

ANÁLISIS DE DATOS MULTIVARIADOS: ANÁLISIS GENERAL

JESSICA FERNÁNDEZ GARZA

T E S I S

Presentada como requisito parcial
para obtener el grado de
Maestro en Ciencias
en Estadística Experimental

Universidad Autónoma Agraria

Antonio Narro

SUBDIRECCIÓN POSTGRADO

Buenavista, Saltillo, Coah.

Diciembre de 2004



UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO

Subdirección Postgrado

ANÁLISIS DE DATOS MULTIVARIADOS: ANÁLISIS GENERAL

TESIS

Por:

JESSICA FERNÁNDEZ GARZA

Elaborada bajo la supervisión del comité particular de asesoría y aprobada como requisito parcial, para optar al grado de

MAESTRO EN CIENCIAS

EN ESTADÍSTICA EXPERIMENTAL

Comité Particular

Asesor principal: _____

Dr. José Antonio Díaz García

Asesor: _____

M.C. Félix de Jesús Sánchez Pérez

Asesor: _____

M.C. Luis Rodríguez Gutiérrez

Dr. Jerónimo Landeros Flores

Subdirector de Postgrado

Buenvista, Saltillo, Coahuila. Diciembre de 2004

AGRADECIMIENTOS

Deseo expresar mis agradecimientos:

- A Dios por la salud y la vida, así como por guiarme a través de caminos inesperados y fortuitos, siempre bajo su protección.
- A CONACYT por brindarme la invaluable oportunidad de superarme profesionalmente.
- A la Universidad Autónoma Agraria Antonio Narro por ser parte de mi formación académica.
- Al Dr. José Antonio Díaz García por la idea de este proyecto, por su calidad profesional y principalmente por su calidad humana.
- A los asesores M.C. Félix de Jesús Sánchez Pérez y M.C. Luis Rodríguez Gutiérrez por sus aportaciones a este trabajo.
- A mis padres que me enseñaron a proponerme metas, pero sobretodo a cumplirlas.
- A Enrique Martínez Alvarado, por ser la constante voz en mi vida, que me alienta a la superación.

DEDICATORIA

A mis padres,

dos personas, parte de la gran mayoría, que no comprenden lo plasmado en este documento, pero que me han brindado su constante apoyo aún a pesar de lo incomprensible que parezcan mis decisiones.

Gracias.

COMPENDIO

Análisis de Datos Multivariados: Análisis General

POR

JESSICA FERNÁNDEZ GARZA

MAESTRÍA

ESTADÍSTICA EXPERIMENTAL

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA. Diciembre de 2004

Dr. José Antonio Díaz García -Asesor-

Palabras claves: Análisis Multivariado, Análisis Multivariado Clásico, Análisis Multivariado Latente, Análisis de Componentes Principales, Análisis de Componentes Principales Normado, Análisis de Correspondencias Simple, Descomposición en Valores Singulares.

Una de las líneas de investigación de mayor desarrollo del Análisis Multivariado, surgida en las últimas décadas, es el Análisis de Datos Multivariados. Sus métodos contribuyen a describir grandes y complicados conjuntos de datos multivariados, permitiendo obtener representaciones simplificadas que facilitan la interpretación.

En la actualidad existe una gran cantidad de métodos multivariados, cada uno de los cuales es estudiado por separado y en la mayoría de los casos sin una conexión entre ellos. Varios de estos métodos comparten un fundamento matemático en común que permite estudiarlos

partiendo desde un mismo punto y no de forma aislada, como se ha hecho hasta ahora. Esta base teórica en común es conocida como el Análisis General (AG) y utiliza criterios de álgebra matricial. En particular el Análisis de Componentes Principales (ACP) y el Análisis de Correspondencias Simple (ACS) pueden ser obtenidos a partir del AG.

Se inicia el documento presentando los conceptos básicos de álgebra requeridos para la comprensión de los métodos multivariados que se abordarán. Después se describe el procedimiento matemático para obtener el ACP y el ACS. Se ilustra cada método con un ejemplo.

Finalmente se presentan las ideas algebraicas en las que se fundamenta el AG y se menciona bajo que circunstancias el ACP y el ACS se obtienen como casos particulares del mismo, además se retoman los ejemplos antes realizados para comprobar que se obtienen los mismos resultados.

ABSTRACT

Multivariate Data Analysis: General Analysis

BY

JESSICA FERNÁNDEZ GARZA

MASTER

EXPERIMENTAL STATISTICS

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA. December, 2004

Ph. D. José Antonio Díaz García -Advisor-

Key Words: Multivariate Analysis, Classic Multivariate Analysis, Latent Multivariate Analysis, Principal Components Analysis, Normalized Principal Components Analysis, Simple Correspondence Analysis, Singular Value Decomposition.

One of the recent investigation line with an enormous development is the Multivariate Data Analysis. Its methods contribute to describe a complicate set of multivariate data, let us get simple representations that help us comprehend the interpretation.

Actually, there are many multivariate methods, each of them is separately study, and in most of the cases they stabilize without connection. Many of these methods have a common mathematical fundament, which let us study them beginning in the same point, instead they go a separate way. It common basis theoretical is known as the General Analysis (AG) and it

uses algebraic ideas. Particular, the Principal Components Analysis (ACP) and the Simple Correspondence Analysis (ACS) can be found by using AG.

We start the document with the basic algebraic concepts, in order to understand the multivariate methods that we will be studying. Then, we describe the mathematical process to achieve ACP and ACS. There is an example for each of the methods.

Finally, we give the fundamental ideas of General Analysis, and show when we find the Principal Components Analysis and Simple Correspondence Analysis using General Analysis. Again, we do the example to show that the results are the same.

ÍNDICE

ÍNDICE DE FIGURAS.....	xi
ÍNDICE DE CUADROS.....	xii
I. INTRODUCCIÓN.....	1
II. CONCEPTOS PRELIMINARES.....	6
Introducción.....	6
Vectores y matrices.....	6
Vectores y matrices ortogonales.....	11
Eigenvalores y eigenvectores.....	13
Derivadas de formas cuadráticas y lineales.....	15
Matrices no negativas.....	17
Esperanza y covarianza de vectores y matrices aleatorias.....	18
III. ANÁLISIS DE COMPONENTES PRINCIPALES.....	21
Introducción.....	21
Descripción del Análisis de Componentes Principales.....	21
Cálculo de las componentes.....	23
Propiedades de las componentes.....	30
Análisis de Componentes Principales Normado.....	33
Interpretación de las componentes.....	34
Ejemplo.....	35
IV. ANÁLISIS DE CORRESPONDENCIAS.....	47
Introducción.....	47
Representación en el Análisis de Correspondencias.....	47
Representación de filas.....	49
Representación de las columnas.....	55
Relación entre ambas representaciones.....	56
Distancia χ^2 y sus propiedades.....	58
Prueba de significación de los ejes factoriales.....	61
Interpretación.....	62

Ejemplo.....	63
V. ANÁLISIS GENERAL.....	68
Introducción.....	68
Descomposición en valores singulares.....	68
Aproximación de una matriz por otra de menor rango.....	75
Aplicación al Análisis de Componentes Principales.....	81
Ejemplo.....	82
Aplicación al Análisis de Correspondencias Simples.....	86
Ejemplo.....	89
Comentarios finales.....	91
LITERATURA CITADA.....	93

ÍNDICE DE FIGURAS

No. de Figura	Pág.
3.1 Proyección de los n puntos.....	33
3.2 Componentes principales y la elipse.....	35
3.3 Porcentaje de varianza explicada al incluir componentes (ACP).....	39
3.4 Coeficientes de las 5 primeras componentes (ACP).....	40
3.5 Porcentaje de varianza explicada al incluir componentes (ACPN).....	43
3.6 Coeficientes de las primeras 5 componentes (ACPN).....	45
4.1 Representación gráfica de la tabla de contingencia.....	67

ÍNDICE DE CUADROS

No. de Cuadro	Pág.
3.1 Comandos para obtener el ACP en S-PLUS.....	36
3.2 Resultados en S-PLUS para el ACP.....	37
3.3 Resumen de las dos componentes principales.....	41
3.4 Orden de los estados de acuerdo a las componentes principales.....	42
3.5 Comandos para obtener el ACPN en S-PLUS.....	42
3.6 Resultados en S-PLUS para el ACPN.....	43
3.7 Resumen de las tres componentes principales.....	46
4.1 Descomposición de la χ^2	62
4.2 Tabla de contingencia.....	63
4.3 Comandos para obtener el ACS en S-PLUS.....	65
4.4 Resultados del ACS en S-PLUS.....	66
5.1 Comandos para obtener el ACP usando el Análisis General.....	83
5.2 Resultados del ACP usando el Análisis General.....	83
5.3 Comandos para obtener el ACPN usando el Análisis General.....	84
5.4 Resultados del ACPN usando el Análisis General.....	85
5.5 Comandos para obtener el ACS usando el Análisis General.....	89
5.6 Resultados del ACS usando el Análisis General.....	90

I. INTRODUCCIÓN

El Análisis Multivariado ha adquirido en la actualidad un desarrollo impresionante, su aplicación en diversos campos científicos ha sido posible gracias al avance en métodos computacionales.

Dado el espectacular desarrollo teórico y práctico de este campo no es claro que puede entenderse hoy como Análisis Multivariado, para delimitar el contenido del Análisis Multivariado es necesario un análisis sobre sus principales líneas de investigación, algunas en las cuales se tiene especial interés en este documento son Análisis Multivariado Clásico, Análisis Multivariado Latente y Análisis de Datos Multivariados.

El Análisis Multivariado surge en lo que hoy se conoce como el Análisis Multivariado Clásico que puede ser bien definido como: el análisis que pretende extender al caso de observaciones de varias variables, las ideas y procedimientos estadísticos que han probado su efectividad en la práctica en el caso univariado (Sibson,1984).

Ejemplo del desarrollo teórico de esta etapa son la distribuciones Normal Multivariada, Wishart y el estadístico T^2 de Hotelling, que pueden considerarse extensiones de la Normal Univariada, χ^2 y del estadístico t de Student, respectivamente. En ocasiones la extensión no se puede dar o se da en varias formas o bien el caso multivariado no tiene equivalencia en el caso univariado.

En años recientes se han desarrollado notablemente los modelos multivariados con variables latentes. La estructura común teórica que liga los modelos que usan variables latentes es conocida como Análisis Multivariado Latente.

Desde un punto de vista estadístico los modelos latentes pretenden explicar las propiedades estadísticas de las variables observables en términos de las variables latentes. Las variables latentes son aquellas que no pueden ser observadas ni siquiera en la población, solamente tenemos su rastro a través del efecto que causan en variables que si son observables.

La estructura que liga las variables originales con las latentes, en los modelos latentes, es lineal. Dos ejemplos de estos modelos son el Análisis de Componentes Principales (ACP) y el Análisis Factorial.

Otra línea de desarrollo de gran interés, surgida en las últimas décadas, es el Análisis de Datos Multivariados, que está contribuyendo en abordar la problemática de trabajar con datos multivariados. Los objetivos del Análisis de Datos Multivariados son:

1. Resumir los datos mediante un pequeño conjunto de nuevas variables, con propiedades deseables, construidas como transformaciones de las originales, con la mínima pérdida de información.
2. Identificar grupos que pueden existir en los datos.
3. Clasificar nuevas observaciones en grupos definidos a priori.
4. Relacionar dos conjuntos de variables.
5. Detectar posibles outliers o datos atípicos.

Los datos multivariados son recogidos a menudo solamente porque es más rápido que observar una variable, que puede ser incluso desconocida no observable o ser causada por otras muchas, pero analizar una base de datos con una gran cantidad de variables no es sencillo.

Los métodos multivariados son muy útiles para describir conjuntos grandes y complicados, permitiendo obtener representaciones simplificadas que facilitan la interpretación de la información.

Los métodos multivariados pueden ser clasificados en varias formas, una de ellas es clasificarlos en exploratorios y de inferencia. En los primeros se extrae la información que contienen los datos disponibles. Los segundos permiten obtener conclusiones sobre la población que ha generado los datos, lo que requiere la construcción de un modelo que explique su generación y permita prever los datos futuros (Peña, 2002).

Entre los métodos multivariados exploratorios se encuentran: ACP, Escalado Multidimensional, Análisis de Correspondencias Simples (ACS), entre otros. Dos de los métodos de inferencia más importantes son el Análisis Factorial y Análisis Discriminante.

El ACP utiliza un procedimiento matemático para transformar el conjunto de variables originales y crear nuevas, conocidas como componentes principales, con la propiedad de no estar correlacionadas.

El ACS se emplea en el análisis de tablas de contingencia, permitiendo obtener una representación gráfica de la misma, este método puede verse de forma análoga al Análisis de Componentes Principales pero para datos cualitativos.

El Análisis Factorial crea nuevas variables que resuman toda la información que podría contenerse en las variables originales, se utiliza para estudiar las relaciones entre variables del conjunto de datos.

Análisis Discriminante, se usa principalmente para clasificar individuos o unidades experimentales en dos o más poblaciones definidas de manera única.

De especial interés son el ACP y el ACS, que por ser métodos multivariados exploratorios tienden a motivar hipótesis más que probarlas.

El ACP y el ACS pueden ser vistos de diferentes puntos de vista, una presentación de estos métodos desde un punto de vista descriptivo puede ser consultada en Peña (2002) y Cuadras

(1991). Una excelente presentación del ACS es realizada por Greenacre (1994). Ambos métodos, enfatizando el aspecto geométrico, son abordados en Escofier (1990). Una forma alternativa de obtener el ACP y el ACS es a través del Análisis General (AG), una aplicación del AG para obtener el ACP puede ser consultada en Lebart (1984).

El AG utiliza criterios de álgebra matricial, tales como la descomposición en valores singulares de una matriz (SVD) y la aproximación por mínimos cuadrados de una matriz por otra de menor rango, con el fin de representar la información observada en una dimensión menor.

En este documento se describen los métodos multivariados ACP y ACS, y se incluye el Análisis General. Este último conforma la base teórica de algunos métodos multivariados, en particular del ACP y del ACS.

Objetivo del presente documento es mostrar de manera formal, para después ejemplificar, que ambos métodos multivariados, ACP y ACS, pueden ser obtenidos como casos particulares del Análisis General.

La estructura del documento consta de cuatro capítulos. En el primero se abordan conceptos preliminares de álgebra que serán útiles para la comprensión de los capítulos posteriores.

En el segundo capítulo se describe de forma detallada el ACP, se presenta el procedimiento matemático para calcular las componentes principales, sus propiedades, identificar la cantidad de componentes a incluir y su interpretación. Además se puntualiza en que momento utilizar el Análisis de Componentes Principales Normado (ACPN). Se concluye con un ejemplo.

El ACS se presenta en el tercer capítulo, se desarrolla hasta llegar al punto en que puede resolverse de forma análoga a Componentes Principales, también se concluye el capítulo con un ejemplo.

Finalmente, el último capítulo presenta el fundamento teórico del AG, la SVD y la aproximación de una matriz por otra, después se indica bajo que circunstancias se obtiene el ACP

y el ACS como casos particulares del Análisis General, se retoman los ejemplos presentados en los capítulos anteriores para observar que se llega a los mismos resultados utilizando el AG. Se concluye el trabajo con algunos comentarios finales.

II. CONCEPTOS PRELIMINARES

Introducción

Al analizar un conjunto de datos multivariados es conveniente representarlos como un arreglo matricial, por lo que en este capítulo se revisan conceptos sobre álgebra, que serán de gran utilidad cuando se aborden los métodos multivariados a estudiar en este documento.

Se inicia el capítulo citando conceptos muy básicos, relacionados con vectores y matrices, para después incluir temas como, vectores y matrices ortogonales, eigenvalores y eigenvectores, derivadas de formas cuadráticas y lineales, matrices no negativas, la última sección del capítulo hace referencia a la esperanza y covarianza de vectores y matrices aleatorias.

Vectores y matrices

Cuando se estudia alguna observación generalmente se consideran varias características de ésta, por ejemplo, si se está interesado en un individuo se pueden recolectar datos de su edad, estatura, peso, entre otros. Algebraicamente estos datos se pueden representar utilizando un vector, el cual se define a continuación.

Definición 1. (*Vector*). *Conjunto de n números ordenados en una columna, esto es, una n -upla de la siguiente forma:*

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (2.1)$$

Los vectores serán denotados por letras minúsculas y negritas como, $\mathbf{v}, \mathbf{x}, \mathbf{y}$, mientras que para las matrices se utilizarán letras mayúsculas y negritas, $\mathbf{A}, \mathbf{B}, \mathbf{X}$.

Geoméricamente un vector puede verse como un segmento dirigido o flecha en el espacio \mathbb{R}^n , gráficamente, un vector representa un punto en el espacio de dimensión n .

Los vectores poseen características geométricas de dirección y longitud, la primera se determina por el ángulo del vector respecto al primer eje, mientras que la longitud se obtiene por el teorema de pitágoras, se define como:

Definición 2. (Norma cuadrática). *La longitud o norma cuadrática de un vector, que inicia en el origen es*

$$L_{\mathbf{x}} = \|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \quad (2.2)$$

La operación de multiplicar dos vectores, de tal forma que se obtiene un escalar, es conocida como producto interno y se presenta a continuación.

Definición 3. (Producto interno). *Sean \mathbf{x} y \mathbf{y} dos vectores $\in \mathbb{R}^n$, el producto interno es*

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x} = \sum_{i=1}^n x_i y_i \quad (2.3)$$

Observe que la longitud de un vector es la raíz cuadrada del producto interno consigo mismo.

Cuando un vector se puede expresar en términos lineales de otro vector se dice que los vectores son linealmente dependientes.

Definición 4. (Dependencia e independencia lineal). *Sea \mathbf{V} un conjunto de n vectores, $\mathbf{V} = \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. Se dice que los vectores son linealmente dependientes si existen n escalares $\alpha_1, \alpha_2, \dots, \alpha_n$, no todos cero tales que*

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_n \mathbf{v}_n = 0 \quad (2.4)$$

Si los vectores no son linealmente dependientes, entonces son linealmente independientes.

Definición 5. (Combinación lineal). *Cualquier vector, \mathbf{y} , de la forma $\mathbf{y} = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \dots + \alpha_n\mathbf{x}_n$ es una combinación lineal de los vectores $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$.*

Durante este documento frecuentemente se hablará de espacios y subespacios vectoriales, por lo que es conveniente dar la definición formal de estos conceptos.

Definición 6. (Espacio vectorial). *Un conjunto no vacío \mathbf{V} de vectores es un espacio vectorial si tenemos definidas dos operaciones:*

1. Operación de adición. *Si $\mathbf{x}, \mathbf{y} \in \mathbf{V}$, entonces $\mathbf{x} + \mathbf{y} \in \mathbf{V}$.*
2. Operación de multiplicación por un escalar. *Si $\mathbf{x} \in \mathbf{V}$, entonces $\alpha\mathbf{x} \in \mathbf{V}$, donde α es un escalar.*

Satisfaciendo los siguientes axiomas:

- *La adición es asociativa, $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{V}$, tal que $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.*
- *La adición es conmutativa, $\mathbf{x}, \mathbf{y} \in \mathbf{V}$, tal que $(\mathbf{x} + \mathbf{y}) = (\mathbf{y} + \mathbf{x})$.*
- *El vector $0 \in \mathbf{V}$.*
- *Si $\mathbf{x} \in \mathbf{V}$, entonces existe $-\mathbf{x} \in \mathbf{V}$.*
- *Si $\mathbf{x}, \mathbf{y} \in \mathbf{V}$, y α es un escalar, entonces $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$.*
- *Si $\mathbf{x} \in \mathbf{V}$, y α, β son escalares, entonces $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$.*
- *Si $\mathbf{x} \in \mathbf{V}$, y α, β son escalares, entonces $(\alpha\beta)\mathbf{x} = \alpha(\beta\mathbf{x})$.*
- *Para cada $\mathbf{x} \in \mathbf{V}$, $1\mathbf{x} = \mathbf{x}$.*

Definición 7. (Subespacio). *Un subconjunto \mathbf{W} de un espacio vectorial \mathbf{V} , se llama subespacio vectorial de \mathbf{V} , si \mathbf{W} es un espacio vectorial bajo las operaciones de adición y multiplicación por un escalar definidas en \mathbf{V} .*

El vector es un caso especial de matrices, como se puede ver en la definición siguiente.

Definición 8. (Matriz). *Es un arreglo rectangular de elementos, llamados escalares y denotados como a_{ij} , donde i se refiere al renglón y j se refiere a la columna de la matriz.*

$$\mathbf{A} = [a_{ij}]$$

Si los renglones de la matriz \mathbf{A} son intercambiados de tal forma que se colocan como las columnas de otra matriz, que denotaremos como \mathbf{A}^t , la matriz resultante es llamada transpuesta de la matriz \mathbf{A} . Si \mathbf{A} tiene tamaño $n \times m$, entonces la matriz \mathbf{A}^t tiene tamaño $m \times n$. Note que el producto interno también puede ser denotado como, $\mathbf{x}^t\mathbf{y}$ ó $\mathbf{y}^t\mathbf{x}$.

Algunas propiedades al transponer una matriz se citan a continuación, sean \mathbf{A} y \mathbf{B} matrices, y α, β escalares, entonces

1. $(\alpha\mathbf{A})^t = (\mathbf{A}\alpha)^t = \mathbf{A}^t\alpha = \alpha\mathbf{A}^t$.
2. $(\mathbf{A}^t)^t = \mathbf{A}$.
3. $(\alpha\mathbf{A} + \beta\mathbf{B})^t = \alpha\mathbf{A}^t + \beta\mathbf{B}^t$.

En la primer propiedad se observa que la transpuesta de un escalar es el mismo escalar.

Cuando una matriz cuenta con la misma cantidad de renglones y columnas es llamada matriz cuadrada, y podemos obtener su traza.

Definición 9. (Traza de una matriz). Sea $\mathbf{A} = [a_{ij}]$ una matriz cuadrada $n \times n$. La traza de la matriz \mathbf{A} , denotada por $\text{tr}(\mathbf{A})$, es la suma de los elementos de la diagonal principal, esto es

$$\sum_{i=1}^n a_{ii} \quad (2.5)$$

Algunas propiedades de la traza de una matriz son:

1. $\text{tr}(c\mathbf{A}) = c \text{tr}(\mathbf{A})$
2. $\text{tr}(\mathbf{A} \pm \mathbf{B}) = \text{tr}(\mathbf{A}) \pm \text{tr}(\mathbf{B})$
3. $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$
4. $\text{tr}(\mathbf{B}^{-1}\mathbf{AB}) = \text{tr}(\mathbf{A})$
5. $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^t)$
6. $\text{tr}(\mathbf{AA}^t) = \text{tr}(\mathbf{A}^t\mathbf{A}) = \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2$

Una matriz puede ser expresada como un vector mediante el operador vec . Veamos como:

Definición 10. (Vectorización de una matriz). Sea $\mathbf{A} \in \mathbb{R}^{m \times n}$, tal que $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_n]$ la vectorización de \mathbf{A} se obtiene colocando bajo la primer columna la segunda y así sucesivamente, el vector resultante es de orden mn , se denota como $\text{vec } \mathbf{A}$.

$$\text{vec } \mathbf{A} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix} \quad (2.6)$$

Este operador tiene varias propiedades, una de ellas es

$$\begin{aligned} \text{tr}(\mathbf{A}^t \mathbf{B}) &= \sum_{i=1}^n (\mathbf{A}^t \mathbf{B})_{ii} = \sum_{i=1}^n \mathbf{e}_i^t (\mathbf{A}^t \mathbf{B}) \mathbf{e}_i = \sum_{i=1}^n (\mathbf{A} \mathbf{e}_i)^t (\mathbf{B} \mathbf{e}_i) \\ &= \begin{bmatrix} \mathbf{a}_1^t & \mathbf{a}_2^t & \dots & \mathbf{a}_n^t \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} = \text{vec } \mathbf{A}^t \text{vec } \mathbf{B} \end{aligned} \quad (2.7)$$

Definición 11. (Norma euclidiana de una matriz). Sea $\mathbf{A} \in \mathbb{R}^{m \times n}$, entonces la $\|\mathbf{A}\|^2 = \text{tr}(\mathbf{A}^t \mathbf{A})$.

Pues observe que

$$\|\text{vec } \mathbf{A}\|^2 = \text{vec } \mathbf{A}^t \text{vec } \mathbf{A} = \text{tr}(\mathbf{A}^t \mathbf{A}) = \|\mathbf{A}\|^2 \quad (2.8)$$

Definición 12. (Matriz simétrica). Sea $\mathbf{A} \in \mathbb{R}^{n \times n}$ tal que $\mathbf{A} = \mathbf{A}^t$, esto es $a_{ij} = a_{ji}$ para todo i, j , entonces se dice que \mathbf{A} es una matriz simétrica.

Un caso especial de una matriz simétrica es la matriz identidad formada por unos en la diagonal principal y el resto de sus elementos son ceros, se denota por la letra \mathbf{I} .

Finalmente se concluye la sección con un concepto más, el rango de una matriz.

Definición 13. (Rango de una matriz). Es el número de líneas o vectores de una matriz (filas o columnas) que son linealmente independientes.

Una matriz es de rango completo cuando el número de renglones o columnas es igual a su rango.

Vectores y matrices ortogonales

Los vectores y matrices ortogonales o perpendiculares son utilizados en importantes aplicaciones en estadística ya que tienen propiedades que facilitan trabajar con ellos. Veamos primero los vectores ortogonales.

Definición 14. (*Vector ortogonal*). Los vectores \mathbf{v} y \mathbf{x} en \mathbb{R}^n son ortogonales si y sólo si su producto interno es igual a cero,

$$\mathbf{v} \cdot \mathbf{x} = \sum_{i=1}^n v_i x_i = 0. \quad (2.9)$$

El hecho de que dos vectores sean ortogonales equivale a que son linealmente independientes, como muestra el siguiente teorema.

