matrices de define si tienen las mismas dimensiones y cada elemento de la matriz suma se obtiene sumando los elementos correspondientes de los sumandos.
  (a) $A+B=B+A$
  (b) $(A+B)'=A'+B'$
- $AB \neq BA$
- Matriz identidad $I_n$
- El producto de una matriz $A_{np}$ por un vector $x_{p1}$ es un nuevo vector cuyas componentes son el producto escalar de las filas de la matriz por el vector: $y=Ax$, se dice que la matriz $A$ transforma un vector $x$ de $R^p$ en otro $y$ en $R^n$.
  (a) $A(B+C)=AB+AC$
  (b) $(AB)'=B'A'$
  (c) $AI=IA=A$
- Producto de Kroenecker, resuelve el problema de construir matrices grandes cuyos elementos son matrices dadas más pequeñas, se define para matrices cualesquiera. El producto de Kroenecker de 2 matrices se obtiene multiplicando cada elemento de la 1ª matriz por todos los elementos de la 2ª:

$$A_{kn}\otimes B_{pq}=\begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{k1}B & \dots & a_{kn}B \end{bmatrix}=C_{(kp)(nq)} \qquad \text{Ejemplo:} \qquad \begin{bmatrix} 1 \\ 2 \end{bmatrix}\otimes\begin{bmatrix} 1 & 0 & 3 \end{bmatrix}=\begin{bmatrix} 1 & 0 & 3 \\ 2 & 0 & 6 \end{bmatrix}$$

  (a) Si $c$ es un escalar $c\otimes A=A\otimes c=cA$
  (b) Si $x$ e $y$ son vectores $x\otimes y'=y'\otimes x$
  (c) $(A\otimes B)'=A'\otimes B'$
  (d) $(A\otimes B)(C\otimes D)=AC\otimes BD$
  (e) En estadística se utiliza si queremos construir una matriz que tenga como elementos diagonales una matriz A, por ejemplo: $I_3\otimes A=\begin{bmatrix} A & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & A \end{bmatrix}$

- Rango de una matriz $A_{np}$:
  (a) $r(A) \leq \min(n, p)$
  (b) Si $r(A) = \min(n, p)$, se dice que $A$ es de rango completo.
  (c) $r(A+B) \leq r(A) + r(B)$
  (d) $r(AB) \leq \min(r(A), r(B))$
  (e) $r(A'A) = r(AA') = r(A)$
  (f) $r(ABC) = r(A)$ para matrices cuadradas del mismo orden, donde $B$ y $C$ son no singulares.
  (g) Analizar el rango de una matriz de datos es la clave para reducir el número de variables sin pérdida de información.

6. **Matrices cuadradas**

- Matriz simétrica $A' = A$
- Matriz diagonal: solo tiene términos no nulos en la diagonal principal.
- $AA'$ y $A'A$ conducen a matrices simétricas.
- La matriz de covarianzas es $A'A/n$ si $A_{np}$ representa los valores de $p$ variables de media cero en $n$ individuos de una población.
- Determinante $|A|$ de una matriz: $|A| = \sum (-1)^r a_{1i_1} a_{2i_2} \dots a_{ni_n}$
  (a) El determinante de orden 2 se interpreta como el área del paralelogramo determinado por los vectores columna. Sea $C$ la matriz cuyas columnas son los vectores $v_1$ y $v_2$:
  $$|C'C| = \left| \begin{bmatrix} v_1' \\ v_2' \end{bmatrix} \begin{bmatrix} v_1 & v_2 \end{bmatrix} \right| = \begin{vmatrix} v_1'v_1 & v_1'v_2 \\ v_2'v_1 & v_2'v_2 \end{vmatrix} = \|v_1\|^2 \|v_2\|^2 (1 - \cos^2 \alpha)$$
  (b) El determinante de orden 3 se interpreta como el volumen del paralelepípedo generado por los vectores columna.
  (c) Para dimensiones mayores de 3 la interpretación correspondería al hipervolumen generado por los vectores columna.
  (d) Menor $m_{ij}$ de un elemento $a_{ij}$ de una matriz de orden $n$, es el determinante de orden $n-1$ que se obtiene al suprimir la fila y la columna a la que pertenece el elemento.
  (e) Adjunto de un elemento $a_{ij}$ es el escalar $(-1)^{i+j} m_{ij}$.
  (f) El determinante de una matriz se puede calcular multiplicando cada elemento de una fila o columna pos sus adjuntos. $|A| = \sum_{j=1}^{n} a_{ij} (-1)^{i+j} m_{ij}$.
  (g) En una matriz diagonal o triangular el determinante es el producto de los valores de la diagonal.
  (h) $|\lambda A| = \lambda^n |A|$, si $A$ es una matriz de orden $n$.
  (i) $|A'| = |A|$
  (j) $|AB| = |A||B|$
  (k) Si se cambian 2 filas o columnas entre sí, el determinante cambia de signo.
  (l) Si una fila o columna de una matriz, es combinación lineal de las restantes, la matriz se llama singular y su determinante es 0.
  (m) $|C|^2 = \|v_1\|^2 \|v_2\|^2 \operatorname{sen}^2 \alpha$, luego el valor absoluto del determinante de la matriz formada por los 2 vectores, es el área del paralelogramo que forman.
  (n) Si los vectores $v_1$ y $v_2$ son variables, el producto $CC$ es proporcional $(1/n)$, a su matriz de varianzas covarianzas, y su determinante que es el área que forman, es una medida global de la independencia entre las variables. Si las variables están incorreladas, su matriz de covarianzas es diagonal y su determinante será máximo, en cambio, si las variables son linealmente dependientes, el determinante de la matriz de covarianzas es cero, luego cuanto mayor sea el determinante, mayor será la independencia entre las variables.
- **Traza de una matriz** $C$, es la suma de los elementos de la diagonal principal $(c_{ii})$
  (a) $t(A+B) = t(A) + t(B)$
  (b) $t(\lambda A) = \lambda t(A)$
  (c) $t(ABC) = t(BCA) = t(CAB)$
  (d) Si $C$ es simétrica $t(C^2) = t(CC) = \sum c_{ij}^2$
  (e) La traza de una matriz de varianzas y covarianzas es la suma de todas las varianzas de las variables, luego es una medida de la variabilidad.

