

Introducción

0.1. Álgebra lineal

Empezaremos tomando la siguiente matriz:

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1j} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2j} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{i1} & X_{i2} & \cdots & X_{ij} & \cdots & X_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nj} & \cdots & X_{np} \end{bmatrix}$$

Entonces cada renglón representa a un vector $\bar{x}_i \in \mathfrak{R}^p$ $i = 1, \dots, n.$, esto puede verse como que tenemos n vectores renglón de dimensión p .

0.1.1. Combinación lineal de vectores

Sean $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ p vectores de dimensión n , si tenemos k_1, k_2, \dots, k_p escalares, entonces

$$k_1\mathbf{x}_1 + k_2\mathbf{x}_2 + \dots + k_p\mathbf{x}_p$$

es una *combinación lineal* de p vectores.

Ejemplo: Encuentra la combinación lineal.

$$[x] + [x] - 3[x].$$

donde

$$\mathbf{x}_1 = \begin{bmatrix} 2 \\ 3 \\ 6 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 10 \\ 1 \\ 2 \end{bmatrix}$$

Solución

$$\mathbf{x}_1 + \mathbf{x}_2 - 3\mathbf{x}_3 = \begin{bmatrix} 2 \\ 3 \\ 6 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} - 3 \begin{bmatrix} 10 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 + 1 - 30 \\ 3 + 1 - 3 \\ 6 + 2 - 6 \end{bmatrix} = \begin{bmatrix} -27 \\ 1 \\ 2 \end{bmatrix}$$

0.1.2. Independencia lineal

Un conjunto de vectores $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ es considerado linealmente independiente si

$$k_1\mathbf{x}_1 + k_2\mathbf{x}_2 + \dots + k_p\mathbf{x}_p = \mathbf{0}$$

tiene como única solución $k_1 = k_2 = \dots = k_p = 0$

Un conjunto de vectores $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ es considerado linealmente dependientes si existe al menos un \mathbf{x}_i tal que

$$k_1\mathbf{x}_1 + \dots + k_{i-1}\mathbf{x}_{i-1} + k_{i+1}\mathbf{x}_{i+1} + \dots + k_p\mathbf{x}_p = \mathbf{x}_i$$

0.1.3. Rango de un conjunto de vectores

En un conjunto de vectores n -dimensionales, el número máximo de vectores linealmente independiente en el conjunto es llamado el *rango del conjunto de vectores*. Debemos notar que el rango del conjunto no puede ser mayor a su dimensión.

Ejemplo

$$\text{Obtener el rango de } \mathbf{x}_1 = \begin{bmatrix} 25 \\ 64 \\ 144 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 5 \\ 8 \\ 12 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Solución: $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ son linealmente independientes, ya que no existen k_1, k_2 y k_3 diferentes de 0 tales que $k_1\mathbf{x}_1 + k_2\mathbf{x}_2 + k_3\mathbf{x}_3 = \bar{\mathbf{0}}$ por lo tanto el rango

del conjunto de vectores $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ es 3.

A continuación se presentan algunos teoremas relacionados con estos temas.

Teorema 1. *Si un conjunto de vectores contiene el vector nulo ($\mathbf{0}$), el conjunto de vectores es linealmente dependiente.*

Sea $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p$ un conjunto de vectores de dimensión n , entonces $k_1\mathbf{A}_1 + k_2\mathbf{A}_2 + \dots + k_p\mathbf{A}_p = \mathbf{A}_0$ es una combinación lineal de p vectores. Si asumimos que \mathbf{A}_1 es el vector 0, cualquier valor de k_1 con $k_2 = k_3 = \dots = k_p = 0$ satisface la ecuación de arriba. De aquí se tiene que el conjunto de vectores es linealmente dependiente ya que existe una solución diferente a $k_1 = k_2 = k_3 = \dots = k_p = 0$

Teorema 2. *Si un conjunto de vectores son linealmente independientes, entonces un subconjunto de los p vectores tiene que ser linealmente independiente.*

Sea este subconjunto $\mathbf{A}_{a1}, \mathbf{A}_{a2}, \dots, \mathbf{A}_{ap}$ donde $m < p$. Suponiendo que este subconjunto es linealmente dependiente, la combinación lineal $k_1\mathbf{A}_{a1} + k_2\mathbf{A}_{a2} + \dots + k_p\mathbf{A}_{ap} = \mathbf{0}$ tiene una solución no-trivial. Entonces $k_1\mathbf{A}_{a1} + k_2\mathbf{A}_{a2} + \dots + k_p\mathbf{A}_{ap} + 0k_1\mathbf{A}_{a(p+1)} + \dots + 0k_p\mathbf{A}_{am} = \mathbf{0}$ también tiene una solución no trivial, donde $\mathbf{A}_{a(p+1)}, \dots, \mathbf{A}_{am}$ son los $(m - p)$ vectores restantes. Esto es una contradicción. Entonces un subconjunto de un conjunto de vectores linealmente independiente no puede ser linealmente dependiente.

Teorema 3. *Si un conjunto de vectores es linealmente dependiente, entonces al menos un vector puede escribirse como combinación lineal de otros.*

Sean $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p$ vectores linealmente dependientes, entonces existe un conjunto de números k_1, \dots, k_m no todos cero para la combinación lineal $k_1\mathbf{A}_1 + k_2\mathbf{A}_2 + \dots + k_p\mathbf{A}_p = \mathbf{0}$ sea m uno de esos valores k_i , $i = 1, \dots, p$ que no son cero, entonces $\mathbf{A}_m = \frac{k_1}{k_m}\mathbf{A}_1 - \dots - \frac{k_{m-1}}{k_m}\mathbf{A}_{m-1} - \frac{k_{m+1}}{k_m}\mathbf{A}_{m+1} - \dots - \frac{k_p}{k_m}\mathbf{A}_p$ lo que prueba el teorema.