Teorema 1. Sea $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r)$ un conjunto de vectores diferentes de cero, en \mathbb{R}^n , tal que cada par de vectores es ortogonal, $\mathbf{v}_i^t \mathbf{v}_j = 0$ para todo $i \neq j$, entonces el conjunto de vectores es linealmente independientes.

Demostración. Suponga que $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_r \mathbf{v}_r = \mathbf{0}$. Entonces para cualquier $i = 1, 2, \dots, r$

$$\begin{aligned} 0 &= \mathbf{0} \cdot \mathbf{v}_i = (\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_r \mathbf{v}_r) \cdot \mathbf{v}_i \\ &= \alpha_1 \mathbf{v}_1 \cdot \mathbf{v}_i + \alpha_2 \mathbf{v}_2 \cdot \mathbf{v}_i + \dots + \alpha_i \mathbf{v}_i \cdot \mathbf{v}_i + \dots + \alpha_r \mathbf{v}_r \cdot \mathbf{v}_i \\ &= \alpha_1 0 + \alpha_2 0 + \dots + \alpha_i \|\mathbf{v}_i\|^2 + \dots + \alpha_r 0 \\ &= \alpha_i \|\mathbf{v}_i\|^2 \end{aligned} \quad (2.10)$$

Como $\mathbf{v}_i \neq \mathbf{0}$ por hipótesis, $\|\mathbf{v}_i\|^2 > 0$, y se tiene que $\alpha_i = 0$. Esto es cierto para $i = 1, 2, \dots, r$. ■

Definición 15. (*Matriz ortogonal*). Sea \mathbf{P} una matriz cuadrada, tal que $\mathbf{P}^t = \mathbf{P}^{-1}$, entonces \mathbf{P} es una matriz ortogonal.

Teorema 2. Sea \mathbf{P} una matriz $n \times n$, tal que $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n]$, donde \mathbf{p}_i es el vector columna $n \times 1$. La condición necesaria y suficiente para que \mathbf{P} sea ortogonal es

$$\mathbf{p}_i^t \mathbf{p}_i = 1 \quad \text{para } i = 1, 2, \dots, n. \quad (2.11)$$

$$\mathbf{p}_i^t \mathbf{p}_j = 0 \quad \text{para } i = 1, 2, \dots, n; j = 1, 2, \dots, n; i \neq j \quad (2.12)$$

Demostración. Note que si \mathbf{P} es ortogonal, entonces $\mathbf{P}^t \mathbf{P} = \mathbf{P} \mathbf{P}^t = \mathbf{I}$, luego

$$\begin{aligned} \mathbf{P}^t \mathbf{P} &= \begin{bmatrix} \mathbf{p}_1^t \\ \mathbf{p}_2^t \\ \vdots \\ \mathbf{p}_n^t \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \dots & \mathbf{p}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{p}_1^t \mathbf{p}_1 & \mathbf{p}_1^t \mathbf{p}_2 & \dots & \mathbf{p}_1^t \mathbf{p}_n \\ \mathbf{p}_2^t \mathbf{p}_1 & \mathbf{p}_2^t \mathbf{p}_2 & \dots & \mathbf{p}_2^t \mathbf{p}_n \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{p}_n^t \mathbf{p}_1 & \mathbf{p}_n^t \mathbf{p}_2 & \dots & \mathbf{p}_n^t \mathbf{p}_n \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \end{aligned} \quad (2.13)$$

Entonces $\mathbf{p}_i^t \mathbf{p}_i = 1$ para $i = 1, 2, \dots, n$ y $\mathbf{p}_i^t \mathbf{p}_j = 0$ para $i, j = 1, 2, \dots, n$, cuando $i \neq j$. \blacksquare

Definición 16. (Matriz diagonalizable). Se dice que una matriz \mathbf{A} de $n \times n$ es diagonalizable si existe una matriz ortogonal \mathbf{P} tal que

$$\mathbf{P}^t \mathbf{A} \mathbf{P} = \mathbf{D} \quad (2.14)$$

donde $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ y $\lambda_1, \lambda_2, \dots, \lambda_n$ son los eigenvalores de \mathbf{A} (ver la siguiente sección).

Teorema 3. Sea \mathbf{A} una matriz $n \times n$, entonces \mathbf{A} es diagonalizable si y sólo si \mathbf{A} es simétrica.

Demostración. Utilizando la definición anterior $\mathbf{P}^t \mathbf{A} \mathbf{P} = \mathbf{D}$, premultiplicando por \mathbf{P} y postmultiplicando por \mathbf{P}^t , se tiene que:

$$\mathbf{A} = \mathbf{P}^t \mathbf{D} \mathbf{P} \quad (2.15)$$

observe que $\mathbf{A}^t = (\mathbf{P}^t \mathbf{D} \mathbf{P})^t = \mathbf{P}^t \mathbf{D} \mathbf{P} = \mathbf{A}$, entonces \mathbf{A} es simétrica. ■

Eigenvalores y eigenvectores

Los eigenvalores son las medidas básicas de tamaño de una matriz, que no se ven alteradas si hacemos un cambio de coordenadas. Las medidas globales de tamaño de una matriz, como la traza o el determinante, son función de los eigenvalores y, en consecuencia, serán también invariantes ante las transformaciones que preservan los eigenvalores.

Definición 17. (Eigenvalor y eigenvector). Sea \mathbf{A} una matriz de orden $n \times n$. El número λ (real o complejo) se llama eigenvalor de \mathbf{A} si existe un vector diferente de cero \mathbf{v} en \mathbb{R}^n tal que:

$$\mathbf{A} \mathbf{v} = \lambda \mathbf{v} \quad (2.16)$$

El vector $\mathbf{v} \neq 0$ se llama eigenvalor de \mathbf{A} correspondiente al eigenvalor λ .

Los eigenvalores también se conocen como valores propios o valores característicos y los eigenvectores reciben el nombre de vectores propios o vectores característicos.

Suponga que λ es un valor propio de \mathbf{A} . Entonces existe un vector diferente de cero \mathbf{v} , tal que $\mathbf{A} \mathbf{v} = \lambda \mathbf{v} = \lambda \mathbf{I} \mathbf{v}$. Esto es

$$(\mathbf{A} - \lambda \mathbf{I}) \mathbf{v} = \mathbf{0} \quad (2.17)$$

Si \mathbf{A} es una matriz $n \times n$, la expresión anterior corresponde a un sistema homogéneo de n ecuaciones, como se ha supuesto que el sistema no tiene soluciones triviales, entonces $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$. Denotando $\mathbf{p}(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$, se tiene que los eigenvalores se encuentran al resolver la siguiente expresión:

$$\mathbf{p}(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}) = 0 \quad (2.18)$$

Definición 18. (Ecuación y polinomio característicos). La ecuación anterior se llama ecuación característica de \mathbf{A} ; $\mathbf{p}(\lambda)$ se conoce como el polinomio característico de \mathbf{A} .

El polinomio $\mathbf{p}(\lambda)$ es de grado n en λ , tiene exactamente n raíces, contando multiplicidades. Una vez encontrados los eigenvalores, las componentes de los eigenvectores se obtienen al sustituir el valor del eigenvalor y resolver el sistema de ecuaciones (2.17).

Los eigenvectores de una matriz no tienen porque ser únicos, en ocasiones existen un número infinito de formas de elegirlos. Además los eigenvalores pueden ser complejos. En matrices simétricas los eigenvalores son siempre reales y los eigenvectores se pueden elegir ortogonales.

Teorema 4. Sea $\mathbf{A} \in \mathbb{R}^n$, tal que $\mathbf{A} = \mathbf{A}^t$, entonces la suma de los eigenvalores de \mathbf{A} es igual a la traza de \mathbf{A} .

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i \quad (2.19)$$

Demostración. La matriz \mathbf{A} puede ser factorizada como

$$\mathbf{A} = \mathbf{P}^t \mathbf{D} \mathbf{P} \quad (2.20)$$

Entonces

$$\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{P}^t \mathbf{D} \mathbf{P}) \quad (2.21)$$

Recordando la tercera propiedad de la traza de una matriz, citada en la sección 2.2.

$$\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{P}^t \mathbf{D} \mathbf{P}) = \text{tr}(\mathbf{P} \mathbf{P}^t \mathbf{D}) = \text{tr}(\mathbf{D}) = \sum_{i=1}^n \lambda_i \quad (2.22)$$

■

Teorema 5. Sean \mathbf{x}_1 y \mathbf{x}_2 eigenvectores de \mathbf{A} correspondientes a los eigenvalores λ_1 y λ_2 , respectivamente, donde $\mathbf{A} \in \mathbb{R}^{n \times n}$, tal que $\mathbf{A} = \mathbf{A}^t$. Si $\lambda_1 \neq \lambda_2$ entonces \mathbf{x}_1 y \mathbf{x}_2 son ortogonales.

Demostración.

$$(\mathbf{A} \mathbf{x}_1)^t \mathbf{x}_2 = (\lambda_1 \mathbf{x}_1)^t \mathbf{x}_2 = \lambda_1 \mathbf{x}_1^t \mathbf{x}_2 \quad (2.23)$$

Dado que $\mathbf{A} = \mathbf{A}^t$.

$$(\mathbf{A} \mathbf{x}_1)^t \mathbf{x}_2 = \mathbf{x}_1^t \mathbf{A} \mathbf{x}_2 = \mathbf{x}_1^t \lambda_2 \mathbf{x}_2 = \lambda_2 \mathbf{x}_1^t \mathbf{x}_2 \quad (2.24)$$

Luego

$$\begin{aligned}\lambda_1 \mathbf{x}_1^t \mathbf{x}_2 &= \lambda_2 \mathbf{x}_1^t \mathbf{x}_2 \\ (\lambda_1 - \lambda_2) \mathbf{x}_1^t \mathbf{x}_2 &= 0\end{aligned}\tag{2.25}$$

El producto interno $\mathbf{x}_1^t \mathbf{x}_2 = 0$, ya que por hipótesis $\lambda_1 \neq \lambda_2$, entonces \mathbf{x}_1 y \mathbf{x}_2 son ortogonales. ■

Derivadas de formas cuadráticas y lineales

Sea \mathbf{x} un vector columna $n \times 1$, $\mathbf{A} \in \mathbb{R}^{n \times n}$, y $f: \mathbb{R}^n \rightarrow \mathbb{R}$ con

$$f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x}\tag{2.26}$$

Donde $\mathbf{A} = [a_{ij}]$ es una matriz de constantes, entonces f es llamada una función cuadrática.

Definición 19. (Forma cuadrática). La función f antes descrita es una forma cuadrática, la matriz \mathbf{A} es conocida como la matriz de la forma cuadrática.

Teorema 6. La matriz de una forma cuadrática siempre se puede seleccionar simétrica.

Demostración.

$$f(\mathbf{x}) = \frac{\mathbf{x}^t \mathbf{A} \mathbf{x} + \mathbf{x}^t \mathbf{A}^t \mathbf{x}}{2} = \mathbf{x}^t \left[\frac{\mathbf{A} + \mathbf{A}^t}{2} \right] \mathbf{x} = \mathbf{x}^t \mathbf{B} \mathbf{x}\tag{2.27}$$

Donde $\mathbf{B} = \frac{\mathbf{A} + \mathbf{A}^t}{2} = \mathbf{B}^t = \frac{\mathbf{A}^t + \mathbf{A}}{2}$, es simétrica. ■

Definición 20. (Gradiente). Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}$. El gradiente de f en el punto $\mathbf{x} \in \mathbb{R}^n$

$$\begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}\tag{2.28}$$

y se denota como $\frac{\partial f}{\partial \mathbf{x}}$. El gradiente de f también es llamado, derivada de f con respecto a \mathbf{x} .

Observe que dada la función $f: \mathbb{R}^n \rightarrow \mathbb{R}$, su diferencial total se define como

$$\begin{aligned}
 df &= \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \dots + \frac{\partial f}{\partial x_n} dx_n \\
 &= \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \dots & \frac{\partial f}{\partial x_n} \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix} \\
 &= \left[\frac{\partial f}{\partial \mathbf{x}} \right]^t d\mathbf{x}
 \end{aligned} \tag{2.29}$$

Si el diferencial total de la función $f: \mathbb{R}^n \rightarrow \mathbb{R}$, se puede expresar como $df = \mathbf{A}d\mathbf{x}$, entonces

$$\frac{\partial f}{\partial \mathbf{x}} = \mathbf{A}^t \tag{2.30}$$

Teorema 7. Sea $f(\mathbf{x})$ una forma cuadrática de n variables aleatorias independientes x_1, x_2, \dots, x_n definida por

$$f(\mathbf{x}) = \mathbf{x}^t \mathbf{A} \mathbf{x},$$

donde $\mathbf{A} = [a_{ij}]$ es una matriz simétrica $n \times n$ de constantes, entonces:

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A} \mathbf{x} \tag{2.31}$$

Demostración.

$$\begin{aligned}
 df &= d\mathbf{x}^t \mathbf{A} \mathbf{x} + \mathbf{x}^t \mathbf{A} d\mathbf{x} \\
 &= (d\mathbf{x}^t \mathbf{A} \mathbf{x})^t + \mathbf{x}^t \mathbf{A} d\mathbf{x} \\
 &= \mathbf{x}^t \mathbf{A}^t d\mathbf{x} + \mathbf{x}^t \mathbf{A} d\mathbf{x} \\
 &= (\mathbf{x}^t \mathbf{A}^t + \mathbf{x}^t \mathbf{A}) d\mathbf{x}
 \end{aligned} \tag{2.32}$$

Entonces

$$\frac{\partial f}{\partial \mathbf{x}} = (\mathbf{x}^t \mathbf{A}^t + \mathbf{x}^t \mathbf{A})^t = \mathbf{A} \mathbf{x} + \mathbf{A}^t \mathbf{x} = 2\mathbf{A} \mathbf{x} \tag{2.33}$$

ya que \mathbf{A} es simétrica. ■

Teorema 8. Sea $q(\mathbf{x})$ una función lineal de n variables independientes definida por $q(\mathbf{x}) = \sum_{i=1}^n a_i x_i = \mathbf{a}^t \mathbf{x} = \mathbf{x}^t \mathbf{a}$, donde $\mathbf{a}^t = [a_1, a_2, \dots, a_n]$, a_i es cualquier constante. Entonces

$$\frac{\partial q(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{a} \quad (2.34)$$

Demostración.

$$q(\mathbf{x}) = \mathbf{a}^t \mathbf{x} = \sum_{i=1}^n a_i x_i = a_1 x_1 + a_2 x_2 + \dots + a_n x_n \quad (2.35)$$

así $\frac{\partial q(\mathbf{x})}{\partial x_i} = a_i$, luego

$$\frac{\partial q(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \mathbf{a} \quad (2.36)$$

■

Matrices no negativas

Definición 21. (Matriz no negativa). Sea $\mathbf{A} \in \mathbb{R}^{n \times n}$, tal que $\mathbf{A} = \mathbf{A}^t$, entonces \mathbf{A} es no negativa si y sólo si es definida positiva o semidefinida positiva. La matriz \mathbf{A} es:

- Definida positiva si y sólo si $\mathbf{x}^t \mathbf{A} \mathbf{x} > 0$ para todo $\mathbf{x} \in \mathbb{R}^n$.
- Semidefinida positiva si y sólo si $\mathbf{x}^t \mathbf{A} \mathbf{x} \geq 0$ para todo $\mathbf{x} \in \mathbb{R}^n$, la igualdad se cumple al menos para un vector $\mathbf{x} \neq 0$.

Teorema 9. Sea \mathbf{A} una matriz simétrica $n \times n$. Una condición necesaria y suficiente para que \mathbf{A} sea:

1. Definida positiva. Los eigenvalores de \mathbf{A} son positivos.
2. Semidefinida positiva. Los eigenvalores de \mathbf{A} son positivos y al menos uno es cero.

Demostración. Si \mathbf{A} es definida positiva entonces $\mathbf{P}^t \mathbf{A} \mathbf{P} > 0$, observe que $\mathbf{A} = \mathbf{P}^t \mathbf{D} \mathbf{P}$, entonces $\mathbf{P}^t \mathbf{A} \mathbf{P} > 0$ equivale a $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) > 0$, esto es, $\lambda_i > 0$ para $i=1,2,\dots,n$.

De forma similar si \mathbf{A} es semidefinida positiva, $\mathbf{P}^t \mathbf{A} \mathbf{P} \geq 0$ equivale a $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \geq 0$, los eigenvalores son positivos y al menos uno es igual a cero. ■

Esperanza y covarianza de vectores y matrices aleatorias

Un vector aleatorio es un vector cuyos elementos son variables aleatorias, similarmente una matriz aleatoria está formada por elementos que son variables aleatorias.

La media o valor esperado de una función de probabilidad es la medida de su tendencia central, la expresión matemática es

$$E(x) = \begin{cases} \int x f(x) dx & \text{si } x \text{ es una variable aleatoria continua.} \\ \sum x p(x) & \text{si } x \text{ es una variable aleatoria discreta.} \end{cases} \quad (2.37)$$

El valor esperado de un vector es conocido como el vector de medias.

Definición 22. (Vector de medias). Sea $\mathbf{x} \in \mathbb{R}^n$ un vector aleatorio, entonces el valor esperado de \mathbf{x} , $E(\mathbf{x})$, es el valor esperado de cada una de sus componentes, donde la i -ésima componente es $E(x_i)$. Entonces, se dice que el vector $E(\mathbf{x})$ es el vector de medias.

El valor esperado de una matriz es la matriz donde el ij -ésimo elemento es el valor esperado del elemento ij . Si $\mathbf{X} \in \mathbb{R}^{n \times m}$, entonces

$$E(\mathbf{X}) = \mu = \begin{bmatrix} E(x_{11}) & E(x_{12}) & \dots & E(x_{1m}) \\ E(x_{21}) & E(x_{22}) & \dots & E(x_{2m}) \\ \vdots & \vdots & \dots & \vdots \\ E(x_{n1}) & E(x_{n2}) & \dots & E(x_{nm}) \end{bmatrix} \quad (2.38)$$

Dos importantes propiedades de la esperanza son:

1. $E(\mathbf{X} + \mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y})$
2. $E(\mathbf{A}\mathbf{X}\mathbf{B}) = \mathbf{A} E(\mathbf{X}) \mathbf{B}$, donde las matrices \mathbf{A} y \mathbf{B} son matrices de constantes.

La covarianza es una medida de asociación lineal entre dos variables, matemáticamente se expresa como:

$$\text{Cov}(x_i, x_j) = E[(x_i - \mu_{x_i})(x_j - \mu_{x_j})] = \sigma_{ij} \quad (2.39)$$

Observe que $\text{Cov}(x_i, x_i) = \sigma_i^2$.

Si dos variables aleatorias, x_i, x_j , son estadísticamente independientes entonces:

$$\text{Cov}(x_i, x_j) = 0 \quad (2.40)$$

Definición 23. (Matriz de covarianzas). Sea \mathbf{y} y \mathbf{x} vectores aleatorios en \mathbb{R}^m y \mathbb{R}^n respectivamente, entonces la matriz de covarianzas de \mathbf{y} y \mathbf{x} es la siguiente matriz $m \times n$.

$$\text{Cov}(\mathbf{x}, \mathbf{y}) = E[(\mathbf{x} - \mu_x)(\mathbf{y} - \mu_y)^t] \quad (2.41)$$

Si \mathbf{X} es una matriz $m \times n$, la matriz de varianza-covarianza es una matriz simétrica de la forma

$$\begin{aligned} \text{Cov}(\mathbf{X}) &= E[(\mathbf{X} - \mu_x)(\mathbf{X} - \mu_x)^t] \\ &= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1m} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \dots & \sigma_m^2 \end{bmatrix} \end{aligned} \quad (2.42)$$

El valor que toma la covarianza depende de la escala de medición de las variables, este problema se elimina estandarizando su valor, utilizando el coeficiente de correlación lineal. La correlación entre dos variables se obtiene como

$$\rho_{ij} = \frac{\text{Cov}(x_i, y_j)}{\sigma_{x_i} \sigma_{y_j}} \quad (2.43)$$

La matriz de correlaciones para los vectores \mathbf{y} y \mathbf{x} es:

$$\begin{aligned} \mathbf{R}[\mathbf{x}, \mathbf{y}] &= [\rho_{ij}] \\ &= \mathbf{D}_{\sigma_{x_i}} \text{Cov}[\mathbf{x}, \mathbf{y}] \mathbf{D}_{\sigma_{y_j}} \end{aligned} \quad (2.44)$$

Donde $\mathbf{D}_{\sigma_{x_i}} = \text{diag}(\frac{1}{\sigma_{x_1}}, \frac{1}{\sigma_{x_2}}, \dots, \frac{1}{\sigma_{x_n}})$ y $\mathbf{D}_{\sigma_{y_j}} = \text{diag}(\frac{1}{\sigma_{y_1}}, \frac{1}{\sigma_{y_2}}, \dots, \frac{1}{\sigma_{y_m}})$.

La correlación de la matriz $\mathbf{X} \in \mathbb{R}^{m \times n}$ se representa con la matriz

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{21} & 1 & \cdots & \rho_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{m1} & \rho_{m2} & \cdots & 1 \end{bmatrix} \quad (2.45)$$

Los siguientes resultados presentan la esperanza y varianza de una combinación lineal.

Teorema 10. Sea $\mathbf{A} \in \mathbb{R}^{m \times n}$ una matriz de constantes, además los vectores $\mathbf{y} \in \mathbb{R}^n$ y $\mathbf{b} \in \mathbb{R}^m$, donde el vector \mathbf{b} es constante, si $\mathbf{W} = \mathbf{A}\mathbf{y} + \mathbf{b}$, entonces

$$\mathbf{E}(\mathbf{W}) = \mathbf{A} \mathbf{E}(\mathbf{y}) + \mathbf{b} \quad (2.46)$$

Demostración. Sea $(a_{i1}, a_{i2}, \dots, a_{in})$ el i -ésimo renglón de la matriz \mathbf{A} y b_i la i -ésima componente de \mathbf{b} , entonces

$$w_i = \sum_j a_{ij} y_j + b_i \quad (2.47)$$

y

$$\mathbf{E}(\mathbf{W}) = \sum_j a_{ij} \mathbf{E}(y_j) + b_i \quad (2.48)$$

■

Teorema 11. Sea $\mathbf{x} \in \mathbb{R}^m$ y $\mathbf{y} \in \mathbb{R}^n$, entonces para las matrices constantes, de orden adecuado, \mathbf{A} y \mathbf{B} , se tiene que

$$\text{Cov}[\mathbf{Ax}, \mathbf{By}] = \mathbf{A} \text{Cov}[\mathbf{x}, \mathbf{y}] \mathbf{B}^t \quad (2.49)$$

Demostración.

$$\begin{aligned} \text{Cov}[\mathbf{Ax}, \mathbf{By}] &= \mathbf{E}[(\mathbf{Ax} - \mathbf{AE}(\mathbf{x}))(\mathbf{By} - \mathbf{BE}(\mathbf{y}))^t] \\ &= \mathbf{AE}[(\mathbf{x} - \mathbf{E}(\mathbf{x}))(\mathbf{y} - \mathbf{E}(\mathbf{y}))^t] \mathbf{B}^t \\ &= \mathbf{A} \text{Cov}[\mathbf{x}, \mathbf{y}] \mathbf{B}^t \end{aligned} \quad (2.50)$$

■

III. ANÁLISIS DE COMPONENTES PRINCIPALES

Introducción

El ACP es de las herramientas más utilizadas en el análisis de datos de tipo multivariado, transforma un conjunto de m variables correlacionadas en un conjunto de p nuevas variables no correlacionadas, donde $p < m$, llamadas componentes principales.

Esta técnica fue descrita por vez primera por Karl Pearson (1901), aunque él no propuso un método práctico de cálculo para más de dos o tres variables, la aportación de métodos computacionales fue dada por Hotelling (1933) y es precisamente a Hotelling a quien se le adjudica la técnica de Componentes Principales.

Los objetivos esenciales del ACP son primordialmente dos, el primero es reducir la dimensión del conjunto de datos a interpretar, el segundo consiste en identificar variables subyacentes.

El ACP es utilizado como paso intermedio para investigaciones posteriores, tal como regresión múltiple o el análisis cluster. Además permite identificar grupos y outliers en los datos.

Descripción del Análisis de Componentes Principales

En este apartado se explica en términos generales lo que la técnica de componentes principales realiza.

Suponga que un investigador obtiene una gran cantidad de observaciones de un número grande de variables, lo cual puede ser representado por la matriz $\mathbf{X} \in \mathbb{R}^{n \times m}$, donde n es el número de individuos en la muestra y m es el número de variables o características medidas u

observadas a cada individuo. Esta situación se puede expresar como:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^t \\ \mathbf{x}_2^t \\ \vdots \\ \mathbf{x}_n^t \end{bmatrix} \quad (3.1)$$

Donde cada $\mathbf{x}_i \in \mathbb{R}^m$ es un elemento de la muestra m -dimensional. Esto es

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{im} \end{bmatrix} \quad (3.2)$$

Para $i = 1, 2, \dots, n$.

Además se asume que a cada variable de la matriz \mathbf{X} se le ha restado su media, de tal forma que $E(\mathbf{X}) = 0$ y la matriz de varianzas y covarianzas, la cual denotaremos como \mathbf{S} , es

$$\mathbf{S} = \text{Cov}(\mathbf{X}) = E(\mathbf{X}^t \mathbf{X}) \quad (3.3)$$

El rango de la matriz \mathbf{S} , $r(\mathbf{S}) = r \leq m$, matemáticamente las matrices \mathbf{S} y \mathbf{X} no tienen porque ser de rango completo, sin embargo por simplicidad se asume que lo son.

En este caso, el investigador está interesado en interpretar o visualizar comportamientos, tendencias, grupos, etc. dentro de la muestra. Dado el tamaño de m , esto sería posible sólo para valores ≤ 3 . El ACP, permite, en cierta medida, lograr que en la mayoría de las situaciones, la dimensión del problema se pueda disminuir a p variables, donde $p < m$.