7. **Formas cuadráticas:** si realizamos una transformación lineal a un vector $x$, ($y = Bx$), el cuadrado de la norma del nuevo vector $y$, será: $y'y = x'B'Bx = x'Ax \geq 0$ donde $A = B'B$ es una matriz cuadrada y simétrica. Forma cuadrática es una expresión escalar del tipo:
$$x'Ax = \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} a_{ij}x_ix_j$$
   - Una matriz $A$ es semidefinida positiva, si cualquier forma cuadrática formada a partir de ella es un número no negativo, para cualquier vector $x \neq 0$. Si es un número positivo la matriz $A$ es definida positiva.
   - El determinante y la traza que se pueden obtener a partir de matrices semidefinidas positivas o definidas positivas, son números no negativos o positivos respectivamente.

8. **Matriz inversa:** sea $A$ una matriz cuadrada no singular de orden $n$, se define $A^{-1}$ como una matriz que cumple: $AA^{-1} = A^{-1}A = I$
   - La matriz inversa resuelve el problema de calcular vectores ortogonales a uno dado, o variables incorreladas con una dada, el espacio ortogonal al vector $a_1$, puede calcularse construyendo una matriz que tenga este vector como primera fila y calculando la inversa de la matriz, los vectores columna de la matriz inversa forman el espacio nulo del vector $a_1$.
   - Si $A = \begin{bmatrix} 2 & 1 \\ 0 & 4 \end{bmatrix}$ y $A^{-1} = \begin{bmatrix} 0.5 & -0.125 \\ 0 & 0.25 \end{bmatrix}$, se comprueba que el primer (segundo) vector columna de la inversa define el espacio ortogonal al segundo (primer) vector fila de la matriz $A$.
   - La necesidad de la matriz inversa aparece a la hora de resolver sistemas de ecuaciones lineales: $Ax = b$.
   - Cálculo de la matriz inversa de $A$:
     (a) Se sustituye cada elemento de $A$ por su adjunto.
     (b) Se transpone la matriz de adjuntos de $A$.
     (c) Se divide cada término de la matriz de adjuntos por $|A|$.
   - $(AB)^-1 = B^{-1}A^{-1}$
   - $(A')^{-1} = (A^{-1})'$
   - $|A^{-1}| = |A|^{-1}$
   - Si $A$ es simétrica también lo es $A^{-1}$.
   - La matriz inversa de la matriz de varianzas covarianzas recoge la información de la dependencia conjunta de todas las variables de forma más completa que la de varianzas covarianzas.
   - Inversa de una suma de matrices: $(A_n + B_{np}C_pD_{pn})^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}$
   - Si $A$ y $C$ tienen el mismo orden $(A+C)^{-1} = C^{-1}(A^{-1}+C^{-1})^{-1}A^{-1}$. Estas expresiones son útiles para estudiar el cambio de la matriz de varianzas covarianzas y otros estadísticos, al eliminar observaciones o variables.