Teorema 4. *Si la dimensión de un conjunto de vectores es menor que el número de vectores en el conjunto, entonces el conjunto es linealmente dependiente.*

0.1.4. Sistemas de ecuaciones lineales

Un conjunto de m ecuaciones lineales con p incógnitas se puede escribir como

$$\begin{array}{rcl} x_{11}k_1 + \dots + x_{1n}k_n & = & c_1 \\ \vdots & & \vdots \\ \vdots & & \vdots \end{array}$$

$$x_{m1}k_1 + \dots + x_{mn}k_n = c_n$$

donde k_1, k_2, \dots, k_n son desconocidas. Podemos reescribir estas ecuaciones con la notación de vectores de la siguiente manera

$$k_1\mathbf{x}_1 + k_2\mathbf{x}_2 + \dots + k_n\mathbf{x}_n = \bar{C}$$

donde

$$\mathbf{x}_1 = \begin{bmatrix} x_{11} \\ \vdots \\ x_{m1} \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} x_{12} \\ \vdots \\ x_{m2} \end{bmatrix} \quad \mathbf{x}_n = \begin{bmatrix} x_{1n} \\ \vdots \\ x_{mn} \end{bmatrix}$$

El problema ahora se convierte en encontrar los escalares k_1, \dots, k_n tales que

$$k_1\mathbf{x}_1 + k_2\mathbf{x}_2 + \dots + k_n\mathbf{x}_n = \bar{C}.$$

0.1.5. Producto punto

Sea $\mathbf{x} = [x_1, x_2, \dots, x_n]$ y $\mathbf{y} = [y_1, y_2, \dots, y_n]$ dos vectores n -dimensionales. Entonces el producto punto entre los vectores $\mathbf{x} = [x_1, x_2, \dots, x_n]$ y $\mathbf{y} = [y_1, y_2, \dots, y_n]$ se define como $\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 + \dots + x_ny_n = \sum_{i=1}^n x_iy_i$. Al producto punto también se le llama producto interior o escalar. También se le define como: $x \cdot y = \|x\| \|y\| \cos(\theta)$

Ejemplo

Encuentra el producto punto de los vectores $\mathbf{x} = (4, 1, 2, 3)$ y $\mathbf{y} = (3, 1, 7, 2)$.

Solución

$$\begin{aligned}\mathbf{x} \cdot \mathbf{y} &= (4, 1, 2, 3) \cdot (3, 1, 7, 2) \\ &= (4)(3) + (1)(1) + (2)(7) + (3)(2) \\ &= 33.\end{aligned}$$

Una propiedad trigonométrica de nuestro interés y que será utilizada más adelante es la siguiente.

$$\cos\theta = \frac{\text{ady}}{\text{hip}} = \frac{P_x}{\|x\|} = \frac{x \cdot y}{\|x\| \|y\|}$$

donde $x \cdot y$ es el tamaño de la proyección de x sobre y , por ejemplo

$x = (2, 3)$ y $y = (5, 3)$ entonces

$$\frac{x \cdot y}{\|y\|} = \frac{19}{\sqrt{34}} = 3,2584$$

$$\|P_x\| = \|x\| |\cos\theta| = \frac{|x \cdot y|}{\|y\|} = \frac{|\sum x_i y_i|}{\sqrt{\sum y_i^2}}$$

Ejemplo

Si tenemos dos vectores aleatorios con media 0, el coseno del ángulo entre ellos es su correlación. $\mathbf{x} = 0$, $\mathbf{y} = 0$

$$r_{xy} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \sqrt{\sum y_i^2}} = \cos\theta$$

0.1.6. Transformaciones lineales

Las transformaciones lineales son de la forma

$$Ax = y$$

algunos ejemplos de éstas son las siguientes.

matriz de rotación en un ángulo θ en sentido inverso a las manecillas del reloj

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

matriz de rotación en el sentido de las manecillas del reloj

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

matriz para cambiar de escala (ya sea amplificar $k > 1$ o reducir $k < 1$)

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} kX & 0 \\ 0 & kY \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Inclinando con respecto al eje x

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

inclinando respecto al eje y

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

reflejando respecto al eje y

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

0.1.7. Eigenvalores y eigenvectores

Para poder hablar de eigenvalores y eigenvectores primero debemos de considerar una matriz cuadrada A de $(p \times p)$ y un vector \mathbf{x} en \mathbb{R}^p . Muchas aplicaciones se resuelven al encontrar vectores \mathbf{x} tales que \mathbf{x} y $A\mathbf{x}$ son paralelos. Para resolver este problema primero se presentarán algunas definiciones y conceptos.

Definición

Sea \mathbf{A} una matriz de $p \times p$. El número real λ es llamado **eigenvalor** de \mathbf{A} ,

si existe un vector $\mathbf{x} \neq 0$ en \mathfrak{R}^p tal que

$$A\mathbf{x} = \lambda\mathbf{x}$$

Cada vector \mathbf{x} diferente de cero que satisface la ecuación de arriba es llamado un **eigenvector de A asociado al eigenvalor λ** . A los eigenvalores también se les llama valores propios, valores característicos o valores latentes. Así mismo, a $\det(\lambda I_n - A)$ se le llama polinomio característico de A .

Procedimientos para calcular valores y vectores propios

1. Encontrar $p(\lambda) = \det(A - \lambda\mathbf{I})$.
2. Calcular las raíces $\lambda_1, \lambda_2, \dots, \lambda_p$ de $p(\lambda) = 0$.
3. Resolver el sistema homogéneo $(A - \lambda_i\mathbf{I})\mathbf{c}_i = \mathbf{0}$ que corresponde a cada valor característico de λ_i .

EJEMPLO

A continuación se presenta un ejemplo para calcular los valores propios de una matriz.

Sea

$$A = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$$

Se desean obtener todos los números reales λ y vectores $\mathbf{x} \neq 0$ que satisfacen $A\mathbf{x} = \lambda\mathbf{x}$, esto es,

$$\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Esta ecuación equivale a:

$$x_1 + x_2 = \lambda x_1$$

$$-2x_1 + 4x_2 = \lambda x_2$$

ó

$$(\lambda - 1)x_1 - x_2 = 0$$

$$2x_1 + (\lambda - 4)x_2 = 0$$

Esta ecuación es un sistema homogéneo de dos ecuaciones con dos incógnitas. Se puede decir que este sistema tiene solución no trivial si y sólo si el determinante de la matriz de coeficientes es cero, esto es

$$\begin{vmatrix} \lambda - 1 & -1 \\ 2 & \lambda - 4 \end{vmatrix} = 0$$

Esto quiere decir que

$$(\lambda - 1)(\lambda - 4) + 2 = 0$$

ó

$$\lambda^2 - 5\lambda + 6 = 0 = (\lambda - 3)(\lambda - 2).$$

por lo que

$$\lambda_1 = 2 \text{ y } \lambda_2 = 3$$

son los eigenvalores de A. De esta forma podemos calcular el eigenvector

$$\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Esta ecuación equivale a:

$$x_1 + x_2 = 2x_1$$

$$-2x_1 + 4x_2 = 2x_2$$

finalmente tenemos

$$x_1 - x_2 = 0$$

$$2x_1 - 2x_2 = 0$$

entonces los vectores que satisfacen estas ecuaciones son de la forma $x_1 = x_2$ para $\lambda = 2$, por ejemplo, el vector $\mathbf{x} = (1, 1)$. De forma análoga procedemos con $\lambda = 3$ obteniendo que los eigenvectores asociados a este valor son de la forma $x_1 = x_2/2$, por ejemplo el vector $\mathbf{x} = (1, 2)$.

