

*Comparación de los métodos Tucker 3 y Análisis Factorial  
Múltiple para el análisis de datos tres vías*

GLORIA ROCÍO BAUTISTA MENDOZA



UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE ESTADÍSTICA  
BOGOTÁ, D.C.  
JULIO DE 2009

*Comparación de los métodos Tucker 3 y Análisis Factorial  
Múltiple para el análisis de datos tres vías*

GLORIA ROCÍO BAUTISTA MENDOZA

TESIS PARA OPTAR AL TÍTULO DE  
MAGISTER EN CIENCIAS - ESTADÍSTICA

DIRECTOR  
CAMPO ELÍAS PARDO TURRIAGO  
M.SC. EN ESTADÍSTICA



UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE ESTADÍSTICA  
BOGOTÁ, D.C.  
JULIO DE 2009

## **Título en español**

Comparación de los métodos Tucker 3 y Análisis Factorial Múltiple para el análisis de datos tres vías

## **Title in English**

The Tucker 3 and Multiple Factor Analysis method comparison for three-way data analysis

**Resumen:** Se presenta la comparación entre el Análisis Factorial Múltiple (AFM) y el modelo de Tucker 3, definiendo como criterios de comparación la función de pérdida, inercia, contribuciones, ayudas a la interpretación, representaciones de individuos, variables, condiciones y trayectorias. Se programó en ambiente R para el modelo de Tucker 3, el algoritmo de estimación de parámetros, las medidas de ajuste y las representaciones de individuos, variables, condiciones y trayectorias para los tres modos y para los componentes. Las metodologías comparadas difieren respecto a su conceptualización, la estructura del modelo y el procedimiento de estimación de parámetros. Son similares respecto a la representación de individuos, variables, condiciones y trayectorias de individuos a través de condiciones. El AFM muestra ventajas sobre el Tucker 3 al generar medidas específicas de contribución y calidad de representación de individuos y variables y medidas de relación entre condiciones. El Tucker 3 presenta ventajas sobre el AFM al generar trayectorias adicionales y medidas de relación entre componentes. Del análisis de la situación de salud en Colombia se extraen mediante las dos metodologías tres factores relacionados con desarrollo y pobreza, natalidad y mortalidad y mortalidad materna y neonatal.

**Abstract:** A comparison between the Multiple Factor Analysis (MFA) and the Tucker 3 model is presented. The criteria for comparison used here was the loss function, the inertia, the contributions, the aids to the interpretation, the individual representation, the variables, the conditions and the trajectories. The Tucker 3 algorithm was coded in the R environment, including the calculation of the adjustment measures and representations of individuals, variables, conditions and trajectories of the three modes and components for the model. Both methods differ in their basic framework, the structure of the model and the parameter estimation procedure. They are similar according to the representation of individuals, variables, conditions and trajectories of individuals throughout conditions. The MFA has shown some advantages over the Tucker 3 method as it is possible to calculate contributions and the quality of representation of individuals and variables and measures of relationship among conditions. The Tucker 3 method has shown some advantages over the MFA method as the generation of additional trajectories and measures of relationship between components. The data analysis of situation health in Colombia with both methods, found three factors: development and poverty, natality, and maternal and neonatal mortality.

**Palabras clave:** Análisis Factorial Múltiple, modelo de Tucker 3, datos tres vías, Análisis de componentes principales.

**Keywords:** Multiple Factor Analysis , Tucker 3 Model, Three-Way Data, Principal Component Analysis.

## Nota de aceptación

---

Jurado  
Leonardo Trujillo O

---

Jurado  
Luis Guillermo Díaz M

---

Director  
Campo Elías Pardo T

Bogota, D.C., Julio de 2009

---

Dedicado a

---

A mi familia por su tiempo y constante apoyo.

---

## Agradecimientos

---

A Dios por la oportunidad de vivir esta experiencia.

A Campo Elías Pardo por su dirección y confianza.

A mi familia por su apoyo.

A todos aquellos que de una u otra forma han colaborado, contribuido o aportado en el desarrollo de este trabajo.

---

# Índice general

---

<b>Índice general</b>	<b>I</b>
<b>Índice de tablas</b>	<b>IV</b>
<b>Índice de figuras</b>	<b>V</b>
<b>Introducción</b>	<b>VI</b>
<b>1. Marco Teórico</b>	<b>1</b>
1.1. Definiciones y notación . . . . .	1
1.2. Técnicas analíticas de datos tres vías . . . . .	2
1.3. Modelo de componentes de dos vías . . . . .	3
1.4. Modelo de Tucker 3 . . . . .	4
1.4.1. Definición y fundamentos del método Tucker 3 . . . . .	4
1.4.2. Estimación de parámetros del modelo Tucker 3 . . . . .	6
1.4.3. Procedimiento para implementación del Modelo de Tucker 3 . . . . .	7
1.5. Análisis factorial múltiple (AFM) . . . . .	9
1.5.1. Definición y fundamentos del AFM . . . . .	9
1.5.2. Factores extraídos por el AFM . . . . .	10
1.5.3. Representaciones del AFM y ayudas a la interpretación . . . . .	10
<b>2. Comparación entre el AFM y el Tucker 3</b>	<b>11</b>
2.1. Preprocesamiento de datos . . . . .	11
2.2. Función de pérdida . . . . .	12
2.3. Inercia . . . . .	13
2.4. Ayudas a la interpretación: contribuciones y calidad de la representación . . .	14

---

2.5. Representaciones . . . . .	16
2.5.1. Representación de individuos . . . . .	16
2.5.2. Representación de variables . . . . .	17
2.5.3. Representación de condiciones . . . . .	17
2.5.4. Trayectorias . . . . .	18
2.6. Medidas de relación . . . . .	20
<b>3. Programación del Tucker 3 en ambiente R</b>	<b>21</b>
3.1. Matriciación del arreglo original (función: <i>matriciacion</i> ) . . . . .	21
3.2. Preprocesamiento de datos (función: <i>preproc</i> ) . . . . .	21
3.3. Estimación del algoritmo de Tucker 3 y selección del mejor modelo (función: <i>tucker3</i> ) . . . . .	22
3.4. Cálculo de ajuste y contribución de los componentes (función: <i>ajuste</i> ) . . . . .	23
3.5. Representación de las entidades de los modos A, B y C (función: <i>conf.mode</i> ) . . . . .	23
3.6. Representación de las trayectorias de la combinación de las entidades de dos modos (función: <i>traj.modes</i> ) . . . . .	23
3.7. Representación de los elementos del arreglo central $\mathbf{G}_a$ (función: <i>repcore</i> ) . . . . .	24
<b>4. Aplicación y comparación</b>	<b>25</b>
4.1. Objetivo y descripción . . . . .	25
4.2. Preprocesamiento de datos . . . . .	25
4.3. Elección de la dimensionalidad de la solución e inercia . . . . .	26
4.4. Contribución de los componentes . . . . .	27
4.5. Estimación de componentes, ayudas a la interpretación y representación de individuos, variables y condiciones . . . . .	28
4.5.1. Individuos . . . . .	28
4.5.2. Variables . . . . .	33
4.5.3. Condiciones . . . . .	37
4.6. Trayectorias . . . . .	39
4.7. Medidas de relación . . . . .	44
<b>5. Conclusiones y Recomendaciones</b>	<b>48</b>
<b>A. Apéndices</b>	<b>49</b>
A.1. Programación del modelo de Tucker 3 en ambiente R . . . . .	49
A.1.1. Matriciación del arreglo de datos original (función: <i>matriciacion</i> ) . . . . .	49



---

A.1.2. Preprocesamiento (función: <i>preproc</i> ) . . . . .	50
A.1.3. Estimación de parámetros del modelo de Tucker 3 (función: <i>tucker3</i> )	51
A.1.4. Cálculo de medidas de ajuste del modelo de Tucker 3 (función: <i>ajuste</i> )	53
A.1.5. Representaciones de las entidades de cada modo (función: <i>conf.mode</i> )	53
A.1.6. Representaciones de las trayectorias de dos modos (función: <i>traj.modes</i> )	54
A.1.7. Representación de los elementos del arreglo central (función: <i>repcore</i> )	55
A.2. Datos de las variables demográficas y socioeconómicas evaluadas en 24 departamentos de Colombia durante el periodo 2004-2007 . . . . .	56
A.3. Definición de las variables demográficas y socioeconómicas evaluadas en 24 departamentos de Colombia durante el periodo 2004-2007 . . . . .	58
<b>Bibliografía</b>	<b>60</b>

---

## Índice de tablas

---

4.1. Diez primeros valores propios del análisis global del AFM . . . . .	26
4.2. Contribución de triadas de componentes a la inercia total . . . . .	28
4.3. Coordenadas de individuos y ayudas a la interpretación (contribuciones y cosenos) en AFM . . . . .	29
4.4. Matriz de componentes para individuos en Tucker 3 . . . . .	30
4.5. Coordenadas (correlaciones) de variables y ayudas a la interpretación (contribuciones y cosenos) en AFM . . . . .	34
4.6. Matriz de componentes para variables en Tucker 3 . . . . .	35
4.7. Coordenadas de condiciones (años 2004-2007) y ayudas a la interpretación en AFM . . . . .	37
4.8. Matriz de componentes para condiciones (años 2004-2007) en Tucker 3 . . . .	37
4.9. Razón de inercia = Inercia inter/Inercia total . . . . .	39
4.10. Contribución de los ejes parciales en AFM . . . . .	44
4.11. Correlación entre ejes parciales en AFM . . . . .	44
4.12. Coeficientes RV en AFM . . . . .	46
4.13. Arreglo central $\mathbf{G}_a$ en Tucker 3 . . . . .	46

---

## Índice de figuras

---

1.1. Representación de un arreglo de datos tres vías . . . . .	1
1.2. Transformación de datos tres vías a datos dos vías. . . . .	2
1.3. Representación del modelo de Tucker 3. . . . .	5
4.1. <i>Scree plot multiway</i> . . . . .	27
4.2. Departamentos de Colombia en plano factorial (1,2) en AFM . . . . .	31
4.3. Departamentos de Colombia en plano factorial (1,2) en Tucker 3 . . . . .	31
4.4. Departamentos de Colombia en plano factorial (2,3) en AFM . . . . .	32
4.5. Departamentos de Colombia en plano factorial (2,3) en Tucker 3 . . . . .	32
4.6. Círculo de correlaciones (coordenadas) para las variables en el plano factorial (1,2) en AFM . . . . .	35
4.7. Variables sobre el plano (1,2) en Tucker 3 . . . . .	36
4.8. Condiciones (años 2004-2007) sobre el plano (1,2) en AFM . . . . .	38
4.9. Condiciones (años 2004-2007) sobre el plano (1,2) en Tucker 3 . . . . .	38
4.10. Trayectorias de los departamentos en el periodo 2004-2007 en AFM, plano (1,2) . . . . .	40
4.11. Trayectorias de los departamentos en el periodo 2004-2007 en Tucker 3, plano (1,2) . . . . .	41
4.12. Trayectorias de los departamentos en el periodo 2004-2007 en AFM, plano (2,3) . . . . .	42
4.13. Trayectorias de los departamentos en el periodo 2004-2007 en Tucker 3, plano (2,3) . . . . .	42
4.14. Trayectorias de las variables en el periodo 2004-2007 en Tucker 3, plano (1,2)	43
4.15. Correlación ejes parciales eje general . . . . .	45
4.16. Trayectorias de los elementos del arreglo central en Tucker 3 . . . . .	47

---

## Introducción

---

El análisis multivariado se encarga del estudio de datos obtenidos de la medición de más de una variable sobre un conjunto de individuos u objetos, representados en un arreglo dos vías denominado tabla o matriz. Estos datos se obtienen de estudios e investigaciones en diferentes áreas del conocimiento como las ciencias ambientales, sociales, del comportamiento, químicas y biomédicas, entre otras.

El análisis de datos tres vías se realiza cuando la información se presenta en arreglos donde cada dato se indexa con tres índices: uno que identifica el individuo  $i$  ( $i = 1, \dots, I$ ), otro la variable  $j$  ( $j = 1, \dots, J$ ) y un tercero que corresponde a la condición  $k$  ( $k = 1, \dots, K$ ); representándose en un arreglo tridimensional: individuos, variables y condiciones. Este tipo de datos ocurren frecuentemente, Bro (1998) destaca algunas aplicaciones en áreas como cromatografía, análisis ambiental y análisis de percepción.

Los métodos utilizados para el análisis de datos tres vías son de carácter exploratorio, porque identifican patrones de la estructura interna presente entre los elementos de las tres vías del arreglo, sin aplicar pruebas a estos patrones porque probar hipótesis supone un supuesto distribucional que en muchas situaciones se desconoce (Kroonenberg 2008). Estos métodos describen la interacción entre las tres vías, prestando atención especial a las diferencias entre individuos, lo cual permite generar conclusiones más completas que las obtenidas al realizar análisis de dos vías aislados (individuos y variables).

Las investigaciones sobre este tipo de datos muestran diferentes métodos de análisis provenientes de escuelas como la anglosajona y la francesa. Dentro de los métodos implementados por la escuela anglosajona se destaca la familia de modelos Tucker: Tucker 1, 2 y 3 (Tucker 1966) y el modelo PARAFAC (PARAllel FACtor analysis, Harshman (1970)). De la escuela francesa por su parte, sobresalen los métodos CANDECOM (CANonical DECOMposition, Carroll & Chang (1970)), STATIS (Structuration des Tableaux A Trois Indices de la Statistique, Lavit et al. (1994)) y AFM (Análisis Factorial Múltiple, Escofier & Pagés (1982)). Entre algunos estudios que realizan comparaciones entre métodos, se destacan: la comparación entre STATIS y AFM realizado por Pagés (1996), la relación jerárquica entre algunos métodos tres vías presentado por Kiers (1991) y una comparación empírica entre algunos métodos tres vías planteado por Mures et al. (2006); estos dos últimos estudios se caracterizan por considerar métodos de las dos escuelas de análisis de datos.

Los estudios de comparación son importantes porque se toman como referencia en la elección del método más apropiado para analizar un conjunto de datos. Este trabajo

pretende dar un aporte a este tipo de estudios, mediante la comparación entre el modelo de Tucker 3 y el AFM pertenecientes a las escuelas anglosajona y francesa respectivamente.

Los métodos Tucker 3 y AFM se fundamentan en el análisis en componentes principales (ACP), el primero corresponde a la generalización del ACP al caso tres vías y el segundo a un ACP ponderado. Esta fundamentación común hace que los dos métodos sean comparables. La estimación de los parámetros en el modelo de Tucker 3 se realiza utilizando algoritmos mínimos cuadrados alternantes, se destaca el algoritmo de Tucker 3 (Bro 1998), el cual está implementado en lenguaje MATLAB y SAS (Amaya & Pacheco 2002). El AFM por su parte, desarrolla una diagonalización y se encuentra implementado, entre otros, en el paquete FactoMineR (Husson et al. 2008) del entorno R (R Development Core Team 2008).