La técnica de Componentes Principales parte de m variables originales correlacionadas y se plantean nuevas variables, conocidas como componentes principales, denotadas por \mathbf{Z}_j , las

cuales son combinaciones lineales de las variables originales, con la propiedad de no estar correlacionadas. Se pueden obtener tantas componentes principales como variables originales existan, además están ordenadas tal que \mathbf{Z}_1 contiene la mayor variación de las variables originales, \mathbf{Z}_2 es la componente de la segunda mayor variación, de manera que si únicamente se toman en cuenta las componentes principales de mayor variabilidad se habrá reducido la dimensión.

Cálculo de las componentes

La presente sección incluye las cuestiones técnicas del ACP. Se obtendrá la primer y segunda componentes principales; después se procede a generalizar el procedimiento de dicho cálculo; además se menciona como determinar el número de componentes a considerar. Finalmente se concluye la sección numerando los pasos a seguir en el ACP.

La primer componente se puede expresar como el siguiente vector

$$\mathbf{Z}_1 = \mathbf{X}\mathbf{a}_1 \quad (3.4)$$

Donde \mathbf{X} es la matriz descrita en la sección anterior y \mathbf{a}_1 es un vector columna $m \times 1$, cuyas componentes indican la dirección e importancia de la contribución de cada una de las variables X_1, X_2, \dots, X_m en la componente \mathbf{Z}_1 ,

$$\mathbf{a}_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} \quad (3.5)$$

Luego

$$\begin{aligned}
\mathbf{Z}_1 &= \mathbf{X}\mathbf{a}_1 \\
&= \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_m \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} \\
&= a_{11}\mathbf{x}_1 + a_{21}\mathbf{x}_2 + \dots + a_{m1}\mathbf{x}_m
\end{aligned} \tag{3.6}$$

Observando, como ya se había mencionado, que las componentes principales son combinaciones lineales de las variables originales.

Utilizando los teoremas de la esperanza y covarianza de una combinación lineal y partiendo de la expresión (3.4)

$$E(\mathbf{Z}_1) = E(\mathbf{X}\mathbf{a}_1) = E(\mathbf{X})\mathbf{a}_1 = 0 \tag{3.7}$$

y la varianza de \mathbf{Z}_1 es

$$\text{Var}(\mathbf{Z}_1) = \text{Var}(\mathbf{X}\mathbf{a}_1) = \mathbf{a}_1^t \mathbf{S} \mathbf{a}_1 \tag{3.8}$$

Donde \mathbf{S} es la matriz de varianzas y covarianzas de las variables observadas.

La primer componente es la de varianza máxima, por lo que es necesario maximizar la expresión anterior, pero para que tenga solución la maximización es sujeta a la condición de $\mathbf{a}_1^t \mathbf{a}_1 = 1$, obteniendo que la ecuación a resolver es

$$\max L = \mathbf{a}_1^t \mathbf{S} \mathbf{a}_1 - \lambda_1 (\mathbf{a}_1^t \mathbf{a}_1 - 1) \tag{3.9}$$

La solución se encuentra aplicando las condiciones de primer orden.

$$\frac{\partial L}{\partial \mathbf{a}_1} = 2\mathbf{S}\mathbf{a}_1 - 2\lambda_1 \mathbf{a}_1 = 0 \tag{3.10}$$

Entonces

$$\mathbf{S}\mathbf{a}_1 = \lambda_1\mathbf{a}_1 \quad (3.11)$$

Donde \mathbf{a}_1 es el eigenvector correspondiente al eigenvalor λ_1 de la matriz \mathbf{S} .

En las condiciones de segundo orden de la ecuación (3.9) se debe encontrar una matriz semidefinida positiva, para confirmar que la solución es un máximo, como se muestra a continuación.

$$\begin{aligned} \frac{\partial^2 L}{\partial^2 \mathbf{a}_1} &= 2(\mathbf{S} - \lambda_1 \mathbf{I}) \\ &= 2(\mathbf{P}^t \mathbf{D}_\lambda \mathbf{P} - \lambda_1 \mathbf{P}^t \mathbf{P}) \\ &= \mathbf{P}^t [2(\mathbf{D}_\lambda - \lambda_1 \mathbf{I})] \mathbf{P} \end{aligned} \quad (3.12)$$

Los eigenvalores de la matriz $2(\mathbf{S} - \lambda_1 \mathbf{I})$ son la diagonal principal de la siguiente matriz

$$2(\mathbf{D}_\lambda - \lambda_1 \mathbf{I}) = 2 \begin{bmatrix} \lambda_1 - \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 - \lambda_1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \lambda_m - \lambda_1 \end{bmatrix} \quad (3.13)$$

El primer eigenvalor es cero, entonces la matriz es semidefinida positiva.

Los coeficientes de la primer componente principal o elementos del eigenvector \mathbf{a}_1 se obtienen resolviendo el siguiente sistema de m ecuaciones.

$$(\mathbf{S} - \lambda_1 \mathbf{I}) \mathbf{a}_1 = 0 \quad (3.14)$$

Observe que la expresión (3.8) se puede reescribir, como

$$\text{Var}(\mathbf{Z}_1) = \mathbf{a}_1^t \mathbf{S} \mathbf{a}_1 = \mathbf{a}_1^t \lambda_1 \mathbf{a}_1 = \lambda_1 \mathbf{a}_1^t \mathbf{a}_1 = \lambda_1 \quad (3.15)$$

Es decir, la maximización de la varianza de la primer componente principal equivale a tomar el máximo eigenvalor de la matriz \mathbf{S} .

Los coeficientes de la segunda componente principal se obtienen maximizando

$$\begin{aligned} \max L &= \mathbf{a}_2^t \mathbf{S} \mathbf{a}_2 & (3.16) \\ \text{s.a.} \quad \mathbf{a}_2^t \mathbf{a}_2 &= 1 \\ \mathbf{a}_1^t \mathbf{a}_2 &= 0 \end{aligned}$$

La primer restricción garantiza que la maximización tenga solución, mientras que la segunda restricción indica que los vectores son ortogonales y linealmente independientes.

Entonces

$$\max L = \mathbf{a}_2^t \mathbf{S} \mathbf{a}_2 - \lambda_2 (\mathbf{a}_2^t \mathbf{a}_2 - 1) - \beta (\mathbf{a}_1^t \mathbf{a}_2) \quad (3.17)$$

Condiciones de primer orden

$$\frac{\partial L}{\partial \mathbf{a}_2} = 2(\mathbf{S} - \lambda_2 \mathbf{I}) \mathbf{a}_2 - \beta \mathbf{a}_1 = 0 \quad (3.18)$$

Condiciones de segundo orden

$$\frac{\partial^2 L}{\partial^2 \mathbf{a}_2} = 2(\mathbf{S} - \lambda_2 \mathbf{I}) \quad (3.19)$$

La matriz $2(\mathbf{S} - \lambda_2 \mathbf{I})$ es semidefinida positiva, su segundo eigenvalor es cero.

Ahora, si se premultiplica la expresión (3.18) por \mathbf{a}_1^t , se tiene que

$$2\mathbf{a}_1^t \mathbf{S} \mathbf{a}_2 - \beta = 0 \quad (3.20)$$

si la ecuación (3.14) es premultiplicada por el vector \mathbf{a}_2^t se encuentra que

$$\mathbf{a}_2^t \mathbf{S} \mathbf{a}_1 = 0 \quad (3.21)$$

se sabe que la transpuesta de un escalar es el mismo escalar, por lo que $(\mathbf{a}_2^t \mathbf{S} \mathbf{a}_1)^t = \mathbf{a}_1^t \mathbf{S} \mathbf{a}_2 = 0$, dado esto, $\beta = 0$. Las componentes del vector \mathbf{a}_2 se encuentran al solucionar

$$(\mathbf{S} - \lambda_2 \mathbf{I}) \mathbf{a}_2 = 0 \quad (3.22)$$

Los coeficientes de la segunda componente son los elementos del eigenvector correspondiente al segundo eigenvalor mayor de la matriz \mathbf{S} .

Generalizando, el cálculo de la j -ésima componente consiste en maximizar la siguiente función objetivo

$$\begin{aligned} \phi &= \mathbf{a}_j^t \mathbf{S} \mathbf{a}_j - \lambda_j (\mathbf{a}_j^t \mathbf{a}_j - 1) - \sum_{k < j} \beta_k (\mathbf{a}_k^t \mathbf{a}_j - 0) \\ \text{s.a.} \quad \mathbf{a}_j^t \mathbf{a}_j &= 1 \\ \mathbf{a}_k^t \mathbf{a}_j &= 0 \quad \text{para todo } k < j \end{aligned} \quad (3.23)$$

donde $j = 1, 2, \dots, m$; $k = 1, 2, \dots, j - 1$.

Aplicar las condiciones de primer orden y encontrar los coeficientes de la j -ésima componente al solucionar el sistema de m ecuaciones de la forma

$$(\mathbf{S} - \lambda_j \mathbf{I}) \mathbf{a}_j = 0 \quad (3.24)$$

La definición siguiente resume de manera formal la j -ésima componente principal.

Definición 24. *La j -ésima componente de la muestra de m variables observadas es una combinación lineal de las variables originales*

$$\mathbf{Z}_j = \mathbf{X} \mathbf{a}_j \quad \text{para } j = 1, 2, \dots, m$$

cuyos coeficientes son las componentes del eigenvector de la matriz de covarianzas \mathbf{S} , correspondiente al j -ésimo eigenvalor mas grande, λ_j . La varianza de la j -ésima componente es λ_j .

La matriz de componentes es

$$\mathbf{Z} = \mathbf{X}\mathbf{A} = \mathbf{X} \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_m \end{bmatrix} = \begin{bmatrix} Z_1 & Z_2 & \dots & Z_m \end{bmatrix} \quad (3.25)$$

La matriz \mathbf{A} tiene como columnas los eigenvectores \mathbf{a}_j , $j = 1, 2, \dots, m$.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mm} \end{bmatrix} \quad (3.26)$$

Observe que \mathbf{A} es una matriz ortogonal, ya que $\mathbf{A}^t\mathbf{A} = \mathbf{I}$. El cálculo de los componentes principales equivale a aplicar una transformación ortogonal \mathbf{A} a las variables \mathbf{X} para obtener nuevas variables \mathbf{Z} no correlacionadas.

Una vez obtenidas las componentes principales, el siguiente paso es saber cuántas de éstas deben ser contempladas para el análisis, obviamente se busca un cantidad menor a las m variables originales, se desearía que fueran dos o tres componentes para facilitar el análisis e incluso visualizar los resultados gráficamente. La prueba de Anderson (1963) permite determinar el número de componentes necesarias.

La prueba consiste en seleccionar hasta la componente p si el resto de las componentes, $p + 1, p + 2, \dots, m$, cuentan con varianzas aproximadamente similares, ya que esto indica que a apartir del eje $p+1$ no hay ejes principales.

La hipótesis a probar es

$$H_o : \lambda_{p+1} = \lambda_{p+2} = \dots = \lambda_m$$

La razón de verosimilitud proporciona el siguiente estadístico

$$\mathbf{W} = -(n-1) \sum_{j=p+1}^m \ln(\lambda_j) + (m-p)(n-1) \ln \frac{\sum_{j=p+1}^m \lambda_j}{(m-p)} \quad (3.27)$$

donde m se refiere al número de variables y n son las observaciones.

El estadístico \mathbf{W} se distribuye χ^2 con $\frac{1}{2}(m-p)(m-p+1) - 1$ grados de libertad, se rechaza la hipótesis nula si el estadístico es significativo. En el momento en que la hipótesis no sea rechazada se habrá encontrado la cantidad de componentes principales a considerar.

Esta prueba se utiliza si se cumple el supuesto de que las variables observadas se distribuyen normal.

Existen otros métodos empíricos ¹ que permiten obtener el número de componentes a incluir, los más utilizados son:

- Incluir a las componentes con varianza mayor a la varianza media, dada por el cociente

$$\frac{\sum_{j=1}^m \lambda_j}{m} \quad (3.28)$$

- Fijar arbitrariamente la variabilidad que se desea explicar, típicamente 90 por ciento, y tomar en cuenta hasta el número de componentes que cubran ese valor. Esta regla debe utilizarse con ciertas reservas en el caso que las variables originales cuenten con escalas de medición muy diferentes, como se verá más adelante.

Resumiendo, los pasos a seguir en el ACP son:

1. Calcule la matriz de varianzas y covarianzas de las m variables originales, \mathbf{S} .
2. Encuentre los eigenvalores y eigenvectores correspondientes a la matriz \mathbf{S} , para estar en condiciones de definir las componentes.
3. Descarte componentes en base a la prueba de Anderson o a alguna de las reglas empíricas citadas.

¹Más pruebas para la determinación del número de componentes principales se pueden consultar en Cuadras (1991).

Propiedades de las componentes

Las nuevas variables cuentan con varias propiedades interesantes, algunas son:

- *Se conserva la variabilidad inicial.* La suma de las varianzas de las componentes es igual a la suma de las varianzas de las variables originales.

Primero observe que

$$\text{tr}(\mathbf{S}) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_m^2 \quad (3.29)$$

de la sección 2.4, del capítulo anterior, se sabe que la traza de una matriz simétrica puede ser expresada como la suma de sus eigenvalores, esto es

$$\text{tr}(\mathbf{S}) = \sum_{j=1}^m \lambda_j \quad (3.30)$$

Obteniendo que

$$\sum_{j=1}^m \lambda_j = \text{tr}(\mathbf{S}) = \sum_{j=1}^m \text{Var}(\mathbf{X}_j) \quad (3.31)$$

Entonces

$$\sum_{j=1}^m \text{Var}(\mathbf{X}_j) = \sum_{j=1}^m \text{Var}(\mathbf{Z}_j) \quad (3.32)$$

- La *proporción de variabilidad* explicada por un componente se obtiene mediante el cociente de la varianza del componente y la varianza total.

$$\frac{\lambda_j}{\sum_j^m \lambda_j} \quad (3.33)$$

- Las *covarianzas* entre la j -ésima componente principal y las variables \mathbf{X} están dadas por el producto de las coordenadas del eigenvector que define al componente y su eigenvalor.

Se tiene que

$$\begin{aligned}
 \text{Cov}(\mathbf{X}, \mathbf{Z}_j) &= \text{E}[(\mathbf{X} - \mu_x)^t (\mathbf{Z}_j - \mu_{Z_j})] \\
 &= \text{E}[(\mathbf{X}^t \mathbf{Z}_j)] = \text{E}[\mathbf{X}^t \mathbf{X} \mathbf{a}_j] \\
 &= \text{E}[\mathbf{X}^t \mathbf{X}] \mathbf{a}_j = \mathbf{S} \mathbf{a}_j
 \end{aligned} \tag{3.34}$$

De la ecuación (3.24) se sabe que $\mathbf{S} \mathbf{a}_j = \lambda_j \mathbf{a}_j$, entonces

$$\text{Cov}(\mathbf{X}, \mathbf{Z}_j) = \lambda_j \mathbf{a}_j \tag{3.35}$$

Por ejemplo, la covarianza de la primer componente y las variables originales es

$$\begin{aligned}
 \text{Cov}(\mathbf{X}, \mathbf{Z}_1) &= \lambda_1 \mathbf{a}_1 \\
 &= \lambda_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}
 \end{aligned} \tag{3.36}$$

Entonces

$$\text{Cov}(X_1, X_2, \dots, X_m, Z_1) = (\lambda_1 a_{11}, \lambda_1 a_{21}, \dots, \lambda_1 a_{m1}) \tag{3.37}$$

Generalizando

$$\text{Cov}(X_1, X_2, \dots, X_m, Z_j) = \lambda_j \mathbf{a}_j^t = (\lambda_j a_{1j}, \lambda_j a_{2j}, \dots, \lambda_j a_{mj}) \tag{3.38}$$

- La *correlación* entre la j -ésima componente principal y la variable X_k es

$$\text{Corr}(Z_j, X_k) = \frac{\text{Cov}(Z_j, X_k)}{\sqrt{\text{Var}(Z_j) \text{Var}(X_k)}} = \frac{\lambda_j a_{kj}}{\sqrt{\lambda_j S_k^2}} = a_{kj} \frac{\sqrt{\lambda_j}}{S_k}, \tag{3.39}$$

mostrando que la correlación es proporcional al coeficiente de esa variable en la definición de la componente.

- Una de las propiedades más importantes es que las nuevas variables son no correlacionadas, para esto se revisa primero la covarianza entre las componentes \mathbf{Z}_j y $\mathbf{Z}_{j'}$.

$$\text{Cov}(\mathbf{Z}_j, \mathbf{Z}_{j'}) = E(\mathbf{Z}_j^t \mathbf{Z}_{j'}) = \mathbf{a}_j^t E(\mathbf{X}^t \mathbf{X}) \mathbf{a}_{j'} = \mathbf{a}_j^t \mathbf{S} \mathbf{a}_{j'} \quad (3.40)$$

Ya antes se observó que el último término es cero, entonces la correlación entre estas variables es

$$\text{Corr}(\mathbf{Z}_j, \mathbf{Z}_{j'}) = \frac{\text{Cov}(\mathbf{Z}_j, \mathbf{Z}_{j'})}{\sqrt{\lambda_j} \sqrt{\lambda_{j'}}} = 0 \quad (3.41)$$

Antes de citar la última propiedad, note que, cada renglón de la matriz \mathbf{X} es un punto en el espacio \mathbb{R}^m , por lo que la matriz se puede visualizar como n puntos en \mathbb{R}^m , se busca encontrar un subespacio p , $p < m$, tal que los n puntos estén próximos al subespacio. Lo necesario para proyectar los puntos en \mathbb{R}^p son las nuevas bases o ejes coordenados y las nuevas coordenadas.

- La elección de los nuevos ejes es tal que la suma de las distancias al cuadrado de cada punto a su proyección son mínimas.

Suponga que se está trabajando en \mathbb{R}^3 y se tienen n puntos que se proyectan hacia la recta, como la Figura (3.1) lo muestra. La proyección del vector \mathbf{a} es el vector \mathbf{b} , se desea minimizar la distancia euclídea entre los puntos \mathbf{a} y \mathbf{b} , esto se logra utilizando el método de mínimos cuadrados.

$$\min D^2(\mathbf{a}, \mathbf{b}) \quad (3.42)$$

Utilizando el teorema de pitágoras

$$\min D^2(\mathbf{a}, \mathbf{b}) = [D^2(\mathbf{o}, \mathbf{a}) - D^2(\mathbf{o}, \mathbf{b})], \quad (3.43)$$

la distancia del origen al punto \mathbf{a} es fija, por lo que minimizar la expresión anterior equivale a maximizar la distancia al cuadrado del origen al punto \mathbf{b} , la longitud de ese vector es el

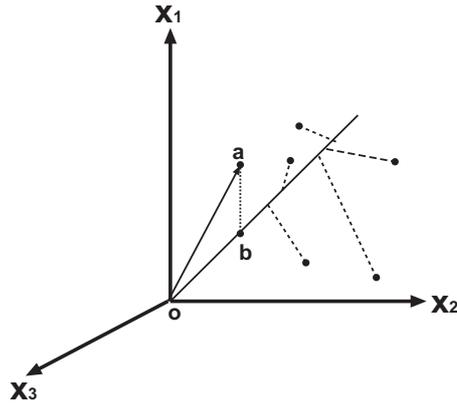


Figura 3.1: Proyección de los n puntos.

producto interno consigo mismo, es decir

$$\max D^2(o,b) = \max \mathbf{b}^t \mathbf{b} = \max \sum_{i=1}^3 b_i^2. \quad (3.44)$$

El último resultado equivale a maximizar la varianza, ya que el vector de medias de \mathbf{b} es cero. Entonces minimizar la distancia entre el valor observado, \mathbf{a} , y su proyección, \mathbf{b} , equivale a maximizar la varianza del vector \mathbf{b} , precisamente la maximización de la varianza es lo que se utiliza para encontrar las componentes principales.

Análisis de Componentes Principales Normado

Al considerar la matriz de varianzas y covarianzas de las variables originales se pueden presentar ciertos problemas. Uno de ellos es que si las variables originales están en unidades muy diferentes, por ejemplo kilómetros, edad, millones, gramos, etc., la maximización de la varianza se verá afectada por las escalas, las variables de valores más grandes tendrán más peso en el análisis.

Otra situación que puede ocurrir es que una de las variables originales cuente con mayor varianza que el resto, de tal forma que la primer componente se vea muy influenciada por esa

variable, recuerde que la primer componente es la de máxima varianza. En las situaciones descritas no es recomendable utilizar la matriz \mathbf{S} .

Con el propósito de evitar que suceda lo mencionado se requiere que las variables originales cuenten con unidades y varianzas similares, cuando no es así se sugiere estandarizar las variables X , de modo que tengan media cero y varianza 1, lo que equivale a utilizar la matriz de correlaciones \mathbf{R} , en lugar de la matriz \mathbf{S} . Bajo estas condiciones el análisis de componentes principales es conocido como Análisis de Componentes Principales Normado.

Las componentes obtenidas de la matriz \mathbf{S} y \mathbf{R} no son las mismas, además no es posible pasar de una solución a otra únicamente utilizando los coeficientes encontrados en cada caso. Todas las propiedades de las componentes principales establecidas en la sección anterior se mantienen, considerando la modificación de que la varianza de cada variable original es uno.

Observe que cuando se enumeraron los pasos a seguir en el Análisis de Componentes Principales, el paso uno debe ser modificado, en lugar de calcular la matriz \mathbf{S} , se calculará la matriz \mathbf{R} . Además en el último paso, al seleccionar el número de componentes, si se utiliza la regla en base a la varianza media, este valor será uno, es decir se considerarán las componentes con eigenvalores o varianzas mayores a la unidad.

Interpretación de las componentes

Tradicionalmente la primer componente se interpreta como un factor de tamaño, o bien un promedio ponderado de las variables, siempre y cuando cuente con todos sus coeficientes de signo positivo. El resto de las componentes se interpretan como factores de forma y generalmente se conforman con coeficientes de signos positivos y negativos, se contraponen el grupo de variables de signo positivo frente al grupo de variables de signo negativo. Sin embargo, esta interpretación debe considerarse con reservas, ya que los coeficientes de la j -ésima componente

son los elementos del eigenvector correspondiente a λ_j de la matriz \mathbf{S} , el cual no tiene porqué ser único.

Otra forma de interpretar a las componentes es observando que las componentes principales son una transformación lineal de las variables originales, geoméricamente una transformación es el traslado, rotación, o bien ambos movimientos de los ejes originales, en el caso de ACP se descarta un traslado de los ejes, ya que $E(\mathbf{X}) = 0$ y $E(\mathbf{Z}) = 0$, entonces, las componentes principales pueden ser vistas como la rotación de los ejes originales.

Suponga que se tienen n observaciones en \mathbb{R}^3 , los puntos tienen la forma de una elipse con centro en el origen, como lo ilustra la Figura (3.2), donde el eje mayor es la variable de mayor varianza, la primer componente \mathbf{Z}_1 , la segunda componente, \mathbf{Z}_2 , representa la variable con la segunda mayor variación de la figura y finalmente el eje menor está dado por la tercer componente, \mathbf{Z}_3 .

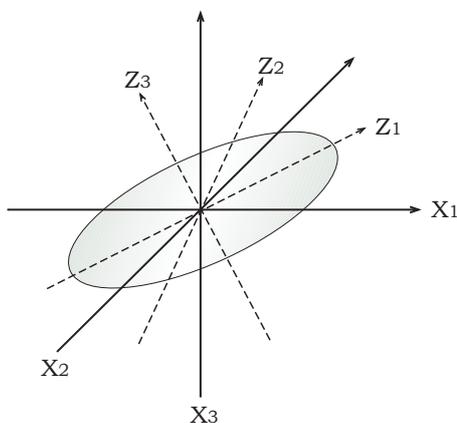


Figura 3.2: Componentes principales y la elipse.

Ejemplo

En México los egresos por entidad federativa se clasifican en 12 capítulos, de acuerdo a la clasificación de la Secretaria de Hacienda y Crédito Publico. Los 12 capítulos son: servicios

personales, SP, materiales y suministro, MS, servicios generales, SG, subsidios transferencias y ayuda, STA, adquisición de bienes, AB, obra pública, OP, inversión financiera, IF, recursos federales, RF, otros egresos, OE, cuenta por terceros, CT, deuda pública, DP y disponibilidad final, DF.

La información de las finanzas estatales para el año 2002 puede ser consultada en el producto Finanzas públicas estatales y municipales de México 1999-2002 de INEGI. Los egresos en los 12 capítulos, ya mencionados, se analizaron por medio del método multivariado ACP, utilizando el paquete S-PLUS, los comandos se encuentran en el Cuadro (3.1) y los resultados en el Cuadro (3.2).