9. **Matriz ortogonal**, es una matriz cuadrada que representa un giro en el espacio o simetría respecto a un plano, si a un vector $x$ le aplicamos una matriz no singular $C$, obtenemos un nuevo vector $Y = Cx$, si esta operación es un giro las normas de $y$ y de $x$ deben ser iguales $y'y = x'C'Cx = x'x$, luego $C'C = I$.
   - Se deduce que $C' = C^{-1}$
   - Una matriz ortogonal debe tener filas o columnas que son vectores ortogonales entre sí y de longitud unidad.
   - $|C| = |C'| = \pm 1$
   - Los vectores fila (columna) de una matriz ortogonal de orden $n$, forman una base ortonormal de $R^n$, ya que son ortogonales y de norma uno.

10. **Matrices particionadas**
    - Una matriz puede subdividirse en elementos que sean a su vez matrices.
    - La inversa y el determinante de una matriz particionada se calcula en pag. 36.

11. **Vectores y valores propios**

- Dada una matriz cuadrada hay determinadas propiedades invariantes ante transformaciones lineales que preservan la información existente en la matriz, al trasponer una matriz las propiedades básicas de los vectores que la forman no varían, y la traza y el determinante no cambian. Si giramos los vectores que la forman, es decir multiplicamos la matriz por una ortogonal, no se alteran ni sus magnitudes ni sus posiciones relativas, luego se espera que las propiedades básicas de la matriz se mantengan.
- Los valores propios son las medidas básicas de tamaño de una matriz, que no se ven alteradas si hacemos un cambio de coordenadas que equivale a una rotación de los ejes. Se demuestra que las medidas globales de tamaño de una matriz, como la traza o el determinante, son solo función de los valores propios y por lo tanto serán también invariantes ante las transformaciones que preservan los valores propios.
- Los vectores propios representan las direcciones características de la matriz y no son invariantes. Al aplicar una matriz cuadrada de orden $n$ a un vector de dimensión $n$, éste se transforma en dirección y magnitud. Sin embargo para cada matriz cuadrada existen ciertos vectores que al transformarlos por la matriz, solo se modifica su longitud (norma), pero no su posición en el espacio, son los vectores propios de la matriz.
- Vectores propios de una matriz cuadrada de orden $n$, son aquellos vectores $u \neq 0$, cuya dirección no se modifica al transformarlos mediante la matriz: $Au = \lambda u$, donde $\lambda$ es un valor propio de la matriz.
- Suponemos que los vectores propios están normalizados y que si $u$ es vector propio $au$ también lo sería, es decir $\|u\| = 1$, aunque el signo queda indeterminado, si $u$ es un vector propio, $-u$ también lo es.
- Para calcular los vectores propios $(A - \lambda I)u = 0$¸ sistema homogéneo que tendrá solución no nula sii la matriz $(A - \lambda I)$ es singular, luego la solución se obtiene de $|A - \lambda I| = 0$ (ecuación característica de la matriz de orden $n$ y sus $n$ raíces se llaman valores propios de la matriz).
- Si una matriz es diagonal, los valores propios son los elementos de la diagonal.
- Una matriz de orden $n$ tiene siempre $n$ valores propios (reales o complejos), aunque éstos pueden aparecer repetidos, (si aparece $r$ veces tiene grado de multiplicidad $r$).
- A cada valor propio $n$ distinto de una matriz cuadrada, podemos asociarle un único vector propio.
- Los vectores propios asociados a valores propios con grado de multiplicidad $r$, no están definidos de manera única, están en un espacio de dimensión $r$, y cualquier vector normalizado de ese espacio es un vector propio de la matriz.
- Si la matriz tiene $n$ valores propios distintos, el conjunto de vectores propios es l.i.
- Si $\lambda$ es un valor propio de $A$, $\lambda^r$ lo es de $A^r$, en particular si $\lambda \neq 0$, $\lambda^{-1}$ es un valor propio de $A^{-1}$.
- Los valores propios de $A$ y de $A'$ son iguales.
- La suma de valores propios de $A$ es igual a la traza.
- El producto de los valores propios de $A$ es igual a $|A|$.
- Si $P$ es no singular, $A$ y $P^{-1}AP$ tienen los mismos valores propios.
- Las matrices $A$ y $A \pm I$ tienen los mismos vectores propios y si $\lambda$ es valor propio de $A$, $\lambda \pm 1$ es valor propio de $A \pm I$.
- Las matrices cuadradas $ABC$, $BCA$ y $CAB$ tienen los mismos valores propios no nulos.
- Si $A$ es triangular, los valores propios son los elementos de la diagonal.
- Si $A_n$ y $B_p$ son matrices cuadradas, los $np$ vectores propios de $A \otimes B$ son el producto de Kronecker de los vectores propios de $A$ y $B$.