Propiedades

Sea A una matriz de orden $p \times p$. Entonces

1. $\sum_{i=1}^p \lambda_i = \text{traza}(A)$.
2. $|A| = \det(A) = \prod_{i=1}^p \lambda_i = \lambda_1 * \lambda_2 * \dots * \lambda_p$.
3. Si A es positiva definida , entonces $\lambda_i > 0$ ($i = 1, \dots, p$)
4. Si A es una matriz de número reales simétrica, entonces sus eigenvalores y eigenvectores son reales.
5. si A es positiva semidefinida de rango r , entonces exactamnete r de los λ_i son positivos y $(p - r)$ son cero.
6. Si $\lambda_i \neq \lambda_j$ entonces los eigenvectores asociados son ortogonales, $\mathbf{x}_i \cdot \mathbf{x}_j = 0$.
Es decir si todos los λ_i son distintos, entonces L la matriz que tiene como columnas a los eigenvectores \mathbf{x}_i es ortogonal $LL' = I$.

0.1.8. Diagonalización de matrices

Se dice que una matriz A es **diagonalizable** si puede escribirse como:

$$A = PDP^{-1}$$

donde P es una matriz invertible cuyos vectores columna son los eigenvectores de A y D es una matriz diagonal formada por los eigenvalores de A .

Teorema 5. *Una matriz A es diagonalizable si todas las raíces de su polinomio característico son reales y diferentes.*

Si además la matriz P es ortogonal se dice entonces que la matriz A es **diagonalizable ortogonalmente**, pudiendo escribirse como

$$A = PDP'$$

Nota:

Sí todas las raíces del polinomio característico de A son reales y no todas diferentes, entonces A puede o no ser diagonalizable. El polinomio característico de A puede escribirse como el producto de n factores, cada uno de la forma $\lambda - \lambda_j$, donde λ_j es una raíz del polinomio característico. Ahora los eigenvalores de A son las raíces reales del polinomio característico de A . De aquí que el polinomio característico se pueda escribir como

$$(\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2} \dots (\lambda - \lambda_r)^{k_r}$$

donde $\lambda_1, \lambda_2, \dots, \lambda_r$ son los distintos eigenvalores de A , y k_1, k_2, \dots, k_r son enteros cuya suma es n . Al entero k_i se le llama la multiplicidad de λ_i . Se puede demostrar que si las raíces del polinomio característico de A son todas reales, entonces A puede ser diagonalizada si y solo si para cada eigenvalor λ_i de multiplicidad k_j podemos encontrar k_j eigenvectores linealmente independientes.

Teorema 6. *Cualquier matriz cuadrada simétrica con coeficientes reales es ortogonalmente diagonalizable. Se le conoce como Teorema Espectral.*

Ejemplo

Sea

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}.$$

Los eigenvalores y eigenvectores de A son : 5, -1 y -1

$$1/\sqrt{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, 1/\sqrt{6} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \text{ y } 1/\sqrt{2} \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}.$$

respectivamente.

Entonces por la descomposición espectral se tiene que

$$A = 5/3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} - 1/6 \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} -2 & 1 & 1 \end{bmatrix} - 1/2 \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 1 \end{bmatrix}$$

0.1.9. Matrices no-negativas definidas

Este tipo de matrices son las que nos interesan principalmente, debido a esto, se tomarán en cuenta las siguientes propiedades.

Definición

Cuando $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ para todo \mathbf{x} diferente a $\mathbf{x} = \mathbf{0}$ se dice que se $\mathbf{x}'\mathbf{A}\mathbf{x}$ es una forma cuadrática positiva definida, y A es una matriz *positiva definida*(p.s.).

Definición

Cuando $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ para todo \mathbf{x} y $\mathbf{x}'\mathbf{A}\mathbf{x} = 0$ para algunos $\mathbf{x} \neq \mathbf{0}$ entonces $\mathbf{x}'\mathbf{A}\mathbf{x}$ es una forma cuadrática semi-positiva definida, y A es una matriz *semi – positiva definida*(p.s.d.).

Definición

Los dos tipos de matrices tomados juntos, positivo definido y positivo semi-definido, son llamados *no – negativos definidos*(n.n.d.).

Si son simétricas tienen las siguientes propiedades:

1. Todos sus eigenvalores son reales.
2. Son diagonalizables.
3. El rango iguala el número de eigenvalores diferentes de cero.

Teorema 7. *Los eigenvalores de una matriz simétrica son todos no-negativos si y solo si la matriz es n.n.d.*

0.2. Distancia

0.2.1. Definición de distancia

Para que una función entre dos puntos sea considerada distancia debe cumplir con las siguientes propiedades:

$$d(x, y) > 0 \quad \forall x \neq y$$

$$d(x, y) = 0 \Leftrightarrow x \equiv y$$

$$d(x, y) \leq d(x, z) + d(z, y) \quad \forall x, y, z$$

Ejemplo

La distancia Euclideana:

$$d^2(x, y) = (x - y)'A(x - y)$$

donde $A > 0$ es decir una matriz positiva definida.

Podemos representar a todos los elipsoides con la siguiente expresión:

$E_d = x \in \mathfrak{R}^p$ tales que $(x - x_0)'A(x - x_0) = d^2$. De esta manera si tenemos que $A = I$ la representación nos da un círculo, de otra forma tenemos elipsoides que dependen directamente de la matriz A .

0.2.2. Norma de un vector

Sean x y y vectores, a norma de un vector (representada $\|x\|$) se define como:

$$\|x\| = d(0, x) = \sqrt{x'x}$$

$$\|x\|_A = \sqrt{x'Ax}$$

La última notación se lee como norma módulo A .

0.3. Conceptos multivariados de Media y Varianza

Retomando la matriz

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1j} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2j} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{i1} & X_{i2} & \cdots & X_{ij} & \cdots & X_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nj} & \cdots & X_{np} \end{bmatrix}$$

Tenemos que cada renglón de la matriz es una realización de nuestro vector aleatorio con p variables, $X' = (X_1, X_2, \dots, X_p)$.

Con estas variables es posible calcular las siguientes funciones

0.3.1. Media Poblacional

La media poblacional, cuando existe, se define como:

$$E[X_1] = \int x_1 dF(x_1) = \mu_1$$

$$\vdots$$

$$E[X_p] = \int x_p dF(x_p) = \mu_p$$

0.3.2. Varianza y Covarianza poblacional

La varianza poblacional, cuando existe, se define como:

$$Var[X_1] = \int (x_1 - \mu_1)^2 f(x_1) dx_1 = \sigma_1^2$$

$$\vdots$$

$$Var[X_p] = \int (x_p - \mu_p)^2 f(x_p) dx_p = \sigma_p^2$$

La covarianza poblacional se define como:

$$\text{cov}(x_1, x_2) = \int \int (x_1 - \mu_1)(x_2 - \mu_2)f(x_1, x_2)dx_1dx_2 = \sigma_{12}$$

0.3.3. Matriz de varianzas y Covarianzas poblacional

En la teoría de la estadística y probabilidad, la matriz de covarianzas es una matriz que tiene como entradas a las covarianzas entre las variables aleatorias del vector. Es la generalización natural a dimensiones más grandes, del concepto de varianza de una variable aleatoria.