Cuadro 3.1: Comandos para obtener el ACP en S-PLUS

```
X <- egresosdatos2
egresos <- princomp(X, covlist=cov.wt(X), cor=F)
Importancia <- summary(egresos)
Loadings <- loadings(egresos)
print(loadings(egresos), cutoff = 0.0)
plot(egresos)
plot (loadings(egresos))
```

Cuadro 3.2: Resultados en S-PLUS para el ACP

```

*** Principal Components Analysis ***

Standard deviations:

  Comp.1  Comp.2  Comp.3  Comp.4  Comp.5  Comp.6  Comp.7
5405.773 3135.61 854.8336 671.457 427.0918 287.5798 218.1182

  Comp.8  Comp.9  Comp.10  Comp.11  Comp.12
186.446 171.8362 125.6533 53.10236 46.34324

The number of variables is 12 and the number of observations is 31
Component names:
"sdev" "loadings" "correlations" "scores" "center" "scale" "n.obs" "terms"
"call" "factor.sdev" "coef"
Call: princomp(x = ., data = egresosdatos, scores = T, cor = F, na.action = na.exclude)

Importance of components:

              Comp.1          Comp.2          Comp.3          Comp.4
Standard deviation 5405.7726863 3135.6095463 854.83356468 671.45698642
Proportion of Variance 0.719739 0.2421633 0.01795327 0.01110296
Cumulative Proportion 0.719739 0.9619023 0.97985557 0.99095853

              Comp.5          Comp.6          Comp.7
Standard deviation 427.091771288 287.579827441 218.118235610
Proportion of Variance 0.004489061 0.002035307 0.001170839
Cumulative Proportion 0.994700863 0.996736170 0.997907008

              Comp.8          Comp.9          Comp.10
Standard deviation 1.864460e+002 1.718362e+002 1.256533e+002
Proportion of Variance 8.554985e-004 7.266789e-004 3.885623e-004
Cumulative Proportion 9.987625e-001 9.994892e-001 9.998777e-001

              Comp.11          Comp.12
Standard deviation 5.310236e+001 4.634324e+001
Proportion of Variance 6.939701e-005 5.285497e-005
Cumulative Proportion 9.999471e-001 1.000000e+000

Loadings:

              Comp.1          Comp.2          Comp.3          Comp.4          Comp.5
SP 0.7404908405 0.6147405871 0.0676503530 -0.0136087133 -0.1963144238
MS 0.0290283696 0.0020514866 -0.0286832393 -0.0339951133 0.0153835359
SG 0.0562983533 0.0202315603 -0.0020776029 0.1033670169 -0.0094837508
STA 0.5237053406 -0.7745778021 0.1945307853 -0.0151913675 -0.2515175718
AB 0.0075098931 0.0079087282 0.0396733949 0.0179838149 -0.0671245090
OP 0.1386485912 -0.0549802101 -0.2602849757 0.8897253607 0.2343494449
IF 0.0037231044 -0.0013690432 0.0513588337 -0.0124935122 -0.0972272521
RF 0.3510677597 -0.0704869959 0.1235317245 -0.1649701535 0.7408494959
OE 0.0041905022 -0.0190161869 0.0045961888 0.1540133798 -0.1389964692
CT 0.0164821425 0.0214473953 0.1352948822 -0.1565411053 -0.0322654213
DP 0.1670532980 -0.1132165600 -0.8954257458 -0.3322521769 0.0751571902
DF 0.0515324268 0.0038021640 -0.2223083418 0.0991009339 -0.5041099167

```

Cuadro 3.2: Resultados en S-PLUS para el ACP (continuación)

*** Principal Components Analysis ***					
	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
SP	-0.1418523260	-0.0829355448	0.0141787104	-0.0426032168	-0.0224651781
MS	0.1675339918	0.0505192591	-0.0185058481	-0.4843213555	-0.4707793560
SG	0.2438074015	0.0880014704	-0.6112730608	-0.3096694621	0.6373558402
STA	-0.1250528056	-0.0836402657	-0.0277815118	-0.0282241851	-0.0072366990
AB	0.1144606025	-0.0245986512	-0.1005866979	-0.0041399903	0.0916727332
OP	-0.0852132187	0.0493795200	-0.1117569211	0.0922613881	-0.1779059751
IF	0.4862646811	-0.0739931855	-0.2573516974	-0.3076835085	-0.4500235297
RF	0.4327531172	0.1721940759	0.1833317754	0.1560923755	0.0724453146
OE	-0.0241767914	0.5964949122	0.5206450439	-0.4896623709	0.2148776457
CT	-0.2192248231	0.7432514416	-0.4428626638	0.2874074730	-0.2752722312
DP	-0.1109900860	0.0526343928	-0.0718000427	-0.0888478371	-0.0007773758
DF	0.6096524123	0.1632821487	0.1738630333	0.4572707126	0.0426091751
	Comp.11	Comp.12			
SP	0.0237480874	-0.0195215926			
MS	-0.6474560359	0.3009586348			
SG	-0.1875832191	-0.0865795107			
STA	-0.0089644084	-0.0096335854			
AB	0.3826649621	0.9026629760			
OP	0.0293122878	0.0162310638			
IF	0.5466552359	-0.2890619565			
RF	-0.0052902599	0.0164941572			
OE	0.2025653893	-0.0460784492			
CT	0.0098414798	0.0182534632			
DP	0.1149632109	0.0096083670			
DF	-0.2108818069	0.0035340232			

Si se calculan las varianzas de cada componente y la varianza media de los resultados se encuentra que el número de componentes principales necesarias para reducir la información es dos, lo cual coincide con el criterio de fijar la varianza acumulada en 90 por ciento. La prueba de Anderson no se puede emplear ya que no existe evidencia empírica de que los datos tengan distribución normal.

Una opción visual de encontrar la cantidad de componentes principales es graficar el histograma de la varianza (λ_j) versus la j -ésima componente, conocido en el idioma inglés como *screeplot*. En la gráfica de la Figura (3.3) las primeras dos componentes explican el 96.1

por ciento de la variación total, a partir de la componente 3 las varianzas son relativamente pequeñas. En lugar de trabajar con las 12 variables originales trabajaremos con las dos primeras componentes principales.

La importancia de las variables originales y la dirección en las componentes principales está dada por los elementos de los eigenvectores (ver Cuadro 3.2), la gráfica de la Figura (3.4) ilustra para las cinco primeras componentes principales las variables de mayor peso y su dirección.

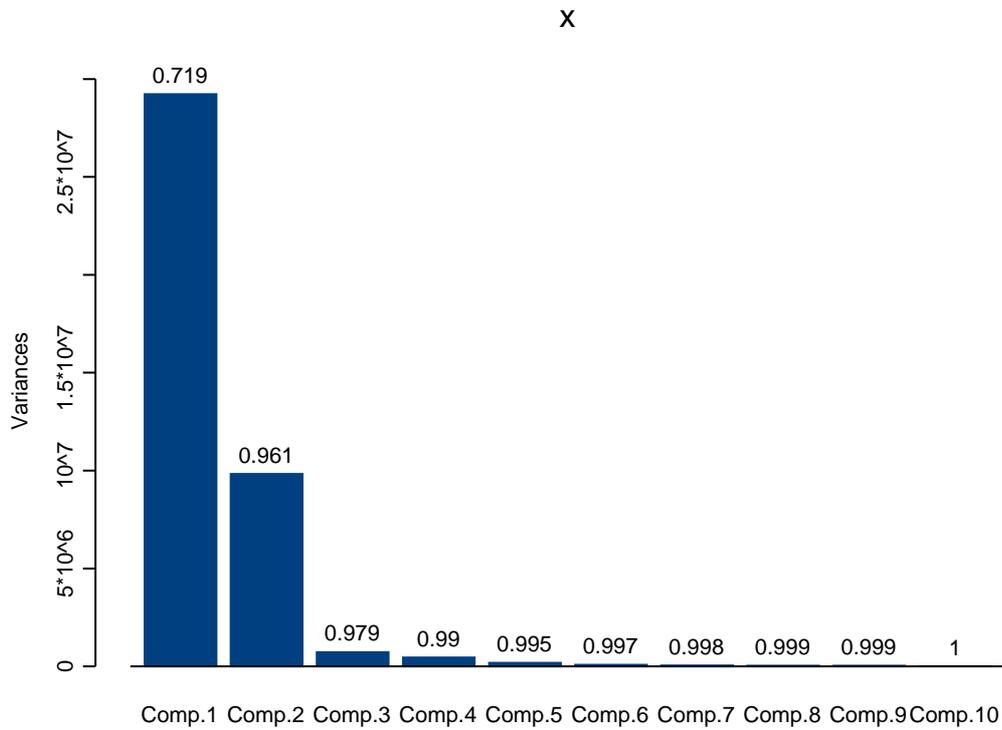


Figura 3.3: Porcentaje de varianza explicada al incluir componentes (ACP).

La primer componente es un promedio ponderado de todas las variables, donde las de mayor peso son SP, STA y RF. La primer componente agrupa a los estados de acuerdo a la cantidad que gastaron en ese año, los que más gastaron de los que menos gastaron, otra forma de interpretarla es como la diferenciación de dos grupos entre los estados, los más ricos y los más pobres, dado que los ingresos y egresos coinciden.

La segunda componente es la comparación de dos medias ponderadas, contraponen las variables de signo positivo, (SP, MS, SG, AB, CT, DF), frente a las de signo negativo, (STA, OP, IF, RF, OE, DP). Nuevamente las variables SP y STA, son las de mayor peso, pero ahora son de signo contrario. La segunda componente principal diferencia a los estados en dos grupos, los que destinan una cantidad mayor de egresos en gasto corriente, nominas, servicios de luz, teléfono, etc. y los que emplean sus ingresos en subsidios, transferencias y ayuda, obra pública, etc.

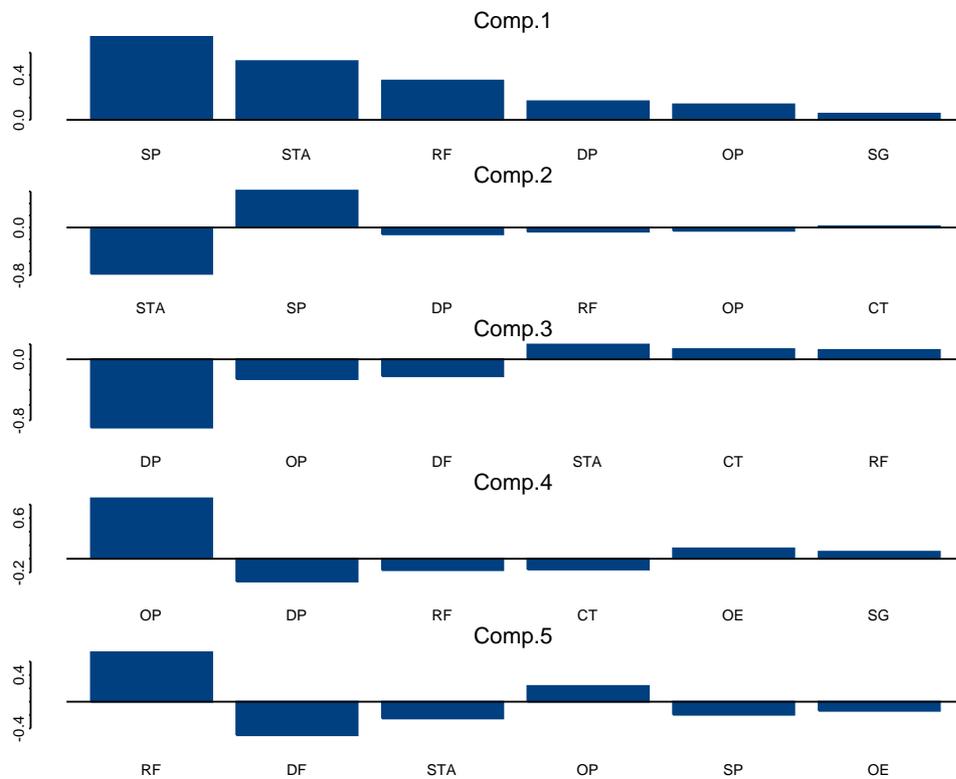


Figura 3.4: Coeficientes de las 5 primeras componentes (ACP).

El Cuadro (3.3) resume para las dos primeras componentes la información más relevante, los valores entre paréntesis son los coeficientes de correlación de la variable original con la componente.

Observe que los coeficientes de correlación de las variables originales con las primeras dos componentes, en general confirman la interpretación de los coeficientes, esto es, las variables de mayor peso cuentan también con coeficientes de correlación altos.

La ordenación de los estados de acuerdo a las componentes principales, 1 y 2, se presenta en el Cuadro (3.4), la primer columna representa los 5 estados que más egresos tuvieron y los 5 estados con menor presupuesto de gasto durante el 2002, la segunda son los primeros 5 estados que emplean mayor parte de su presupuesto en gasto corriente, mientras que los últimos cinco estados son los que emplean mayor parte de sus egresos en subsidios, transferencias y ayuda, obra pública, etc.

Cuadro 3.3: Resumen de las dos componentes principales

	Comp. 1	Comp.2
SP	0.740 (0.90)	0.615 (0.43)
MS	0.029 (0.78)	0.002 (0.03)
SG	0.056 (0.85)	0.020 (0.18)
STA	0.528 (0.76)	-0.775 (-0.65)
AB	0.008 (0.45)	0.008 (0.27)
OP	0.139 (0.75)	-0.055 (-0.17)
IF	0.004 (0.11)	-0.001 (-0.02)
RF	0.351 (0.97)	-0.070 (-0.11)
OE	0.004 (0.10)	-0.019 (-0.26)
CT	0.016 (0.32)	0.021 (0.24)
DP	0.167 (0.72)	-0.113 (-0.28)
DF	0.052 (0.62)	0.004 (0.03)
λ	29222378	9832047
Proporción de varianza	0.7197	0.2422
Varianza acumulada	0.7197	0.9619

Las variables de mayor peso en las primeras dos componentes, del ejemplo anterior, son también las de mayor varianza, para evitar el efecto de grandes varianzas, se estandarizan las variables y se realiza el ACPN. Los comandos se muestran en el Cuadro (3.5) y los resultados en el Cuadro (3.6).

El ACPN, siguiendo el criterio de la varianza media, sugiere que las primeras tres componentes principales son suficientes. En la Figura (3.5) se observa que estas componentes explican 73.5 por ciento de la variación total, si se fija la cantidad de varianza a explicar en 90 por ciento se requiere trabajar con 6 componentes, utilizaremos el resultado del criterio de la varianza media.

Cuadro 3.4: Orden de los estados de acuerdo a las componentes principales

Comp.1	Comp.2
México	Tabasco
Veracruz	Puebla
Jalisco	Coahuila
Puebla	Michoacán
Nuevo León	Veracruz
:	:
Quintana Roo	Oaxaca
Tlaxcala	Guerrero
Aguascalientes	Nuevo León
Colima	México
Baja California Sur	Guanajuato

Cuadro 3.5: Comandos para obtener el ACPN en S-PLUS

```
X <- egresosdatos2
egresos.corr <- princomp(X, cor=T)
Importancia.corr <- summary(egresos.corr)
Loadings.corr <- loadings(egresos.corr)
print(loadings(egresos.corr),cutoff=0.0)
plot(loadings(egresos.corr))
plot(egresos.corr)
```

La primer componente cambia considerablemente, la mayoría de las variables son de igual peso, ver la gráfica de la Figura (3.6), los signos de las variables originales se mantienen, sigue siendo un promedio ponderado, se interpreta de igual forma, aunque ahora el orden de los estado de acuerdo a la componente uno cambiará.

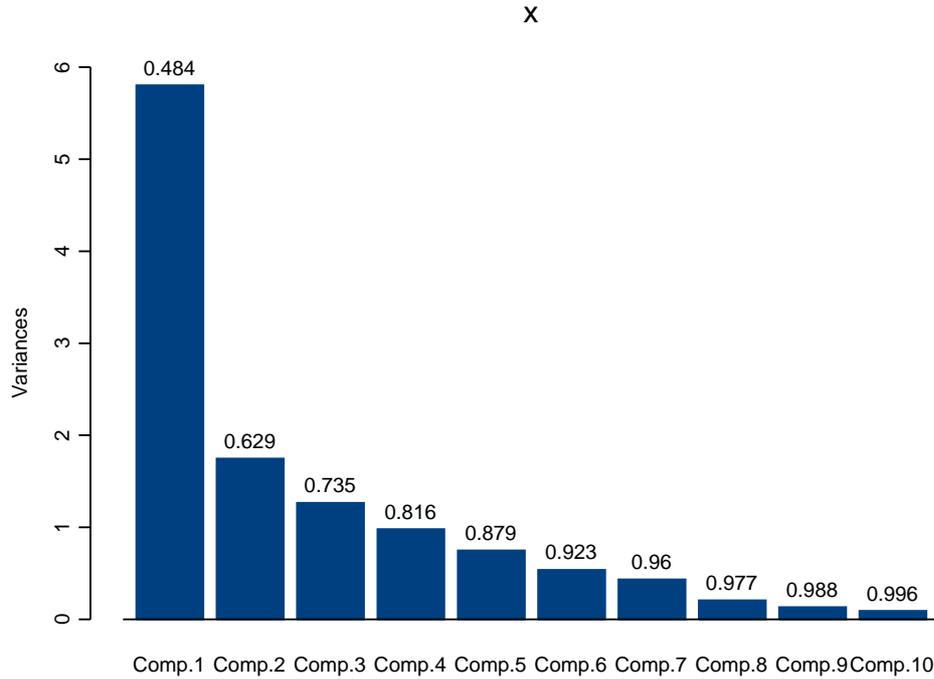


Figura 3.5: Porcentaje de varianza explicado al incluir componentes (ACPN).

Cuadro 3.6: Resultados en S-PLUS para el ACPN

```

*** Principal Components Analysis (normado) ***

Standard deviations:

Comp.1  Comp.2  Comp.3  Comp.4  Comp.5  Comp.6  Comp.7  Comp.8
2.40876  1.32149  1.125414  0.9900632  0.8649852  0.7332368  0.6588046  0.4539065

Comp.9  Comp.10  Comp.11  Comp.12
0.3648661  0.3042059  0.1906764  0.1297741

The number of variables is 12 and the number of observations is 31 Component names:
"sdev" "loadings" "correlations" "scores" "center" "scale" "n.obs" "terms"
"call" "factor.sdev" "coef"
Call: princomp(x = ., data = egresosdatos, scores = T, cor = T, na.action = na.exclude)

Importance of components:

Standard deviation      Comp.1  Comp.2  Comp.3  Comp.4  Comp.5
Proportion of Variance  0.4835103  0.1455281  0.1055464  0.08168544  0.06234995
Cumulative Proportion  0.4835103  0.6290384  0.7345848  0.81627021  0.87862016
    
```

Cuadro 3.6: Resultados en S-PLUS para el ACPN (continuación)

*** Principal Components Analysis (normado) ***					
		Comp.6	Comp.7	Comp.8	Comp.9
Standard deviation		0.73323678	0.65880457	0.45390650	0.36486612
Proportion of Variance		0.04480301	0.03616862	0.01716926	0.01109394
Cumulative Proportion		0.92342318	0.95959180	0.97676106	0.98785500
		Comp.10	Comp.11	Comp.12	
Standard deviation		0.304205851	0.190676438	0.129774055	
Proportion of Variance		0.007711767	0.003029792	0.0014034424	
Cumulative Proportion		0.995566766	0.998596558	1.000000000	
Loadings:					
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
SP	0.3523866343	-0.1684703939	0.2514605479	0.0306735439	-0.3044249892
MS	0.3485420775	-0.0639415489	-0.0338095253	-0.1462152598	0.3271356155
SG	0.3717823765	-0.1534830882	0.0173715664	0.0562326179	-0.2209006350
STA	0.3116907006	0.2513302745	0.0163510694	0.0026426071	0.4351286914
AB	0.2125082617	-0.5102530059	-0.2088037382	0.1773844372	-0.2769864936
OP	0.3285590447	0.2854957399	-0.1076902163	0.1208460440	-0.2670721686
IF	0.1087930536	-0.4581495560	-0.5166729977	-0.2270087588	0.4404815605
RF	0.3864989294	0.0344348408	0.2106378499	-0.0318213334	0.1260275527
OE	0.0781361769	0.2769540899	-0.3726508556	0.7783641390	0.1789957932
CT	0.0966069482	-0.3536383059	0.5688338538	0.4021425702	0.3067360323
DP	0.3096604294	0.3515891798	0.1098308194	-0.3209623274	0.0877263715
DF	0.3054716805	0.0497387925	-0.3066536867	-0.0728722644	-0.2653209991
	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
SP	0.1312366357	-0.2693781033	-0.1830446269	0.3778933321	0.2606610313
MS	0.2142337245	-0.5366069197	-0.2819211107	-0.0250473774	-0.3684196948
SG	-0.1353818017	-0.1733024355	0.3478791876	-0.6204162128	0.4096684345
STA	-0.4138781983	0.4320387229	-0.0931946874	0.0844803808	0.1227872193
AB	-0.2572361925	0.3169458438	-0.4692325122	-0.1862736742	-0.3350041001
OP	-0.3277220141	-0.1406405711	0.4910302373	0.2142857646	-0.5330545720
IF	-0.0001367757	-0.0542137045	0.3328853061	0.1088561628	0.0973783849
RF	-0.2247154957	-0.0090577740	-0.1504574311	0.2825179378	0.3117413685
OE	0.1839812052	-0.1490881208	-0.1548452731	-0.0847255393	0.1481847793
CT	0.2691523086	0.2286730433	0.3273661963	-0.0103572429	-0.2314935306
DP	0.2979904877	0.1785328021	-0.1190296983	-0.4655039933	-0.1688348082
DF	0.5727982469	0.4394872716	0.1313632699	0.2612547977	0.0795267084
	Comp.11	Comp.12			
SP	0.1807841253	-0.5690958994			
MS	-0.4279933883	0.1096760259			
SG	-0.2494903743	0.0398496073			
STA	-0.3474710567	-0.3748106713			
AB	0.0859019353	0.0398409711			
OP	0.1206644226	-0.0097464378			
IF	0.3370481350	-0.1306508966			
RF	0.2863056671	0.6752647164			
OE	0.1753873647	-0.0219316941			
CT	0.0198401951	0.0153733412			
DP	0.5107348564	-0.1297316899			
DF	-0.3027231361	0.1724759572			

En la segunda componente las variables de mayor peso son AB y IF, se contraponen dos medias ponderadas, las variables SP, MS, SG, AB, IF, CT contra el resto, note que son las mismas variables obtenidas en ACP, con excepción del intercambio DF por la variable IF.

En la tercer componente la variable IF es la de mayor peso, le sigue CT, pero con signo contrario, esta componente diferencia los estados que invierten y ahorran de los que no lo hacen.

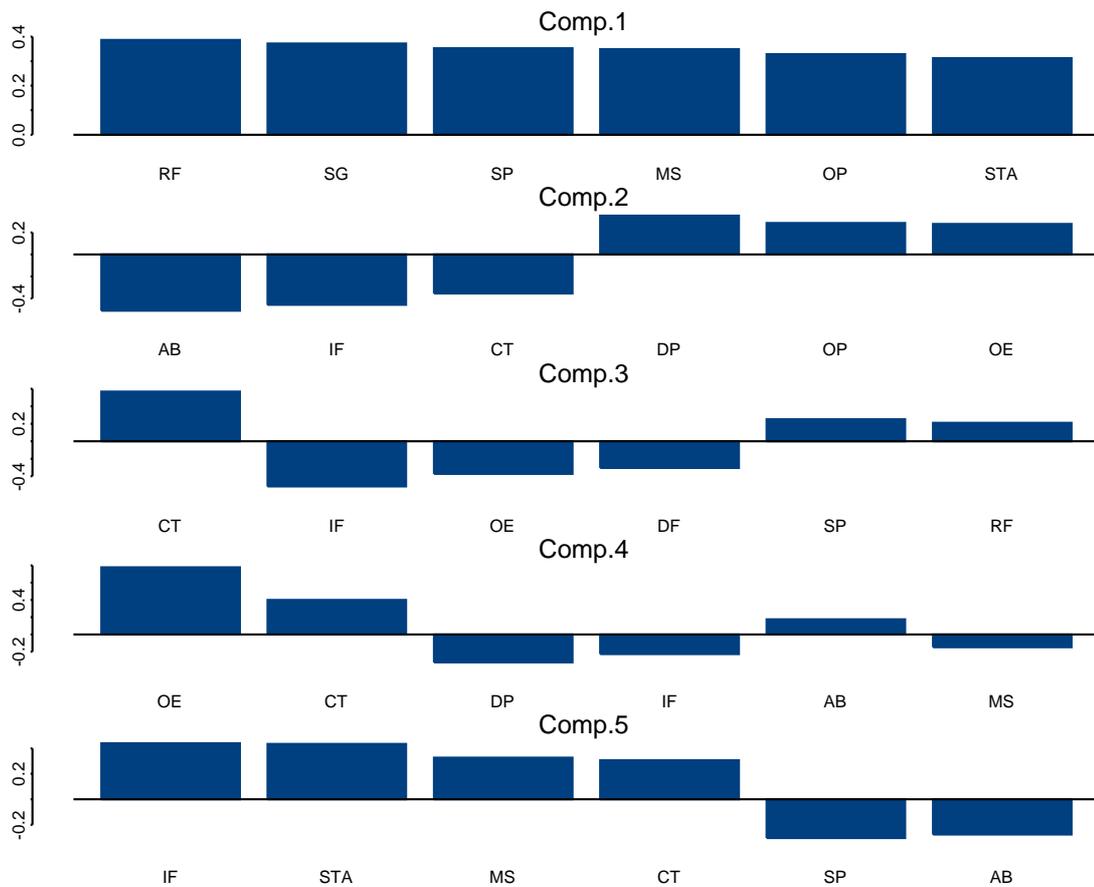


Figura 3.6: Coeficientes de las primeras 5 componentes (ACPN).

El Cuadro (3.7) resume la información más relevante de las primeras tres componentes principales.

Cuadro 3.7: Resumen de las tres componentes principales

	Comp.1	Comp.2	Comp.3
SP	0.352 (0.85)	-0.168 (-0.22)	0.251 (0.28)
MS	0.349 (0.84)	-0.064 (-0.08)	-0.034 (-0.04)
SG	0.372 (0.90)	-0.153 (-0.20)	0.017 (0.02)
STA	0.312 (0.75)	0.251 (0.33)	0.016 (0.02)
AB	0.216 (0.51)	-0.510 (-0.67)	-0.209 (-0.23)
OP	0.329 (0.79)	0.285 (0.38)	-0.108 (-0.12)
IF	0.109 (0.26)	-0.458 (-0.61)	-0.517 (-0.58)
RF	0.386 (0.93)	0.034 (0.05)	0.211 (0.24)
OE	0.078 (0.19)	0.277 (0.37)	-0.373 (-0.42)
CT	0.097 (0.23)	-0.354 (-0.47)	0.569 (0.64)
DP	0.310 (0.75)	0.352 (0.46)	0.110 (0.12)
DF	0.305 (0.74)	0.050 (0.07)	-0.307 (-0.35)
λ	5.8021	1.7463	1.2666
Proporción de varianza	0.4835	0.1455	0.1055
Varianza acumulada	0.4835	0.6290	0.7346

IV. ANÁLISIS DE CORRESPONDENCIAS

Introducción

El Análisis de Correspondencias (AC) es un método estadístico utilizado para representar asociaciones entre variables categóricas o cualitativas, con el propósito de establecer, si es posible, patrones o estructuras en los datos obtenidos.