12. **Valores y vectores propios de matrices simétricas:**
    - Los valores propios son siempre reales.
    - Los vectores propios son ortogonales.
    - Interpretación de los valores y vectores propios:
      (a) Sea la matriz $A = \begin{bmatrix} a & b \\ b & a \end{bmatrix}$, los valores propios son $\lambda = a \pm b$ y sus vectores propios asociados son $(1,1)$ y $(1,-1)$ y normalizados $(1/\sqrt{2}, 1/\sqrt{2})$ y $(1/\sqrt{2}, -1/\sqrt{2})$ respectivamente. Si tomamos valores concretos, por ejemplo $a = 3$ y $b = 1$, los valores propios son 4 y 2.
      (b) Si construimos una elipse con centro en el origen y que pase por los extremos de los 2 vectores de la matriz $(3,1)$ y $(1,3)$, los valores propios representan las distancias de los extremos de cada eje de la elipse al origen, es decir 8 sería el eje mayor y 4 el menor.

(c) Los vectores propios asociados representan las direcciones de los ejes de la elipse, el asociado al mayor $\lambda$ es un vector unitario en la dirección de la diagonal principal, y el otro ortogonal a él.
(d) Si cambiamos los valores de $a$ y $b$, los vectores propios no cambian, aunque sí los valores propios, si aumentamos $a$ con $b$ fijo, la elipse tiende a una circunferencia y si es al revés apuntamos más la elipse, es decir si $a = 100$ y $b = 1$, los valores propios son 101 y 99 y los mismos vectores propios, pero si $a = 1.2$ y $b = 1$, los valores propios son 2.2 y 0.2.
(e) Si en la matriz $A$ los elementos diagonales son iguales, los ejes de la elipse van según las bisectrices de los ejes de coordenadas. Si los elementos diagonales son distintos, el eje mayor de la elipse estará mucho más cerca del vector de módulo mayor.
(f) Como conclusión los valores propios de una matriz simétrica representan los semiejes del elipsoide con centro el origen y determinado por los extremos de los vectores que definen la matriz. Los vectores propios indican las direcciones de los ejes del elipsoide.

13. **Diagonalización de matrices simétricas:**

- Una matriz simétrica se puede convertir en diagonal mediante una transformación ortogonal. Sea $A_n$ cuadrada y simétrica, si $\lambda_1, \ldots, \lambda_n$ y $u_1, \ldots, u_n$ son los valores y vectores propios respectivamente, podemos poner:

$$A[u_1, \ldots, u_n] = [\lambda_1 u_1, \ldots, \lambda_n u_n]$$

donde los valores propios son reales y pueden no ser todos distintos, e incluso algunos pueden ser nulos, y los vectores propios ortonormales, si $\mathbb{D}_\lambda$ es una matriz diagonal de los valores propios, podemos poner: $AU = U\mathbb{D}_\lambda$ donde $U$ es ortonormal ($U' = U^{-1}$), luego $U'AU = \mathbb{D}_\lambda$ y hemos transformado la matriz original $A$ en una diagonal $\mathbb{D}_\lambda$, mediante una matriz ortogonal $U$.

- La interpretación geométrica de esta expresión se puede realizar en los siguientes términos: $U'A = \mathbb{D}_\lambda U'$, donde $U'A$ representa una rotación de los vectores que forman la matriz $A$, y estos vectores rotados son iguales a $\mathbb{D}_\lambda U'$ que supone multiplicar por los términos de $\mathbb{D}_\lambda$ una base de vectores ortonormales $U'$, como $A = U\mathbb{D}_\lambda U'$, el proceso de generar una matriz simétrica es:
  (a) Se parte de una base ortonormal de vectores $U'$.
  (b) Se modifica la norma de cada vector de esta base, multiplicándolo por una matriz diagonal $\mathbb{D}_\lambda$.
  (c) Por último se rotan los vectores obtenidos.
  (d) Los valores propios representan las constantes por las que se multiplican los vectores ortonormales iniciales, y los vectores propios definen el giro realizado.
- Se cumple que $|U'||A||U| = |\mathbb{D}_\lambda|$ y como $|U| = 1$, el determinante de $A$, será el producto de sus raíces características, es decir de sus valores propios.
- El rango de $A$ es igual al de $\mathbb{D}_\lambda$ ya que $U$ y $U'$ son no singulares, luego el rango de una matriz simétrica es igual al nº de valores propios distintos de cero.
- Descomposición espectral o autodescomposición de una matriz $A$: $A = U\mathbb{D}_\lambda U'$ que se puede escribir como:

$$A = [u_1, \ldots, u_n] \begin{bmatrix} \lambda_1 u_1' \\ \vdots \\ \lambda_n u_n' \end{bmatrix} \text{ de donde } A = \sum_{i=1}^{n} \lambda_i u_i u_i'$$

que descompone la matriz $A$ como suma de $n$ matrices de rango uno $u_i u_i'$ con coeficientes $\lambda_i$. La importancia de esta descomposición es que si algunos valores propios son muy pequeños, podemos reconstruir aproximadamente la matriz $A$ utilizando el resto de valores y vectores propios.

- La autodescomposición de $A^{-1}$ es $A^{-1} = \sum_{i=1}^{n} \lambda_i^{-1} u_i u_i'$, ya que $A^{-1}$ tiene los mismos vectores propios que $A$ y valores propios $\lambda_i^{-1}$
- La descomposición espectral de una matriz cuadrada y simétrica, separa sus fuentes de variación intrínsecas, en las direcciones de los vectores propios con coeficientes que dependen de los valores propios.

14. **Raíz cuadrada de una matriz semidefinida positiva**

- Una matriz cuadrada, simétrica y semidefinida positiva $A$, siempre se puede descomponer como producto de una matriz por su traspuesta: $A = HH'$. En efecto por la descomposición espectral, $A = (U\mathbb{D}_\lambda^{1/2})(\mathbb{D}_\lambda^{1/2}U')$ y tomando $H = U\mathbb{D}_\lambda^{1/2}$ obtenemos que $A = HH'$

- A la matriz $H$ se le llama raíz cuadrada de la matriz $A$, y no es única ya que también se cumple que $A = H^*H'^*$ donde $H^* = HC$ siendo $C$ cualquier matriz ortogonal.
- Una forma de definir una raíz única es exigir que $H$ sea simétrica, con lo que $A = HH$ y $H = U\mathbb{D}_\lambda^{1/2}U'$.
- La raíz cuadrada única de una matriz cuadrada simétrica y definida positiva se puede obtener mediante la descomposición de Cholesky, donde $A = TT'$ siendo $T$ una matriz triangular inferior con términos diagonales positivos.
- Diagonalización de dos matrices simétricas: Sean $A$ y $B$ dos matrices simétricas de la misma dimensión y $A$ definida positiva, entonces la matriz $H = A^{-1/2}C$ donde $C$ es la matriz de vectores propios de la matriz simétrica $A^{-1/2}BA^{-1/2}$ cumple:
$$H'AH = I \qquad\qquad H'BH = D$$
donde $D$ es la matriz diagonal que contiene los valores propios de $A^{-1/2}BA^{-1/2}$