Este estudio presenta la comparación entre los dos métodos mediante la expresión del modelo y la función de pérdida, la inercia, las contribuciones y ayudas a la interpretación y las representaciones de individuos, variables, condiciones y trayectorias, desarrollando la temática en cinco capítulos. En el primero, se presenta una revisión teórica general de los métodos para análisis de datos tres vías y de manera específica para los métodos bajo estudio. Los capítulos dos, tres y cuatro presentan los resultados obtenidos, los cuales se estructuran definiendo la comparación entre los dos métodos, la programación del modelo de Tucker 3, la aplicación y comparación de los dos métodos estudiados en una situación con datos reales, que describe características sociodemográficas y de salud de algunos departamentos de Colombia en el periodo 2004-2007. Finalmente, el capítulo cinco presenta las conclusiones y recomendaciones generadas del estudio.

# CAPÍTULO 1

---

## Marco Teórico

---

### 1.1. Definiciones y notación

-**Arreglo de datos tres vías.** Es cualquier conjunto de datos que se puede representar en un bloque tridimensional como el que se presenta en la figura 1.1, donde  $i$  ( $i = 1, \dots, I$ ),  $j$  ( $j = 1, \dots, J$ ) y  $k$  ( $k = 1, \dots, K$ ) representan los modos individuo, variable y condición, respectivamente. En adelante se nota como  $\underline{\mathbf{X}}$ , con elemento  $x_{ijk}$  (Kiers 2000b).

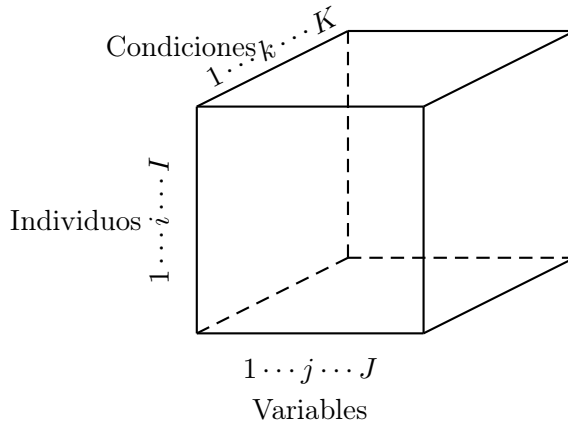


Figura 1.1: Representación de un arreglo de datos tres vías

-**Modo (vía).** Es la unidad básica utilizada en la construcción de un arreglo. Una matriz tiene dos modos: individuo y variable. Un arreglo *tres vías* tiene tres modos: individuo (modo A), variable (modo B) y condición (modo C).

-**Transformación de arreglos de datos tres vías a arreglos de datos dos vías.** La figura 1.2 presenta algunas opciones para transformar datos tres vías (a) en datos dos vías (b), (c), (d) y (e). La matriciación (b) es la transformación de un arreglo *tres vías* en una matriz utilizando el modo de combinación ancho. La matriciación del modo A, conserva el primer modo (individuos) en filas, mientras el segundo y el tercer modo

(variables y condiciones) se combinan en columnas. Las matriciaciones de los modos B y C se pueden desarrollar de forma equivalente conservando los modos B y C respectivamente en filas y confundiendo los modos restantes en columnas, según sea el caso. Otros arreglos dos vías generados a partir de un arreglo tres vías son: las matrices de productos cruzados de cada condición o matrices de covarianza (c), la matriz promedio (d) y la matriz del modo combinación alto (e) (figura 1.2) (Kroonenberg 2008).

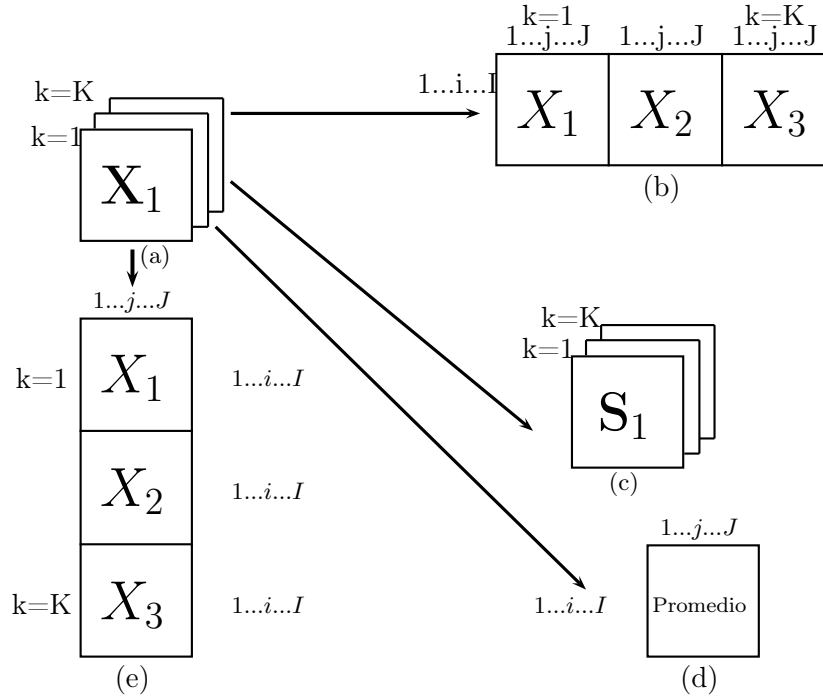


Figura 1.2: Transformación de datos tres vías a datos dos vías.

- **Función de pérdida.** Es la función que cuantifica la diferencia entre los datos observados y los ajustados por el modelo, también llamada función de discrepancia.

## 1.2. Técnicas analíticas de datos tres vías

El análisis de datos *tres vías* tiene su origen en la década del 60 con los trabajos de Ledyard R. Tucker, con el análisis de factor de tres modos, actualmente conocido como el análisis de componentes de tres modos (Tucker 1966).

Kroonenberg (2008) identifica el análisis de componentes *tres vías* como una técnica de interdependencia. Dentro de estos, el método de Tucker 3 se define como un modelo de análisis de datos. Aunque la clasificación recientemente presentada por Kroonenberg no incluye los métodos propuestos por la escuela francesa (STATIS Y AFM), estos se definen y estudian como métodos de análisis de datos, los cuales son procedimientos ejecutados en varias etapas que aunque no se centran directamente en el ajuste de un modelo, puede asignársele uno subyacente.

Kiers (1991) divide los métodos para análisis de datos tres vías en asimétricos y simétricos. La asimetría asignada a los primeros está dada porque uno de los modos (generalmente las condiciones) tiene un manejo diferente a los otros dos (individuos y variables). Los métodos simétricos por su parte, manejan por igual los tres modos del arreglo.

Los métodos asimétricos se ubican dentro del enfoque Interestructura- Compromiso- Intraestructura (ICI) y corresponde con las etapas en las que se desarrollan los métodos STATIS y AFM. Tienen como objetivo principal la definición de tipologías de objetos. La interestructura estudia la relación entre las tablas que forman la tabla global, a través de los operadores que las representan. El compromiso busca obtener una nube media que represente la estructura común de esas tablas. La intraestructura analiza las relaciones entre individuos y variables en la tabla compromiso. Los métodos simétricos se identifican como los modelos de componentes, entre estos se destacan los modelos Tucker 1, Tucker 2, Tucker 3 y PARAFAC. Se basan en la idea que existe una estructura factorial subyacente asociada a cada modo. Es decir, los tres modos observados de los datos están asociados cada uno con otro modo (latente) que puede considerarse como un conjunto de factores (Mures, Vallejo & García 2006).

### 1.3. Modelo de componentes de dos vías

Kroonenberg (2008) presenta el modelo de análisis en componentes principales (ACP) para una matriz de datos de dos vías  $\mathbf{X} = (x_{ij})$  como un modelo multiplicativo. La descomposición de  $\mathbf{X}$  se expresa como:

$$x_{ij} = \hat{x}_{ij} + e_{ij} = \sum_{s=1}^S a_{is}f_{js} + e_{ij}, \quad i = 1, \dots, I; j = 1, \dots, J \quad (1.1)$$

donde  $s$  representa un componente,  $S$  corresponde al total de componentes extraídos en la solución,  $a_{is}$  es el coeficiente del  $i$ -ésimo individuo sobre el  $s$ -ésimo componente,  $f_{js}$  corresponde al coeficiente de la  $j$ -ésima variable sobre el  $s$ -ésimo componente y  $e_{ij}$  es el residual para el  $i$ -ésimo individuo en la  $j$ -ésima variable.

En notación matricial:

$$\mathbf{X} = \hat{\mathbf{X}} + \mathbf{E} = \mathbf{A}\mathbf{F}' + \mathbf{E}$$

donde  $\mathbf{A} = (\mathbf{a}_{is})$  es una matriz normalizada  $I \times S$  que contiene los coeficientes de los individuos sobre los  $S$  componentes,  $\mathbf{F} = (f_{js})$  es una matriz  $J \times S$  que contiene los coeficientes de las variables sobre los  $S$  componentes,  $\mathbf{E}$  es la matriz de residuales  $I \times J$ . La matriz  $\hat{\mathbf{X}}$  se conoce como base o imagen estructural de la matriz de datos (Bro 1998).

El modelo (1.1) se descompone en valores singulares, lo que permite escribir los coeficientes de las variables ( $f_{js}$ ) como el producto de los coeficientes normalizados de las variables ( $b_{js}$ ) y la variabilidad de los componentes ( $g_{ss}$ ), en la forma:

$$x_{ij} = \sum_{s=1}^S a_{is}f_{js} + e_{ij} = \sum_{s=1}^S a_{is} [g_{ss}b_{js}] + e_{ij}$$

o, en notación matricial:

$$\mathbf{X} = \mathbf{A}\mathbf{G}\mathbf{B}' + \mathbf{E}$$



donde  $\mathbf{B} = (b_{js})$  es una matriz normalizada  $J \times S$  que contiene los puntajes de las variables sobre los  $S$  componentes y  $\mathbf{G} = (g_{ss})$  es una matriz diagonal  $S \times S$  que contiene los valores singulares de la matriz  $\mathbf{X}$ . Los componentes de  $\mathbf{A}$  se llaman **vectores singulares izquierdos** y los de  $\mathbf{B}$ , **vectores singulares derechos**.

La descomposición en valores singulares se puede escribir como la suma de matrices de rango uno, de la forma:

$$\mathbf{X} = \sum_{s=1}^S g_{ss} (\mathbf{a}_s \otimes \mathbf{b}_s) + \mathbf{E}$$

donde  $g_{ss}$  corresponde al  $s$ -ésimo valor singular,  $\mathbf{a}_s$  es el vector singular izquierdo y  $\mathbf{b}_s$  es el vector singular derecho. Cada producto  $(\mathbf{a}_s \otimes \mathbf{b}_s)$  genera una matriz del mismo tamaño que la matriz de datos pero de rango uno. Para obtener la representación completa de la matriz de datos original se requieren  $S$  aproximaciones, cada una ponderada por su valor singular ( $g_{ss}$ ).

En términos de la notación presentada por Lebart et al. (1995) este modelo se escribe como:

$$\mathbf{X} = \sum_{s=1}^S \sqrt{\lambda_s} \mathbf{v}_s \mathbf{u}'_s + \mathbf{E}$$

donde  $\sqrt{\lambda_s}$  son los valores singulares obtenidos de la descomposición,  $\mathbf{v}_s$  es un vector del conjunto de vectores singulares izquierdos y  $\mathbf{u}_s$  es un vector del conjunto de vectores singulares derechos.

## 1.4. Modelo de Tucker 3

### 1.4.1. Definición y fundamentos del método Tucker 3

El método Tucker 3 es una generalización del modelo de componentes dos vías al modelo de componentes tres vías. La generalización se puede representar de manera análoga al modelo (1.1) como:

$$x_{ijk} = \hat{x}_{ijk} + e_{ijk}; \quad i = 1, \dots, I; \quad j = 1, \dots, J; \quad k = 1, \dots, K$$

En este modelo, se realiza una descomposición completa del arreglo tridimensional  $\mathbf{X}$  en tres matrices ortonormales: una matriz  $\mathbf{A}$  de dimensión  $I \times P$  que contiene los coeficientes para los componentes del modo individuos, una matriz  $\mathbf{B}$  de tamaño  $J \times Q$  que contiene los coeficientes para los componentes del modo variables, una matriz  $\mathbf{C}$  de tamaño  $K \times R$  que contiene los coeficientes para los componentes del modo ocasiones, y un arreglo central  $\mathbf{G}$  de dimensión  $P \times Q \times R$  (figura 1.3). El objetivo principal del modelo es estimar los parámetros  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  y  $\mathbf{G}$ .

En su expresión general, utilizando sumatoria, el modelo de Tucker 3 se representa como:

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} (a_{ip} b_{jq} c_{kr}) + e_{ijk}; \quad i = 1, \dots, I; \quad j = 1, \dots, J; \quad k = 1, \dots, K$$

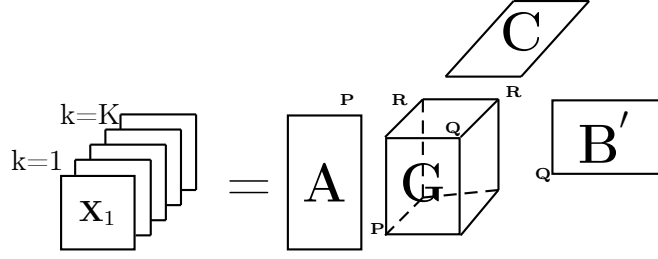


Figura 1.3: Representación del modelo de Tucker 3.

$a_{ip}, b_{jq}, c_{kr}$  denotan los elementos de  $\mathbf{A}$ ,  $\mathbf{B}$  y  $\mathbf{C}$ , respectivamente y  $g_{pqr}$  denota los elementos del arreglo central  $\underline{\mathbf{G}}$ .

Utilizando la matriciación del modo A, el modelo toma la forma:

$$x_{ijk} = \sum_{p=1}^P a_{ip} \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} (b_{jq} c_{kr}) + e_{ijk}$$

que en notación matricial se expresa como:

$$\mathbf{X}_a = \mathbf{A} \mathbf{G}_a (\mathbf{C}' \otimes \mathbf{B}') + \mathbf{E}_a$$

donde  $\mathbf{X}_a$  es una matriz de dimensión  $I \times JK$  obtenida por matriciación del modo A del arreglo  $\underline{\mathbf{X}}$ ,  $\mathbf{G}_a$  es una matriz de dimensión  $P \times QR$  obtenida por matriciación por el modo A del arreglo central y  $\mathbf{E}_a$  es la matriz de residuales de dimensión  $I \times JK$  (Tucker 1966).