La técnica matemáticamente fue introducida por Hirschfeld en 1935, el estadístico francés Jean Paul Benzecri la redescubrió en 1962, como una herramienta de análisis exploratorio de datos, agregando la representación geométrica.

Este método se ha empleado en diferentes campos de aplicación no relacionados, utilizando diferentes nombres, tales como Análisis de componentes principales para datos cualitativos, Método de los promedios recíprocos, Análisis factorial de correspondencias, Escalado dual; entre otros.

El AC es de tipo exploratorio, en el sentido que describe los datos, además es complementario para otros análisis de tipo inferencial tales como el análisis de modelos log-lineales o de regresión logística.

El objetivo del AC es, al igual que en el ACP, la reducción de la dimensión del conjunto de datos observados, transforma una tabla de información referida a variables p -dimensionales de forma que se puedan visualizar gráficamente, facilitando la interpretación notablemente.

Representación en el Análisis de Correspondencias

El AC es empleado para representar tablas de contingencia, una tabla de contingencia es la que contiene las frecuencias de aparición de la combinación de dos o más variables cualitativas

de un conjunto de observaciones.

Cuando el AC es dirigido hacia tablas de contingencia con dos variables se conoce como Análisis de Correspondencias Simples. Suponga que se tiene información para dos variables, la variable fila, X y la variable columna, Y, como se presenta a continuación.

	Y_1	Y_2	\dots	Y_J	Total
X_1	n_{11}	n_{12}	\dots	n_{1J}	$n_{1.}$
X_2	n_{21}	n_{22}	\dots	n_{2J}	$n_{2.}$
\vdots	\vdots	\vdots	\dots	\vdots	\vdots
X_I	n_{I1}	n_{I2}	\dots	n_{IJ}	$n_{I.}$
Total	$n_{.1}$	$n_{.2}$	\dots	$n_{.J}$	$n_{..}$

Donde:

n_{ij} = frecuencia absoluta de X_i y Y_j .

$n_{i.}$ = frecuencia marginal de la variable X_i

$n_{.j}$ = frecuencia marginal de la variable Y_j

Cada celda, n_{ij} , indica la frecuencia absoluta de aparición del total de las observaciones que cumplen con la combinación de la variable X_i y Y_j . La frecuencia marginal $n_{i.}$ es la cantidad de observaciones que cumplen con la variable X_i , para $i = 1, 2, \dots, I$, mientras que el total de observaciones con la categoría j de la variable Y es $n_{.j}$, $j = 1, 2, \dots, J$. El número total de observaciones es $n_{..}$.

Las frecuencias absolutas de la tabla de contingencia se pueden representar en una matriz $\mathbf{N} \in \mathbb{R}^{I \times J}$, donde la variable cualitativa X, representada en las filas de \mathbf{N} , toma I modalidades

o categorías, la variable Y , contenida en las columnas, tiene J modalidades.

$$\mathbf{N} = \begin{bmatrix} n_{11} & n_{12} & \dots & n_{1J} \\ n_{21} & n_{22} & \dots & n_{2J} \\ \vdots & \vdots & \dots & \vdots \\ n_{I1} & n_{I2} & \dots & n_{IJ} \end{bmatrix} \quad (4.1)$$

El análisis de correspondencias pretende presentar las variables en un espacio de dimensión menor, análogo a componentes principales, pero para variables cualitativas. Observe que en el ACP la proyección se realizó por filas, ya que las variables se encontraban representadas en las columnas de la matriz \mathbf{X} , pero en el caso de ACS se puede hacer la representación tanto de las filas, como de las columnas o bien la proyección de ambas. Estudiemos primero el caso de las filas.

Representación de las filas

Defina la matriz \mathbf{F} como la matriz de frecuencias relativas, se obtiene dividiendo cada frecuencia absoluta por el total de observaciones. Esto es

$$\mathbf{F} = \frac{\mathbf{N}}{n_{..}} \quad (4.2)$$

La nueva tabla de contingencia es

	Y_1	Y_2	\dots	Y_J	Total
X_1	f_{11}	f_{12}	\dots	f_{1J}	$f_{1.}$
X_2	f_{21}	f_{22}	\dots	f_{2J}	$f_{2.}$
\vdots	\vdots	\vdots	\dots	\vdots	\vdots
X_I	f_{I1}	f_{I2}	\dots	f_{IJ}	$f_{I.}$
Total	$f_{.1}$	$f_{.2}$	\dots	$f_{.J}$	$f_{..}$

Donde:

f_{ij} = frecuencia relativa de X_i y Y_j .

$f_{i.}$ = frecuencia marginal relativa de la variable X_i

$f_{.j}$ = frecuencia marginal relativa de la variable Y_j

Los elementos f_{ij} de la matriz \mathbf{F} indican el porcentaje del total de las observaciones que cumplen simultáneamente la combinación de dos variables, la frecuencia marginal relativa de la fila i , $f_{i.}$, es el porcentaje del total de las observaciones que cuentan con la categoría i de la variable X , de forma similar, el porcentaje de observaciones que cumplen con la variable Y_j es $f_{.j}$. Observe que $\sum_{i=1}^I f_{i.} = \sum_{j=1}^J f_{.j} = \sum_{i=1}^I \sum_{j=1}^J f_{ij} = f_{..} = 1$.

La suma de las frecuencias marginales relativas de las filas de \mathbf{F} suman uno, esto no implica que deban ser iguales, cuando son diferentes, las filas no tienen igual peso, debido a que no contienen el mismo porcentaje de las observaciones. Realizar comparaciones entre las filas, bajo esta circunstancia, no tiene sentido, para que las comparaciones sean válidas deben tener la misma base, lo que se logra condicionando cada fila a su frecuencia marginal relativa, $f_{i.}$, de manera que cada fila suma uno. La matriz que incluye esta operación se denotará por \mathbf{P} .

$$\mathbf{P} = \mathbf{D}_f^{-1} \mathbf{F} = \begin{bmatrix} \frac{f_{ij}}{f_{i.}} \end{bmatrix} \quad (4.3)$$

Donde \mathbf{D}_f es una matriz diagonal $I \times I$ con las frecuencias marginales relativas de las filas, $f_{i.}$, en la diagonal principal.

Las filas de la matriz \mathbf{P} son funciones de probabilidad condicionadas a la variable X_i , los elementos de cada fila suman uno e indica el porcentaje de las observaciones que cumplen la modalidad j de la variable Y sabiendo que se posee la condición i de la variable fila. A cada fila de la matriz \mathbf{P} se le conoce como perfil fila.

El vector conformado por el último renglón de la tabla de frecuencias relativas, cuyas componentes son las frecuencias marginales relativas de las columnas, es conocido como el perfil medio, $\mathbf{G} = [f_{.1} \ f_{.2} \ \dots \ f_{.J}]$

Geoméricamente cada perfil fila de la matriz \mathbf{P} es un punto en \mathbb{R}^{J-1} , la matriz es el conjunto de I puntos en \mathbb{R}^{J-1} , se busca proyectar los I puntos en un espacio de dimensión menor r , tal que $r < J - 1$, de manera que las filas de estructura semejante se encuentren muy cercanas y viceversa. Dos filas se consideran de estructura semejante si se asocian de igual modo al conjunto de las modalidades de la variable columna, es decir, si se asocian mucho o poco a las mismas columnas.

Establecer la semejanza entre dos perfiles fila, p_i y $p_{i'}$, no se obtendrá por la distancia euclídea, como en el ACP, debido a que no da mayor ponderación a las filas con mayor cantidad de observaciones, la manera de equilibrar los perfiles fila es dando igual importancia a cada uno de ellos, ponderándolos con $\frac{1}{f_{.j}}$. En lugar de obtener la distancia euclídea

$$D^2(p_i, p_{i'}) = \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 \quad (4.4)$$

se utiliza la distancia euclídea ponderada

$$D^2(p_i, p_{i'}) = \frac{(p_{ij} - p_{i'j})^2}{f_{.j}} = \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 \frac{1}{f_{.j}} \quad (4.5)$$

La expresión anterior es conocida como la distancia χ^2 , se estudiará más adelante.

Matricialmente se obtiene como

$$D^2(p_i, p_{i'}) = (p_i - p_{i'})^t \mathbf{D}_c^{-1} (p_i - p_{i'}) \quad (4.6)$$

Donde $\mathbf{D}_c \in \mathbb{R}^{J \times J}$ es una matriz diagonal con las frecuencias marginales relativas de las columnas, $f_{.j}$, en la diagonal principal.

La distancia χ^2 equivale a la distancia euclídea entre los vectores transformados $\mathbf{w}_i = \mathbf{D}_c^{-1/2} \mathbf{p}_i$, como se muestra a continuación.

La matriz \mathbf{W} puede ser expresada como

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_1^t \\ \mathbf{w}_2^t \\ \vdots \\ \mathbf{w}_I^t \end{bmatrix} \quad (4.7)$$

Esto es

$$\mathbf{W} = \mathbf{P} \mathbf{D}_c^{-1/2} = \mathbf{D}_f^{-1} \mathbf{F} \mathbf{D}_c^{-1/2} = \left[\frac{f_{ij}}{f_{i.} \sqrt{f_{.j}}} \right] \quad (4.8)$$

La distancia euclídea entre las filas k y k' de la matriz \mathbf{W} es

$$D^2(\mathbf{w}_k, \mathbf{w}_{k'}) = \left(\frac{f_{kj}}{f_{k.} \sqrt{f_{.j}}} - \frac{f_{k'j}}{f_{k'.} \sqrt{f_{.j}}} \right)^2 = \left(\frac{f_{kj}}{f_{k.}} - \frac{f_{k'j}}{f_{k'.}} \right)^2 \frac{1}{f_{.j}} \quad (4.9)$$

Igual a lo encontrado en la ecuación (4.5), por lo que se utilizará la matriz \mathbf{W} ya que la distancia euclídea entre sus filas es equivalente a la distancia χ^2 entre las filas de \mathbf{P} .

Los elementos de \mathbf{W} son las frecuencias relativas condicionadas por filas, pero divididos por la raíz cuadrada de la frecuencias marginales de las columnas, ni las filas, ni las columnas suman uno.

La matriz \mathbf{W} puede ser tratada como la matriz \mathbf{X} del ACP, con observaciones en filas y variables en columnas, entonces el interés es proyectar las filas. Las distancias entre las filas de \mathbf{W} deben mantenerse en la proyección, de tal manera que las filas con estructura similar estén cercanas, esto implica encontrar una dirección de norma uno.

$$\mathbf{a}^t \mathbf{a} = 1 \quad (4.10)$$

El vector proyectado sobre esta dirección será denotado como $w_p(\mathbf{a}) = \mathbf{W}\mathbf{a}$ y es el de mayor varianza, entonces

$$\max L = \mathbf{a}^t \mathbf{W}^t \mathbf{W} \mathbf{a} - \lambda(\mathbf{a}^t \mathbf{a} - 1) \quad (4.11)$$

El problema con la expresión anterior es que no toma en cuenta que las filas no tiene el mismo peso, para considerarlo se maximiza la suma de cuadrados ponderada:

$$\max L = \mathbf{a}^t \mathbf{W}^t \mathbf{D}_f \mathbf{W} \mathbf{a} - \lambda(\mathbf{a}^t \mathbf{a} - 1) \quad (4.12)$$

Sustituyendo \mathbf{W} , se obtiene

$$\max L = \mathbf{a}^t \mathbf{D}_c^{-1/2} \mathbf{F}^t \mathbf{D}_f^{-1} \mathbf{F} \mathbf{D}_c^{-1/2} \mathbf{a} - \lambda(\mathbf{a}^t \mathbf{a} - 1) \quad (4.13)$$

Ahora defina la matriz $\mathbf{H} = \mathbf{D}_f^{-1/2} \mathbf{F} \mathbf{D}_c^{-1/2}$, el ij -ésimo elemento es

$$h_{ij} = \left(\frac{f_{ij}}{\sqrt{f_{i.}} \sqrt{f_{.j}}} \right) \quad (4.14)$$

Entonces, la expresión de la maximización se reescribe como

$$\max L = \mathbf{a}^t \mathbf{H}^t \mathbf{H} \mathbf{a} - \lambda(\mathbf{a}^t \mathbf{a} - 1) \quad (4.15)$$

La solución ya es conocida

$$\mathbf{H}^t \mathbf{H} \mathbf{a} = \lambda \mathbf{a} \quad (4.16)$$

Donde \mathbf{a} es el eigenvector correspondiente al eigenvalor λ de la matriz $\mathbf{H}^t \mathbf{H}$.

La matriz $\mathbf{H}^t \mathbf{H}$ tiene como mayor eigenvalor al uno y como eigenvector a $\mathbf{D}_c^{1/2}$. Si sustituimos \mathbf{H} , se encuentra que

$$\begin{aligned} (\mathbf{D}_f^{-1/2} \mathbf{F} \mathbf{D}_c^{-1/2})^t (\mathbf{D}_f^{-1/2} \mathbf{F} \mathbf{D}_c^{-1/2}) \mathbf{a}_1 &= \lambda_1 \mathbf{a}_1 \\ \mathbf{D}_c^{-1/2} \mathbf{F}^t \mathbf{D}_f^{-1/2} \mathbf{D}_f^{-1/2} \mathbf{F} \mathbf{D}_c^{-1/2} \mathbf{a}_1 &= \lambda_1 \mathbf{a}_1 \\ \mathbf{D}_c^{-1/2} \mathbf{F}^t \mathbf{D}_f^{-1} \mathbf{F} \mathbf{D}_c^{-1/2} \mathbf{a}_1 &= \lambda_1 \mathbf{a}_1 \end{aligned} \quad (4.17)$$

Premultiplicando por $\mathbf{D}_c^{-1/2}$

$$\mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{D}_f^{-1}\mathbf{F}(\mathbf{D}_c^{-1/2}\mathbf{a}_1) = \lambda_1(\mathbf{D}_c^{-1/2}\mathbf{a}_1) \quad (4.18)$$

La matriz $\mathbf{D}_f^{-1}\mathbf{F} = \mathbf{P}$, sus filas suman uno, entonces si $\mathbf{D}_c^{-1/2}\mathbf{a}_1 = \mathbf{1}$, implica que $\mathbf{a}_1 = \mathbf{D}_c^{1/2}\mathbf{1}$, tenemos que

$$\begin{aligned} \mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{P}\mathbf{1} &= \lambda_1\mathbf{1} \\ \mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{1} &= \lambda_1\mathbf{1} \\ \mathbf{1} &= \lambda_1\mathbf{1} \end{aligned} \quad (4.19)$$

Nuevamente se obtiene otro vector de unos, ya que la matriz $\mathbf{D}_c^{-1}\mathbf{F}^t$ es la matriz de frecuencias relativas condicionada al total de la columna, sus filas suman uno, entonces $\lambda_1 = 1$.

La matriz $\mathbf{H}^t\mathbf{H}$ tiene el mismo eigenvalor que la matriz $\mathbf{D}_c^{-1/2}\mathbf{H}^t\mathbf{H}$, como se demuestra a continuación

$$\mathbf{D}_c^{-1/2}\mathbf{H}^t\mathbf{H}\mathbf{a}_1 = \mathbf{D}_c^{-1/2}\mathbf{a}_1 \quad (4.20)$$

premultiplicando por $\mathbf{D}_c^{1/2}$

$$\begin{aligned} \mathbf{D}_c^{1/2}\mathbf{D}_c^{-1/2}\mathbf{H}^t\mathbf{H}\mathbf{a}_1 &= \mathbf{D}_c^{1/2}\mathbf{D}_c^{-1/2}\mathbf{a}_1 \\ \mathbf{H}^t\mathbf{H}\mathbf{a}_1 &= \mathbf{a}_1 \end{aligned} \quad (4.21)$$

El mayor eigenvalor de $\mathbf{H}^t\mathbf{H}$ es $\lambda_1 = 1$ con el eigenvector $\mathbf{a}_1 = \mathbf{D}_c^{1/2}\mathbf{1}$.

No se considera la solución anterior ya que no aporta información sobre la estructura de las filas, sólo se toman los eigenvalores más altos de la matriz $\mathbf{H}^t\mathbf{H}$, pero menores a la unidad, después se calculan los eigenvectores, $\mathbf{a}_i \in \mathbb{R}^J$.

Finalmente, proyectando la matriz \mathbf{W} sobre la dirección \mathbf{a}_i , se obtienen las nuevas coordenadas de la i -ésima fila de la tabla de contingencia.

$$\mathbf{C}_{f_i} = \mathbf{W}\mathbf{a}_i = \mathbf{D}_f^{-1}\mathbf{F}\mathbf{D}_c^{-1/2}\mathbf{a}_i \quad (4.22)$$

Generalizando, las coordenadas de la proyección de cada fila están dadas por las filas de la matriz.

$$\mathbf{C}_f = \mathbf{W}\mathbf{A} = \mathbf{D}_f^{-1}\mathbf{F}\mathbf{D}_c^{-1/2}\mathbf{A} \quad (4.23)$$

La matriz $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_J]$, sus columnas son los eigenvectores de la matriz $\mathbf{H}^t\mathbf{H}$, la primer fila no se contempla porque es la correspondiente al mayor eigenvalor, $\lambda_1 = 1$.

Las coordenadas son vectores en \mathbb{R}^{J-1} , más adelante se verá como identificar los ejes factoriales significativos, r , para determinar la dimensión en la que se trabajará, se espera que $r < J-1$, sea igual a dos o tres, para interpretar fácilmente los resultados.

Representación de las columnas

La proyección de las columnas de la tabla de contingencia es simétrica a la representación de las filas, observe que si se trabaja con la matriz \mathbf{F}^t las filas de esta matriz contiene las columnas de \mathbf{F} y el análisis es idéntico a lo presentado en la sección anterior.

Ahora la matriz \mathbf{D}_c contiene las frecuencias relativas de las filas de \mathbf{F}^t y la matriz \mathbf{D}_f contiene las frecuencias relativas de las columnas de \mathbf{F}^t , haciendo estas modificaciones, la matriz $\tilde{\mathbf{P}} = \mathbf{D}_c^{-1}\mathbf{F}^t$ y el vector $\tilde{\mathbf{W}} = \tilde{\mathbf{P}}\mathbf{D}_f^{-1/2}$, entonces la expresión a maximizar es

$$m = \mathbf{b}^t\tilde{\mathbf{W}}^t\mathbf{D}_c\tilde{\mathbf{W}}\mathbf{b} \quad (4.24)$$

En términos de la matriz \mathbf{H} , se obtiene

$$\begin{aligned} m &= \mathbf{b}^t\tilde{\mathbf{W}}^t\mathbf{D}_c\tilde{\mathbf{W}}\mathbf{b} \\ &= \mathbf{b}^t(\mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{D}_f^{-1/2})^t\mathbf{D}_c\mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{D}_f^{-1/2}\mathbf{b} \\ &= \mathbf{b}^t\mathbf{D}_f^{-1/2}\mathbf{F}\mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{D}_f^{-1/2}\mathbf{b} \\ &= \mathbf{b}^t\mathbf{D}_f^{-1/2}\mathbf{F}\mathbf{D}_c^{-1/2}\mathbf{D}_c^{-1/2}\mathbf{F}^t\mathbf{D}_f^{-1/2}\mathbf{b} \\ &= \mathbf{b}^t\mathbf{H}\mathbf{H}^t\mathbf{b} \end{aligned} \quad (4.25)$$

Entonces

$$\max L = \mathbf{b}^t \mathbf{H} \mathbf{H}^t \mathbf{b} - \lambda (\mathbf{b}^t \mathbf{b} - 1) \quad (4.26)$$

La solución es

$$\mathbf{H} \mathbf{H}^t \mathbf{b} = \lambda \mathbf{b} \quad (4.27)$$

Los eigenvalores no nulos de la matriz $\mathbf{H}^t \mathbf{H}$ son iguales a los de matriz $\mathbf{H} \mathbf{H}^t$, lo que implica que la matriz $\mathbf{H} \mathbf{H}^t$ también tendrá al uno como máximo eigenvalor. Los eigenvectores correspondiente al eigenvalor λ_i de la matriz $\mathbf{H} \mathbf{H}^t$ se denotarán como \mathbf{b}_i .

Las coordenadas de la proyección de las columnas de la tabla de contingencia son las filas de la siguiente matriz

$$\mathbf{C}_c = \tilde{\mathbf{W}} \mathbf{B} = \mathbf{D}_c^{-1} \mathbf{F} \mathbf{D}_f^{-1/2} \mathbf{B} \quad (4.28)$$

Donde $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_I]$, nuevamente la primer fila de la matriz \mathbf{C}_c no se considera.

Relación entre ambas representaciones

Las proyecciones de las filas y columnas de la tabla de contingencia están estrechamente relacionadas, dada la simetría del análisis, algunas propiedades entre ambas representaciones se mencionan a continuación.

1. Las matrices $\mathbf{H}^t \mathbf{H}$ y $\mathbf{H} \mathbf{H}^t$ tienen los mismos eigenvalores no nulos.

Demostración. Si \mathbf{a}_i es el eigenvector de $\mathbf{H}^t \mathbf{H}$ correspondiente al eigenvalor λ_i , entonces

$$\mathbf{H}^t \mathbf{H} \mathbf{a}_i = \lambda_i \mathbf{a}_i \quad (4.29)$$

premultiplicando por \mathbf{H}

$$\mathbf{H} \mathbf{H}^t (\mathbf{H} \mathbf{a}_i) = \lambda_i (\mathbf{H} \mathbf{a}_i) \quad (4.30)$$

Entonces λ_i es eigenvalor de la matriz $\mathbf{H} \mathbf{H}^t$. ■

2. Los eigenvectores de cada matriz que pertenecen al mismo eigenvalor están relacionados, si \mathbf{a}_i es el eigenvector correspondiente al eigenvalor λ_i de $\mathbf{H}^t\mathbf{H}$, entonces el vector $\mathbf{b}_i = \mathbf{H}\mathbf{a}_i$, es el eigenvector correspondiente a λ_i de la matriz $\mathbf{H}\mathbf{H}^t$. Generalizando, la matriz $\mathbf{B} = \mathbf{H}\mathbf{A}$ es la matriz cuyas columnas son los eigenvectores de la matriz $\mathbf{H}\mathbf{H}^t$, se supone $I < J$.

Demostración. Inmediata a partir de la propiedad anterior. ■

3. Se cumple lo siguiente:

$$\mathbf{H}^t\mathbf{H} = \mathbf{W}^t\mathbf{D}_f\mathbf{W} \quad \text{y} \quad \mathbf{H}\mathbf{H}^t = \tilde{\mathbf{W}}^t\mathbf{D}_f\tilde{\mathbf{W}} \quad (4.31)$$

4. Las coordenadas de las columnas se expresan en función de las coordenadas de las filas y viceversa.

Demostración. Primero note que $\mathbf{H}^t\mathbf{H} = \mathbf{Q}\mathbf{D}_\lambda\mathbf{Q}^t$, partiendo de esto la matriz de eigenvectores \mathbf{B} es

$$\mathbf{B} = \mathbf{H}\mathbf{Q} \quad (4.32)$$

De la segunda propiedad se sabe que $\mathbf{B} = \mathbf{H}\mathbf{A}$, esto es, $\mathbf{A} = \mathbf{Q}$. Para normalizar los eigenvectores ubicados en las columnas de \mathbf{B} se postmultiplica por $\mathbf{D}_\lambda^{-1/2}$, la matriz resultante es

$$\tilde{\mathbf{B}} = \mathbf{H}\mathbf{Q}\mathbf{D}_\lambda^{-1/2} \quad (4.33)$$

Donde $\tilde{\mathbf{B}}^t\tilde{\mathbf{B}} = \mathbf{D}_\lambda^{-1/2}\mathbf{Q}^t\mathbf{H}^t\mathbf{H}\mathbf{Q}\mathbf{D}_\lambda^{-1/2} = \mathbf{D}_\lambda^{-1/2}\mathbf{D}_\lambda\mathbf{D}_\lambda^{-1/2} = \mathbf{I}$

Utilizando los eigenvectores normalizados de $\mathbf{H}^t\mathbf{H}$, se tiene que

$$\begin{aligned}
\mathbf{C}_c &= \mathbf{W}\tilde{\mathbf{B}} = \mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{D}_f^{-1/2}\tilde{\mathbf{B}} \\
&= \mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{D}_f^{-1/2}\mathbf{H}\mathbf{Q}\mathbf{D}_\lambda^{-1/2} \\
&= \mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{D}_f^{-1/2}\mathbf{D}_f^{-1/2}\mathbf{F}\mathbf{D}_c^{-1/2}\mathbf{Q}\mathbf{D}_\lambda^{-1/2} \\
&= \mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{D}_f^{-1}\mathbf{F}\mathbf{D}_c^{-1/2}\mathbf{Q}\mathbf{D}_\lambda^{-1/2} \\
&= \mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{C}_f\mathbf{D}_\lambda^{-1/2}
\end{aligned} \tag{4.34}$$

De forma análoga las coordenadas de las filas se expresan como

$$\mathbf{C}_f = \mathbf{D}_f^{-1}\mathbf{F}\mathbf{C}_c\mathbf{D}_\lambda^{-1/2} \tag{4.35}$$

■

Lo encontrado conduce a poder representar las coordenadas de filas y columnas en los mismo ejes factoriales.

Distancia χ^2 y sus propiedades

La distancia χ^2 proviene del contraste de independencia entre las variables fila y columna en una tabla de contingencia, $I \times J$, se realiza por el siguiente estadístico.

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \tag{4.36}$$

Donde f_o = frecuencia observada y f_e = frecuencia esperada.