15. **Descomposición en valores singulares** Toda matriz rectangular $A_{np}$ de rango $r$ puede expresarse como el producto de tres matrices, dos de ellas ortogonales y la otra diagonal: $A = V_{nr}\mathbb{D}_\lambda^{1/2}U'_{rp}$
    - La matriz $\mathbb{D}_\lambda^{1/2}$es diagonal y contiene las raíces cuadradas de los valores propios no nulos de las matrices $AA'$ o $A'A$, que son positivos y se llaman valores singulares de la matriz $A$.
    - La matriz $V$ contiene en columnas los vectores propios de norma uno asociados a los valores propios no nulos de $AA'$ y $U$ los de $A'A$.
    - Se puede expresar $A = \sum_{i=1}^r d_i v_i u_i'$.
    - Como $AA'v_i = \lambda_i v_i$ si multiplicamos por $A'$, tenemos $(A'A)A'v_i = \lambda_i A'v_i = \lambda_i w_i$ siendo $w_i = A'v_i$ un vector propio de $AA'$, aunque no es de norma unidad, ya que $w_i'w_i = v_i'AA'v_i = \lambda_i$.
    - Los vectores propios de $A'A$ de norma unidad vienen dados por $U = A'VD_\lambda^{-1/2}$
16. **Diagonalización de matrices generales**
    - Si $A_n$ es una matriz cuadrada de orden $n$, es diagonalizable sii sus $n$ vectores propios son l.i. siendo $A = UDU^{-1}$.
    - Una condición suficiente para que $A$ sea diagonalizable es que tenga $n$ valores propios distintos.
17. **Inversa generalizada** de una matriz $A_{np}$ es otra matriz $A^-_{pn}$ que cumple $AA^-A = A$
    - Si $A^-AA^-$, $A^-A$ y $AA^-$ son simétricas, se demuestra que $A^-$ es única y se llama matriz inversa generalizada de Moore-Penrose (MP) y se comprueba que si $n > p$ y $A$ tiene rango completo $r(A) = p$, la inversa MP es $A^- = (A'A)^{-1}A'$.
    - Si $A$ no tiene rango completo, la expresión de la inversa generalizada como resultado de invertir las tres matrices de la descomposición en valores singulares de $A$ es $A^- = UD_\lambda^{-1/2}V'$
18. **Proyección ortogonal**
    - **Matrices idempotentes**: en un modelo lineal la estimación por mínimos cuadrados equivale a la proyección ortogonal del vector de datos sobre el espacio generado por las variables explicativas. La proyección ortogonal se realiza multiplicando el vector que se desea proyectar, por una matriz idempotente.
      (a) Matriz idempotente es una matriz cuadrada simétrica que verifica $AA = A = A'A$, una matriz idempotente puede ser no simétrica aunque vamos a considerar aquí las que sean simétricas.
      (b) Una matriz idempotente que no sea $I$ es singular, es decir $|A| = 0$.
      (c) Los valores propios de una matriz idempotente son 0 o 1, luego si se diagonaliza una matriz idempotente, lo que siempre se puede hacer por ser simétrica, obtendremos en la diagonal principal, un nº de unos igual al rango de la matriz y el resto ceros.
      (d) Por lo tanto, una matriz idempotente es siempre semidefinida positiva.
      (e) Si $A$ es idempotente también lo es $I - A$.
      (f) Si $A$ es idempotente simétrica, su rango es igual a su traza.
    - **Proyección ortogonal** Dado un vector $y \in \mathbb{R}^n$, diremos que $v$ es la proyección ortogonal de $y$ sobre un subespacio $E_p \subset \mathbb{R}^n$ si: $y = v + w$ con $v \in E_p$ siendo $v'w = 0$ $\forall v \in E_p$

(a) Es decir $y$ se puede descomponer en suma de 2 vectores perpendiculares, el primero $v$ es la proyección ortogonal de $y$ sobre $E_p$ y el segundo $w$ es ortogonal a todos los vectores de $E_p$.

(b) Consideremos un subespacio $E_p$ de dimensión 1, generado por el vector $x$, la proyección de $y$ sobre $x$ vendrá dada por $v = xc$, donde $c$ es una constante. Para determinar $c$ imponemos la condición de que $w$ es ortogonal a $v$ y por lo tanto a $x$: $x'(y-v) = 0$, es decir $x'y = x'xc$ lo que implica que $c = (x'x)^{-1}x'y$ de donde $v = x(x'x)^{-1}x'y = Ay$.

(c) Es decir la proyección $v$ de un vector $y$ sobre otro $x$ se obtiene multiplicando el vector por la matriz $A$, es decir $v = Ay$ donde $A = x(x'x)^{-1}x'$. Esta matriz $A$ es idempotente y de rango igual a la dimensión del espacio sobre el que proyectamos.

(d) Sea $y \in \mathbb{R}^n$ y $X_{np}$ una matriz cuyas columnas son una base de un cierto subespacio $E_p$. La proyección del vector $y$ sobre el espacio $E_p$ es $Ay$ donde $A$ es una matriz cuadrada, simétrica e idempotente de rango $p$ y tal que $A = X(X'X)^{-1}X'$

(e) La condición necesaria y suficiente para que $v = Ay$ sea la proyección ortogonal de $y \in \mathbb{R}^n$ sobre un cierto espacio $E_p$ es que $A$ sea una matriz cuadrada idempotente y simétrica de rango $p$.

(f) Se cumple que $||y||^2 = ||v||^2 + ||y-v||^2$ donde $||y||$ es la norma del vector $y$.

(g) La proyección ortogonal de un vector $y$ sobre un espacio $E_p$ es aquel vector $v \in E_p$ tal que $||y-v||$ es mínimo.

(h) Si $y \in \mathbb{R}^n$, el cuadrado de la norma de su proyección sobre un espacio $E_p$ definido por las columnas de la matriz $X$ viene dado por la expresión $y'Ay$ donde $A$ es idempotente.