Análogamente, utilizando la matriciación del modo B, el modelo tiene la forma:

$$x_{ijk} = \sum_{q=1}^Q b_{jq} \sum_{p=1}^P \sum_{r=1}^R g_{pqr} (a_{ip} c_{kr}) + e_{ijk}$$

y en notación matricial:

$$\mathbf{X}_b = \mathbf{B} \mathbf{G}_b (\mathbf{C}' \otimes \mathbf{A}') + \mathbf{E}_b$$

donde  $\mathbf{X}_b$  es una matriz de dimensión  $J \times IK$  obtenida por matriciación del modo B del arreglo  $\underline{\mathbf{X}}$ ,  $\mathbf{G}_b$  es una matriz de dimensión  $Q \times PR$  obtenida por matriciación por el modo B del arreglo central y  $\mathbf{E}_b$  es la matriz de residuales de dimensión  $J \times IK$ .

Utilizando la matriciación del modo C, el modelo es:

$$x_{ijk} = \sum_{r=1}^R c_{kr} \sum_{p=1}^P \sum_{q=1}^Q g_{pqr} (a_{ip} b_{jq}) + e_{ijk}$$

y en notación matricial:

$$\mathbf{X}_c = \mathbf{C} \mathbf{G}_c (\mathbf{B}' \otimes \mathbf{A}') + \mathbf{E}_c$$

donde  $\mathbf{X}_c$  es una matriz de dimensión  $K \times IJ$  obtenida por matriciación del modo C del arreglo  $\underline{\mathbf{X}}$ ,  $\mathbf{G}_c$  es una matriz de dimensión  $R \times PQ$  obtenida por matriciación por el modo C del arreglo central y  $\mathbf{E}_c$  es la matriz de residuales de dimensión  $K \times IJ$ .

El modelo de Tucker 3 es una descomposición en valores singulares de orden superior, que para construir la imagen estructural del arreglo  $\underline{\mathbf{X}}$  necesita  $P \times Q \times R$  términos de la forma  $(\mathbf{a}_p \otimes \mathbf{b}_q \otimes \mathbf{c}_r)$ , cada uno ponderado por su propio  $g_{pqr}$ . Utilizando la notación de rango uno, el modelo de Tucker 3 tiene la forma:

$$\underline{\mathbf{X}} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} (\mathbf{a}_p \otimes \mathbf{b}_q \otimes \mathbf{c}_r) + \mathbf{E}$$

donde  $a_p$ ,  $b_q$  y  $c_r$  son los vectores singulares izquierdos obtenidos de las descomposiciones en valores singulares.

El término  $(\mathbf{a}_p \otimes \mathbf{b}_q \otimes \mathbf{c}_r)$  se denomina arreglo de rango uno y  $g_{pqr}$  corresponde a la ponderación para las combinaciones de componentes (Kroonenberg 2008).

### 1.4.2. Estimación de parámetros del modelo Tucker 3

Tucker (1966) propuso métodos algebraicos para resolver la estimación de los parámetros del modelo Tucker 3: correspondientes a las matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  y al arreglo central  $\underline{\mathbf{G}}$ . Estos métodos se utilizan cuando la solución es exacta. Para los casos en que existen soluciones aproximadas, Kroonenberg & De Leeuw (1980) describen algoritmos vía mínimos cuadrados alternantes para estimar los parámetros.

El algoritmo de mínimos cuadrados alternantes se basa en dividir los parámetros en grupos. Cada conjunto de parámetros se estima en el sentido de mínimos cuadrados condicionalmente sobre los parámetros restantes. La estimación de los parámetros se repite iterativamente hasta que los valores de los parámetros estimados no difieran significativamente (Bro 1998).

El algoritmo Tucker 3 descrito por Bro (1998) genera las matrices  $\mathbf{B}$  y  $\mathbf{C}$  como punto de partida del algoritmo, estima el arreglo central, minimiza la función de pérdida e itera hasta su convergencia. Incluye los siguientes pasos:

1. Determinar el criterio de parada o convergencia del algoritmo
2. Inicializar  $\mathbf{B}$  y  $\mathbf{C}$
3.  $[\mathbf{A}, \mathbf{S}, \mathbf{V}] = \text{dvs}(\mathbf{X}_{(I \times JK)}(\mathbf{C} \otimes \mathbf{B}), P)$
4.  $[\mathbf{B}, \mathbf{S}, \mathbf{V}] = \text{dvs}(\mathbf{X}_{(J \times IK)}(\mathbf{C} \otimes \mathbf{A}), Q)$
5.  $[\mathbf{C}, \mathbf{S}, \mathbf{V}] = \text{dvs}(\mathbf{X}_{(K \times IJ)}(\mathbf{B} \otimes \mathbf{A}), R)$
6. Ir al paso 3 hasta cumplir con el criterio de convergencia
7. Calcular  $\mathbf{G}_a = \mathbf{A}'\mathbf{X}_a(\mathbf{C} \otimes \mathbf{B})$

En este algoritmo,  $[\mathbf{A}, \mathbf{S}, \mathbf{V}] = \text{dvs}(\mathbf{X}_{(I \times JK)}(\mathbf{C} \otimes \mathbf{B}), P)$  significa que se realiza descomposición en valores singulares de  $(\mathbf{X}_{(I \times JK)}(\mathbf{C} \otimes \mathbf{B}))$ , generando la matriz  $\mathbf{A}$  con los primeros  $P$  vectores singulares izquierdos, la matriz  $\mathbf{V}$  con los primeros  $P$  vectores singulares derechos y la matriz diagonal  $\mathbf{S}$  cuyos elementos diagonales corresponden a los primeros  $P$  valores singulares.  $\mathbf{X}_{(I \times JK)}$  es la matriz  $\mathbf{X}_a$  generada por matriciación del arreglo tridimensional por el modo A.

El algoritmo Tucker 3 requiere un mecanismo de inicialización de las matrices componentes  $\mathbf{B}$  y  $\mathbf{C}$  para obtener una estimación de la matriz  $\mathbf{A}$ . Las matrices  $\mathbf{B}$  y  $\mathbf{C}$  iniciales corresponden a los vectores singulares izquierdos de la descomposición en valores singulares de la matriciación del arreglo por el modo B ( $\mathbf{X}_{(J \times IK)}$ ) y por el modo C ( $\mathbf{X}_{(K \times IJ)}$ ), reteniendo los primeros  $Q$  y  $R$  vectores singulares izquierdos, respectivamente.

El número de vectores singulares retenidos para la solución  $P$ ,  $Q$  y  $R$  se seleccionan considerando la regla del producto mínimo presentada por Kroonenberg (2008), expresada como:

$$P \times Q \geq R, \quad P \times R \geq Q, \quad Q \times R \geq P$$

### 1.4.3. Procedimiento para implementación del Modelo de Tucker 3

En la implementación de un análisis tres vías utilizando el modelo de Tucker 3, se consideran los siguientes pasos:

1. Matriciación del arreglo de datos original

2. Realizar un preprocesamiento de datos: Bro & Smilde (2003) presentan los dos tipos básicos de preprocesamiento: el centrado o ajuste aditivo y la escalado o ajuste multiplicativo.

El centrado elimina un término constante en todo elemento tal que los datos resultantes tengan media 0, debe desarrollarse a través de un único modo y no a través de modos combinados. El tipo de centrado se indica especificando a través de qué modo se centran los datos. De esta forma, el centrado a través del modo A corresponde a promediar los datos sobre los individuos, se calcula como:

$$\tilde{x}_{ijk} = x_{ijk} - \bar{x}_{.jk}$$

El escalado divide todo elemento por un término constante permitiendo eliminar diferencias de escala entre las variables. Se sugiere escalar cualquier ocurrencia de una variable por la misma cantidad.

Escalar dentro de los elementos del modo B (variables) lleva a calcular el factor:

$$\nu_j = \sqrt{\sum_{i=1}^I \sum_{k=1}^K x_{ijk}^2 / IK}$$

y los datos escalados se calculan como:

$$\tilde{x}_{ijk} = \frac{x_{ijk}}{\nu_j}$$

3. Extracción del número de componentes y selección del modelo más adecuado: el modelo de Tucker 3 puede extraer un número de componentes diferente por cada modo del arreglo. Para elegir el número de componentes adecuado, se selecciona el modelo que tenga la menor suma de cuadrados del error (SCE) entre las múltiples soluciones posibles, usando el número de componentes para cada modo que cumpla con la restricción o regla del producto mínimo presentada anteriormente.

Kiers & Mechelen (2001), consideran que la combinación adecuada de componentes es aquella para la cual la SCE presenta poca variación al considerar un componente adicional en alguno de los modos. Al respecto, Timmerman & Kiers (2000) especifican que dentro de los modelos Tucker 3 con igual número de componentes se debe seleccionar el modelo que tenga la SCE más pequeña.

Para comparar modelos con número de componentes  $S = P + Q + R$  diferente, se define el criterio *DiffFit* en términos de la SCE como el cociente  $b_S = \frac{dif_S}{dif_{S^*}}$ , para calcular este criterio se utiliza el siguiente procedimiento:

- Definir la SCE de un modelo con  $S$  componentes como  $V_S$  y la SCE de un modelo con  $S-1$  componentes como  $V_{S-1}$
  - Obtener el cálculo de  $dif_S = V_{S-1} - V_S$
  - Calcular  $b_S = \frac{dif_S}{dif_{S^*}}$
- $dif_{S^*}$  es la diferencia de mayor valor encontrada después de  $dif_S$ .
- Seleccionar el modelo  $S$  con el valor  $b_s$  más bajo.

Para visualizar este procedimiento, se construye el *scree-plot multiway* (SPM), en el cual se grafica la SCE de cada dimensionalidad *versus* la suma del número total de componentes  $S = P + Q + R$ . Se consideran como adecuados los modelos ubicados en la parte convexa de la gráfica, en donde la adición de un componente adicional no disminuye significativamente la SCE.

4. Realizar estudio de ajuste: sobre las bases de ajuste y parsimonia se pueden obtener una o más soluciones. Es útil estudiar el ajuste del modelo completo, el cual se realiza con base en los elementos del arreglo central.

Kiers & Mechelen (2001) y Kroonenberg (2008) evalúan el ajuste del modelo completo utilizando expresiones que contienen los elementos del arreglo central  $g_{pqr}$ . Se define:

$$SC(Ajustada) = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr}^2$$

La proporción de variabilidad total explicada por el modelo completo es:

$$\frac{\sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr}^2}{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{ijk}^2}$$

5. Interpretar y reportar la solución: el modelo de Tucker 3 se interpreta considerando el arreglo central y las matrices de componentes.

a. Interpretación del arreglo central: el arreglo central es la parte más importante de la descomposición de los datos por medio del modelo de Tucker 3, porque proporciona la descripción de las interacciones tres modos. Los valores del arreglo central se grafican utilizando trayectorias de los componentes del modo A (sujetos), a través de los componentes del modo C (ocasiones) en cada uno de los componentes del modo B (variables) (Amaya & Pacheco 2002).

b. Interpretación de las matrices **A**, **B** y **C**: Kiers & Mechelen (2001) consideran que estas matrices presentan estructura simple si se obtienen puntajes bien diferenciados por elemento sobre un componente específico y en este sentido, su interpretación depende solamente de las diferencias relativas de los valores respecto a cada componente.

Kiers (2000a) sugiere realizar su interpretación utilizando representaciones, de los elementos de un modo particular y las configuraciones para combinaciones de los elementos de dos modos (trayectorias).

La conceptualización respecto a estas representaciones es objeto de este trabajo y se desarrolla en el capítulo 2.

## 1.5. Análisis factorial múltiple (AFM)

### 1.5.1. Definición y fundamentos del AFM

El AFM es un análisis factorial ponderado que estudia grupos de variables definidas sobre el mismo conjunto de individuos. A cada grupo de variables le corresponde una tabla, todas las tablas pueden ser yuxtapuestas y formar así una tabla única estructurada en subtablas (Escofier & Pagés 1982, Husson & Pagés 2006) (figura 1.2 -b)).

Para efectos de este estudio se configura una tabla yuxtapuesta, donde cada subtabla contiene el registro de las variables evaluadas bajo una condición.

El método se desarrolla en dos etapas: inicialmente se realiza ACP separado para cada condición, obteniendo el primer valor propio  $\lambda_k^1$ , en cada caso. En la segunda etapa se realiza un ACP a la tabla completa en la cual las variables evaluadas sobre una misma condición se ponderan por el inverso del primer valor propio  $\frac{1}{\lambda_k^1}$  extraído del ACP de la condición (Escofier & Pagés 1994).

Se denota: **X** la tabla completa que corresponde a **X<sub>a</sub>**, *I* el conjunto de individuos, *J* el conjunto de variables, *K* el conjunto de condiciones y **X<sub>k</sub>** la tabla asociada a la condición *k*. Se supone que los individuos y las variables están provistos de un peso: se denomina  $p_i$  al peso que afecta al individuo *i* ( $\sum_i p_i = 1$ ) y  $m_{jk}$  al peso asociado a la variable  $v_{jk}$ . Las matrices diagonales de pesos de los individuos y de las variables se representan por **D** y **M<sub>k</sub>** (para la condición *k*, donde el peso de las variables es 1) y **M** (para el análisis global o de todas las condiciones, donde el peso de las variables de la condición *k* corresponde a  $\frac{1}{\lambda_k^1}$ , cuando las ponderaciones iniciales de las variables son todas iguales a 1).

El método se fundamenta en realizar un ACP al conjunto completo de variables, en todas las condiciones, en el que cada condición se pondera por el inverso del primer valor propio extraído de su ACP separado:  $\lambda_k^1$  (Escofier & Pagés 1994).

Según Escofier & Pagés (1992), este análisis factorial induce un análisis global de condiciones en el que cada condición se representa, en el espacio  $\mathbb{R}^{I^2}$ , por la matriz **W<sub>k</sub>D**, donde **D** = *diag*( $p_i$ ) y **W<sub>k</sub>** corresponde a una matriz  $I \times I$  de productos escalares establecidos

entre individuos para el grupo  $k$ :

$$\mathbf{W}_k = \mathbf{X}_k \mathbf{M}_k \mathbf{X}_k'; \quad \mathbf{M}_k = \text{diag} \left( \frac{1}{\lambda_1^k} \right)$$

### 1.5.2. Factores extraídos por el AFM

En el AFM un factor o eje extraído del análisis global para individuos se define como:

$$F_s = \mathbf{X} \mathbf{M} \mathbf{u}_s$$

Para variables se define como:

$$G_s = \mathbf{X}' \mathbf{D} \mathbf{v}_s$$

donde  $\mathbf{X}$  es la matriciación por el modo A:  $\mathbf{Xa}$  del arreglo tridimensional,  $\mathbf{M}$  es la matriz diagonal de pesos de variables,  $\mathbf{D}$  es la matriz diagonal de pesos de individuos,  $u_s$  es el vector propio obtenido por diagonalización de la matriz  $\mathbf{X} \mathbf{M} \mathbf{X}' \mathbf{D}$  y  $v_s$  es el vector propio

obtenido por diagonalización de la matriz  $\mathbf{X}' \mathbf{D} \mathbf{X} \mathbf{M}$ .