El estadístico tiene una distribución χ^2 con $(I-1)(J-1)$ grados de libertad. La frecuencia esperada de cada celda, dado que se asume independencia, se obtiene multiplicando $n_{..}f_{i..}f_{.j}$, entonces el estadístico puede ser escrito como

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{..}f_{ij} - n_{..}f_{i..}f_{.j})^2}{n_{..}f_{i..}f_{.j}} \tag{4.37}$$

Desarrollando

$$\begin{aligned}
\chi^2 &= \sum_{i=1}^I \sum_{j=1}^J \frac{n_{..} f_{i.} (f_{ij} - f_{i.} f_{.j})^2}{f_{.j} f_{i.}^2} \\
&= n_{..} \sum_{i=1}^I f_{i.} \sum_{j=1}^J \left[\frac{f_{ij}}{f_{i.}} - f_{.j} \right]^2 \frac{1}{f_{.j}} \\
&= n_{..} \sum_{i=1}^I f_{i.} \sum_{j=1}^J \left[\frac{f_{ij}}{f_{i.} \sqrt{f_{.j}}} - \frac{f_{.j}}{\sqrt{f_{.j}}} \right]^2 \\
&= n_{..} \sum_{i=1}^I f_{i.} \sum_{j=1}^J \left[\frac{f_{ij}}{f_{i.} \sqrt{f_{.j}}} - \sqrt{f_{.j}} \right]^2 \\
&= n_{..} \left[\sum_{i=1}^I \sum_{j=1}^J \frac{f_{ij}}{f_{i.} f_{.j}} - 1 \right] \tag{4.38}
\end{aligned}$$

Una de las razones por la que se trabaja con la distancia χ^2 es porque cuenta con la propiedad de *equivalencia distribucional*, la cual garantiza que no se modifiquen las distancias al unir o separar filas de igual estructura.

La distancia χ^2 entre las filas p_i y $p_{i'}$ es

$$D^2(p_i, p_{i'}) = \sum_{j=1}^J \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2 \frac{1}{f_{.j}} \tag{4.39}$$

Si la fila $p_i = (p_a \cup p_b)$, donde p_a y p_b son de igual estructura, entonces la distancia de p_i a $p_{i'}$ no se modifica ya que $\frac{f_{ij}}{f_{i.}}$ y $f_{.j}$ permanecen sin cambio. Suponga que

$$\frac{f_{aj}}{f_{a.}} = \frac{f_{bj}}{f_{b.}} = \alpha \tag{4.40}$$

Esto implica que $D^2(p_a, p_b) = 0$ y ambos perfiles fila se representan por el mismo punto en \mathbb{R}^{J-1} . Ahora si se unen las filas a y b.

$$\frac{f_{aj}}{f_{a.}} + \frac{f_{bj}}{f_{b.}} = \frac{f_{aj} + f_{bj}}{f_{a.} + f_{b.}} = \alpha \tag{4.41}$$

y la distancia hacia los otros perfiles fila no se modifica.

Antes de ver la siguiente propiedad es necesario primero revisar el concepto de inercia. Cada objeto tiene un centro de gravedad o centroide, cada partícula del objeto tiene masa, r , y distancia, d , hacia el centroide, G_i . La inercia es una medida de dispersión, se define como la sumatoria de las cantidades rd^2 . En nuestro caso la masa o densidad de los perfiles fila es f_i y la distancia de cada punto hacia su centroide (perfil medio) esta dada por la distancia χ^2 . Entonces

$$\begin{aligned}
 \mathbf{IT} &= \sum_{i=1}^I \text{Inercia}(i) = \sum_{i=1}^I f_i \cdot D^2(p_i, G_i) \\
 &= \sum_{i=1}^I f_i \cdot \sum_{j=1}^J \left(\frac{f_{ij}}{f_i} - f_{.j} \right)^2 \frac{1}{f_{.j}} \\
 &= \frac{\chi^2}{n_{..}}
 \end{aligned} \tag{4.42}$$

Note que

$$\text{tr}(\mathbf{H}^t \mathbf{H}) = \sum_{i=1}^I \sum_{j=1}^J \left[\frac{f_{ij}}{f_i \cdot f_{.j}} \right] \tag{4.43}$$

Luego

$$\begin{aligned}
 \chi^2 &= n_{..} [\text{tr}(\mathbf{H}^t \mathbf{H}) - 1] \\
 &= n_{..} [1 + \lambda_2 + \lambda_3 + \dots + \lambda_J - 1] \\
 &= n_{..} [\lambda_2 + \lambda_3 + \dots + \lambda_J]
 \end{aligned} \tag{4.44}$$

Entonces

$$\mathbf{IT} = \lambda_2 + \lambda_3 + \dots + \lambda_J \tag{4.45}$$

La inercia total es la suma de los eigenvalores menores a la unidad de la matriz $\mathbf{H}^t \mathbf{H}$, entonces la proyección de las filas puede verse como la descomposición del estadístico χ^2 en sus fuentes de variación.

Prueba de significación de los ejes factoriales

Las pruebas que existen para determinar el número de ejes factoriales necesarios en el ACS, al igual que en el ACP, son pruebas con fundamento empírico mas que estadístico, de hecho algunos de los criterios establecidos en el ACP se pueden utilizar en el ACS.¹

El método más utilizado es el propuesto por Williams (1952), consiste en descomponer la χ^2 total en suma de χ^2 correspondientes a los eigenvalores. La descomposición se muestra en el Cuadro (4.1).

Si $\chi_r^2 = n_{..}\lambda_r$ con $(I+J-2r+1)$ grados de libertad no es significativa, entonces el eje factorial r no es significativo.

En general, si

$$\chi_r^2 = n_{..}(\lambda_2 + \lambda_3 + \dots + \lambda_r) \quad (4.46)$$

con $[(I+J-3)+ \dots + (I+J-2r+1)]$ es significativa y la χ^2 residual

$$\chi_{J-r}^2 = n_{..}(\lambda_{r+1} + \lambda_{r+2} + \dots + \lambda_J) \quad (4.47)$$

con $[(I+J-2r-1)+ \dots + (I-J+1)]$ grados de libertad no es significativa, entonces r es la cantidad de factores suficientes para la representación.

La tasa de inercia asociada al eje r indica la parte de la inercia total explicada por ese eje factorial.

$$\gamma_r = \frac{\lambda_r}{\lambda_2 + \lambda_3 + \dots + \lambda_J} \quad (4.48)$$

¹Más pruebas de significación de los ejes factoriales se pueden consultar en Cuadras (1991).

Cuadro 4.1: Descomposición de la χ^2

Eigenvalor	χ^2	Grados de libertad
λ_2	$\chi_2^2 = n_{..}\lambda_2$	$I+J-3$
λ_3	$\chi_3^2 = n_{..}\lambda_3$	$I+J-5$
\vdots	\vdots	\vdots
λ_J	$\chi_J^2 = n_{..}\lambda_J$	$I+J-(2J-1) = I-J+1$
$\text{tr}(\mathbf{H}^t\mathbf{H}) - 1$	χ^2	$(I-1)(J-1)$

El porcentaje de la inercia explicada por los ejes factoriales 2, 3,...,r, es:

$$\gamma = \frac{\lambda_2 + \lambda_3 + \dots + \lambda_r}{\lambda_2 + \lambda_3 + \dots + \lambda_J} \quad (4.49)$$

Finalmente, se pueden enumerar los pasos a seguir en el ACS.

1. Calcular cualquiera de las matrices $\mathbf{H}^t\mathbf{H}$ o $\mathbf{H}\mathbf{H}^t$.
2. Obtener sus eigenvalores (son los mismos), recuerde no tomar en cuenta el uno, para después calcular los eigenvectores de cada matriz.
3. Encontrar las coordenadas de la proyección de las filas y columnas.
4. Identificar cuantos factores o ejes son significativos para realizar la representación gráfica.

Interpretación

El ACS no mide la intensidad de la relación entre las variables fila y columna, sino que permite apreciar si existen asociaciones entre las categorías o modalidades de las variables.

A menudo suele ser suficiente dos ejes factoriales ($r=2$), para la representación simultánea de las filas y columnas de la tabla de contingencia. Suponga que dos ejes son suficientes, entonces se pueden considerar los siguientes aspectos a la hora de interpretar.

- Si existen dos filas o columnas con estructura parecida, su ubicación en el plano será cercana.

- Si la estructura es igual entonces el mismo punto en el plano representará ambas filas o columnas.
- La situación cercana de un punto fila y un punto columna sólo se puede interpretar si están alejados del origen.
- Cuando una fila o columna tiene un comportamiento medio se encontrará cerca del origen.

Ejemplo

El Cuadro (4.2) contiene la tabla de contingencia que resume información de la clasificación de los 300 municipios más representativos de México, determinados por el volumen de captación y aplicación de recursos, de acuerdo al porcentaje de sus egresos destinados a obra pública durante el año 2001 y a su región geográfica.

Cuadro 4.2: Tabla de contingencia

Región	Porcentaje de egresos destinado a obra pública ^a			Total
	Bajo (0-24.40)%	Medio (24.41-48.80)%	Alto (48.81-73.20)%	
Norte ^b	54	21	5	80
Centro ^c	78	56	10	144
Sur ^d	38	28	10	76
Total	170	105	25	300

^aFuente:Elaboración propia de acuerdo a las Finanzas públicas estatales y municipales de México (2001), INEGI.

^bMunicipios de Sinaloa, Sonora, Baja California, Baja California Sur, Chihuahua, Durango, Zacatecas, Coahuila, Nuevo León, Tamaulipas.

^cMunicipios de Jalisco, Michoacán, Aguascalientes, Nayarit, Colima, San Luis Potosí, Guanajuato, Querétaro, México, Morelos, Hidalgo, Tlaxcala.

^dMunicipios de Puebla, Oaxaca, Guerrero, Veracruz, Tabasco, Chiapas, Campeche, Yucatán, Quintana Roo.

El paquete estadístico S-PLUS no tiene incorporado dentro de su ambiente desarrollar el ACS, por lo que en el Cuadro (4.3) se presentan los comandos necesarios, para obtener el ACS, los resultados se encuentran en el Cuadro (4.4).

Para calcular el ACS se obtiene primero la matriz \mathbf{H} , para después calcular los eigenvectores de $\mathbf{H}^t\mathbf{H}$ y $\mathbf{H}\mathbf{H}^t$, una vez obtenido esto, se pueden encontrar las matrices \mathbf{C}_f y \mathbf{C}_c , cuyos renglones son las coordenadas de las filas y columnas de la tabla de contingencia, respectivamente.

El primer eje factorial explica 74.72 por ciento de la inercia total, el segundo 25.28 por ciento, ambos ejes explican el total de la inercia.

Las coordenadas de las categorías de las dos variables se muestran en la Figura (4.1), se observa que los municipios de la región norte se asocian más con un bajo porcentaje de egresos destinado a obra pública en el año 2001, mientras que los municipios del centro se asocian con porcentajes medios, los municipios de la región sur se asocian a porcentajes altos y medios.

Cuadro 4.3: Comandos para obtener el ACS en S-PLUS

```

Alto <- c(5,10,10)
Medio <- c(21,56,28)
Bajo <- c(54,78,38)
N<-cbind(Bajo,Medio,Alto)
tabla de contingencia absoluta <- print.crosstabs(N)
Inames <- dimnames(N)[1]
Jnames <- dimnames(N)[2]
I <- nrow(N)
J <- ncol(N)
total <- sum(N)
Fr <- matrix (N*(1/total), nrow=I, ncol=J)
Imass <- as.matrix(apply(Fr,1,sum))
Jmass <- as.matrix(apply(Fr,2,sum))
dI <-diag(Imass,nrow=I,ncol=I)
dJ <-diag(Jmass,nrow=J,ncol=J)
Ih <- Imass^-0.5
Jh <-Jmass^-0.5
dIh <- diag(Ih,nrow=I,ncol=I)
dJh <- diag(Jh,nrow=J,ncol=J)
PF <- dIh^2 %*% Fr
PC <- Fr%*%dJh^2
H <- dIh%*%Fr%*%dJh
eigen.HtH <- eigen(t(H)%*%H)
eigen.HHt <- eigen(H%*%t(H))
A2 <- cbind(c(eigen.HtH$vectors[, 2]), c(eigen.HtH$vectors[, 3]))
Cf <- dIh%*%H%*%A2
B2 <- cbind(c(eigen.HHt$vectors[, 2]), c(eigen.HHt$vectors[, 3]))
Cc <- dJh%*%t(H)%*%B2
cord <- cbind(c(Cf[, 1],Cc[, 1]),c(Cf[, 2],Cc[, 2]))
plot(cord,pch=16,xlab=paste("Principal axis 1"), ylab=paste("Principal axis 2"))
text(cord[1,1],cord[1,2]-0.0045, "Norte")
text(cord[2,1],cord[2,2]-0.0045, "Centro")
text(cord[3,1],cord[3,2]-0.0045, "Sur")
text(cord[4,1],cord[4,2]-0.0045, "Bajo")
text(cord[5,1],cord[15,2]-0.0045, "Medio")
text(cord[6,1],cord[6,2]-0.0045, Alto")
Inercia total <- sum(H^2)-1
Chi.squared <- total *Inercia total
Porcentaje.inercia<-eigen.HtH$values[-1]/Inercia total
Acumulada.inercia <-cumsum(Porcentaje.inercia)
Inercia <-cbind(Porcentaje.inercia, Acumulada.inercia)
dimnames(Inercia)[1]<-list(paste("Eje", 2:min(I,J)))

```

Cuadro 4.4: Resultados del ACS en S-PLUS

```

> H
      [,1]      [,2]      [,3]
[1,] 0.4630462 0.2291288 0.1118034
[2,] 0.4985272 0.4554200 0.1666667
[3,] 0.3343123 0.3134416 0.2294157

> eigen.HtH

$values:
[1] 1.000000000 0.019254778 0.006513483

$vectors:
      [,1]      [,2]      [,3]
[1,] 0.7527727 0.6398695 0.1545974
[2,] 0.5916080 -0.5546195 -0.5851472
[3,] 0.2886751 -0.5319438 0.7960543

> eigen.HHt

$values:
[1] 1.000000000 0.019254778 0.006513483

$vectors:
      [,1]      [,2]      [,3]
[1,] 0.5163978 0.7908292 0.3285156
[2,] 0.6928203 -0.1603451 -0.7030572
[3,] 0.5033223 -0.5906595 0.6307044

> Cf
      [,1]      [,2]
[1,] 0.21250424 0.05134265
[2,] -0.03211474 -0.08189864
[3,] -0.16283968 0.10113148

> Cc
      [,1]      [,2]
[1,] 0.1179497 0.01657467
[2,] -0.1300859 -0.07982478
[3,] -0.2556970 0.22255633

> Inercia total
[1] 0.02576826

> Chi.squared
[1] 7.730478

> Inercia
      Porcentaje.inercia  Acumulada.inercia
Eje 2      0.7472285      0.7472285
Eje 3      0.2527715      1.0000000

```

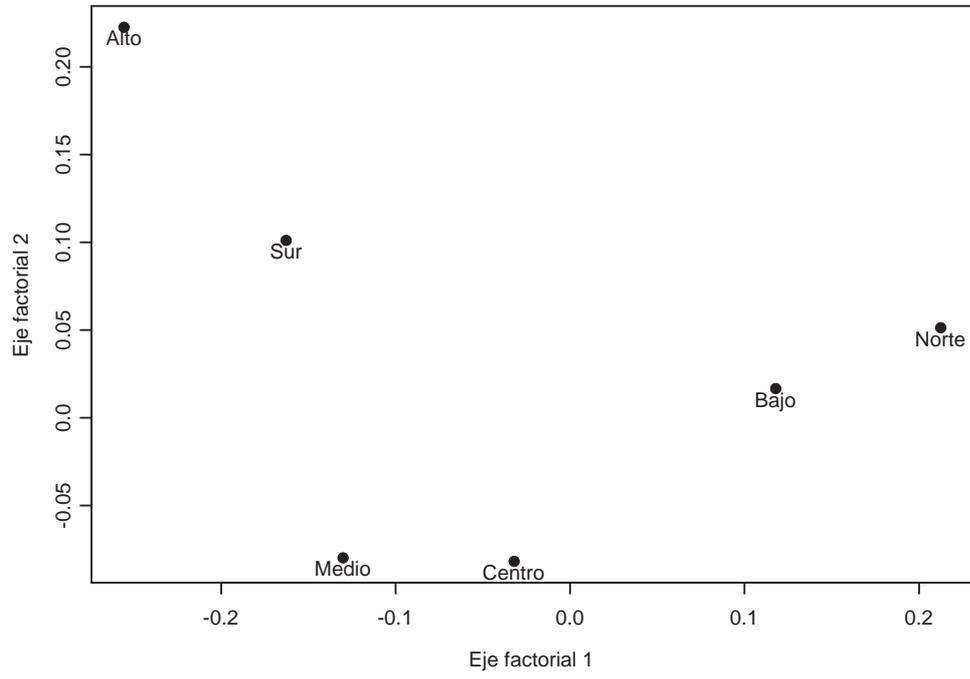


Figura 4.1: Representación gráfica de la tabla de contingencia

V. ANÁLISIS GENERAL

Introducción

La base teórica en común, el Análisis General, de los dos métodos multivariados estudiados en los capítulos anteriores, se origina en las ideas desarrolladas por Eckart y Young (1936), quienes mostraron que una matriz puede ser aproximada por otra de menor rango.

Un conjunto de datos multivariados puede ser representado en un arreglo matricial, como ya se a mencionado, el AG utiliza criterios de álgebra matricial, tales como la descomposición en valores singulares de una matriz y la aproximación por mínimos cuadrados de una matriz por otra de menor rango, con el fin de representar la información observada en una dimensión menor.

En este capítulo se muestra que el ACP y el ACPN pueden ser obtenido a partir del AG aplicando la SVD a las matrices adecuadas, que se estudiarán más adelante. El ACS también es un caso particular del Análisis General, aplicando la SVD a la matriz de residuales estandarizados que se definirá en secciones posteriores.

Descomposición en valores singulares

Iniciemos la sección revisando el teorema de Eckart-Young.

Teorema 12. *Para cualquier matriz real \mathbf{A} , dos matrices ortogonales \mathbf{U} y \mathbf{V} pueden ser encontradas, tal que $\mathbf{U}^t \mathbf{A} \mathbf{V} = \Delta$ sea una matriz diagonal de elementos no negativos.*

La *matriz diagonal*, Δ , (rectangular o cuadrada) fue definida por Eckart y Young como la matriz con el ij -ésimo elemento igual a cero, excepto en el caso de que $i=j$.

Demostración. Sea $\mathbf{A} \in \mathbb{R}^{m \times n}$, suponga que $m \geq n$, el rango de \mathbf{A} es $r(\mathbf{A}) = k \leq n$, recuerde que $r(\mathbf{A}) = r(\mathbf{A}\mathbf{A}^t)$, entonces ambas matrices tienen a lo más k eigenvalores diferentes y distintos de cero. Sea \mathbf{U} una matriz ortogonal de orden $m \times m$ con los eigenvectores de la matriz $\mathbf{A}\mathbf{A}^t$ en sus columnas, y $\Lambda \in \mathbb{R}^{n \times n}$ la matriz diagonal con los k eigenvalores de $\mathbf{A}\mathbf{A}^t$ en la diagonal principal, entonces

$$\mathbf{U}^t \mathbf{A}\mathbf{A}^t \mathbf{U} = \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \quad (5.1)$$

Ahora defina el siguiente producto como

$$\mathbf{U}^t \mathbf{A} = \begin{bmatrix} \mathbf{B} \\ \mathbf{T} \end{bmatrix} \quad (5.2)$$

Donde \mathbf{B} y \mathbf{T} son de orden $n \times n$ y $(m - n) \times n$, respectivamente. Entonces

$$\mathbf{U}^t \mathbf{A}\mathbf{A}^t \mathbf{U} = \begin{bmatrix} \mathbf{B} \\ \mathbf{T} \end{bmatrix} \begin{bmatrix} \mathbf{B}^t & \mathbf{T}^t \end{bmatrix} = \begin{bmatrix} \mathbf{B}\mathbf{B}^t & \mathbf{B}\mathbf{T}^t \\ \mathbf{T}\mathbf{B}^t & \mathbf{T}\mathbf{T}^t \end{bmatrix} \quad (5.3)$$

Obteniendo que

$$\mathbf{B}\mathbf{B}^t = \Lambda \quad \text{y} \quad \mathbf{T}\mathbf{T}^t = 0 \quad (5.4)$$

Los renglones de \mathbf{B} son ortogonales, $\mathbf{b}_i \mathbf{b}_j^t = 0$, para $i, j = 1, 2, \dots, n$ con $i \neq j$ y el producto interno de cada renglón con sí mismo es el elemento ii de la diagonal principal de Λ , $\mathbf{b}_i \mathbf{b}_i^t = \Lambda_{ii}$. La matriz $\mathbf{T}\mathbf{T}^t = 0$, lo que implica que $\mathbf{T} = 0$.

La matriz \mathbf{B} se puede plantear como

$$\mathbf{B} = \Lambda^{1/2} \mathbf{V}^t \quad (5.5)$$

Donde \mathbf{V} es una matriz ortogonal de orden $n \times n$. Sustituyendo lo encontrado en (5.2)

$$\mathbf{U}^t \mathbf{A} = \begin{bmatrix} \Lambda^{1/2} \mathbf{V}^t \\ 0 \end{bmatrix} = \begin{bmatrix} \Lambda^{1/2} \\ 0 \end{bmatrix} \mathbf{V}^t \quad (5.6)$$

Postmultiplicando por \mathbf{V}

$$\mathbf{U}^t \mathbf{A} \mathbf{V} = \begin{bmatrix} \Lambda^{1/2} \\ 0 \end{bmatrix} = \Delta \quad (5.7)$$

Esto muestra que cualquier matriz real \mathbf{A} puede ser transformada en una matriz diagonal. ■

Varios resultados se pueden extraer del teorema anterior, algunos son:

1. La matriz ortogonal \mathbf{U} de orden $m \times m$, contiene en sus columnas los eigenvectores de $\mathbf{A} \mathbf{A}^t$ y los eigenvalores no nulos son los elementos de la diagonal principal de Λ .
2. Las columnas de la matriz ortogonal $\mathbf{V} \in \mathbb{R}^{n \times n}$ son los eigenvectores de $\mathbf{A}^t \mathbf{A}$, los eigenvalores están contenidos en Λ . Observe que

$$\begin{aligned} (\mathbf{U}^t \mathbf{A} \mathbf{V})^t (\mathbf{U}^t \mathbf{A} \mathbf{V}) &= \begin{bmatrix} \Lambda^{1/2} & 0 \end{bmatrix} \begin{bmatrix} \Lambda^{1/2} \\ 0 \end{bmatrix} \\ \mathbf{V}^t \mathbf{A}^t \mathbf{A} \mathbf{V} &= \Lambda \end{aligned} \quad (5.8)$$

3. Los eigenvalores no nulos de $\mathbf{A}^t \mathbf{A}$ y de $\mathbf{A} \mathbf{A}^t$ son los mismos, cuando $m > n$ $\mathbf{A} \mathbf{A}^t$ tiene $(m-n)$ eigenvalores que son ceros.
4. Cuando \mathbf{A} es cuadrada la matriz \mathbf{T} de elementos nulos se elimina en la ecuación (5.7), entonces $\mathbf{U}^t \mathbf{A} \mathbf{V} = \Lambda^{1/2}$.

La importancia del teorema de Eckart-Young radica en que partiendo de su resultado, la matriz \mathbf{A} puede ser factorizada como el producto de tres matrices, una diagonal y dos ortogonales, esta factorización es conocida como la descomposición en valores singulares y se denota como $\text{SVD}(\mathbf{A})$.

Teorema 13. Sea $\mathbf{A} \in \mathbb{R}^{m \times n}$, el rango de la matriz \mathbf{A} es $r(\mathbf{A}) = k \leq n$. Entonces

$$\mathbf{A} = \mathbf{U}_1 \mathbf{D} \mathbf{V}_1^t = \sum_{i=1}^k \alpha_i \mathbf{u}_i \mathbf{v}_i^t, \quad (5.9)$$

donde $\mathbf{U}_1 \in \mathbb{R}^{m \times k}$ y $\mathbf{V}_1 \in \mathbb{R}^{n \times k}$, son matrices semi-ortogonales. La matriz $\mathbf{D} = \text{diag}(\alpha_1, \alpha_2, \dots, \alpha_k)$, tal que $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_k > 0$.

Demostración. Primero note que si $r(\mathbf{A}) = k = n$, la matriz $\Lambda^{1/2}$, de la ecuación (5.7), es de orden $n \times n$, pero si $r(\mathbf{A}) = k < n$, entonces la expresión puede ser escrita como

$$\mathbf{U}^t \mathbf{A} \mathbf{V} = \begin{bmatrix} \Lambda_k^{1/2} & 0 \\ 0 & 0 \end{bmatrix} \quad (5.10)$$

Donde $\Lambda_k^{1/2}$ es la matriz diagonal de orden $k \times k$.

Ahora, premultiplicando por $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$, donde $\mathbf{U}_1 \in \mathbb{R}^{m \times k}$, $\mathbf{U}_2 \in \mathbb{R}^{m \times (m-k)}$ y postmultiplicando por $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2]$, donde $\mathbf{V}_1 \in \mathbb{R}^{n \times k}$, $\mathbf{V}_2 \in \mathbb{R}^{n \times (n-k)}$, entonces

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \Lambda_k^{1/2} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{V}^t \quad (5.11)$$

$$\begin{aligned} &= \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \Lambda_k^{1/2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^t \\ \mathbf{V}_2^t \end{bmatrix} \\ &= \mathbf{U}_1 \mathbf{D} \mathbf{V}_1^t, \end{aligned} \quad (5.12)$$

donde $\mathbf{D} = \Lambda_k^{1/2} = \text{diag}(\alpha_1, \alpha_2, \dots, \alpha_k)$.

La factorización (5.11) es conocida como descomposición en valores singulares completa, mientras que el resultado (5.12) se llama SVD incompleta.