19. **Derivadas matriciales**

- Dada una función $f$ que depende de $n$ variables $x_1, \dots, x_n$ que pueden considerarse las componentes de un vector $x$, la derivada de $f$ respecto de $x$ es un vector de componentes, la derivada de $f$ respecto a cada componente de $x$:

$$\text{Si } f = 2x_1 + 3x_2 - 4x_3 \qquad \frac{\partial f}{\partial x} = \begin{bmatrix} 2 \\ 3 \\ -4 \end{bmatrix}$$

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

  (b) $\frac{\partial(x'Ax)}{\partial x} = 2Ax$ donde $A$ es una matriz cuadrada y simétrica.

- Dada una función $f$ que depende de $np$ variables $x_{11}, \dots, x_{np}$ que son las componentes de una matriz $X_{np}$, la derivada de $f$ respecto a $X$ se define como la matriz cuyos componentes son la derivada de $f$ respecto a cada componente de $X'$, luego su dimensión es $pn$.

  (a) Si $X_{np}$ es una matriz de orden $n$ por $p$, $\frac{\partial(a'Xb)}{\partial x} = ba'$

  (b) Si $f = a'X'Xb$ se cumple: $\frac{\partial(a'X'Xb)}{\partial x} = (ab' + ba')X'$

- Si $y' = (f_1(x), \dots, f_n(x))$ se cumple:

$$\frac{\partial y}{\partial x} = \left[\frac{\partial f_1}{\partial x}, \dots, \frac{\partial f_n}{\partial x}\right] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial f_1}{\partial x_n} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

  que es la matriz Jacobiana.

- Si $y = Ax$ siendo $A$ una matriz cualquiera $\frac{\partial(Ax)}{\partial x} = A'$

**Descripción de datos multivariantes. Cap-3**

1. **Datos multivariantes**

- **Tipos de variables**, las variables pueden ser cuantitativas o cualitativas, las cuantitativas pueden ser continuas o de intervalo y discretas, y las cualitativas pueden ser binarias o generales.

  (a) Las variables binarias se suelen codificar como numéricas, (género, hombre-mujer se codifica con 0 y 1), las variables cualitativas generales se pueden también convertir en binarias, por ejemplo el color de los ojos, azul (A), verde (V), castaño (C) y negro (N), podemos representarlo a través de 3 variables binarias:

| $CO$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|
| $A$ | 1 | 0 | 0 |
| $V$ | 0 | 1 | 0 |
| $C$ | 0 | 0 | 1 |
| $N$ | 0 | 0 | 0 |

(b) Si el nº de clases posibles es muy grande, conviene agrupar las clases para evitar tener variables que casi siempre tomen el mismo valor.

(c) También se podría haber codificado el color con 1, 2, 3 y 4, pero podría sugerir una graduación de valores que puede no existir. En cambio para codificar empresas pequeñas, medianas y grandes sí sería aceptable la codificación, 1, 2 y 3, que indicaría un sentido de orden.

- **La matriz de datos**, es una matriz $X_{np}$ con $p$ variables en un conjunto de $n$ elementos (individuos).
- En 100 estudiantes medimos la edad, el género, la calificación media, el municipio de residencia, (codificado en 4 categorías según tamaño), y el curso más alto en el que se encuentra matriculado. Los datos se representan en una tabla de:

  (a) 100 filas y 5 columnas, 3 de ellas cuantitativas, una binaria y otra cualitativa general.

  (b) 100 filas y 7 columnas si el municipio de residencia lo codificamos con 3 variables binarias.

2. **Análisis univariante**, describir datos multivariantes supone estudiar cada variable aisladamente y además las relaciones entre ellas. Resumimos algunos estadísticos descriptivos univariantes:

   (a) La media de $x_j$ es $\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ que para una variable binaria es la frecuencia relativa de aparición del atributo y para una numérica es el centro de gravedad de los datos.

   (b) La desviación típica viene dada por
   $$s_j = \sqrt{\frac{\sum_{i=1}^{n} d_{ij}}{n}} = \sqrt{\frac{\sum_{i=1}^{n}(x_{ij}-\overline{x}_j)^2}{n}}$$

   (c) Para comparar la variabilidad de distintas variables utilizamos una variablidad relativa que no depende de las unidades de medida como el coeficiente de variación $CV_j = \sqrt{\frac{s_j^2}{\overline{x}_j^2}}$

   (d) Para medir la simetría de los datos respecto a su centro, utilizamos el coeficiente de asimetría $A_j = \frac{1}{n}\frac{\sum(x_{ij}-\overline{x}_j)^3}{s_j^3}$ que vale cero para una variable simétrica. Si es mayor que uno en valor absoluto la distribución de los datos es asimétrica.