Si X es un vector columna con p variables aleatorias con media μ_k i.e. $\mu_k = E[X_k]$, entonces la matriz de covarianza se define como:

$$\Sigma = E[(X - E[X])(X - E[X])'] = \text{var}(x) \quad (1)$$

$$= \begin{bmatrix} E[(X_1 - \mu_1)(X_1 - \mu_1)] & E[(X_1 - \mu_1)(X_2 - \mu_2)] & \cdots & E[(X_1 - \mu_1)(X_n - \mu_p)] \\ E[(X_2 - \mu_2)(X_1 - \mu_1)] & E[(X_2 - \mu_2)(X_2 - \mu_2)] & \cdots & E[(X_2 - \mu_2)(X_n - \mu_p)] \\ \vdots & \vdots & & \vdots \\ E[(X_p - \mu_p)(X_1 - \mu_1)] & E[(X_p - \mu_p)(X_2 - \mu_2)] & \cdots & E[(X_p - \mu_p)(X_p - \mu_p)] \end{bmatrix}$$

El elemento (i, j) de la matriz es la covarianza de X_i y X_j esto es $(\text{cov}(x_i, x_j))$.

Las nomenclaturas de esta matriz difieren según el autor, algunas veces se le llama matriz de varianza del vector aleatorio X , debido a que es la generalización natural a dimensiones más grandes de la varianza uni-dimensional. Otras veces es llamada la matriz de covarianzas debido a que es la matriz de covarianzas entre los componentes del vector X .

$$\Sigma = \begin{bmatrix} \sigma_{x_1x_1} & \cdots & \sigma_{x_1x_p} \\ \sigma_{x_2x_1} & \cdots & \sigma_{x_2x_p} \\ \vdots & \vdots & \vdots \\ \sigma_{x_px_1} & \cdots & \sigma_{x_px_p} \end{bmatrix}$$

0.3.4. Matriz de Correlaciones Poblacional

Otra matriz de nuestro interés es la matriz R que contiene las correlaciones $\rho_{x_i x_j} = \text{cov}(x_i, x_j) / \sqrt{\text{var}(x_i) \text{var}(x_j)}$:

$$R = \begin{bmatrix} \rho_{x_1 x_1} & \cdots & \rho_{x_1 x_p} \\ \rho_{x_2 x_1} & \cdots & \rho_{x_2 x_p} \\ \vdots & \vdots & \vdots \\ \rho_{x_p x_1} & \cdots & \rho_{x_p x_p} \end{bmatrix}$$

0.3.5. Media Muestral

La media muestral de la j -ésima variable está dada por

$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ Denotaremos al conjunto de las medias en un vector de medias muestrales

$$\bar{\mathbf{x}}' = (\bar{x}_1, \dots, \bar{x}_p)$$

0.3.6. Varianza y Covarianza muestral

La varianza muestral de la j -ésima variable se calcula como:

$$s_{jj}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$

La covarianza entre la j -ésima variable y la k -ésima variable esta dada por

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

La matriz de covarianzas muestral denotada por \mathbf{S} , contiene a las varianzas y covarianzas.

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' = \begin{bmatrix} s_1^2 & \cdots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{p1} & \cdots & s_p^2 \end{bmatrix}.$$

0.3.7. Propiedades

Para $\Sigma = E[(X - E[X])(X - E[X])']$ y $\mu_k = E[X_k]$ se tienen las siguientes propiedades básicas:

1. $\Sigma = E(XX') - \mu\mu'$
2. $var(a'X) = a'var(X)a$
3. Σ es una matriz positiva semi-definida
4. $var(AX + a) = Avar(X)A'$
5. $cov(X, Y) = cov(Y, X)'$
6. $cov(X_1 + X_2, Y) = cov(X_1, Y) + cov(X_2, Y)$
7. Si $p=q$, entonces $cov(X + Y) = var(X) + cov(X, Y) + cov(Y, X) + var(Y)$
8. $cov(AX, BX) = Acov(X, X)B'$
9. Si X y Y son independientes, entonces $cov(X, Y) = 0$

donde X_1 y X_2 son vectores aleatorios de $(p \times 1)$, Y es un vector aleatorio de $(q \times 1)$, A y B son matrices de $(p \times q)$ y a es un vector de $p \times 1$.

Teorema 8. Si X y Y son variables independientes entonces $\rho(X, Y) = cov(X, Y) = 0$

nota: el converso no siempre es cierto, por ejemplo sean $X \sim N(0, 1)$ y $Y = X^2$ variables aleatorias no independientes.

$$E(X) = 0 \text{ y } E(X^2) = 1 \text{ ya que } var(x) = E(X - 0)^2 = E(X^2) = 1$$

$Cov(X, Y) = E(XY) - E(X)E(Y) = E(X^3) = 0$ la última igualdad se obtiene por la simetría de la función de densidad normal.

NOTA: Si X y Y son normales y $cov(X, Y) = 0$ entonces las variables son independientes.

0.4. Normal multivariada

Sea $\Sigma_{p \times p} = (\sigma_{ij})$ una matriz simétrica, positiva semidefinida y $\mu_{p \times 1}$. $\mathbf{x} \sim N_p(\mu, \Sigma)$ si y solo si $t'\mathbf{x} \sim N(t'\mu, t'\Sigma t) \quad \forall t \in \mathfrak{R}^p$.

Proposición 1. $\mathbf{x} \sim N_p(\mu, \Sigma)$, $E(\mathbf{x}) = \bar{\mu}$, $var(\mathbf{x}) = \Sigma$. si t es el vector canónico se tiene $x_i \sim N(\mu_i, \sigma_{ii})$ si $t = (0, \dots, 1, \dots, 1, \dots, 0)$ donde $t_i = 1$ y $t_j = 1$

$$x_i + x_j \sim N(\mu_i + \mu_j, \sigma_{ii} + \sigma_{jj} + 2\sigma_{ij}), \quad i \neq j$$

$$\text{por otro lado sabemos } var(x_i + x_j) = \sigma_{ii} + \sigma_{jj} + 2cov(x_i, x_j)$$

$$\text{por lo tanto } cov(x_i, x_j) = \sigma_{ij}.$$

Proposición 2. $A_{m \times p}$ y $\mathbf{x}_{p \times 1} \Rightarrow A\mathbf{x} \sim N_p(A\mu, A\Sigma A')$