Observe que

$$\begin{aligned}
 \mathbf{A} &= \mathbf{U}_1 \mathbf{D} \mathbf{V}_1^t \\
 &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_k \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \alpha_k \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^t \\ \mathbf{v}_2^t \\ \vdots \\ \mathbf{v}_k^t \end{bmatrix} \\
 &= \sum_{i=1}^k \alpha_i \mathbf{u}_i \mathbf{v}_i^t
 \end{aligned} \tag{5.13}$$

■

Los vectores \mathbf{u}_i se conocen como vectores singulares por la izquierda, los \mathbf{v}_i son llamados vectores singulares por la derecha. Los valores en la diagonal principal de la matriz \mathbf{D} son los valores singulares de \mathbf{A} .

Observe que los vectores singulares por la izquierda son los eigenvectores de $\mathbf{A}\mathbf{A}^t$, mientras que los eigenvectores de $\mathbf{A}^t\mathbf{A}$ son los vectores singulares por la derecha de \mathbf{A} , los eigenvalores de $\mathbf{A}\mathbf{A}^t$ y $\mathbf{A}^t\mathbf{A}$ son el cuadrado de los valores singulares de \mathbf{A} .

Corolario 1. *Si \mathbf{A} es simétrica ($n=m$), entonces los valores singulares corresponden a los eigenvalores de \mathbf{A} y los vectores singulares son los eigenvectores.*

Demostración.

$$\mathbf{A} = \mathbf{U}_1 \mathbf{D} \mathbf{V}_1^t = \mathbf{A}^t = \mathbf{V}_1 \mathbf{D} \mathbf{U}_1^t \tag{5.14}$$

Entonces, $\mathbf{U}_1 = \mathbf{V}_1$, por lo que $\mathbf{A} = \mathbf{U}_1 \mathbf{D} \mathbf{U}_1^t$.

■

Cuando la SVD se aplica a una matriz simétrica la factorización es llamada descomposición espectral.

Teorema 14. *Sea $\mathbf{F} \in \mathbb{R}^{m \times k}$, tal que $\mathbf{F} = \mathbf{U}_1 \mathbf{D}$, además defina la matriz $\mathbf{C} \in \mathbb{R}^{n \times k}$, tal que $\mathbf{C} = \mathbf{V}_1 \mathbf{D}$, entonces las filas de las matrices \mathbf{F} y \mathbf{C} son las coordenadas de las filas y columnas de \mathbf{A} con respecto a las bases \mathbf{V}_1 y \mathbf{U}_1 , respectivamente.*

Demostración.

$$\mathbf{A} = \mathbf{U}_1 \mathbf{D} \mathbf{V}_1^t = \mathbf{F} \mathbf{V}_1^t \quad (5.15)$$

Si se expresa la matriz \mathbf{A} por vectores fila, se tiene que

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^t \\ \mathbf{a}_2^t \\ \vdots \\ \mathbf{a}_m^t \end{bmatrix} = \mathbf{F} \mathbf{V}_1^t = \begin{bmatrix} \mathbf{f}_1^t \\ \mathbf{f}_2^t \\ \vdots \\ \mathbf{f}_m^t \end{bmatrix} \mathbf{V}_1^t = \begin{bmatrix} \mathbf{f}_1^t \mathbf{V}_1^t \\ \mathbf{f}_2^t \mathbf{V}_1^t \\ \vdots \\ \mathbf{f}_m^t \mathbf{V}_1^t \end{bmatrix} \quad (5.16)$$

Entonces

$$\mathbf{a}_i^t = \mathbf{f}_i^t \mathbf{V}_1^t = \begin{bmatrix} f_{i1} & f_{i2} & \dots & f_{ik} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^t \\ \mathbf{v}_2^t \\ \vdots \\ \mathbf{v}_k^t \end{bmatrix} = \sum_{j=1}^k f_{ij} \mathbf{v}_j^t \quad (5.17)$$

Entonces $\mathbf{a}_i = \sum_{j=1}^k f_{ij} \mathbf{v}_j$, para $i = 1, 2, \dots, m$, los coeficientes f_{ij} son las coordenadas de \mathbf{a}_i con respecto a la base \mathbf{V}_1 . Generalizando, las filas de \mathbf{F} contienen las coordenadas de las filas de \mathbf{A} .

Similarmente si $\mathbf{C} = \mathbf{V}_1 \mathbf{D}$, entonces $\mathbf{A} = \mathbf{U}_1 \mathbf{C}^t$, luego si se expresa la matriz \mathbf{A}^t por filas

$$\mathbf{A}^t = \begin{bmatrix} \mathbf{a}_1^t \\ \mathbf{a}_2^t \\ \vdots \\ \mathbf{a}_n^t \end{bmatrix} = \mathbf{C} \mathbf{U}_1^t = \begin{bmatrix} \mathbf{c}_1^t \\ \mathbf{c}_2^t \\ \vdots \\ \mathbf{c}_n^t \end{bmatrix} \mathbf{U}_1^t = \begin{bmatrix} \mathbf{c}_1^t \mathbf{U}_1^t \\ \mathbf{c}_2^t \mathbf{U}_1^t \\ \vdots \\ \mathbf{c}_n^t \mathbf{U}_1^t \end{bmatrix} \quad (5.18)$$

La i -ésima fila de \mathbf{A}^t es

$$\mathbf{a}_i^t = \mathbf{c}_i^t \mathbf{U}_1^t = \begin{bmatrix} c_{i1} & c_{i2} & \dots & c_{ik} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^t \\ \mathbf{u}_2^t \\ \vdots \\ \mathbf{u}_k^t \end{bmatrix} = \sum_{j=1}^k c_{ij} \mathbf{u}_j^t \quad (5.19)$$

En este caso, $\mathbf{a}_i = \sum_{j=1}^k c_{ij} \mathbf{u}_j$, para $i = 1, 2, \dots, n$, las filas de \mathbf{C} contiene las coordenadas de las filas de \mathbf{A}^t , o de las columnas de \mathbf{A} , con respecto a la base \mathbf{U}_1 . ■

Una generalización de la SVD es la descomposición en valores singulares generalizada, presentada en el siguiente teorema.

Teorema 15. Si $\Omega \in \mathbb{R}^{m \times m}$ y $\phi \in \mathbb{R}^{n \times n}$ son matrices definidas positivas, entonces cualquier matriz $\mathbf{A} \in \mathbb{R}^{m \times n}$, se puede factorizar como

$$\mathbf{A} = \mathbf{N} \mathbf{D} \mathbf{M}^t = \sum_{i=1}^k \alpha_i \mathbf{n}_i \mathbf{m}_i^t, \quad (5.20)$$

donde $r(\mathbf{A})=k$, $\mathbf{N} \in \mathbb{R}^{m \times k}$ y $\mathbf{M} \in \mathbb{R}^{n \times k}$. Además las columnas de \mathbf{N} y \mathbf{M} están ortonormalizadas con respecto a Ω y ϕ , esto es, $\mathbf{N}^t \Omega \mathbf{N} = \mathbf{M}^t \phi \mathbf{M} = \mathbf{I}_k$.

Demostración. Defina las matrices $\mathbf{N} = \Omega^{-1/2} \mathbf{U}_1$ y $\mathbf{M} = \phi^{-1/2} \mathbf{V}_1$. Considere la SVD de la matriz $\Omega^{1/2} \mathbf{A} \phi^{1/2}$.

Luego

$$\Omega^{1/2} \mathbf{A} \phi^{1/2} = \mathbf{U}_1 \mathbf{D} \mathbf{V}_1^t \quad (5.21)$$

Entonces

$$\mathbf{N} \mathbf{D} \mathbf{M}^t = \Omega^{-1/2} \mathbf{U}_1 \mathbf{D} \mathbf{V}_1^t \phi^{-1/2} = \Omega^{-1/2} (\Omega^{1/2} \mathbf{A} \phi^{1/2}) \phi^{-1/2} = \mathbf{A} \quad (5.22)$$

■

Observe que

$$\mathbf{N}^t \Omega \mathbf{N} = \mathbf{U}_1^t \Omega^{-1/2} \Omega \Omega^{-1/2} \mathbf{U}_1 = \mathbf{U}_1^t \mathbf{U}_1 = \mathbf{I}_k \quad (5.23)$$

$$\mathbf{M}^t \phi \mathbf{M} = \mathbf{V}_1^t \phi^{-1/2} \phi \phi^{-1/2} \mathbf{V}_1 = \mathbf{V}_1^t \mathbf{V}_1 = \mathbf{I}_k \quad (5.24)$$

En este caso, los elementos de la diagonal principal de \mathbf{D} son los valores singulares generalizados de \mathbf{A} , a las columnas de la matriz \mathbf{N} se les llaman vectores singulares generalizados

por la izquierda y son las bases ortonormales para las columnas de \mathbf{A} , de forma anloga, a las columnas de la matriz \mathbf{M} se les llaman vectores singulares generalizados por la derecha y conforman una base ortonormal para las filas de \mathbf{A} , respectivamente.

Aproximación de una matriz por otra de menor rango

Una interesante aplicación de la SVD es la aproximación de una matriz por otra de menor rango. Considere nuevamente el resultado del teorema de Eckart y Young, la expresión (5.7) y defina lo siguiente:

- Sea $\Lambda_r^{1/2}$ la submatriz diagonal que contiene los primeros r elementos de la matriz $\Lambda_k^{1/2}$, ($r < k$), y sea $\Lambda_s^{1/2}$ la matriz diagonal con $(k-r)$ elementos de $\Lambda_k^{1/2}$.
- Particione la matriz ortogonal \mathbf{U} en dos submatrices, \mathbf{U}_r es la submatriz que contiene las primeras r columnas y \mathbf{U}_s contiene las columnas restantes, $(m-r)$.
- La matriz $\mathbf{V} = [\mathbf{V}_r \ \mathbf{V}_s]$, donde \mathbf{V}_r es la submatriz con las primeras r columnas y \mathbf{V}_s con las últimas $(n-r)$ columnas.

Entonces la descomposición en valores singulares de la matriz \mathbf{A} , es de la siguiente forma

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} \mathbf{U}_r & \mathbf{U}_s \end{bmatrix} \begin{bmatrix} \Lambda_r^{1/2} & 0 \\ 0 & \Lambda_s^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{V}_r^t \\ \mathbf{V}_s^t \end{bmatrix} \\ &= \mathbf{U}_r \Lambda_r^{1/2} \mathbf{V}_r^t + \mathbf{U}_s \Lambda_s^{1/2} \mathbf{V}_s^t \end{aligned} \quad (5.25)$$

Si los últimos valores singulares son muy pequeños, $(k-r)$, de tal forma que puedan ser omitidos, entonces se obtiene la aproximación a la matriz \mathbf{A} con otra matriz de menor rango, ($r < k$), dicha aproximación es de mínimos cuadrados, ya que se minimizan las distancias al cuadrado entre los elementos de ambas matrices, como lo muestra el siguiente teorema.

Teorema 16. Sea $\mathbf{A}_{[r]}$ de orden $m \times n$ de rango r , tal que

$$\mathbf{A}_{[r]} = \mathbf{U}_r \Lambda_r^{1/2} \mathbf{V}_r^t = \sum_{i=1}^r \alpha_i \mathbf{u}_i \mathbf{v}_i^t, \quad (5.26)$$

entonces $\mathbf{A}_{[r]}$ es la matriz de aproximación de mínimos cuadrados, que minimiza la siguiente expresión.

$$\sum_{i=1}^m \sum_{j=1}^n (a_{ij} - x_{ij})^2 = \text{tr}(\mathbf{A} - \mathbf{X})(\mathbf{A} - \mathbf{X})^t = \|\mathbf{A} - \mathbf{X}\|^2, \quad (5.27)$$

donde $r(\mathbf{X}) \leq r$.

Demostración.

$$\begin{aligned} \|\mathbf{A} - \mathbf{X}\|^2 &= \text{tr}[(\mathbf{A} - \mathbf{X})(\mathbf{A} - \mathbf{X})^t] \\ &= \text{tr}[\mathbf{U}\mathbf{U}^t(\mathbf{A} - \mathbf{X})\mathbf{V}\mathbf{V}^t(\mathbf{A} - \mathbf{X})^t] \end{aligned} \quad (5.28)$$

Sea $\mathbf{G} = \mathbf{U}^t \mathbf{X} \mathbf{V}$ y recordando la SVD completa $\mathbf{A} = \mathbf{U} \Delta \mathbf{V}^t$, entonces, $\Delta = \mathbf{U}^t \mathbf{A} \mathbf{V}$, sustituyendo se tiene que

$$\|\mathbf{A} - \mathbf{X}\|^2 = \text{tr}[\mathbf{U} \Delta \mathbf{V}^t \mathbf{A}^t - \mathbf{U} \mathbf{G} \mathbf{V}^t \mathbf{A}^t - \mathbf{U} \Delta \mathbf{V}^t \mathbf{X}^t + \mathbf{U} \mathbf{G} \mathbf{V}^t \mathbf{X}^t] \quad (5.29)$$

La transpuesta de \mathbf{A} es $\mathbf{A}^t = \mathbf{V} \Delta^t \mathbf{U}^t$, ahora si se premultiplica por \mathbf{V}^t se encuentra que $\mathbf{V}^t \mathbf{A}^t = \mathbf{V}^t \mathbf{V} \Delta^t \mathbf{U}^t = \Delta^t \mathbf{U}^t$, además recordando que la $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$, luego la expresión a minimizar es

$$\begin{aligned} \|\mathbf{A} - \mathbf{X}\|^2 &= \text{tr}[\mathbf{U} \Delta \Delta^t \mathbf{U}^t - \mathbf{U} \mathbf{G} \Delta^t \mathbf{U}^t - \mathbf{U} \Delta \mathbf{V}^t \mathbf{X}^t + \mathbf{U} \mathbf{G} \mathbf{V}^t \mathbf{X}^t] \\ &= \text{tr}(\Delta \Delta^t) - \text{tr}(\mathbf{G} \Delta^t) - \text{tr}(\Delta \mathbf{G}^t) + \text{tr}(\mathbf{G} \mathbf{G}^t) \\ &= \text{tr}(\Delta - \mathbf{G})(\Delta - \mathbf{G})^t \\ &= \|\Delta - \mathbf{G}\|^2 \end{aligned} \quad (5.30)$$

Entonces

$$\begin{aligned} \min_x \|\mathbf{A} - \mathbf{X}\|^2 &= \min_x [\text{tr}(\Delta - \mathbf{G})(\Delta - \mathbf{G})^t] \\ &= \min_x \sum_{i=1}^k (\alpha_i - g_{ii})^2 + \sum_{i=1}^m \sum_{\substack{1 \leq j \leq n \\ i \neq j}} g_{ij}^2 \end{aligned} \quad (5.31)$$

La sumatoria anterior será mínima si $g_{ij} = 0$ para todo i, j , tal que $i \neq j$, se minimiza si \mathbf{G} es diagonal. Observe que $r(\mathbf{G}) = r(\mathbf{U}^t \mathbf{X} \mathbf{V}) = r(\mathbf{X}) = r$, entonces $(k - r)$ elementos de la diagonal principal de la matriz \mathbf{G} son ceros. Esto es

$$\min_x \|\mathbf{A} - \mathbf{X}\|^2 = \min_x \sum_{i=1}^r (\alpha_i - g_{ii})^2 + \sum_{i=r}^k \alpha_i^2 \quad (5.32)$$

Además si en (5.32) se cumple que $\alpha_i = g_{ii}$ para $i = 1, 2, \dots, r$ la expresión se minimiza.

Luego $\mathbf{X} = \mathbf{U} \mathbf{G} \mathbf{V}^t$

$$\begin{aligned} \mathbf{X} &= \mathbf{U} \begin{bmatrix} \alpha_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \alpha_2 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \alpha_r & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \mathbf{V}^t \\ &= \mathbf{U}_r \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \alpha_r \end{bmatrix} \mathbf{V}_r^t \end{aligned}$$

$$\begin{aligned}
\mathbf{X} &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_r \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \alpha_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^t \\ \mathbf{v}_2^t \\ \vdots \\ \mathbf{v}_r^t \end{bmatrix} \\
&= \sum_{i=1}^r \alpha_i \mathbf{u}_i \mathbf{v}_i^t \\
&= \mathbf{A}_{[r]}
\end{aligned} \tag{5.33}$$

■

El error de la aproximación de la matriz $\mathbf{A}_{[r]}$ a \mathbf{A} esta dada por

$$\mathbf{E} = \mathbf{A} - \mathbf{U}_r \Lambda_r^{1/2} \mathbf{V}_r^t = \mathbf{U}_s \Lambda_s^{1/2} \mathbf{V}_s^t \tag{5.34}$$

La suma de cuadrados del error es

$$\begin{aligned}
\text{tr}[\mathbf{E}^t \mathbf{E}] &= \text{tr}[\mathbf{V}_s \Lambda_s^{1/2} \mathbf{U}_s^t \mathbf{U}_s \Lambda_s^{1/2} \mathbf{V}_s^t] = \text{tr}[\mathbf{V}_s \Lambda_s \mathbf{V}_s^t] \\
&= \text{tr}[\Lambda_s] = \sum_{i=r+1}^s \alpha_i^2
\end{aligned} \tag{5.35}$$

La suma de cuadrados del error de la aproximación son los eigenvalores de la matriz $\mathbf{A}\mathbf{A}^t$ que no son utilizados en la aproximación. Además observe que

- $\text{tr}[\mathbf{A}\mathbf{A}^t] = \text{tr}[\mathbf{U}\Delta\mathbf{V}^t\mathbf{V}\Delta\mathbf{U}^t] = \text{tr}[\mathbf{U}\Delta^2\mathbf{U}^t] = \text{tr}[\Delta^2] = \sum_{i=1}^k \alpha_i^2$
- $\text{tr}[\mathbf{A}_{[r]}\mathbf{A}_{[r]}^t] = \text{tr}[\mathbf{U}_r\Lambda_r^{1/2}\mathbf{V}_r^t\mathbf{V}_r\Lambda_r^{1/2}\mathbf{U}_r^t] = \text{tr}[\Lambda_r] = \sum_{i=1}^r \alpha_i^2$

Así una medida de calidad de la aproximación de \mathbf{A} por $\mathbf{A}_{[r]}$ es

$$\tau = \left[1 - \frac{\sum_{i=r+1}^s \alpha_i^2}{\sum_{i=1}^k \alpha_i^2} \right] 100 = \left[\frac{\sum_{i=1}^r \alpha_i^2}{\sum_{i=1}^k \alpha_i^2} \right] 100 \tag{5.36}$$

La medida anterior es llamada porcentaje de varianza relativo a los primeros r factores. El valor de r se obtiene fijando la cantidad de aproximación que se desea alcanzar.

Teorema 17. Sea $\mathbf{A} \in \mathbb{R}^{m \times n}$, $r(\mathbf{A})=k$, y su descomposición en valores singulares generalizada es

$$\mathbf{A} = \sum_{i=1}^k \alpha_i \mathbf{n}_i \mathbf{m}_i^t, \quad (5.37)$$

entonces la matriz de aproximación por mínimos cuadrados generalizados o ponderados es

$$\tilde{\mathbf{A}}_{[r]} = \sum_{i=1}^r \alpha_i \mathbf{n}_i \mathbf{m}_i^t = \mathbf{N}_r \mathbf{D}_r \mathbf{M}_r^t \quad (5.38)$$

Demostración. Defina $\tilde{\mathbf{G}} = \mathbf{N}^t \mathbf{X} \mathbf{M}$ y observe que $\mathbf{N}^t \mathbf{A} \mathbf{M} = \Delta$. Entonces minimizar la siguiente norma euclidiana ponderada equivale a

$$\begin{aligned} \|\Omega^{-1/2}(\mathbf{A} - \mathbf{X})\phi^{-1/2}\|^2 &= \text{tr}[\Omega^{-1}(\mathbf{A} - \mathbf{X})\phi^{-1}(\mathbf{A} - \mathbf{X})^t] \\ &= \text{tr}[(\Omega^{-1}\mathbf{A}\phi^{-1} - \Omega^{-1}\mathbf{X}\phi^{-1})(\mathbf{A} - \mathbf{X})^t] \\ &= \text{tr}[\Delta\Delta^t - \tilde{\mathbf{G}}\Delta^t - \Delta\tilde{\mathbf{G}}^t - \tilde{\mathbf{G}}\tilde{\mathbf{G}}^t] \\ &= \text{tr}(\Delta - \tilde{\mathbf{G}})(\Delta - \tilde{\mathbf{G}})^t \\ &= \|\Delta - \tilde{\mathbf{G}}\|^2 \end{aligned} \quad (5.39)$$

Observe que

$$\begin{aligned} \text{tr}[\Omega^{-1}\mathbf{A}\phi^{-1}\mathbf{A}^t] &= \text{tr}[\Omega^{-1/2}\mathbf{A}\phi^{-1/2}\phi^{-1/2}\mathbf{A}^t\Omega^{-1/2}] \\ &= \text{tr}[\mathbf{U}\mathbf{U}^t\Omega^{-1/2}\mathbf{A}\phi^{-1/2}\mathbf{V}\mathbf{V}^t\phi^{-1/2}\mathbf{A}^t\Omega^{-1/2}] \\ &= \text{tr}[\mathbf{U}^t\Omega^{-1/2}\mathbf{A}\mathbf{M}\mathbf{M}^t\mathbf{A}^t\Omega^{-1/2}\mathbf{U}] \\ &= \text{tr}[\mathbf{N}^t\mathbf{A}\mathbf{M}\mathbf{M}^t\mathbf{A}^t\mathbf{N}] \\ &= \text{tr}[\Delta\Delta^t] \end{aligned} \quad (5.40)$$

$$\begin{aligned}
\text{tr}[\Omega^{-1} \mathbf{A} \phi^{-1} \mathbf{X}^t] &= \text{tr}[\Omega^{-1/2} \mathbf{A} \phi^{-1/2} \phi^{-1/2} \mathbf{X}^t \Omega^{-1/2}] \\
&= \text{tr}[\mathbf{U} \mathbf{U}^t \Omega^{-1/2} \mathbf{A} \phi^{-1/2} \mathbf{V} \mathbf{V}^t \phi^{-1/2} \mathbf{X}^t \Omega^{-1/2}] \\
&= \text{tr}[\mathbf{U}^t \Omega^{-1/2} \mathbf{A} \mathbf{M} \mathbf{M}^t \mathbf{X}^t \Omega^{-1/2} \mathbf{U}] \\
&= \text{tr}[\mathbf{N}^t \mathbf{A} \mathbf{M} \mathbf{M}^t \mathbf{X}^t \mathbf{N}] \\
&= \text{tr}[\Delta \mathbf{G}^t]
\end{aligned} \tag{5.41}$$

$$\begin{aligned}
\text{tr}[\Omega^{-1} \mathbf{X} \phi^{-1} \mathbf{A}^t] &= \text{tr}[\Omega^{-1/2} \mathbf{X} \phi^{-1/2} \phi^{-1/2} \mathbf{A}^t \Omega^{-1/2}] \\
&= \text{tr}[\mathbf{U} \mathbf{U}^t \Omega^{-1/2} \mathbf{X} \phi^{-1/2} \mathbf{V} \mathbf{V}^t \phi^{-1/2} \mathbf{A}^t \Omega^{-1/2}] \\
&= \text{tr}[\mathbf{U}^t \Omega^{-1/2} \mathbf{X} \mathbf{M} \mathbf{M}^t \mathbf{A}^t \Omega^{-1/2} \mathbf{U}] \\
&= \text{tr}[\mathbf{N}^t \mathbf{X} \mathbf{M} \mathbf{M}^t \mathbf{A}^t \mathbf{N}] \\
&= \text{tr}[\mathbf{G} \Delta^t]
\end{aligned} \tag{5.42}$$

$$\begin{aligned}
\text{tr}[\Omega^{-1} \mathbf{X} \phi^{-1} \mathbf{X}^t] &= \text{tr}[\Omega^{-1/2} \mathbf{X} \phi^{-1/2} \phi^{-1/2} \mathbf{X}^t \Omega^{-1/2}] \\
&= \text{tr}[\mathbf{U} \mathbf{U}^t \Omega^{-1/2} \mathbf{X} \phi^{-1/2} \mathbf{V} \mathbf{V}^t \phi^{-1/2} \mathbf{X}^t \Omega^{-1/2}] \\
&= \text{tr}[\mathbf{U}^t \Omega^{-1/2} \mathbf{X} \mathbf{M} \mathbf{M}^t \mathbf{X}^t \Omega^{-1/2} \mathbf{U}] \\
&= \text{tr}[\mathbf{N}^t \mathbf{X} \mathbf{M} \mathbf{M}^t \mathbf{X}^t \mathbf{N}] \\
&= \text{tr}[\mathbf{G} \mathbf{G}^t]
\end{aligned} \tag{5.43}$$

Nuevamente se obtiene que la expresión se minimiza cuando

$$\min \|\Omega^{-1/2}(\mathbf{A} - \mathbf{X})\phi^{-1/2}\|^2 = \min \sum_{i=1}^k (\alpha_i - \tilde{g}_{ii})^2 + \sum_{i=1}^m \sum_{j=1}^n \tilde{g}_{ij}^2 \tag{5.44}$$

Por lo tanto

$$\mathbf{X} = \sum_{i=1}^r \alpha_i \mathbf{n}_i \mathbf{m}_i^t = \tilde{\mathbf{A}}_{[r]} \quad (5.45)$$

■

Aplicación al Análisis de Componentes Principales

Sea $\mathbf{X} = (x_{ij})$ la matriz definida en el capítulo 3, como la matriz donde los renglones representan n individuos y las columnas son $j = 1, 2, \dots, m$ variables observadas.