   (e) La homogeneidad de los datos es otra característica importante, si las desviaciones $d_{ij} = (x_{ij}-\overline{x}_j)^2$ son muy distintas, significa que hay datos que se separan mucho de la media y que hay por tanto alta heterogeneidad, si los datos se separan en dos mitades correspondientes a dos distribuciones muy alejadas entre sí, la heterogeneidad será muy pequeña, la medimos por la varianza de las desviaciones $\frac{1}{n}\sum_{i=1} n(d_{ij}-s_j^2)^2$ ya que $\overline{d}_j = s^2$. Una medida adimensional es el coeficiente de homogeneidad
   $$H_j = \frac{\frac{1}{n}\sum_{i=1}^{n}(d_{ij}-s_j^2)^2}{s_j^4} = \frac{1}{n}\frac{\sum(x_{ij}-\overline{x}_j)^4}{s_j^4} - 1 = K_j - 1 \geq 0$$
   $K_j \geq 1$ es el coeficiente de kurtosis y es otra forma de medir la homogeneidad.

   (f) Un objetivo importante en la descripción de datos es decidir si los datos son una muestra homogénea de una población o corresponden a una mezcla de poblaciones distintas que deben estudiarse por separado. Un caso importante de heterogeneidad es cuando hay una pequeña proporción de observaciones atípicas, (outliers), con el resto. Detectar estas observaciones es fundamental para una correcta descripción de los datos, ya que estos valores extremos distorsionan los valores descriptivos del conjunto, y el coeficiente de kurtosis puede ayudar en este punto, ya que tomará un valor alto, mayor que 7 u 8. Por ejemplo si contaminamos datos de una distribución normal con un 1% de atípicos generados por otra distribución normal de igual media pero varianza 20 veces mayor, el coeficiente de kurtosis estará cerca de 10.

   (g) En el caso de una mezcla de dos poblaciones muy distintas a partes iguales, el coeficiente de kurtosis es muy bajo y al aumentar la separación entre las poblaciones su valor tiende a uno.

   (h) La posible presencia de datos atípicos sugiere calcular medidas robustas de centralización y dispersión como la mediana y la MEDA, (mediana de las desviaciones absolutas respecto a la mediana).

   (i) Conviene también representar las variables continuas gráficamente por histogramas o diagramas de caja.

   (j) Al calcular en el análisis inicial, la media y la mediana de cada variable, si ambas son similares, la media es un buen indicador del centro de los datos. Sin embargo si difieren mucho, la media puede no serlo debido a: una distribución asimétrica, presencia de valores atípicos que afectan mucho a la media y poco a la mediana o heterogeneidad en los datos.

3. **Medidas de centralización: el vector de medias** es un vector de dimensión $p$ cuyas componentes son las medias de cada una de las $p$ variables.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \frac{1}{n}X'\mathbf{1} = \frac{1}{n}[x_1 \dots x_n]\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

   - El vector de medias se encuentra en el centro de los datos, es decir: $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

4. **La matriz de varianzas y covarianzas**, la variabilidad respecto a la media se mide por la varianza y la relación lineal entre dos variables se mide por la covarianza. Por tanto la covarianza mide la dependencia lineal entre $x_j$ y $x_k$ a través de la expresión: $s_{jk} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$.
   Para una variable multidimensional definimos la matriz de varianzas y covarianzas como $S = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'$, que es una matriz cuadrada y simétrica que contiene, en la diagonal las varianzas y fuera de la diagonal las covarianzas.

   - Cálculo a partir de la matriz de datos centrados: si la matriz de datos centrados es $\tilde{X} = X - \mathbf{1}\bar{x}' = X - \frac{1}{n}\mathbf{1}\mathbf{1}'X = (I - \frac{1}{n}\mathbf{1}\mathbf{1}')X = PX$ donde la matriz $P = I - \frac{1}{n}\mathbf{1}\mathbf{1}'$ es simétrica e idempotente.
   - La matriz $P$ tiene rango $p-1$ ya que es ortogonal al vertor $\mathbf{1}$, y proyecta los datos ortogonalmente al espacio definido por el vector constante, (todas las coordenadas iguales). La matriz $S$ se puede expresar como $S = \frac{1}{n}\tilde{X}'\tilde{X} = \frac{1}{n}X'PX$.
   - Algunos autores dividen por $n-1$ en lugar de por $n$ para obtener un estimador insesgado de la matriz de la población, ya que no hay $n$ desviaciones independientes sino $n-1$, al estar las desviaciones ligadas por la ecuación $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$, nosotros la llamaremos matriz de varianzas corregida $\hat{S} = \frac{1}{n-1}\tilde{X}'\tilde{X}$