Un factor o eje parcial se define como:

$$F_s^k = \tilde{\mathbf{X}}_k \mathbf{M} \mathbf{u}_s$$

$\tilde{\mathbf{X}}_k$  es una matriz de tamaño  $I \times JK$  que contiene las coordenadas de los puntos de la nube  $N_I^k$  sobre  $\mathbf{X}_k$  y el resto se completa con ceros. Tiene la forma:

$$\tilde{\mathbf{X}}_k = [\mathbf{0} \ \cdots \ \mathbf{0} \ \cdots \ \mathbf{X}_k \ \cdots \ \mathbf{0} \ \cdots \ \mathbf{0}]$$

Las coordenadas parciales se multiplican por  $K$  para que el centro de gravedad de las que corresponden a cada individuo sea la respectiva coordenada global (Escofier & Pagés 1992).

### 1.5.3. Representaciones del AFM y ayudas a la interpretación

Según Escofier & Pagés (1994) el método AFM presenta:

a. Los resultados clásicos de un ACP: la representación de individuos, variables y las ayudas a la interpretación como calidad de la representación y las contribuciones de individuos y variables a la formación de un eje.

b. Los resultados específicos de la estructura tres vías. Entre estos se destacan: la representación de las condiciones (grupos de variables) como puntos, la importancia de los factores comunes en cada grupo, la representación simultánea de las nubes de individuos y las ayudas a la interpretación para evaluar la calidad de las representaciones desde un punto de vista global.

Las representaciones se generan en tres espacios:  $\mathcal{R}^J$ ,  $\mathcal{R}^I$  y  $\mathcal{R}^{I^2}$ , en los cuales se localiza la nube de individuos, variables y condiciones respectivamente (Escofier & Pagés 1992).

Entre otras ayudas a la interpretación el AFM permite extraer factores comunes al conjunto de las condiciones, a algunas condiciones y los factores específicos de una condición.

## CAPÍTULO 2

---

### Comparación entre el AFM y el Tucker 3

---

Los criterios de comparación entre los métodos AFM y Tucker 3 se definieron considerando: el preprocesamiento de datos, la función de pérdida, la inercia, las ayudas a la interpretación: contribuciones y calidad de representación, las representaciones de individuos, variables y condiciones y las medidas de relación.

#### 2.1. Preprocesamiento de datos

En este criterio se consideraron los procedimientos realizados previamente sobre el arreglo de datos original, correspondientes a matriciación, centrado y escalado de datos.

**AFM:** realiza un único procedimiento de matriciación por el modo individuos (modo A), la matriz obtenida se denota  $\mathbf{X}_a$  de dimensión  $I \times JK$ .

Respecto al proceso de estandarización realiza en su orden centrado y escalado sobre la matriz  $\mathbf{X}_a$ . El procedimiento de centrado consiste en restar a cada columna su media y el escalado a dividir cada columna por su desviación estándar; esto es, centra y escala (estandariza) a través de individuos.

**Tucker 3:** requiere en su etapa previa un único procedimiento de matriciación, que corresponde al modo por el cual se desee realizar el centrado. En la mayoría de los casos corresponde al modo A, generando la matriz  $\mathbf{X}_a$  de dimensión  $I \times JK$ .

El Tucker 3 puede realizar centrado de cualquiera de las matrices  $\mathbf{X}_a$ ,  $\mathbf{X}_b$  o  $\mathbf{X}_c$ . Sobre la matriz centrada el método realiza matriciación por el modo variables (matriz de dimensión  $J \times IK$ ) y escala dividiendo cada fila por la desviación estándar correspondiente. Partiendo de la matriz  $\mathbf{X}_a$ , esto corresponde a centrar a través de individuos y escalar a través de variables.

**Comparación:** los dos métodos son similares en el procedimiento de centrado, realizándolo a través del modo individuos, llevan a cero la media de cada combinación variable-condición.



Se diferencian en el procedimiento de escalado. El AFM lo realiza a través del modo individuos, dividiendo cada columna de la matriz de dimensión  $I \times JK$  por su desviación estándar, que corresponde a escalar por separado la variable evaluada en condiciones diferentes. El Tucker 3 escala a través del modo variables, dividiendo cada fila de la matriz de dimensión  $J \times IK$  por la desviación estándar, que corresponde a dividir por la desviación estándar de la variable evaluada en todas las condiciones.

## 2.2. Función de pérdida

**AFM:** es un análisis factorial ponderado, específicamente es un ACP normado y ponderado. Utilizando la notación del diagrama de dualidad de la escuela francesa, el AFM es un ACP( $\mathbf{X}, \mathbf{M}, \mathbf{D}$ ) (ver sección 1.5).

Por el teorema de descomposición en valores singulares, en un ACP( $\mathbf{X}, \mathbf{M}, \mathbf{D}$ ) la matriz  $\mathbf{X}$  de dimensión  $I \times J$  se estima como:

$$\hat{\mathbf{X}} = \mathbf{V}\mathbf{G}\mathbf{U}'$$

donde  $\mathbf{V}$  es una matriz de dimensión  $I \times S$  cuyas columnas son los vectores propios de  $\mathbf{X}\mathbf{M}\mathbf{X}'\mathbf{D}$  asociados a los  $S$  primeros valores propios no nulos.  $\mathbf{G}$  es una matriz diagonal

de orden  $S \times S$ , con entradas sobre la diagonal correspondiente a los valores singulares.  $\mathbf{U}$  es una matriz de dimensión  $J \times S$  cuyas columnas son los vectores propios de  $\mathbf{X}'\mathbf{D}\mathbf{X}\mathbf{M}$  asociados a los  $S$  primeros valores propios no nulos.

Haciendo  $\mathbf{A}=\mathbf{V}$ ,  $\mathbf{B}=\mathbf{U}$

$$\hat{\mathbf{X}} = \mathbf{A}\mathbf{G}\mathbf{B}'$$

En forma general, el AFM minimiza una función de pérdida que tiene la forma:

$$\|\mathbf{X} - \mathbf{A}\mathbf{G}\mathbf{B}'\|^2$$

**Tucker 3:** el modelo se ajusta en el sentido de mínimos cuadrados alternantes, minimizando:

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} - \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr}(a_{ip}b_{jq}c_{kr}))^2$$

sobre  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  y  $\mathbf{G}$ .

Es decir, el ajuste se calcula como la suma de cuadrados de las aproximaciones del modelo a los datos.

Considerando la expresión del modelo de Tucker 3 según sea la matriciación del arreglo utilizada del modo A, B o C, respectivamente, la función de pérdida toma la forma:

$$\|\mathbf{X}_a - \mathbf{A}\mathbf{G}_a(\mathbf{C}' \otimes \mathbf{B}')\|^2$$

$$\|\mathbf{X}_b - \mathbf{B}\mathbf{G}_b(\mathbf{C}' \otimes \mathbf{A}')\|^2$$

$$\|\mathbf{X}_c - \mathbf{C}\mathbf{G}_c(\mathbf{B}' \otimes \mathbf{A}')\|^2$$

donde  $\mathbf{A}$ ,  $\mathbf{B}$  y  $\mathbf{C}$  son matrices ortonormales,  $\mathbf{G}_a$ ,  $\mathbf{G}_b$  y  $\mathbf{G}_c$ , representan las matriciaciones del arreglo central por los modos A, B y C respectivamente.

**Comparación:** los métodos AFM y Tucker 3 difieren en la expresión de la función de pérdida que minimizan en términos de:

1. El modelo ajustado: el AFM ajusta un modelo correspondiente al de un ACP dos vías, el Tucker 3 al de un ACP generalizado al caso tres vías.
2. Los parámetros estimados, el número y dimensión del sistema de componentes elegido: el AFM estima las matrices  $\mathbf{A}$  y  $\mathbf{B}$  sobre un sistema de componentes único, mientras que el modelo de Tucker3 estima las matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  y el arreglo central  $\underline{\mathbf{G}}$  utilizando tres sistemas de componentes diferentes, donde cada uno de los sistemas puede tener dimensión igual o diferente. El AFM calcula la matriz de correlación componente -condición que contiene las coordenadas de ubicación de las condiciones, en este aspecto puede considerarse equivalente a la matriz  $\tilde{\mathbf{C}}$  generada en Tucker 3. Sin embargo, por tratarse de un ACP dos vías, esta no se hace explícita en el modelo, porque en este caso, el modo condiciones está representado únicamente en las ponderaciones. A diferencia de este, el Tucker 3 estima e incluye la matriz  $\mathbf{C}$  en la expresión del modelo.
3. La estructura del parámetro  $\mathbf{G}$ , en AFM se trata de una matriz diagonal de valores singulares, mientras en Tucker 3 es un arreglo tridimensional que relaciona los componentes extraídos de los tres modos del arreglo.
4. El procedimiento de estimación de parámetros que permite minimizar la función de pérdida: el AFM estima los parámetros  $\mathbf{A}$  y  $\mathbf{B}$  por medio de una diagonalización. El modelo de Tucker 3 estima los parámetros  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  y  $\underline{\mathbf{G}}$ , utilizando algoritmos mínimos cuadrados alternantes.

## 2.3. Inercia

**AFM:** la inercia retenida por la solución seleccionada, corresponde a:

$$\frac{\sum_{s=1}^S \lambda_s}{\sum_{s=1}^{JK} \lambda_s}$$

**Tucker3:** la inercia retenida o variabilidad del arreglo original explicada por el modelo de Tucker 3 corresponde al porcentaje de ajuste del modelo, definido en términos de los elementos del arreglo central como:

$$\frac{\sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr}^2}{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{ijk}^2}$$

**Comparación:** los dos métodos se asemejan al expresar la inercia como un cociente de la variabilidad explicada por la solución respecto a la variabilidad total.

## 2.4. Ayudas a la interpretación: contribuciones y calidad de la representación

**AFM:** genera la contribución de cada componente a la inercia total, definida como la proporción de inercia explicada por cada componente extraído. La contribución de cada componente a la inercia total se define como:

$$\frac{\lambda_s}{\sum_{s=1}^S \lambda_s}$$

El AFM genera a partir del análisis global, las ayudas a la interpretación comunes de un ACP (contribuciones y calidad de representación) para individuos y variables.

La contribución absoluta del individuo  $i$  al eje  $s$  es:

$$Con_s(i) = \frac{p_i F_s^2(i)}{\lambda_s}$$

La calidad de representación del individuo  $i$  sobre el eje  $s$  es:

$$Cos_s^2(i) = \frac{F_s^2(i)}{\|\mathbf{i}\|^2}$$

Para las variables se definen las mismas expresiones:

La contribución absoluta de la variable  $j$  al eje  $s$  es:

$$Con_s(j) = \frac{m_{jk} G_s^2(j)}{\lambda_s}$$

La calidad de representación de la variable  $j$  sobre el eje  $s$  es:

$$Cos_s^2(j) = \frac{G_s^2(j)}{\|\mathbf{j}\|^2}$$

El AFM permite evaluar la calidad de representación de las  $K$  nubes de individuos mediante el cociente inercia-inter/inercia total. Al aproximarse a 1 indica que estas tienen suficientes características comunes, que justifica un estudio detallado de sus diferencias.

Por el contrario, si la razón es próxima a 0, las diferencias de forma entre las nubes son tan importantes que su estudio profundo no es útil (Escofier & Pagés 1992).

La contribución de una condición a la inercia retenida por un eje global se expresa con la medida de relación  $L(\mathbf{F}_s, \mathbf{J}_k)$  que determina la inercia proyectada de las variables de la ocasión  $k$  denotadas  $\mathbf{J}_k$  sobre un factor del AFM global ( $\mathbf{F}_s$ ), en la forma:

$$L(F_s, J_k) = \sum_{j \in J_k} m_j \langle F_s, v_j \rangle^2$$

La medida  $L$  representa la importancia del factor común en la condición  $k$  y la contribución de las variables evaluadas en una condición  $\mathbf{J}_k$  a la construcción del factor. Su valor es a lo más 1, porque al ponderar la condición (por  $\frac{1}{\lambda_1^k}$ ), esta se equilibra haciendo que su contribución máxima al eje sea 1. Toma el valor de 1 cuando la componente se confunde con la primera componente principal del ACP de cada condición, constituyéndose en una dirección de inercia importante para esa condición.

**Tucker 3:** genera las contribuciones de los factores o componentes que explican cada uno de los modos y la contribución de la combinación de componentes de los diferentes modos.

La contribución de los componentes extraídos por uno de los modos se obtiene considerando la suma de cuadrados de los elementos del arreglo central relacionados con este componente particular respecto a la suma de cuadrados total.

La contribución de los componentes del modo A se expresa como:

$$\frac{\sum_{q=1}^Q \sum_{r=1}^R g_{qr}^2}{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{ijk}^2}$$

La contribución de los componentes del modo B se expresa como:

$$\frac{\sum_{p=1}^P \sum_{r=1}^R g_{pr}^2}{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{ijk}^2}$$

La contribución de los componentes del modo C se expresa como:

$$\frac{\sum_{p=1}^P \sum_{q=1}^Q g_{pq}^2}{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{ijk}^2}$$

La proporción de variabilidad explicada por cada combinación de componentes particular se obtiene tomando el cuadrado del valor central  $g_{pqr}$  que identifica la combinación del componente  $p$  del modo A, el componente  $q$  del modo B y el componente  $r$  del modo C, dividido por la suma de cuadrados de todos los valores. Su expresión es dada por:

$$\frac{g_{pqr}^2}{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{ijk}^2}$$

**Comparación:** los dos métodos generan la contribución de cada componente a la inercia total a través de la inercia retenida o porcentaje de variabilidad explicado.

El AFM genera medidas específicas de calidad de representación y contribución de las entidades de los diferentes modos a la inercia de cada factor, aspecto que se constituye en una ventaja del método sobre el Tucker 3, al permitir identificar las entidades que contribuyen y se representan de manera relevante en el análisis.

El Tucker 3 por extraer tres sistemas de ejes, requiere generar la contribución de la combinación de componentes a la inercia total, a partir de los elementos del arreglo central; esto le permite identificar e interpretar las combinaciones de componentes más influyentes. El AFM, al extraer un sistema de ejes único, no requiere generar medidas de contribución de una combinación de componentes.

Los métodos AFM y Tucker 3 se pueden definir como complementarios respecto a las ayudas a la interpretación generadas, al calcular cada uno medidas adicionales que permiten complementar la información.

## 2.5. Representaciones

En esta sección, se presenta la fundamentación y comparación respecto a las representaciones de individuos, variables, condiciones y trayectorias generadas por los dos métodos.