Ahora defina la matriz \mathbf{Y} como

$$\mathbf{Y} = \left[\frac{x_{ij} - \bar{x}_j}{\sqrt{n}} \right] \quad (5.46)$$

Donde $\bar{x}_j = \sum_{i=1}^n \frac{x_{ij}}{n}$

Pues observe que

$$\mathbf{Y}^t \mathbf{Y} = \left[\frac{x_{ij} - \bar{x}_j}{\sqrt{n}} \right]^t \left[\frac{x_{ij} - \bar{x}_j}{\sqrt{n}} \right] = \mathbf{S} \quad (5.47)$$

Es la matriz de varianzas y covarianzas, entonces si se descomponen en valores singulares la matrices \mathbf{Y} , se tiene que

$$\text{SVD}(\mathbf{Y}) = \mathbf{U} \mathbf{D} \mathbf{V}^t \quad (5.48)$$

El interés ahora es encontrar la matriz de menor rango de aproximación a la matriz \mathbf{Y} , esta es

$$\mathbf{A}_{[r]} = \sum_{i=1}^r \alpha_i \mathbf{u}_i \mathbf{v}_i^t = \sum_{i=1}^r \sqrt{\lambda_i} \mathbf{u}_i \mathbf{v}_i^t \quad (5.49)$$

La calidad de la aproximación es

$$\tau = \left[\frac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^k \lambda_i} \right] 100 \quad (5.50)$$

El ACPN se obtiene si se considera la matriz

$$\tilde{\mathbf{Y}} = \left[\frac{x_{ij} - \bar{x}_j}{s_j \sqrt{n}} \right] \quad (5.51)$$

Donde s_j es la desviación estándar de la j -ésima variable. Entonces

$$\tilde{\mathbf{Y}}^t \tilde{\mathbf{Y}} = \left[\frac{x_{ij} - \bar{x}_j}{s_j \sqrt{n}} \right]^t \left[\frac{x_{ij} - \bar{x}_j}{s_j \sqrt{n}} \right] = \mathbf{R} \quad (5.52)$$

Se obtiene la matriz de correlaciones de las variables observadas, de manera que si se aplica la descomposición en valores singulares a $\tilde{\mathbf{Y}}$ se obtendrán los mismos resultados que en el caso del ACP normado.

La matriz de aproximación por mínimos cuadrados de $\tilde{\mathbf{Y}}$, es

$$\mathbf{A}_{[r]} = \sum_{i=1}^r \sqrt{\lambda_i} \mathbf{u}_i \mathbf{v}_i^t \quad (5.53)$$

La calidad de la aproximación es

$$\tau = \left[\frac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^k \lambda_i} \right] 100 \quad (5.54)$$

Ejemplo

Utilizando los datos del ejemplo de la sección 3.7, se muestra a continuación que se obtienen los mismos resultados utilizando el AG. Los comandos en S-PLUS para obtener el ACP a partir del AG se encuentran en el Cuadro (5.1).

Los resultados más útiles de S-PLUS se muestran en el Cuadro (5.2). Se observa que de la SVD de la matriz \mathbf{Y} los valores singulares coinciden con las desviaciones estándar de las componentes, los vectores singulares por la derecha son similares a los eigenvectores o coeficientes de las componentes.

La matriz de aproximación $\mathbf{A}_{[r]}$ de \mathbf{Y} , tal que la calidad de la aproximación sea mayor a 90 por ciento es $\mathbf{A}_{[2]}$, la aproximación es de 96.11 por ciento, el rango $r(\mathbf{A}_{[2]})=2$, reduciendo así la dimensión del conjunto de datos.

Cuadro 5.1: Comandos para obtener el ACP usando el Análisis General

```
X <- egresosdatos2
n <- 31
m <- 12
y1 <- matrix(0,n,m)
a <-sqrt(n)
for(i in 1:m)
y1 [, i] <- (x[, i]-colMeans (x[, i]))/a
sva1 <- svd(y1)
Armenor <- sva1$d[1]*sva1$u[, 1] %*%t(sva1$v[, 1])+
sva1$d[2]*sva1$v[, 2]%*%t(sva1$v[, 2])
Aproximacion <- 100*(sum(Armenor^2)/sum(y1^2))
```

Cuadro 5.2: Resultados del ACP usando el Análisis General

```
> sva1
$d:

[1] 5405.773  3135.61  854.8336  671.457  427.0918  287.5798
[7] 218.1182  186.446  171.8362  125.6533  53.10236  46.34324

$v:

      [, 1]      [, 2]      [, 3]      [, 4]      [, 5]
[1,] 0.7404908405  0.6147405871  0.0676503530 -0.0136087133 -0.1963144238
[2,] 0.0290283696  0.0020514866 -0.0286832393 -0.0339951133  0.0153835359
[3,] 0.0562983533  0.0202315603 -0.0020776029  0.1033670169 -0.0094837508
[4,] 0.5237053406 -0.7745778021  0.1945307853 -0.0151913675 -0.2515175718
[5,] 0.0075098931  0.0079087282  0.0396733949  0.0179838149 -0.0671245090
[6,] 0.1386485912 -0.0549802101 -0.2602849757  0.8897253607  0.2343494449
[7,] 0.0037231044 -0.0013690432  0.0513588337 -0.0124935122 -0.0972272521
[8,] 0.3510677597 -0.0704869959  0.1235317245 -0.1649701535  0.7408494959
[9,] 0.0041905022 -0.0190161869  0.0045961888  0.1540133798 -0.1389964692
[10,] 0.0164821425  0.0214473953  0.1352948822 -0.1565411053 -0.0322654213
[11,] 0.1670532980 -0.1132165600 -0.8954257458 -0.3322521769  0.0751571902
[12,] 0.0515324268  0.0038021640 -0.2223083418  0.0991009339 -0.5041099167
```

Cuadro 5.2: Resultados del ACP usando el Análisis General (continuación)

	[, 6]	[, 7]	[, 8]	[, 9]	[, 10]
[1,]	-0.1418523260	-0.0829355448	0.0141787104	-0.0426032168	-0.0224651781
[2,]	0.1675339918	0.0505192591	-0.0185058481	-0.4843213555	-0.4707793560
[3,]	0.2438074015	0.0880014704	-0.6112730608	-0.3096694621	0.6373558402
[4,]	-0.1250528056	-0.0836402657	-0.0277815118	-0.0282241851	-0.0072366990
[5,]	0.1144606025	-0.0245986512	-0.1005866979	-0.0041399903	0.0916727332
[6,]	-0.0852132187	0.0493795200	-0.1117569211	0.0922613881	-0.1779059751
[7,]	0.4862646811	-0.0739931855	-0.2573516974	-0.3076835085	-0.4500235297
[8,]	0.4327531172	0.1721940759	0.1833317754	0.1560923755	0.0724453146
[9,]	-0.0241767914	0.5964949122	0.5206450439	-0.4896623709	0.2148776457
[10,]	-0.2192248231	0.7432514416	-0.4428626638	0.2874074730	-0.2752722312
[11,]	-0.1109900860	0.0526343928	-0.0718000427	-0.0888478371	-0.0007773758
[12,]	0.6096524123	0.1632821487	0.1738630333	0.4572707126	0.0426091751

	[, 11]	[, 12]
[1,]	0.0237480874	-0.0195215926
[2,]	-0.6474560359	0.3009586348
[3,]	-0.1875832191	-0.0865795107
[4,]	-0.0089644084	-0.0096335854
[5,]	0.3826649621	0.9026629760
[6,]	0.0293122878	0.0162310638
[7,]	0.5466552359	-0.2890619565
[8,]	-0.0052902599	0.0164941572
[9,]	0.2025653893	-0.0460784492
[10,]	0.0098414798	0.0182534632
[11,]	0.1149632109	0.0096083670
[12,]	-0.2108818069	0.0035340232

Para este mismo ejemplo veamos como obtener el ACPN como caso especial del AG. Los comandos necesarios se listan a continuación.

Cuadro 5.3: Comandos para obtener el ACPN usando el Análisis General

```
X <- egresosdatos2
y2 <- matrix(0,n,p)
a <-sqrt(n)
for(i in 1:p)
  y2[, i] <- ((x[, i]-colMeans(x[, i]))/(colStdevs(x[, i],unbiased=F)*a))
sva1cor <- svd(y2)
Arcorm <- sva1cor$d[1]*sva1cor$u[, 1]%*%t(sva1cor$v[, 1])+
sva1cor$d[2]*sva1cor$u[, 2]%*%t(sva1cor$v[, 2]) +
sva1cor$d[3]*sva1cor$u[, 3]%*%t(sva1cor$v[, 3])
Aproximacioncor <- 100*(sum(Arcorm^2)/sum((t(y2)%*%y2)^2))
```

Los principales resultados de S-PLUS se muestran en el Cuadro (5.4). Los vectores singulares

por la derecha de $\tilde{\mathbf{Y}}$ coinciden con los eigenvectores encontrados en el ejemplo de la sección 3.7, para el caso ACPN. De forma similar, los eigenvalores obtenidos anteriormente son las raíces cuadradas de los valores singulares de $\tilde{\mathbf{Y}}$. Entonces nuevamente el AG proporcionará las mismas componentes principales que en el ACPN.

La matriz $\tilde{\mathbf{A}}_{[3]}$ se aproxima a la matriz $\tilde{\mathbf{Y}}$ en 73.46 por ciento .

Cuadro 5.4: Resultados del ACPN usando el Análisis General

\$ d:						
[1]	2.40876	1.32149	1.125414	0.9900632	0.8649852	0.7332368
[7]	0.6588046	0.4539065	0.3648661	0.3042059	0.1906764	0.1297741
\$ v:						
	[, 1]	[, 2]	[, 3]	[, 4]	[, 5]	
[1,]	0.3523866343	-0.1684703939	0.2514605479	0.0306735439	-0.3044249892	
[2,]	0.3485420775	-0.0639415489	-0.0338095253	-0.1462152598	0.3271356155	
[3,]	0.3717823765	-0.1534830882	0.0173715664	0.0562326179	-0.2209006350	
[4,]	0.3116907006	0.2513302745	0.0163510694	0.0026426071	0.4351286914	
[5,]	0.2125082617	-0.5102530059	-0.2088037382	0.1773844372	-0.2769864936	
[6,]	0.3285590447	0.2854957399	-0.1076902163	0.1208460440	-0.2670721686	
[7,]	0.1087930536	-0.4581495560	-0.5166729977	-0.2270087588	0.4404815605	
[8,]	0.3864989294	0.0344348408	0.2106378499	-0.0318213334	0.1260275527	
[9,]	0.0781361769	0.2769540899	-0.3726508556	0.7783641390	0.1789957932	
[10,]	0.0966069482	-0.3536383059	0.5688338538	0.4021425702	0.3067360323	
[11,]	0.3096604294	0.3515891798	0.1098308194	-0.3209623274	0.0877263715	
[12,]	0.3054716805	0.0497387925	-0.3066536867	-0.0728722644	-0.2653209991	
	[, 6]	[, 7]	[, 8]	[, 9]	[, 10]	
[1,]	0.1312366357	-0.2693781033	-0.1830446269	0.3778933321	0.2606610313	
[2,]	0.2142337245	-0.5366069197	-0.2819211107	-0.0250473774	-0.3684196948	
[3,]	-0.1353818017	-0.1733024355	0.3478791876	-0.6204162128	0.4096684345	
[4,]	-0.4138781983	0.4320387229	-0.0931946874	0.0844803808	0.1227872193	
[5,]	-0.2572361925	0.3169458438	-0.4692325122	-0.1862736742	-0.3350041001	
[6,]	-0.3277220141	-0.1406405711	0.4910302373	0.2142857646	-0.5330545720	
[7,]	-0.0001367757	-0.0542137045	0.3328853061	0.1088561628	0.0973783849	
[8,]	-0.2247154957	-0.0090577740	-0.1504574311	0.2825179378	0.3117413685	
[9,]	0.1839812052	-0.1490881208	-0.1548452731	-0.0847255393	0.1481847793	
[10,]	0.2691523086	0.2286730433	0.3273661963	-0.0103572429	-0.2314935306	
[11,]	0.2979904877	0.1785328021	-0.1190296983	-0.4655039933	-0.1688348082	
[12,]	0.5727982469	0.4394872716	0.1313632699	0.2612547977	0.0795267084	

Cuadro 5.4: Resultados del ACPN usando el Análisis General (continuación)

	[, 11]	[, 12]
[1,]	0.1807841253	-0.5690958994
[2,]	-0.4279933883	0.1096760259
[3,]	-0.2494903743	0.0398496073
[4,]	-0.3474710567	-0.3748106713
[5,]	0.0859019353	0.0398409711
[6,]	0.1206644226	-0.0097464378
[7,]	0.3370481350	-0.1306508966
[8,]	0.2863056671	0.6752647164
[9,]	0.1753873647	-0.0219316941
[10,]	0.0198401951	0.0153733412
[11,]	0.5107348564	-0.1297316899
[12,]	-0.3027231361	0.1724759572

Aplicación al Análisis de Correspondencias Simples

Considere la SVD generalizada de la siguiente matriz

$$\tilde{\mathbf{A}} = \mathbf{F} - \mathbf{f} \mathbf{c}^t = \mathbf{N} \mathbf{D} \mathbf{M}^t \quad (5.55)$$

Donde los elementos del vector $\mathbf{f} = \mathbf{D}_f \mathbf{1}$ son las frecuencias marginales relativas de las filas de la tabla de contingencia, el vector $\mathbf{c} = \mathbf{D}_c \mathbf{1}$ contiene las frecuencias marginales relativas de las columnas. Además $\mathbf{N}^t \mathbf{D}_f^{-1} \mathbf{N} = \mathbf{I}$ y $\mathbf{M}^t \mathbf{D}_c^{-1} \mathbf{M} = \mathbf{I}$.

Ahora defina la matriz $\mathbf{A} = \mathbf{D}_f^{-1/2} \tilde{\mathbf{A}} \mathbf{D}_c^{-1/2}$, entonces

$$\text{SVD}(\mathbf{A}) = \mathbf{D}_f^{-1/2} (\mathbf{F} - \mathbf{f} \mathbf{c}^t) \mathbf{D}_c^{-1/2} = \mathbf{U} \mathbf{D} \mathbf{V}^t \quad (5.56)$$

La matriz anterior incluye la representación de las filas y las columnas de la tabla de contingencia, como se muestra a continuación.

La representación por filas de la tabla de contingencia se obtiene a partir de la matriz \mathbf{P} definida en el capítulo 4 como la matriz de perfiles fila, el perfil medio es \mathbf{c} , entonces la distancia de cualquier punto hacia el centro es

$$(\mathbf{D}_f^{-1}\mathbf{F} - \mathbf{1c}^t) \quad (5.57)$$

Si se postmultiplica por $\mathbf{D}_c^{-1/2}$

$$(\mathbf{D}_f^{-1}\mathbf{F} - \mathbf{1c}^t)\mathbf{D}_c^{-1/2} = \mathbf{W} - \mathbf{D}_c^{1/2} \quad (5.58)$$

Obteniendo la matriz \mathbf{W} menos su baricentro, \mathbf{W} es la matriz que puede ser trabajada con observaciones en los renglones y variables en las columnas. Sin embargo no considera la diferencia entre el peso de las filas, luego premultiplicando por $\mathbf{D}_f^{1/2}$

$$\begin{aligned} \mathbf{D}_f^{1/2}(\mathbf{D}_f^{-1}\mathbf{F} - \mathbf{1c}^t)\mathbf{D}_c^{-1/2} &= \mathbf{D}_f^{-1/2}(\mathbf{F} - \mathbf{fc}^t)\mathbf{D}_c^{-1/2} \\ &= \mathbf{H} - \mathbf{f}^{1/2}(\mathbf{c}^{1/2})^t \end{aligned} \quad (5.59)$$

Se encuentra la matriz \mathbf{H} definida en el capítulo 4, entonces la matriz \mathbf{A} es

$$\mathbf{A} = \mathbf{D}_f^{-1/2}(\mathbf{F} - \mathbf{fc}^t)\mathbf{D}_c^{-1/2} = \mathbf{H} - \mathbf{f}^{1/2}(\mathbf{c}^{1/2})^t \quad (5.60)$$

La representación por columnas se obtiene utilizando la matriz $\tilde{\mathbf{P}}$, entonces la distancia de cualquier punto hacia el centro, \mathbf{f} , es

$$(\mathbf{D}_c^{-1}\mathbf{F}^t - \mathbf{1f}^t) \quad (5.61)$$

Postmultiplicando por $\mathbf{D}_f^{-1/2}$ y premultiplicando por $\mathbf{D}_c^{1/2}$, se encuentra

$$\begin{aligned} \mathbf{D}_c^{1/2}(\mathbf{D}_c^{-1}\mathbf{F}^t - \mathbf{1f}^t)\mathbf{D}_f^{-1/2} &= \mathbf{D}_c^{-1/2}(\mathbf{F}^t - \mathbf{cf}^t)\mathbf{D}_f^{-1/2} \\ &= \mathbf{H}^t - \mathbf{c}^{1/2}(\mathbf{f}^{1/2})^t = \mathbf{A}^t \end{aligned} \quad (5.62)$$

Es la transpuesta de la matriz obtenida en la representación de las filas, ambas proyecciones pueden ser resueltas por la SVD(\mathbf{A}) ya que los vectores singulares por la derecha de \mathbf{A} son los vectores singulares por la izquierda de \mathbf{A}^t y viceversa. La matriz \mathbf{A} es conocida como matriz de residuales estandarizados.

Las coordenadas de las filas de \mathbf{A} son las filas de la matriz

$$\mathbf{C}_f = \mathbf{N}\mathbf{D} = \mathbf{D}_f^{-1/2}\mathbf{U}\mathbf{D} \quad (5.63)$$

Las coordenadas de las columnas de \mathbf{A} , las filas de \mathbf{A}^t , son

$$\mathbf{C}_c = \mathbf{M}\mathbf{D} = \mathbf{D}_c^{-1/2}\mathbf{V}\mathbf{D} \quad (5.64)$$

El ij -ésimo elemento de la matriz \mathbf{A} es

$$a_{ij} = \frac{(f_{ij} - f_i c_j)}{\sqrt{f_i c_j}} \quad (5.65)$$

Entonces la χ^2 de la tabla de contingencia es

$$\chi^2 = n \sum_i \sum_j \frac{(f_{ij} - f_i c_j)^2}{f_i c_j} = n \sum_i \sum_j a_{ij}^2 \quad (5.66)$$

La inercia total es

$$\mathbf{IT} = \frac{\chi^2}{n} = \sum_i \sum_j a_{ij}^2 \quad (5.67)$$

La matriz de aproximación de \mathbf{A} es

$$\mathbf{A}_{[r]} = \sum_{i=1}^r \alpha_i \mathbf{u}_i \mathbf{v}_i^t \quad (5.68)$$

La calidad de la aproximación es

$$\tau = \left[\frac{\sum_{i=1}^r \alpha_i^2}{\sum_{i=1}^k \alpha_i^2} \right] 100 \quad (5.69)$$

Ejemplo

Aplicamos lo expuesto al ejemplo de la sección 4.6. Los comandos para obtener el ACS como caso particular del AG se presentan en el Cuadro (5.5), los resultados en S-PLUS se presentan en el Cuadro (5.6).

De la SVD de la matriz de residuales, $\mathbf{D}_f^{-1/2}(\mathbf{F} - \mathbf{f}\mathbf{c}^t)\mathbf{D}_c^{-1/2}$, (vea sv1 en el Cuadro 5.6), se tiene que los valores singulares son las raíces cuadradas de los eigenvalores de la matriz $\mathbf{H}^t\mathbf{H}$, observe que entre los valores singulares no se encuentra la solución trivial, $\lambda = 1$, sino que uno de los valores singulares es cero, el rango de la matriz de residuales es 2.

Cuadro 5.5: Comandos para obtener el ACS usando el Análisis General

```

Alto <- c(5,10,10)
Medio <- c(21,56,28)
Bajo <- c(54,78,38)
N <- cbind(Bajo,Medio,Alto)
total <- sum(N)
Fr <- N/total
Imass <- as.matrix(apply(Fr,1,sum))
Jmass <- as.matrix(apply(Fr,2,sum))
Ih <- Imass^-0.5
Jh <- Jmass^-0.5
dIh <- diag(Ih,nrow=I,ncol=I)
dJh <- diag(Jh,nrow=J,ncol=J)
ItC <- Imass%*%t(Jmass)
Y <- Fr - ItC
mres <- dIh%*%Y%*%dJh
sv1 <- svd(mres)
valoresingulares al cuadrado <- sv1$d^2
coordenadas de las filas <- dIh%*%sv1$u%*%diag(sv1$d)
coordenadas de las columnas <- dJh%*%sv1$v%*%diag(sv1$d)
Inercia total.svd <- sum(mres^2)
Arm <- sv1$d[1]*(sv1$u[, 1])%*%t(sv1$v[, 1])+ sv1$d[2]*(sv1$u[, 2])%*%t(sv1$v[, 2])
aproximacion <- 100*(sum((Arm)^2)/sum((mres)^2))

```

Los vectores singulares por la derecha, correspondientes a los valores singulares diferentes de cero, son iguales a los eigenvectores de $\mathbf{H}^t\mathbf{H}$, correspondientes a los eigenvalores diferentes

de uno, mientras que los vectores singulares por la izquierda correspondientes a los valores singulares diferentes de cero coinciden con los eigenvectores de $\mathbf{H}\mathbf{H}^t$.

Los vectores singulares por la derecha de la matriz de residuales, y por lo tanto los eigenvectores de $\mathbf{H}^t\mathbf{H}$, son los vectores singulares por la izquierda de la SVD de la matriz $\mathbf{D}_c^{-1/2}(\mathbf{F} - \mathbf{f}\mathbf{c}^t)^t\mathbf{D}_f^{-1/2}$.

Las coordenadas de las filas de la tabla de contingencia, contenidas en la matriz $\mathbf{D}_f^{-1/2}\mathbf{U}\mathbf{D}$ son equivalentes a las obtenidas por la matriz \mathbf{C}_f . De forma similar coinciden las coordenadas de las columnas.

La suma de los elementos al cuadrado de la matriz de residuales nos da como resultado la misma inercia total obtenida anteriormente como la traza de la matriz $\text{tr}(\mathbf{H}^t\mathbf{H}) - 1$.

La aproximación de la matriz $\mathbf{A}_{[2]}$ a la matriz de residuales es de 100 por ciento.

Cuadro 5.6: Resultados del ACS usando el Análisis General

```

> mres
      [,1]      [,2]      [,3]
[1,] 0.07431605 -0.07637626 -0.03726780
[2,] -0.02300895  0.04554200 -0.03333333
[3,] -0.04457497  0.01567208  0.08411910

> sv1
$d:
[1] 1.387616e-001  8.070615e-002  1.196741e-017
$v:
      [,1]      [,2]      [,3]
[1,] -0.6398695  0.1545974  0.7527727
[2,]  0.5546195 -0.5851472  0.5916080
[3,]  0.5319438  0.7960543  0.2886751

$u:
      [,1]      [,2]      [,3]
[1,] -0.7908292  0.3285156  0.5163978
[2,]  0.1603451 -0.7030572  0.6928203
[3,]  0.5906595  0.6307044  0.5033223

> valoresingulares al cuadrado
[1] 1.925478e-002  6.513483e-003  1.432189e-034

```

Cuadro 5.6: Resultados del ACS usando el Análisis General (continuación)

```

> coordenadas de las filas

      [, 1]      [, 2]      [, 3]
[1,] -0.21250424  0.05134265  1.196741e-0177
[2,]  0.03211474 -0.08189864  1.196741e-017
[3,]  0.16283968  0.10113148  1.196741e-017

> coordenadas de las columnas

      [, 1]      [, 2]      [, 3]
[1,] -0.1179497  0.01657467  1.196741e-017
[2,]  0.1300859 -0.07982478  1.196741e-0177
[3,]  0.2556970  0.22255633  1.196741e-017

> Inercia total.svd
[1] 0.02576826

> ji cuadrada.svd
[1] 7.730478

> aproximacion
[1] 100

```

Comentarios finales

De lo expuesto en este documento se desprende que los métodos multivariados aquí presentados, el ACP, el ACPN y el ACS, pueden ser obtenidos partiendo del AG, aplicando la SVD a la matriz adecuada, según el método que se busque calcular.

Al retomar los ejemplos utilizados para ilustrar cada uno de los métodos multivariados, se encuentran los mismos resultados si se parte del AG, pero en este último caso se obtienen de manera más directa y sencilla.

Po ejemplo, para obtener el ACS se realizan varias operaciones con la finalidad de expresar la tabla de contingencia como una matriz con observaciones en las filas y variables en las columnas, para después trabajarla como en el ACP. Al utilizar el AG no es necesario realizar

todas esas operaciones, ya que los resultados serán los mismos si se aplica la SVD a la matriz de residuales estandarizados, simplificando notablemente el procedimiento.

El AG es una poderosa herramienta que facilita considerablemente el estudio de los métodos multivariados aquí presentados, puesto que permite abordarlos de forma más homogénea, desde un mismo punto de vista, utilizando la SVD.

Obtener más métodos multivariados por medio del AG puede ser considerado para futuros estudios de investigación.

LITERATURA CITADA

- Cuadras, C. M. (1991), *Métodos de análisis multivariantes*, Mc Graw Hill.
- Escofier, B. y Pagès, J. (1990), *Análisis factoriales simples y múltiples*, Bordas.
- Graybill, F. A. (1969), *Matrices with applications in statistics*, Duxbury.
- Graybill, F. A. (1976), *Theory and application of the linear model*, Duxbury.
- Greenacre, M. J. y Blasius (1994), *Correspondence analysis in the social science*, Academic Press.
- Johnson, D. E. (2000), *Métodos multivariados aplicados al análisis de datos*, International Thomson Editores.
- Johnson, R. A. (1998), *Applied multivariate statistical analysis*, Prentice Hall.
- Johnson, R. M. (1963), *On a theorem stated by Eckart and Young*, Psychometrika, Vol. 28, No. 3.
- Kshirsagar, A. M. (1972), *Multivariate analysis*, Marcel Dekker Inc.
- Lebart, L., Morineau, A. y Warwick, K. (1984), *Multivariate descriptive statistical analysis*, John Wiley and Sons.
- Morrison, D. F. (1976), *Multivariate statistical methods*, Mc Graw Hill.
- Peña, D. (2002), *Análisis de datos multivariantes*, Mc Graw Hill.