0.4.1. Función de densidad

La función de densidad probabilística para x normal multivariada con media μ y matriz varianzas covarianzas Σ .

$$f(x) = (2\pi)^{-p/2} \det(\Sigma)^{-1/2} \exp\left[-\frac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu)\right]$$

0.4.2. Transformación de Mahalanobis

Tomemos la transformación $z_i = S^{-\frac{1}{2}}(x_i - \mu)$ con $i = 1, \dots, n$. A esta transformación se le llama «transformación de Mahalanobis». Aplicando la proposición 2 podemos obtener:

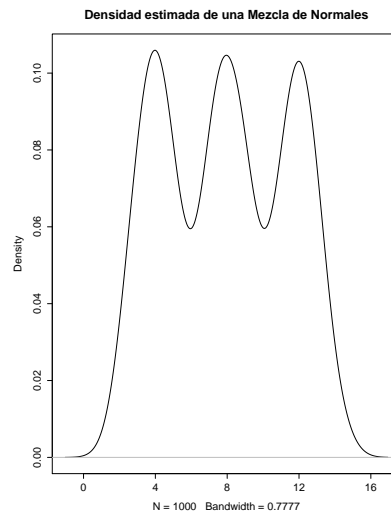
$$var(z) = S_z = S^{-\frac{1}{2}}SS^{-\frac{1}{2}} = I_p$$

Con esta transformación se elimina la correlación entre las variables y estandariza la varianza.

0.4.3. Mezcla de Normales

Por ejemplo si se extraen al azar observaciones de tres distintas poblaciones normales. Entonces formalmente se tiene una variable $Y \sim \text{discreta}(p_1, p_2, p_3)$ y según el valor que tome Y se muestrea de N_i . Si $Y = i$ se elige una muestra de una $N(\mu_i, \sigma_i^2)$. La variable aleatoria resultante tiene función de densidad:

$$f(x) = \sum_y p(y) f_{Normal}(x; \mu_y, \sigma_y^2)$$



Esta es una distribución multimodal y por tanto no normal.

Componentes principales

Para tener una idea del análisis de componentes principales, daré un ejemplo en el plano. Si se tienen mediciones x_1 como alturas y x_2 como pesos de un grupo de personas. Abajo aparece la gráfica de dispersión de los datos. El análisis de componentes principales tiene como objetivo reducir la dimensión y conservar en lo posible su estructura, es decir la "forma" de la nube de datos, que por cierto no depende de los ejes utilizados para dar las coordenadas.

En este ejemplo se buscará proyectar los datos sobre un eje que reproduzca de la mejor manera la "forma" de la nube de datos.

El primer paso es centrar los datos en el centroide (\bar{X}_1, \bar{X}_2) y después se hace una rotación, de manera que la "las proyecciones" sean lo más parecidas posibles a los vectores originales. En la figura de arriba puede verse que los individuos que quedan a la izquierda en el eje OY_1 , son los más pequeños en **talla**, y a la derecha los más grandes. Tomado en cuenta el otro eje OY_2 , los sujetos que quedan por encima de este son los aquellos que tienen un peso mayor dada la estatura, y por debajo los que tienen poco peso dada su estatura, es decir este eje habla de habla de la **forma** de los sujetos. Para estos sujetos ocurre que varían mucho en talla y hay poca variación en la forma.

Para este procedimiento se requiere girar los ejes, esto se consigue aplicando una transformación lineal a los datos, esto es

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} X_1 \cos \alpha + X_2 \sin \alpha \\ -X_1 \sin \alpha + X_2 \cos \alpha \end{bmatrix}$$

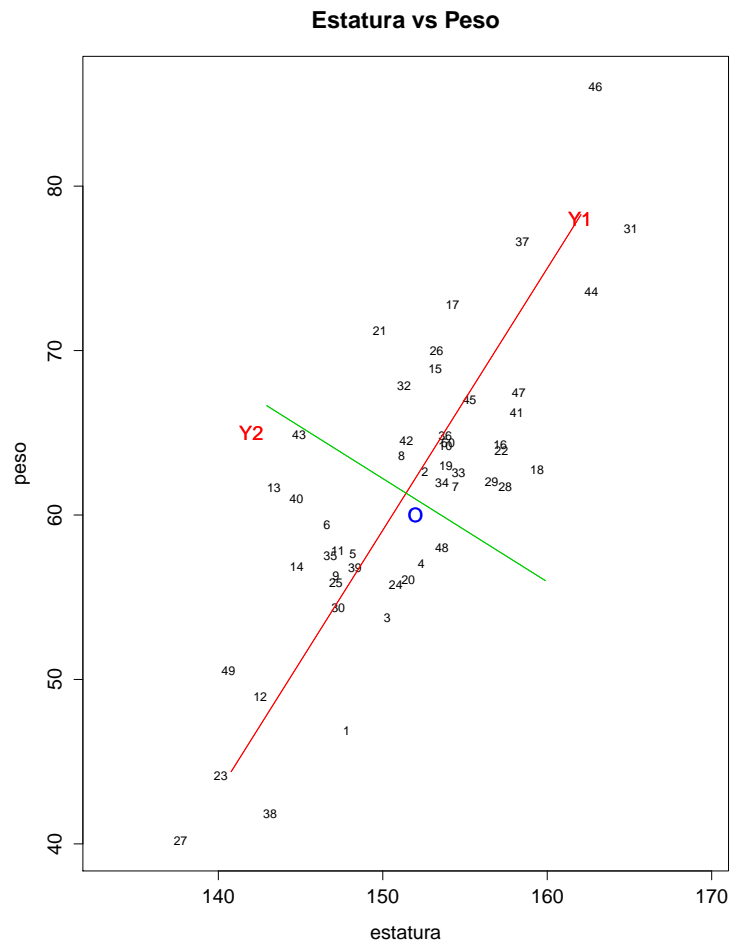


Figura 1: Primera y segunda componentes

Las proyecciones de los puntos sobre el eje OY_1 son una buena aproximación a los datos, ya que en la otra dirección hay poca variación.

Entonces se puede usar únicamente

$Y_1 = X_1 \cos \alpha + X_2 \sin \alpha$, y así la nueva variable Y_1 resume a las otras dos.

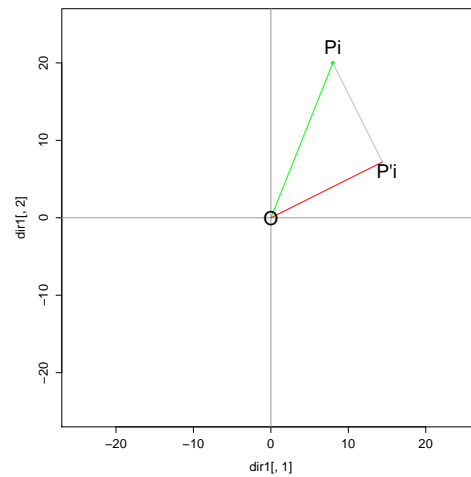


Figura 2: Proyección de un punto sobre un nuevo eje

Por Teorema de Pitágoras se tiene que:

$$(OP_i)^2 = (OP'_i)^2 + (P_iP'_i)^2.$$

La cantidad $(P_iP'_i)^2$ puede ser vista como error.