### 2.5.1. Representación de individuos

**AFM:** genera en el espacio  $\mathbb{R}^{JK}$  la representación de la nube de individuos, denotada nube  $N_I$ . Las coordenadas para los  $I$  individuos se ubican en un subespacio  $S$ -dimensional de  $\mathbb{R}^{JK}$  y corresponden a las filas de la matriz de componentes para individuos.

**Tucker 3:** genera en el espacio  $\mathbb{R}^{JK}$  la representación de la nube de individuos o entidades del modo A. Las coordenadas para los  $I$  puntos se ubican en un subespacio  $P$ -dimensional de  $\mathbb{R}^{JK}$  y corresponden a las filas de la matriz  $\tilde{\mathbf{A}}$ .

Para obtener  $\tilde{\mathbf{A}}$  se escribe:

$$\hat{\mathbf{X}}_a = \mathbf{A}((\mathbf{C} \otimes \mathbf{B})\mathbf{G}'_a)'$$

y se expresa:

$$\mathbf{F} = (\mathbf{C} \otimes \mathbf{B})\mathbf{G}'_a$$

$\hat{\mathbf{X}}_a$  se expresa en términos de una matriz  $\mathbf{A}$  ortonormal y una matriz  $\mathbf{F}$  con columnas ortogonales, esta última se ortonormaliza haciendo:  $\tilde{\mathbf{F}} = \mathbf{F}\mathbf{T}$ .  $\mathbf{T}$  es una matriz diagonal cuyos elementos corresponden al inverso de los valores singulares generados por *dvs* de la matriz  $\mathbf{F}$ , obtenida a partir de las estimaciones de la última iteración del algoritmo. Aplicando la transformación inversa sobre  $\mathbf{A}$ , finalmente, las coordenadas para las entidades del modo A se ubican en las filas de la matriz  $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{T}^{-1}$ .

**Comparación:** los dos métodos son similares respecto a la dimensión del espacio en el cual generan la representación de individuos: ambos lo hacen en  $\mathbb{R}^{JK}$ .

Difieren en la dimensión del subespacio en el cual se ubican las coordenadas de los individuos: en AFM corresponde a un subespacio  $S$ -dimensional de  $\mathbb{R}^{JK}$ , con coordenadas sobre la matriz de componentes para individuos (matriz cuyas columnas corresponden a los factores  $F_s$ ), mientras en Tucker 3 corresponde a un subespacio  $P$ -dimensional de  $\mathbb{R}^{JK}$ , con coordenadas sobre la matriz  $\tilde{\mathbf{A}}$ ; siendo  $S$  y  $P$  subespacios de dimensión similar o diferente.

### 2.5.2. Representación de variables

**AFM:** genera en el espacio  $\mathbb{R}^I$  la representación de las variables, denotada nube  $N_J$ . Las coordenadas para las  $J$  variables se ubican en un subespacio  $S$ -dimensional de  $\mathbb{R}^I$  y corresponden a las filas de la matriz de componentes para variables.

**Tucker 3:** genera en el espacio  $\mathbb{R}^{IK}$  la representación de la nube de variables o entidades del modo B. Las coordenadas para los  $J$  puntos se ubican en un subespacio  $Q$ -dimensional de  $\mathbb{R}^{IK}$  y corresponden a las filas de la matriz  $\tilde{\mathbf{B}}$ .

Para obtener  $\tilde{\mathbf{B}}$  se escribe:

$$\hat{\mathbf{X}}_b = \mathbf{B}((\mathbf{A} \otimes \mathbf{C})\mathbf{G}'_b)'$$

expresando:

$$\mathbf{F} = (\mathbf{A} \otimes \mathbf{C})\mathbf{G}'_b$$

$\hat{\mathbf{X}}_b$  se expresa en términos de una matriz  $\mathbf{B}$  ortonormal y una matriz  $\mathbf{F}$  con columnas ortogonales, esta última se ortonormaliza haciendo:  $\tilde{\mathbf{F}} = \mathbf{F}\mathbf{T}$ .  $\mathbf{T}$  es una matriz diagonal cuyos elementos corresponden al inverso de los valores singulares generados por *dvs* de la matriz  $\mathbf{F}$ , obtenida a partir de las estimaciones de la última iteración del algoritmo. Aplicando la transformación inversa sobre  $\mathbf{B}$ , finalmente, las coordenadas para las entidades del modo B se ubican en las filas de la matriz  $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{T}^{-1}$ .

**Comparación:** los dos métodos difieren respecto a:

-La dimensión del espacio en el cual generan la representación de variables: el AFM lo hace en  $\mathbb{R}^I$  y Tucker 3 en  $\mathbb{R}^{IK}$ .

-La dimensión del subespacio en el cual se ubican las coordenadas de las variables: en AFM corresponde a un subespacio  $S$ -dimensional de  $\mathbb{R}^I$ , con coordenadas sobre la matriz de componentes para variables (matriz cuyas columnas corresponden a los factores  $G_s$ ), mientras en Tucker 3 corresponde a un subespacio  $Q$ -dimensional de  $\mathbb{R}^{IK}$ , con coordenadas sobre la matriz  $\tilde{\mathbf{B}}$ ; siendo  $S$  y  $Q$  subespacios de dimensión similar o diferente.

### 2.5.3. Representación de condiciones

**AFM:** define en el espacio  $\mathbb{R}^{I^2}$  la representación de las condiciones, denotada nube  $N_K$ . Esta nube se parece más a una nube de variables que a una nube de individuos porque el

producto escalar entre los vectores que representan a dos condiciones se interpreta como una medida de la relación entre las  $J$  variables evaluadas en las dos condiciones.

La matriz  $\mathbf{W}_k \mathbf{D}$  es un elemento del espacio vectorial de dimensión  $\mathbb{R}^{I^2}$ . En este espacio, las variables evaluadas en la condición  $k$  se representan por la matriz u operador  $\mathbf{W}_k \mathbf{D}$ , porque la diagonalización de  $\mathbf{W}_k \mathbf{D}$  permite una reconstrucción exacta de la estructura de la nube de individuos  $N_I^k$  definida por el grupo de variables  $J_k$ .

Las coordenadas para las  $K$  condiciones se ubican en un subespacio  $S$ -dimensional de  $\mathbb{R}^{I^2}$  y corresponden a las filas de la matriz de correlación condición-componente.

**Tucker 3:** genera en el espacio  $\mathbb{R}^{IJ}$  la representación de la nube de condiciones o entidades del modo  $C$ . Las coordenadas para los  $K$  puntos se ubican en un subespacio  $R$ -dimensional de  $\mathbb{R}^{IJ}$  y corresponden a las filas de la matriz  $\tilde{\mathbf{C}}$ .

Para obtener  $\tilde{\mathbf{C}}$  se escribe:

$$\hat{\mathbf{X}}_c = \mathbf{C}((\mathbf{B} \otimes \mathbf{A})\mathbf{G}'_c)'$$

expresando:

$$\mathbf{F} = (\mathbf{B} \otimes \mathbf{A})\mathbf{G}'_c$$

$\hat{\mathbf{X}}_c$  se expresa en términos de una matriz  $\mathbf{C}$  ortonormal y una matriz  $\mathbf{F}$  con columnas ortogonales, esta última se ortonormaliza haciendo:  $\tilde{\mathbf{F}} = \mathbf{F}\mathbf{T}$ .  $\mathbf{T}$  es una matriz diagonal cuyos elementos corresponden al inverso de los valores singulares generados por  $dvs$  de la matriz  $\mathbf{F}$ , obtenida a partir de las estimaciones de la última iteración del algoritmo. Aplicando la transformación inversa sobre  $\mathbf{C}$ , finalmente, las coordenadas para las entidades del modo  $C$  se ubican en las filas de la matriz  $\tilde{\mathbf{C}} = \mathbf{C}\mathbf{T}^{-1}$ .

**Comparación:** los dos métodos difieren respecto a:

-La dimensión del espacio en el cual generan la representación de condiciones: el AFM lo hace en  $\mathbb{R}^{I^2}$  y Tucker 3 en  $\mathbb{R}^{IJ}$ .

-La dimensión del subespacio en el cual se ubican las coordenadas de las condiciones: en AFM corresponde a un subespacio  $S$ -dimensional de  $\mathbb{R}^{I^2}$ , con coordenadas que representan la relación condición-componente, mientras en Tucker 3 corresponde a un subespacio  $R$ -dimensional de  $\mathbb{R}^{IJ}$ , con coordenadas sobre la matriz  $\tilde{\mathbf{C}}$ ; siendo  $S$  y  $R$  subespacios de dimensión similar o diferente.

#### 2.5.4. Trayectorias

**AFM:** genera una trayectoria correspondiente a la representación simultánea de las  $K$  nubes parciales de  $I$  individuos  $N_I^k$ , en el espacio  $\mathbb{R}^{JK}$  y la representación de la nube media o nube de los centros de gravedad  $N_I^*$ . La representación simultánea es posible porque todas las nubes  $N_I^k$  y su unión denotada  $N_I^K$  se sitúan en  $\mathbb{R}^{JK}$ .

La representación se realiza proyectando sobre un subespacio de  $\mathbb{R}^{JK}$  que posea inercia inter máxima. Este subespacio corresponde al engendrado por los primeros ejes de inercia obtenidos del ACP de la tabla completa. Como las coordenadas de  $N_I^k$  están en la tabla  $\tilde{X}_k$ ,

se introducen en el ACP las tablas  $\tilde{X}_k$  como suplementarias y se genera la representación simultánea de las  $N_I^K$ . Las coordenadas se amplifican por  $K$ .

El AFM permite mediante esta representación, caracterizar a cada individuo en cada tabla gracias a la representación simultánea de todos los individuos (individuos parciales), detectar la condición responsable de situar el centroide de cada individuo, identificar su estructura y estudiar las semejanzas entre individuos.

Desde el punto de vista de los  $IK$  puntos parciales vistos como la unión de  $I$  nubes de  $K$  puntos cada una,  $N_i^K$ . La razón  $IerciaInter/IerciaTotal$  se constituye un buen indicador de la utilidad de la representación simultánea; si es próxima a 1, todas las nubes tienen suficientes caracteres comunes para justificar un estudio detallado de sus diferencias; si es próxima a 0, las diferencias de forma entre las nubes son tan importantes que su estudio profundo no es útil (Escofier & Pagés 1992).

**Tucker 3:** genera tres tipos de trayectorias: individuo-condición, variable-condición e individuo-variable.

La trayectoria individuo-condición permite observar el cambio de localización de los individuos a través de las diferentes condiciones, en el espacio generado por las variables. Esto es, representar todas las combinaciones de los modos A y C.

Para obtener esta representación se toman las columnas de  $\hat{\mathbf{X}}_b$  porque cada una corresponde a una combinación individuo-condición. Las coordenadas para estas columnas se ubican respecto a un espacio variable completo  $\mathcal{R}^J$ , proyectándose en un subespacio de  $\mathcal{R}^J$  de menor dimensión, cuya base está en  $\mathbf{B}$ .

Las coordenadas de cada combinación individuo-condición se ubican en las filas de  $\mathbf{F}$  con respecto a la base generada por las columnas de la matriz  $\mathbf{B}$ . Como en el modelo de Tucker 3,  $\mathbf{B}$  es ortonormal, las coordenadas se usan directamente para graficar las combinaciones individuo-condición con respecto a ejes cartesianos.

Lo anterior se expresa como:

$$\hat{\mathbf{X}}_b = \mathbf{B}(\mathbf{A} \otimes \mathbf{C})\mathbf{G}_b'$$

con:

$$\mathbf{F} = (\mathbf{A} \otimes \mathbf{C})\mathbf{G}_b'$$

Posterior a la localización de cada punto de la combinación sujeto- condición se conectan los puntos del mismo individuo para encontrar su trayectoria.

Las representaciones de las trayectorias de las entidades de los modos B y C (variable-condición) o A y B (individuo -variable), tienen un procedimiento análogo al anterior, basado en  $\hat{\mathbf{X}}_c$  y  $\hat{\mathbf{X}}_a$ , respectivamente.

**Comparación:** el AFM genera las trayectorias de los individuos en el espacio  $\mathcal{R}^{JK}$  mientras que en Tucker 3 lo hace en  $\mathcal{R}^J$ .

Tucker 3 genera trayectorias adicionales: variables a través de condiciones e individuos a través de variables.



## 2.6. Medidas de relación

**AFM:** determina medidas de relación entre ejes parciales y ejes generales así como la relación entre condiciones.

La contribución de un eje parcial al eje global extraído del análisis se expresa mediante una medida de correlación entre ellos, permitiendo evaluar su grado de semejanza. Este coeficiente puede tomar un valor elevado para el conjunto de las condiciones, para algunas de ellas o incluso para una sola, permitiendo evidenciar la presencia de factores comunes al conjunto de las condiciones, algunas condiciones y factores específicos de una condición.

La correlación entre condiciones se determina a través del coeficiente RV, el cual toma valores entre 0 (no existe relación entre las variables de las condiciones consideradas) y 1 (existe relación total).

**Tucker 3:** estima el arreglo central que permite establecer relaciones entre los componentes extraídos por los tres modos del arreglo e identificar interacciones tres-vías.

**Comparación:** el AFM proporciona medidas de relación entre grupos mientras el Tucker 3 permite la lectura de las interacciones entre componentes, aportando con esto cada uno información adicional valiosa al analizar datos de estructura tres vías.

## CAPÍTULO 3

---

### Programación del Tucker 3 en ambiente R

---

El modelo de Tucker 3 ha sido programado por Pieter Kroonenberg y puesto a disposición comercialmente en el paquete 3WayPack, disponible en <http://three-mode.leidenuniv.nl/> y en lenguaje SAS por Amaya & Pacheco (2002).

En este trabajo se programaron siete funciones en ambiente R, correspondientes a los procesos de: matriciación, preprocesamiento, estimación del algoritmo de Tucker3, cálculo de ajuste y contribución de los componentes, representación de las entidades de cada modo, representación de trayectorias de la combinación de dos modos y representación del arreglo central  $\underline{\mathbf{G}}$ ; las cuales se describen a continuación de forma general y se presentan en el apéndice A.1.

#### 3.1. Matriciación del arreglo original (función: *matriciacion*)

El insumo de la función es un objeto tipo lista que contiene las matrices correspondientes a las diferentes condiciones.

La función despliega el arreglo tres modos a una matriz manteniendo constante alguno de los modos del arreglo en las filas de la matriz y confundiendo los modos restantes en las columnas. La función genera un objeto tipo lista que contienen las tres matriciaciones generadas por los modos A, B y C respectivamente, denominadas  $\mathbf{X}_a$  de dimensión  $I \times JK$ ,  $\mathbf{X}_b$  de dimensión  $J \times IK$  y  $\mathbf{X}_c$  de dimensión  $K \times IJ$ .

La estructura de programación de la función *matriciacion* se presenta en el apéndice A.1.1

#### 3.2. Preprocesamiento de datos (función: *preproc*)

La función *preproc* toma un objeto tipo lista de clase matriciación, realizando sobre este el centrado o centrado y escalado.

Los procedimientos de centrado y escalado se desarrollaron siguiendo la conceptualiza-

ción sugerida por Bro & Smilde (2003). Se programó el centrado por el modo de individuos por ser el procedimiento más ampliamente utilizado, partiendo de la matriz  $\mathbf{X}_a$  y restando a cada columna su media. Sobre la matriz centrada la función realiza matriciación por el modo variables (matriz  $\mathbf{X}_b$  de dimensión  $J \times IK$ ) y matriciación por el modo condiciones (matriz  $\mathbf{X}_c$  de dimensión  $K \times IJ$ ) y escala por variables y condiciones dividiendo cada fila por la desviación estándar correspondiente. Es decir, se programó lo sugerido por Bro & Smilde (2003) centrando a través de individuos y escalando a través de variables y condiciones.

La función genera un objeto tipo lista que contiene las matrices centradas o centradas y escaladas, desplegadas por los modos A, B y C.

La estructura de programación de la función *preproc* se presenta en el apéndice A.1.2

### 3.3. Estimación del algoritmo de Tucker 3 y selección del mejor modelo (función: *tucker3*)

El insumo de la función es un objeto tipo lista de arreglos matricizados (preprocesados o no).

La función *tucker3* implementa el algoritmo mínimos cuadrados alternantes de Tucker 3 dado por (Bro 1998) y presentado en la sección 1.4.2. Mediante esta función se realiza estimación de los parámetros del modelo de Tucker 3 correspondiente a las matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  y al arreglo central  $\mathbf{G}$ , con base en un procedimiento de mínimos cuadrados alternantes.

El algoritmo estima  $\mathbf{B}$ ,  $\mathbf{C}$ ,  $\mathbf{A}$  y  $\mathbf{G}_a$  en un sentido mínimo cuadrático condicionalmente sobre los parámetros restantes. La estimación de los parámetros se repite iterativamente hasta obtener uno de los siguientes criterios de convergencia: - la obtención de una diferencia no superior a  $1\text{E-}10$  entre la función de pérdida (SCE) evaluada en dos iteraciones consecutivas, o -la ejecución de  $N=100$  iteraciones.

La función permite generar todos los modelos posibles aplicando la regla del producto mínimo presentada en el numeral 3 de la sección 1.4.3. En cada modelo se calcula el número de componentes seleccionado por cada modo, el número de componentes total y el valor para la función de pérdida minimizada (SCE).

Por medio de esta función es posible seleccionar el modelo más adecuado considerando dos criterios: - el gráfico SPM y - el criterio DiffFit, cuya fundamentación teórica se presentó en el numeral 3 de la sección 1.4.3.

La función genera un objeto tipo lista que contiene las estimaciones de los parámetros  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ ,  $\mathbf{G}_a$ ,  $\mathbf{G}_b$  y  $\mathbf{G}_c$ .

La estructura de programación de la función *tucker3* se presenta en el apéndice A.1.3

### 3.4. Cálculo de ajuste y contribución de los componentes (función: *ajuste*)

La función toma como parámetro el objeto tipo lista generado de la función `tucker3` que contiene la estimación del arreglo central con sus matriciaciones  $\mathbf{G}_a$ ,  $\mathbf{G}_b$  y  $\mathbf{G}_c$ .

Las medidas de ajuste se calcularon a partir de los elementos del arreglo central  $\underline{\mathbf{G}}$ , considerando que este se define con base en las matrices  $\mathbf{A}$ ,  $\mathbf{B}$  y  $\mathbf{C}$ , las cuales se estabilizan cuando el arreglo central  $\underline{\mathbf{G}}$  alcanza su estabilidad.

Las medidas de ajuste programadas corresponden a:

- El ajuste total del modelo o proporción de variabilidad explicada (AJtotal): se definió según lo presentado en el numeral 5 de la sección 1.4.3.
- La contribución de los componentes de cada modo al ajuste del modelo (`contrib.compsA`, `contrib.compsB`, `contrib.compsC`).
- La contribución de cada triada de componentes al ajuste del modelo (`contrib.triadacomps`)

La estructura de programación de la función *ajuste* se presenta en el apéndice A.1.4

### 3.5. Representación de las entidades de los modos A, B y C (función: *conf.mode*)

Los parámetros de la función *conf.mode* corresponden a: un objeto lista generado al ejecutar la función `tucker3`, el modo a representar y la definición del plano a visualizar por medio de los ejes x y y.

La programación de la función *conf.mode* se realizó considerando la conceptualización teórica presentada por (Kiers 2000a), según la cual se calculan las matrices de componentes que contienen las coordenadas para los individuos, variables y condiciones, a partir de las matrices  $\mathbf{A}$ ,  $\mathbf{B}$  y  $\mathbf{C}$  denotadas  $\tilde{\mathbf{A}}$ ,  $\tilde{\mathbf{B}}$  y  $\tilde{\mathbf{C}}$ .

La función *conf.mode* representa en los diferentes planos de proyección las entidades de los modos A, B y C, según la dimensión del modelo seleccionado y proyecta en el mismo plano los ejes originales utilizando las funciones `s.label` y `s.arrow` de la librería `ade4` (Dray et al. 2007).

La estructura de programación de la función *conf.mode* se presenta en el apéndice A.1.5

### 3.6. Representación de las trayectorias de la combinación de las entidades de dos modos (función: *traj.modes*)

Los parámetros de la función *traj.modes* corresponden a: un objeto tipo lista con las estimaciones generadas mediante la función `tucker3`, la combinación de componentes y los ejes del plano de interés a representar.

La función *traj.modes* permite representar la configuración de las trayectorias de un

modo a través de otro modo. Para obtener esta representación se utilizó el procedimiento sugerido por (Kiers 2000a), ubicando las coordenadas de una combinación de modos en las filas de la matriz  $\mathbf{F}$ , calculada a partir de las estimaciones de los parámetros del modelo.

La función permite generar representaciones de las trayectorias combinando los diferentes modos, utilizando la función *s.traject* de *ade4*.

La estructura de programación de la función *traj.modes* se presenta en el apéndice A.1.6.

### 3.7. Representación de los elementos del arreglo central $\mathbf{G}_a$ (función: *repcore*)

La función *repcore* depende del objeto lista que contiene las estimaciones del arreglo central  $\mathbf{G}_a$ ,  $\mathbf{G}_b$  y  $\mathbf{G}_c$ , así como de la definición del número de componente de los modos A y B que se deseen representar.

La función permite representar los valores del arreglo central, trazando las trayectorias de los componentes del modo A (sujetos), a través de los componentes del modo C (ocasiones), en cada uno de los componentes del modo B (variables).

La estructura de programación de la función *repcore* se presenta en el apéndice A.1.7.

---

## Aplicación y comparación

---

### 4.1. Objetivo y descripción

El ejemplo utilizado en esta sección para realizar la comparación entre los métodos, pretende describir la situación sociodemográfica y de salud de 24 departamentos de Colombia en el periodo 2004-2007, identificando los aspectos más relevantes asociados a los departamentos, variables y periodo de estudio.

Los datos corresponden a un arreglo de dimensión  $(24 \times 12 \times 4)$ . Contienen la evaluación en  $I=24$  departamentos de Colombia, de  $J=12$  variables sociodemográficas: tasa bruta de natalidad (TBN), esperanza de vida al nacer (EVN), tasa bruta de mortalidad (TBM), mortalidad neonatal precoz (MNP), mortalidad materna (MM), población con necesidades básicas insatisfechas (PNBI), población bajo línea de indigencia (PBLI), población bajo línea de pobreza (PBLP), índice de desarrollo humano (IDH), índice de condiciones de vida (ICV), población en miseria (PM) y años promedio de educación (APE), a través de  $K=4$  condiciones correspondientes a los años 2004 a 2007 y se presentan en el apéndice A.2.

Las definiciones de las variables se tomaron del documento: "Situación de salud en Colombia: indicadores básicos", disponible en <http://www.minproteccionsocial.gov.co> (acceso marzo 20 de 2009) y se presentan en el apéndice A.3.

Los resultados de la comparación entre los métodos AFM y Tucker 3 aplicados en la base de datos descrita anteriormente, se estructuran considerando los criterios de comparación definidos previamente. Se presenta, en su orden, la comparación respecto a los criterios: preprocesamiento de datos, elección de la dimensionalidad de la solución e inercia retenida, estimación de parámetros, contribuciones, ayudas a la interpretación: contribuciones y calidad de la representación, representaciones y medidas de relación.

### 4.2. Preprocesamiento de datos

Los datos se estandarizan (centrado y escalamiento) previo a la implementación de ambos métodos. Se consideró la necesidad de estandarizar los datos porque según la definición

establecida en el apéndice A.3, cada variable evaluada tiene unidades de medida diferente.

En AFM se utiliza el procedimiento establecido en la función MFA de la librería FactoMineR en ambiente R (Husson et al. 2008), que realiza estandarización (centrado y escalado) de las columnas de la matriz de dimensión  $24 \times 48$  para este caso.

En Tucker 3 los datos se estandarizan utilizando el procedimiento programado en la función *preproc* en ambiente R, la cual según se especificó previamente, realiza centrado a través de individuos sobre la matriz de dimensión  $24 \times 48$  y posteriormente escalado sobre la matriz centrada desplegada por el modo B (variables) de dimensión  $12 \times 96$ , que corresponde a escalar dentro de variables, tomando la sugerencia de Bro & Smilde (2003).

Los dos métodos son similares en el centrado al realizarlo por individuos; difieren en el escalado, el AFM lo realiza a través de individuos y el Tucker 3 a través de variables y condiciones. Esta diferencia se origina en la forma de operar de las dos metodologías, el AFM desarrolla en el análisis global un ACP dos vías sobre una gran tabla considerando cada ocurrencia de una variable en las distintas condiciones como una variable diferente, mientras el Tucker 3 desarrolla un ACP tres vías en el que las diferentes ocurrencias de una variable se consideran como una variable única, esto hace que para el AFM sea posible estandarizar por individuos, mientras para el Tucker 3 no se recomienda (Bro & Smilde 2003).

### 4.3. Elección de la dimensionalidad de la solución e inercia

El paso inicial, al abordar el análisis de la situación planteada, consiste en la elección de la dimensionalidad de la solución (número de componentes a retener).

En AFM, según la información proporcionada por los valores propios del análisis global presentados en la tabla 4.1, una solución de 2 componentes explica un porcentaje importante de variabilidad (71,3 %) y asocia valores propios que cumplen con el criterio de ser superiores a 1.0. Sin embargo, la observación del plano (1,3) permitió detectar una distribución para los departamentos no observable en el primer plano (1,2). Considerando que el eje adicional podía aportar tanto en interpretación como en inercia, se optó por una solución de un sistema de ejes conformado por tres componentes, denotado  $S=3$ .

Tabla 4.1: Diez primeros valores propios del análisis global del AFM

No.	Valor propio	Porcentaje	P.acumulado
comp 1	3.88	54.69	54.69
comp 2	1.18	16.59	71.28
comp 3	0.66	9.31	80.59
comp 4	0.42	5.96	86.55
comp 5	0.26	3.66	90.21
comp 6	0.18	2.51	92.72
comp 7	0.12	1.74	94.46
comp 8	0.09	1.26	95.73
comp 9	0.06	0.89	96.61
comp 10	0.04	0.57	97.19

Se encontró que los valores propios asociados a estos tres primeros factores, explican

el 80,6 % de la información inicial, adicionalmente, el valor del primer valor propio (3.88) cercano a  $K=4$  refleja la presencia del mismo primer factor en todos los años evaluados. (tabla 4.1).

En Tucker 3 la elección de la dimensionalidad de la solución, se realizó utilizando el gráfico *scree plot multiway* (figura 4.1) y el criterio *Dif-Fit* de Timmerman y Kiers citados por Kroonenberg (2008). Según estos, el mejor modelo corresponde a la combinación  $P=3$ ,  $Q=3$  y  $R=2$  ( $S=8$ ), con un ajuste correspondiente al 80,1 %.

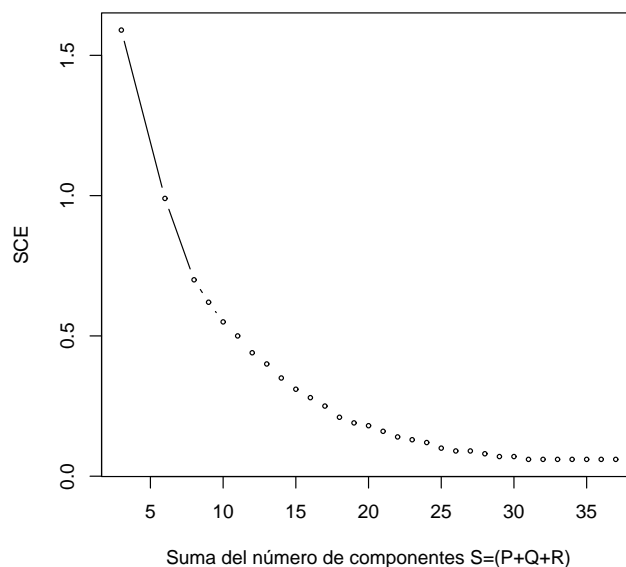


Figura 4.1: *Scree plot multiway*

Respecto a la solución seleccionada, las dos metodologías difieren respecto a:

- los criterios utilizados para seleccionar la solución. El AFM lo realiza mediante el número de ejes que sean interpretables y expliquen un porcentaje adecuado de variabilidad y el Tucker 3 utiliza criterios de ajuste.

- el número de sistemas de ejes que conforman la solución. Para el caso tratado, el AFM seleccionó un sistema de tres ejes, mientras el Tucker 3 escogió tres sistemas de 3, 3 y 2 ejes respectivamente.

Respecto a la inercia retenida o variabilidad explicada por la solución seleccionada, para el caso tratado, las dos metodologías mostraron ser similares, presentando una diferencia mínima de 0.5 unidades porcentuales.

#### 4.4. Contribución de los componentes

En relación a la contribución general de cada factor a la inercia total; en AFM se observa que el primer factor extraído contribuye con 54,7 % a la inercia total, el segundo



factor lo hace con 16,6 % y el tercero con 9,3 % (tabla 4.1).

El modelo de Tucker 3 permite generar la contribución de los componentes extraídos por cada uno de los modos. Con la solución seleccionada ( $P=3$ ,  $Q=3$ ,  $R=2$ ), la contribución de los componentes del modo A fue de 55,5 %, 16,8 % y 7,8 %; los componentes del modo B contribuyeron en 55,5 %, 16,5 % y 8,0 % y los del modo C en 78,6 % y 1,4 %, respectivamente. En todos los modos se destaca el primer componente por contribuir en un porcentaje importante a explicar la variabilidad total.