Si se hace la sumatoria sobre todos los sujetos desde $i = 1, \dots, n$ y se divide entre $n - 1$ se tiene

$$C = \frac{\Sigma(OP_i)^2}{n-1} = \frac{\Sigma(OP'_i)^2}{n-1} + \frac{\Sigma(P_iP'_i)^2}{n-1}$$

El objetivo es entonces minimizar la cantidad $\frac{\Sigma(P_iP'_i)^2}{n-1}$

Ahora $\Sigma(OP_i)^2$ es una cantidad fija, no depende de los ejes de coordenadas, y por tanto minimizar $\frac{\Sigma(P_iP'_i)^2}{n-1}$ es equivalente a maximizar $\frac{\Sigma(OP'_i)^2}{n-1}$, esta última cantidad coincide con la **varianza de las proyecciones sobre eje OY_1** , es decir que el ángulo de rotación que se busca es aquel que MAXIMIZE la

varianza de las proyecciones.

En general se tienen datos en un espacio de p dimensiones entonces se busca la transformación lineal $a'x = y_1$ de manera que tenga máxima varianza, se conoce como primer componente principal. Si el vector x tiene matriz de varianzas y covarianzas Σ entonces $var(y_1) = var(a'x) = a'\Sigma a$

Se buscan también otras combinaciones lineales Y_i de las variables originales

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$$

que sean ortogonales entre si (y_i ortogonal a y_j) y que de manera sucesiva vayan maximizando la varianza.

Digamos

$$y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p.$$

Elegir a_1 de manera que la $var(y_1)$ sea de máxima varianza.

Si a se toma de norma muy grande, entonces la varianza de y_1 puede ser tan grande como se quisiera, de manera que se deben imponer condiciones a a_1 para acotar el tamaño de varianza, la condición que se impone es

$$\|a_1\| = a'_1 \cdot a_1 = 1$$

De esta manera obtenemos

$$var(y_1) = var(a'_1 \bar{x}) = a'_1 \Sigma a_1$$

Esta es la función objetivo, es decir, debemos encontrar que vector a_1 es el que maximiza $a'_1 \Sigma a_1$ y además $a'_1 \cdot a_1 = 1$.

Para lograr maximizar estas condiciones se usan multiplicadores de Lagrange:

$$f(x_1, x_2, \dots, x_p) \text{ sujeto a } g(x_1, x_2, \dots, x_p) = c,$$

donde f es una función diferenciable. Existe una λ tal que $\frac{\partial f}{\partial x_i} - \lambda \frac{\partial g}{\partial x_i} = 0$.

en este caso es

$$L(a_1) = a'_1 \Sigma a_1 - \lambda(a'_1 a_1 - 1)$$

$$\frac{\partial L}{\partial a_1} = 2\Sigma a_1 - 2\lambda a_1$$

al derivar e igualar a cero, se tiene $(\Sigma - \lambda I)a_1 = 0$ entonces resulta que a_1 es eigenvector de Σ y λ su eigenvalor, que equivale a decir que

$$|\Sigma - \lambda I| = 0.$$

Sean $\lambda_1, \lambda_2, \dots, \lambda_p$ los eigenvalores que satisfacen

$$\lambda_1 > \lambda_2 > \dots > \lambda_p \geq 0 \text{ y } \Sigma = A\Lambda A' \text{ donde } \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{bmatrix}$$

NOTA. Esto se da pues la matriz es simétrica y semi positiva definida.

¿Cual de ellos determina a la primer componente?

$$\text{var}(a_1'X) = a_1'\Sigma a_1$$

$$= a_1'\lambda I a_1$$

$$= \lambda$$

Como interesa que sea el que maximize la varianza, λ es el mayor de los λ_i , digamos λ_1 .

Entonces a_1 es el eigenvector asociado a λ_1 .

Para buscar la segunda componente $y_2 = a_2'X$, se impone también la condición $a_2'a_2 = 1$, y además y_2 debe ser no correlacionada a y_1 , entonces su covarianza debe ser cero. Haciendo el desarrollo se tiene

$$\text{cov}(y_2, y_1) = \text{cov}(a_2'X, a_1'X)$$

$$= E[a_2'(x - \mu)(x - \mu)'a_1]$$

$$= a_2'\Sigma a_1$$

Se sabe que

$\Sigma a_1 = \lambda_1 a_1$ y sustituyendo en la expresión anterior se tiene

$$\lambda_1 a_2'a_1 = a_2'\lambda_1 a_1 = 0$$

$$\Leftrightarrow a_2'a_1 = 0$$

es decir que a_1 es perpendicular a a_2 .

$$L(a_2) = a_2'\Sigma a_2 - \lambda(a_2'a_2 - 1) - \delta a_2'a_1$$

$$\frac{\partial L}{\partial a_2} = 2(\Sigma - \lambda I)a_2 - \delta a_1 = 0$$

Premultiplicando esto por a_1' y operando

$$2a_1'\Sigma a_2 - \delta = 0$$

como $a_1'a_2 = 0$ y como también $a_1'\Sigma a_2 = 0$ (no correlacionado), se tiene que δ debe ser cero, entonces la ecuación que interesa es $(\Sigma - \lambda I)a_2 = 0$ y de acuerdo a esto λ_2 corresponde al segundo eigenvalor y a_2 al segundo eigenvector.

Cuando hay eigenvalores iguales se eligen eigenvectores ortogonales.

$$\text{Sea } A = \begin{bmatrix} \bar{a}_1 & \bar{a}_2 & \dots & \bar{a}_p \end{bmatrix}$$

Sea $Y_{p \times 1}$, el vector de las componentes principales.

$$Y = A'X$$

La matriz de covarianzas de Y es Λ y esta dada por

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ & & \ddots & & \\ 0 & 0 & 0 & \dots & \lambda_p \end{bmatrix}$$

$$\text{var}(Y) = A'\Sigma A = A'\Lambda A' A = \Lambda$$

$$\text{traza}(\Lambda) = \sum_{i=1}^p \lambda_i = \sum_{i=1}^p \text{var}(y_i)$$

$$\text{traza}(\Lambda) = \text{traza}(A'\Sigma A) = \text{traza}(\Sigma A A') = \text{traza}(\Sigma) = \sum_{i=1}^p \text{var}(x_i)$$

$$\sum_{i=1}^p \text{var}(y_i) = \sum_{i=1}^p \text{var}(x_i)$$

Esto es útil para determinar el número de componentes a utilizar.