La tabla 4.2 presenta la contribución de cada triada de componentes a la inercia total. La combinación de los primeros componentes de cada modo se destaca por ser la de mayor contribución con 55 %, seguida por la combinación correspondiente al segundo componente del modo A y B y el primer componente del modo C con 16,2 % y por último la combinación de los terceros componentes del modo A y B y el primer componente del modo C con 7,48 %. Aunque la contribución del segundo componente del modo C es una medida pesimista de contribución, se consideró conservarlo respetando la elección del mejor modelo a través de los criterios utilizados y la interpretación suministrada por este componente.

Tabla 4.2: Contribución de triadas de componentes a la inercia total

	B1C1	B2C1	B3C1	B1C2	B2C2	B3C2
A1	55.00	0.00	0.01	0.017	0.31	0.26
A2	0.00	16.19	0.038	0.52	0.00	0.01
A3	0.00	0.034	7.48	0.12	0.00	0.20

Las dos metodologías permiten determinar la contribución de los componentes, llegando a concluir en ambas que el primer componente explica la mayor variabilidad. Debido a la generación de componentes independientes por cada modo, el Tucker 3 genera una medida porcentual de contribución de la interacción entre componentes que le permite identificar la combinación de estos que explica el mayor porcentaje de variabilidad, el AFM no genera esta medida debido a que extrae un sistema de ejes dual sobre el que se representa individuos, variables y condiciones.

## 4.5. Estimación de componentes, ayudas a la interpretación y representación de individuos, variables y condiciones

### 4.5.1. Individuos

En AFM con la solución  $S = 3$ , las estimaciones generadas a partir del análisis global corresponden a los elementos de la matriz de componentes para individuos de dimensión  $24 \times 3$  y contiene las coordenadas de los 24 departamentos sobre los tres factores extraídos del análisis global (tabla 4.3). Se presentan adicionalmente las contribuciones y calidad de representación de los departamentos a los tres factores seleccionados. El departamento de Chocó y la ciudad de Bogotá se destacan por su contribución y calidad de representación en el primer factor. Los departamentos de la costa caribe (Guajira, Bolívar, Córdoba, Magdalena y Sucre) muestran un aporte importante sobre el segundo factor. Finalmente, en el tercer factor se destacan Chocó por ser muy contributivo y estar moderadamente

representado y Sucre por estar bien representado y ser contributivo.

Tabla 4.3: Coordenadas de individuos y ayudas a la interpretación (contribuciones y cosenos) en AFM

Departamento	Coordenadas			Contribuciones			Cosenos		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
Antioquia	-1.46	-0.43	0.05	2.30	0.66	0.02	0.75	0.07	0.00
Atlántico	-1.69	1.01	0.23	3.07	3.58	0.32	0.53	0.18	0.01
Bogotá	-4.39	0.94	1.11	20.74	3.15	7.83	0.84	0.04	0.00
Bolívar	0.30	1.75	0.01	0.01	10.89	0.00	0.02	0.75	0.00
Boyacá	0.81	-1.35	-0.97	0.70	6.44	5.91	0.14	0.41	0.21
Caldas	-0.95	-1.09	-0.12	0.97	4.17	0.09	0.28	0.37	0.00
Caquetá	1.89	-0.21	0.93	3.82	0.15	5.43	0.44	0.00	0.10
Cauca	2.04	-0.74	-0.58	4.48	1.92	2.11	0.62	0.08	0.05
Cesar	0.75	1.03	-0.15	0.61	3.74	0.14	0.16	0.31	0.01
Chocó	5.96	-0.78	2.05	38.16	2.13	26.50	0.84	0.01	0.10
Córdoba	1.94	1.49	-1.24	4.02	7.85	9.70	0.41	0.24	0.16
Cundinamarca	-0.86	-1.17	-0.65	0.80	4.84	2.70	0.18	0.34	0.11
Huila	-0.06	0.02	-0.73	0.00	0.00	3.38	0.00	0.00	0.42
Guajira	0.14	2.18	0.55	0.02	16.83	1.90	0.00	0.68	0.00
Magdalena	0.60	1.49	0.55	0.38	7.84	1.93	0.09	0.59	0.08
Meta	-0.31	-0.98	0.79	0.11	3.36	3.91	0.02	0.19	0.12
Nariño	1.85	-0.63	-0.83	3.69	1.42	4.32	0.62	0.07	0.12
N-Santander	-0.23	-0.18	0.33	0.06	0.11	0.69	0.02	0.01	0.05
Quindio	-1.53	-1.18	0.60	2.51	4.90	2.24	0.47	0.28	0.07
Risaralda	-1.65	-0.30	-0.01	2.90	0.31	0.00	0.73	0.02	0.00
Santander	-1.53	-0.49	-0.38	2.50	0.83	0.89	0.73	0.07	0.04
Sucre	1.01	1.48	-1.35	1.11	7.71	11.14	0.16	0.35	0.29
Tolima	-0.09	-1.26	-0.91	0.01	5.63	5.18	0.00	0.47	0.25
Valle	-2.53	-0.63	0.71	6.91	1.41	3.22	0.84	0.05	0.07

En Tucker 3 la tabla 4.4, presenta la matriz de componentes para individuos, de dimensión  $24 \times 3$  y contiene las coordenadas de los 24 departamentos sobre los tres factores extraídos.

A partir de la interpretación de estas matrices, las dos metodologías permiten identificar tres factores comunes a los departamentos, el primero determina una diferencia entre departamentos, destacando una oposición marcada entre los departamentos de Chocó y Bogotá. El segundo deja clara la diferencia entre departamentos de la costa atlántica (Guajira, Bolívar, Córdoba y Sucre) y departamentos de la zona centro (Tolima y Boyacá). En el tercer factor se marca nuevamente diferencia entre departamentos, esta vez destacando la oposición entre Chocó y el grupo de Sucre y Córdoba.

Las representaciones de los departamentos generadas por los dos métodos, permiten corroborar la interpretación anterior. En el primer plano, los departamentos de Bogotá y Chocó mantienen una posición opuesta y distante del centro de la nube (ver figuras 4.2 y 4.3).

En relación con el segundo eje se observa oposición entre departamentos de la costa caribe (Guajira, Bolívar y Magdalena) y departamentos de la zona central del país (Tolima, Boyacá y Quindio). El tercer componente por su parte, opone claramente el departamento de Chocó a Sucre y Córdoba (figuras 4.4 y 4.5).

Tabla 4.4: Matriz de componentes para individuos en Tucker 3

Departamento	Dim.1	Dim.2	Dim.3
Antioquia	-0.21	-0.04	0.00
Atlántico	-0.26	0.13	-0.03
Bogotá	-0.63	0.20	-0.13
Bolívar	0.04	0.20	-0.00
Boyacá	0.10	-0.18	0.13
Caldas	-0.16	-0.17	0.02
Caquetá	0.28	-0.08	-0.15
Cauca	0.27	-0.13	0.04
Cesar	0.12	0.15	0.04
Chocó	0.85	-0.08	-0.25
Córdoba	0.28	0.16	0.16
Cundinamarca	-0.14	-0.14	0.06
Huila	-0.00	-0.018	0.10
Guajira	0.15	0.41	-0.05
Magdalena	0.10	0.20	-0.06
Meta	-0.06	-0.11	-0.12
Nariño	0.24	-0.10	0.10
N-Santander	-0.02	-0.05	-0.04
Quindio	-0.24	-0.15	-0.09
Risaralda	-0.25	-0.07	-0.01
Santander	-0.23	-0.05	0.05
Sucre	0.16	0.16	0.21
Tolima	-0.02	-0.19	0.10
Valle	-0.37	-0.06	-0.10

Las dos metodologías generaron representaciones muy similares para la representación de los departamentos.

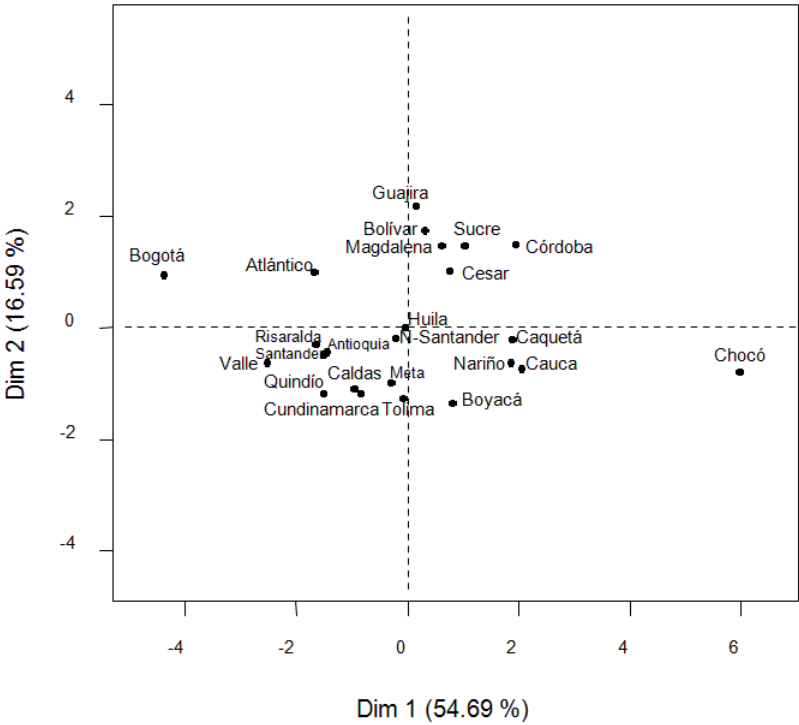


Figura 4.2: Departamentos de Colombia en plano factorial (1,2) en AFM

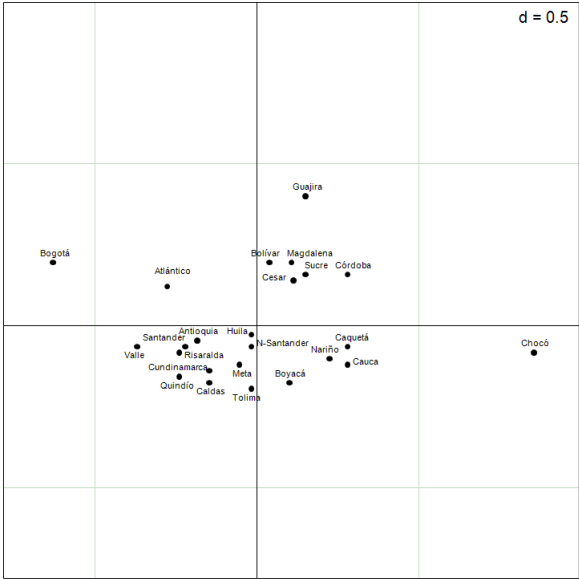


Figura 4.3: Departamentos de Colombia en plano factorial (1,2) en Tucker 3

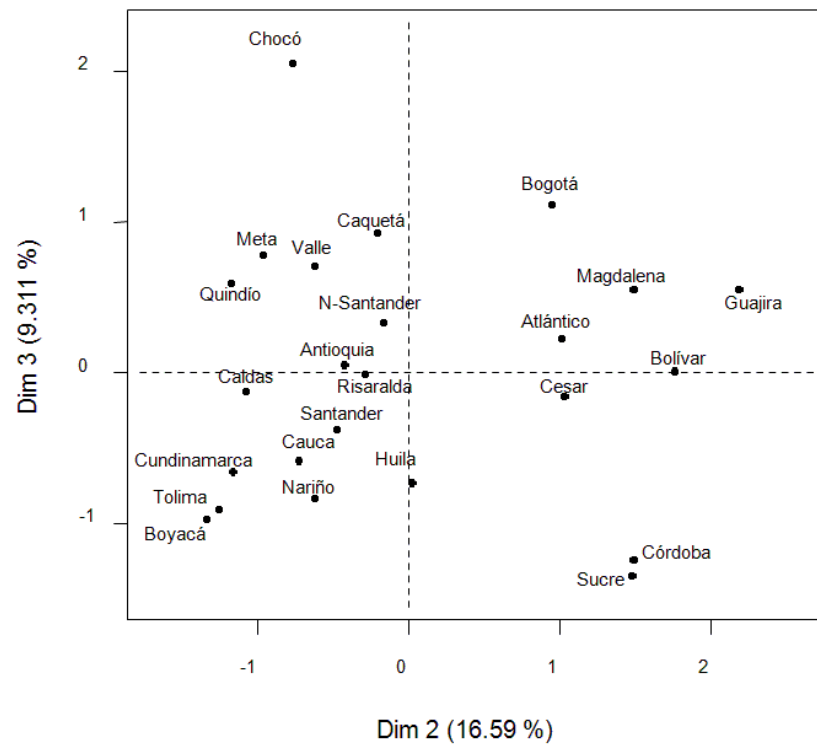


Figura 4.4: Departamentos de Colombia en plano factorial (2,3) en AFM

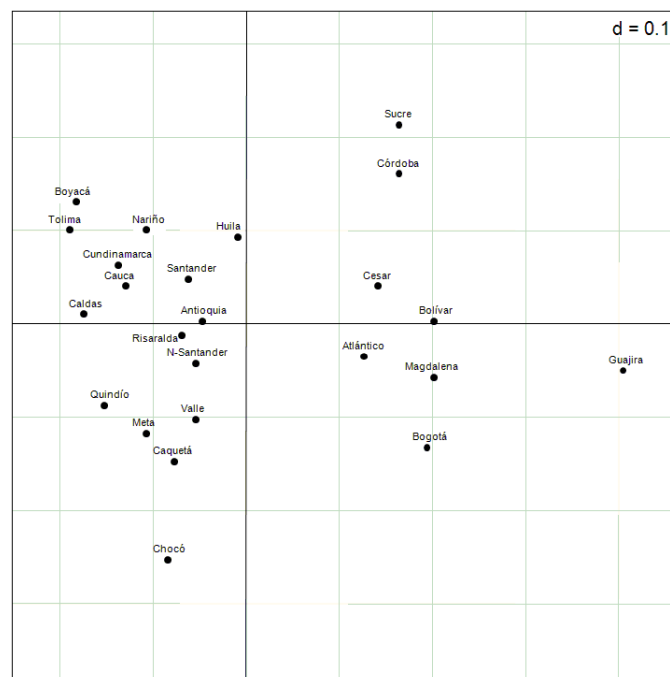


Figura 4.5: Departamentos de Colombia en plano factorial (2,3) en Tucker 3

### 4.5.2. Variables

En AFM con la solución  $S = 3$ , las estimaciones generadas a partir del análisis global corresponden a los elementos de la matriz de componentes para variables de dimensión  $48 \times 3$  y contiene las coordenadas de las 12 variables evaluadas en cada uno de los 4 años estudiados, sobre los tres factores extraídos (tabla 4.5). Se presentan adicionalmente las contribuciones y calidad de representación. Se destacan las variables que evalúan desarrollo y pobreza ICV, IDH, PNBI, por su contribución y calidad de representación sobre el primer factor. La variable TBM que evalúa mortalidad general sobresale por su contribución y calidad de representación sobre el segundo factor. Las mortalidades materna y neonatal (MM, MNP) se destacan por estar bien representadas y contribuir a la formación del tercer factor.