Si se considera como **varianza generalizada** a $\sum_{i=1}^p \sigma^2_i = \text{traza}(\Sigma)$, entonces

$$\text{traza}(\Sigma) = \sum_{i=1}^p \lambda_i.$$

De esta forma tenemos que

$\frac{\lambda_j}{\sum_{i=1}^p \lambda_i}$ nos dice el porcentaje de la varianza generalizada que es explicado por la componente j -ésima y

$\frac{\sum_{i=1}^j \lambda_i}{\sum_{i=1}^p \lambda_i}$ nos da el porcentaje de la varianza generalizada dado por las primeras j componentes principales.

La covarianza entre x_i y y_j es el vector $\Sigma a_j = \lambda_j a_j$ entonces

$$\begin{aligned} \text{cov}(x_i, y_j) &= \lambda_j a_{ij} \text{ y} \\ \text{corr}(x_i, y_j) &= \frac{\lambda_j a_{ij}}{\sqrt{\lambda_j} \sqrt{\sigma^2_i}} \\ &= \frac{\sqrt{\lambda_j} a_{ij}}{\sigma_i} \end{aligned}$$

NOTA.- Como se desconoce Σ , todo en la práctica se hace con su estimador S .

Cuando se trabaja con distintas unidades conviene hacer el análisis con la matriz R de correlaciones. Los eigenvectores de R y S no coinciden, y no hay una forma de obtener unos a partir de los otros.

Análisis de factores

Nuevamente el vector $x' = (x_1, \dots, x_p)$ contiene las p variables que interesan estudiar los n individuos de la muestra. Las variables consideradas tienen una escala de medición **continua**.

La correlación r_{ij} entre dos variables x_i y x_j puede deberse a que ambas tienen una relación con otra variable z_k que se considera como no observable, es decir que si x_i está relacionada con z_k y x_j está relacionada con z_k , entonces debe haber una relación entre x_i y x_j .

El coeficiente de correlación parcial $r_{ij.k}$ mide la asociación que hay entre x_i y x_j luego de remover el efecto que tiene z_k en cada una. Se puede pensar que si el valor de $r_{ij.k}$ se acerca a cero, sucede que z_k ha explicado la correlación existente entre x_i y x_j ; pueden considerarse varias variables z_1, \dots, z_m para explicar la asociación entre x_i y x_j .

La idea en análisis de factores es explicar las correlaciones contenidas en $R = [r_{ij}]$ a través de un conjunto de variables no observables f_1, \dots, f_m de manera que las **correlaciones parciales** $r_{ij.f_1, \dots, f_m}$ sean muy pequeñas, se busca también que esa m sea pequeña.

Por ejemplo las calificaciones que los estudiantes puedan sacar en cálculo mental, manejo de vocabulario, comprensión de lectura, dibujo, etc, están correlacionadas entre sí, ya que cada estudiante posee cualidades como inteligencia, memoria, habilidad espacial, etc, todas estas cualidades no pueden medirse directamente por lo que se les considera como **no observables**.

0.5. Modelo Básico

En análisis de factores se quiere explicar las relaciones entre las variables manifiestas x_1, \dots, x_p a través de m variables latentes f_1, \dots, f_m que en el modelo se llaman factores comunes, de la siguiente forma:

$$x_j = \mu_j + \lambda_{j1}f_1 + \dots + \lambda_{jm}f_m + e_j$$

Las μ_j y las λ_{ji} son constantes, mientras que e_j y los f_i (con $j = 1 \dots p$ y $i = 1 \dots m$) son variables aleatorias. Se considera a los e_j **no** correlacionados entre sí y los e_j son **no** correlacionados con f_i . A los e_j se les llama factores específicos.

De manera matricial se puede escribir como

$x = \mu + \Lambda f + e$ o centrando la variable en cero $x - \mu = \Lambda f + e$ donde $x = (x_1, x_2, \dots, x_p)$, $e' = [e_1, e_2, \dots, e_p]$, $f' = [f_1, f_2, \dots, f_m]$ y la matriz Λ contiene sólo constantes

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1m} \\ \vdots & & & \vdots \\ \lambda_{p1} & \dots & \dots & \lambda_{pm} \end{bmatrix}$$

Se supone que e y f tienen media cero, esto es $E(e) = 0$ y $E(f) = 0$.

La matriz de varianzas y covarianzas del vector de factores específicos

$cov(e) = \Psi$ es diagonal

$$\Psi = \begin{bmatrix} \psi_1^2 & 0 & \dots & 0 \\ 0 & \psi_2^2 & \dots & 0 \\ & & \ddots & \\ 0 & \dots & 0 & \psi_p^2 \end{bmatrix}$$

Suponiendo que la matriz de varianzas y covarianzas de f es Γ , por ser simétrica y semipositiva definida se puede escribir como $\Gamma = LL'$, si se considera el vector $f^* = L^{-1}f$, este vector mantiene su media en cero y tiene matriz de varianzas y covarianzas $var(f^*) = L^{-1}\Gamma L^{-1'} = L^{-1}LL'L^{-1'} = I$.

Ahora escribiendo $\Lambda^* = \Lambda L$ entonces $\Lambda^* f^* = \Lambda L L^{-1} f = \Lambda f$, haciendo esto que el modelo

$$x = \mu + \Lambda^* f^* + e = \mu + \Lambda L^{-1} f + e$$

resulte indistinguible del original, entonces sin pérdida de generalidad se consideran a los factores comunes como **no** correlacionados entre sí.

Como se han considerado los factores comunes con correlación cero al calcular la varianza de x_j se tiene que:

$$var(x_j) = \lambda_{j1}^2 + \lambda_{j2}^2 + \dots + \lambda_{jm}^2 + var(e_j) = \underbrace{\sum \lambda_{jk}^2}_{\text{comunalidad}} + \underbrace{\psi_j^2}_{\text{especificidad}}$$

Y las covarianzas quedan como:

$$cov(x_i, x_j) = \sum_{k=1}^m \lambda_{ik} \lambda_{jk}$$

Entonces el modelo de análisis de factores también permite expresar a la matriz de varianzas y covarianzas de x , la Σ como:

$$\Sigma = \Lambda \Lambda' + \Psi$$

Al calcular la covarianza entre los factores comunes y las variables originales se tiene que:

$$cov(x, f) = E((\Lambda f + e)f') = \Lambda$$

esto porque $E(ff') = I$ y $E(ef') = 0$, de manera que las cargas λ_{ij} son la covarianza entre la variable original x_i y el factor f_j .

Entonces el análisis de factores es adecuado para explicar las covarianzas y el análisis de componentes principales es adecuado para explicar las varianzas.

Los factores se calculan por tres métodos:

1. Extracción inicial de factores: factores principales
2. factores principales iterados

3. factores de máxima verosimilitud

0.5.1. Factores principales

$$y_1 = \alpha_{11}x_1 + \dots + \alpha_{1p}x_p$$

⋮

$$y_p = \alpha_{p1}x_1 + \dots + \alpha_{pp}x_p$$

como los eigenvectores α_j son ortogonales se tiene que

$$x_1 = \alpha_{11}y_1 + \dots + \alpha_{1p}y_p$$

⋮

$$x_p = \alpha_{p1}y_1 + \dots + \alpha_{pp}y_p$$

de tal forma que en los últimos términos se explica poco pues las componentes principales van reduciendo su varianza.

Podemos escribir lo anterior como

$$x_i = \alpha_{1i}y_1 + \alpha_{2i}y_2 + \dots + \alpha_{mi}y_m + \eta_i$$

que se parece a

$$x_i = \lambda_{1i}f_1 + \lambda_{2i}f_2 + \dots + \lambda_{mi}f_m + e_i$$

donde $\eta_i = \alpha_{m+1i}y_{m+1} + \dots + \alpha_{pi}y_p$.

Hay que fijarse que los η_i no cumplen con el supuesto de ser no correlacionados. Como $\eta_i = \alpha_{m+1i}y_{m+1} + \dots + \alpha_{pi}y_p$ para cada i aparecen las mismas componentes principales y_j , entonces las η_i sí están correlacionadas.

Sin embargo esto se toma como una aproximación.

Para asegurar que los factores comunes f_j tengan varianza 1 se redefinen como

$$f_j = \frac{y_j}{\sqrt{\text{var}(y_j)}}$$

Tomando esto en cuenta cada x_i queda definida como:

$$x_i = \alpha_{1i}f_1\sqrt{\text{var}(y_1)} + \alpha_{2i}f_2\sqrt{\text{var}(y_2)} + \dots + \alpha_{pi}f_p\sqrt{\text{var}(y_p)}$$

Las cargas quedan definidas así

$$\lambda_{ij} = \alpha_{ji}\sqrt{\text{var}(y_i)}$$

y los factores específicos

$$\eta_i = \alpha_{m+1i}f_{m+1}\sqrt{\text{var}(y_{m+1})} + \dots + \alpha_{pi}f_p\sqrt{\text{var}(y_p)}$$

Como las componentes principales son no correlacionadas, entonces las η_i no tienen correlación con f_j quedando finalmente

$$x_i = \lambda_{i1}f_1 + \lambda_{i2}f_2 + \dots + \lambda_{im}f_m + \eta_i$$

Sin embargo las η_i si son correlacionadas entre sí.

0.5.2. Factores Principales Iterados

Si la matriz de varianzas y covarianzas corresponde a datos estandarizados ocurre que

$1 = \text{var}(x_i) = \sum_{k=1}^m \lambda_{ik}^2 + \psi_j^2$, de aquí se puede obtener un estimador de la matriz Ψ

Se hace un método iterativo para hallar los factores comunes de la manera siguiente:

1. encontrar una comunalidad inicial.
2. hacer un componentes principales de la matriz $S - \hat{\Psi}$, usar las cargas de los primeros m componentes como columnas de $\hat{\Lambda}$
3. recalcular $\hat{\Psi}$ como la diagonal de $S - \hat{\Lambda}\hat{\Lambda}'$
4. regresar al paso 2 con este nuevo estimador de Ψ

Se itera hasta que las diferencias entre los pares $\hat{\Lambda}$ y Ψ de dos pasos consecutivos sean prácticamente cero.

0.5.3. Factores por máxima verosimilitud

Suponiendo normalidad multivariada en los datos observados y suponiendo el modelo de factores se tiene que la verosimilitud L es función de los datos

y de los parámetros es decir, $L(x; \mu, \Sigma) = L(x; \mu, \Lambda, \Psi)$. En este caso la μ se considera como parámetro de ruido. Y se puede proceder de dos formas, una sustituyendo a μ por el estimador máximo verosímil, y otra descomponiendo la verosimilitud en dos factores, usando verosimilitud condicional dada la \bar{x} . En ambos casos resulta que los estimadores máximo verosímiles $\hat{\Lambda}$ y $\hat{\Psi}$ son los valores que maximizan :

$$F = -\{ \ln|\Lambda\Lambda' + \Psi| + \text{traza}(S|\Lambda\Lambda' + \Psi|^{-1}) - \ln|S| - m \}$$

La función F toma valor de cero si $S = \Lambda\Lambda' + \Psi$ y valores negativos de otra manera. Se usa realmente un método que hace la maximización de la verosimilitud en pasos, ver capítulo 16 de Krzanowski. El método de máxima verosimilitud es el único que es independiente de la escala de medición, da los mismos resultados usando la matriz de correlaciones que la de covarianzas.

0.6. Consideraciones a tomar en cuenta

Los datos están contenidos en la matriz S de tamaño $p \times p$ que contiene $p(p+1)/2$ elementos no redundantes.

Para cada x_i hay que determinar m elementos de Λ_{ij} y un elemento ψ_i

Tambien se introducen $m * (m - 1)$ condiciones para que la matriz $\Lambda'\Psi^{-1}\Lambda$ tenga ceros como elementos fuera de la diagonal.

Entonces para que el modelo sea estimable debe ocurrir que

$p(p+1)/2 \geq pm + p - 1/2 * m * (m - 1)$ En la tabla se muestra que ocurre con distintos valores para p .

Número de variables	3	4	5	6	7	8	9	10	15	20
Máximo número de factores	1	1	2	3	3	4	5	6	10	14

0.6.1. Rotación de factores

Si se tiene una matriz de rotación H esta cumple con $HH' = I$.

Si se rota el factor f como $f^* = Hf$ y se hace $\Lambda^* = \Lambda H'$, resulta que $\Lambda^* f^* = \Lambda H' H f = \Lambda f$, esta f^* cumple con $E(f^*) = H E(f) = 0$ y $Var(f^*) = H Var(f) H' = H I H' = H H' = I$ de manera que el nuevo modelo $x = \Lambda^* f^* + e$ es indistinguible del original $x = \Lambda f + e$.

En muchas ocasiones se usa la rotación para lograr una mejor interpretación de los factores comunes, y sucede que las comunalidades y especificidades se MANTIENEN SIN CAMBIO.

0.6.2. Inconvenientes

1. Hay que decidir cuantos factores hay.
2. Hay limitaciones en cuanto al número de factores según el número de variables originales.
3. Los factores comunes NO son únicos.
4. Si aumenta el tamaño de muestra también aumenta el número de estimadores de factores comunes f_{ij} , m para cada sujeto.