En Tucker 3, la tabla 4.6 presenta la matriz de componentes para variables, de dimensión  $12 \times 3$  y contiene las coordenadas de las 12 variables sobre los tres factores extraídos.

Los métodos difieren en la dimensión de la matriz de componentes para variables, la estimada por el AFM es de mayor dimensión debido a que en el análisis global se consideran  $K$  subtablas de  $J$  variables originando la estimación de las  $JK$  variables, mientras el Tucker 3 considera el modo variables constituido por  $J$  y realiza la estimación considerando esta dimensión.

Ambos métodos permiten identificar tres factores comunes en relación con las variables, el primero marca la diferencia entre las variables que evalúan el desarrollo humano (IDH, ICV) de aquellas indicadoras de pobreza (PNBI, PBLP, PBLI) en todos los años. El segundo factor se asocia con un componente de mortalidad general, marcando la diferencia entre TBM y EVN. El tercer componente se asocia a variables relacionadas con la gestación, indicando una diferencia entre las variables que evalúan mortalidad materna y neonatal (MNP, MM) y esperanza de vida (EVN).

Las representaciones de las variables generadas por los dos métodos permiten observar el comportamiento mencionado, en AFM la figura 4.6 presenta la configuración de las variables activas del análisis global, donde por tratarse de un ACP estandarizado, la coordenada para cada variable corresponde a la correlación entre esta y el eje global del análisis.

En Tucker 3 la figura 4.7 presenta la configuración de las 12 variables (entidades del modo B), en el primer plano, en ambos casos existe similitud en la distribución de las variables en el plano. La diferencia en ubicación obedece a una rotación.

Las dos representaciones permiten generar conclusiones similares. El primer eje marca la diferencia entre las variables que evalúan el desarrollo humano (IDH, ICV) de aquellas indicadoras de pobreza (PNBI, PBLP, PBLI) en todos los años, por lo que puede definirse como un componente que evalúa calidad de vida. El segundo eje se asocia con mortalidad general, marcando la diferencia entre TBM y EVN. El tercer componente se relaciona con variables que evalúan mortalidad relacionada con la gestación, indicando una diferencia entre estas variables y la esperanza de vida (EVN). Considerando las variables que aportan a este último eje, puede relacionarse con calidad de atención en salud.

Tabla 4.5: Coordenadas (correlaciones) de variables y ayudas a la interpretación (contribuciones y cosenos) en AFM

Variable	Coordenadas			Contribuciones			Cosenos		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
TBN-2004	0.56	0.37	0.09	1.22	1.80	0.21	0.31	0.14	0.01
EVN-2004	-0.46	0.69	-0.42	0.82	6.2	4.10	0.21	0.48	0.18
TBM-2004	0.35	-0.90	0.09	0.49	10.48	0.18	0.13	0.81	0.00
MNP-2004	0.60	0.37	0.39	1.39	1.79	3.59	0.35	0.14	0.16
MM-2004	0.74	-0.26	0.47	2.17	0.86	4.99	0.55	0.07	0.21
PNBI-2004	0.85	0.33	-0.01	2.87	1.43	0.00	0.73	0.11	0.00
PBLI-2004	0.77	-0.06	-0.21	2.34	0.05	1.04	0.59	0.00	0.05
PBLP-2004	0.81	-0.02	-0.26	2.58	0.01	1.56	0.66	0.00	0.07
IDH-2004	-0.87	0.06	0.11	2.96	0.05	0.27	0.75	0.00	0.01
ICV-2004	-0.94	-0.08	0.27	3.48	0.08	1.72	0.89	0.01	0.07
PM-2004	0.75	0.53	-0.10	2.21	3.60	0.24	0.56	0.27	0.01
APE-2004	-0.77	0.43	0.33	2.35	2.43	2.54	0.60	0.18	0.11
TBN-2005	0.53	0.48	0.02	1.19	3.26	0.01	0.28	0.23	0.00
EVN-2005	-0.58	0.62	-0.30	1.42	5.37	2.28	0.34	0.38	0.09
TBM-2005	0.32	-0.89	-0.06	0.44	11.02	0.09	0.10	0.79	0.00
MNP-2005	0.71	0.28	0.46	2.14	1.09	5.23	0.51	0.08	0.21
MM-2005	0.80	-0.20	0.40	2.71	0.55	3.98	0.64	0.04	0.16
PNBI-2005	0.84	0.32	-0.02	2.97	1.44	0.01	0.71	0.10	0.00
PBLI-2005	0.53	0.03	-0.56	1.16	0.01	7.64	0.27	0.00	0.31
PBLP-2005	0.52	0.05	-0.63	1.15	0.04	9.98	0.27	0.00	0.40
IDH-2005	-0.91	0.00	-0.01	3.51	0.00	0.00	0.83	0.00	0.00
ICV-2005	-0.93	0.02	0.21	3.61	0.00	1.06	0.86	0.00	0.04
PM-2005	0.75	0.53	-0.10	2.37	3.87	0.26	0.56	0.28	0.01
APE-2005	-0.77	0.43	0.33	2.52	2.61	2.73	0.60	0.19	0.11
TBN-2006	0.54	0.40	0.11	1.07	1.98	0.26	0.29	0.16	0.01
EVN-2006	-0.58	0.62	-0.31	1.23	4.71	2.06	0.33	0.39	0.09
TBM-2006	0.32	-0.89	-0.06	0.38	9.65	0.08	0.10	0.79	0.00
MNP-2006	0.67	0.08	0.51	1.63	0.07	5.65	0.44	0.01	0.26
MM-2006	0.79	-0.03	0.50	2.28	0.01	5.48	0.62	0.00	0.25
PNBI-2006	0.89	0.27	0.16	2.93	0.86	0.55	0.79	0.07	0.03
PBLI-2006	0.77	-0.23	-0.35	2.21	0.65	2.66	0.60	0.05	0.12
PBLP-2006	0.82	-0.04	-0.47	2.46	0.02	24.77	0.67	0.00	0.22
IDH-2006	-0.93	0.06	-0.03	3.19	0.05	0.02	0.86	0.00	0.00
ICV-2006	-0.94	-0.11	0.16	3.29	0.14	0.55	0.89	0.01	0.02
PM-2006	0.78	0.50	0.04	2.27	3.06	0.03	0.61	0.25	0.01
APE-2006	-0.78	0.34	0.33	2.27	1.38	2.42	0.62	0.11	0.11
TBN-2007	0.76	0.43	0.29	1.95	2.07	1.62	0.58	0.19	0.08
EVN-2007	-0.74	0.34	-0.34	1.86	1.29	2.29	0.55	0.11	0.11
TBM-2007	0.36	-0.81	-0.03	0.44	7.29	0.02	0.13	0.65	0.00
MNP-2007	0.64	0.17	0.57	1.40	0.31	6.39	0.41	0.02	0.32
MM-2007	0.59	-0.11	0.64	1.19	0.15	8.13	0.35	0.01	0.41
PNBI-2007	0.86	0.45	0.03	2.52	2.29	0.01	0.74	0.20	0.00
PBLI-2007	0.88	-0.21	-0.06	2.63	0.51	0.08	0.77	0.05	0.00
PBLP-2007	0.90	0.01	-0.27	2.74	0.00	1.49	0.81	0.00	0.08
IDH-2007	-0.93	0.06	-0.03	2.93	0.04	0.01	0.86	0.00	0.00
ICV-2007	-0.94	-0.16	0.18	2.97	0.28	0.64	0.87	0.02	0.03
PM-2007	0.67	0.62	-0.00	1.53	4.31	0.00	0.45	0.39	0.00
APE-2007	-0.85	0.26	0.23	2.47	0.73	1.04	0.73	0.06	0.05

Tabla 4.6: Matriz de componentes para variables en Tucker 3

Variable	Dim.1	Dim.2	Dim.3
TBN	-0.40	-0.24	-0.07
EVN	0.33	-0.27	0.19
TBM	-0.16	0.46	0.00
MNP	-0.34	-0.08	-0.27
MM	-0.32	0.09	-0.27
PNBI	-0.50	-0.20	0.00
PBLI	-0.40	0.10	0.14
PBLP	-0.43	0.04	0.21
IDH	0.46	-0.09	0.01
ICV	0.50	0.01	-0.10
PM	-0.45	-0.34	0.04
APE	0.42	-0.22	-0.14

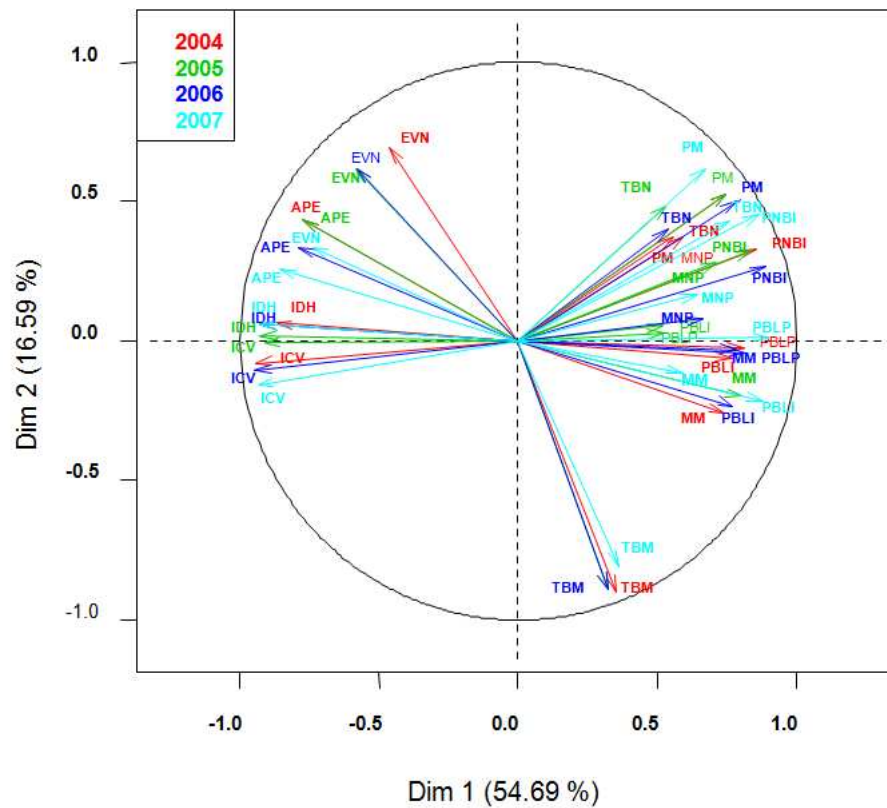


Figura 4.6: Círculo de correlaciones (coordenadas) para las variables en el plano factorial (1,2) en AFM



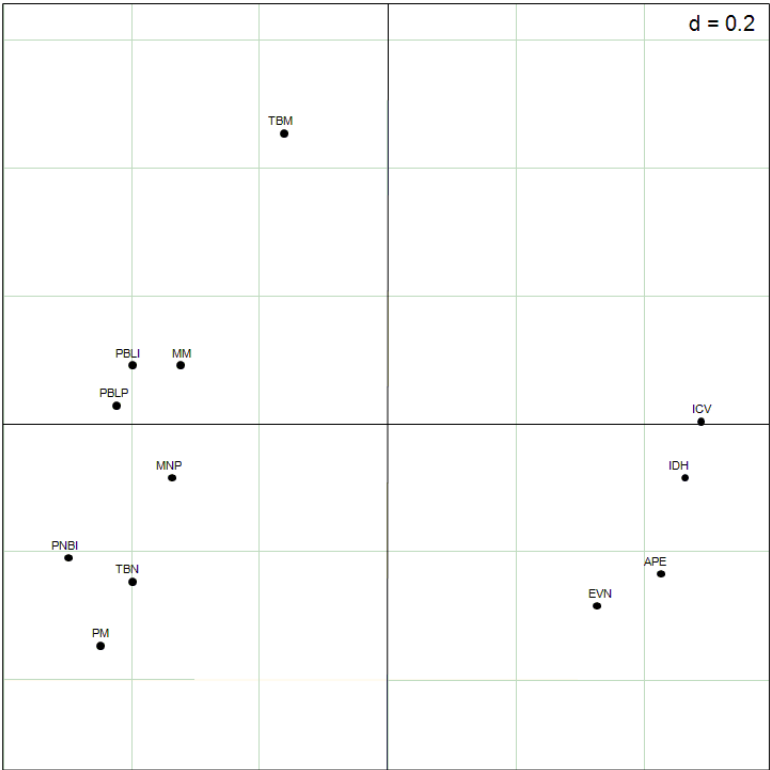


Figura 4.7: Variables sobre el plano (1,2) en Tucker 3

### 4.5.3. Condiciones

A partir del análisis de condiciones, el AFM estima la matriz de correlación componente-condición presentada en la tabla 4.7. Estas correlaciones corresponden a las coordenadas para las condiciones sobre los tres factores extraídos. En todos los casos, se mantiene una relación positiva entre los años y el componente. En la primera componente, los valores son bastante próximos del valor máximo 1, constituyéndose en una dirección de inercia importante para cada año. La segunda y tercera componentes están relacionadas moderada y levemente con los diferentes años. Se presenta además información referente a contribuciones y calidad de representación, el primer factor está igualmente afectado por los cuatro años evaluados, es decir, el conjunto de variables evaluadas en cada uno de los años contribuye prácticamente al 25 % de la inercia recogida por el primer factor. Respecto al factor 2, la inercia está explicada principalmente por las variables evaluadas en los años 2004 y 2005 y en menor porcentaje por las evaluadas durante el 2006 y 2007. El año 2005 le contribuye de manera importante al factor 3. Todos los periodos se encuentran muy bien representados en el factor 1 (con más del 78 %).

Tabla 4.7: Coordenadas de condiciones (años 2004-2007) y ayudas a la interpretación en AFM

Año	Coordenadas			Contribuciones			Cosenos		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
2004	0.97	0.34	0.14	24.90	28.81	20.45	0.79	0.10	0.02
2005	0.98	0.34	0.22	25.21	29.30	33.27	0.78	0.10	0.04
2006	0.98	0.27	0.16	25.25	22.60	24.52	0.84	0.06	0.02
2007	0.96	0.23	0.14	24.63	19.29	21.74	0.82	0.05	0.02

En Tucker 3 la tabla 4.8 presenta la matriz de componentes para las condiciones de dimensión  $4 \times 2$ , la cual contiene las coordenadas de los 4 años estudiados sobre los dos factores extraídos por el modo C. El primer componente es un promedio de los cuatro años bajo estudio. El segundo componente representa la diferencia entre los años 2004-2006 y el año 2007.

Tabla 4.8: Matriz de componentes para condiciones (años 2004-2007) en Tucker 3

Año	Dim.1	Dim.2
2004	-0.63	-0.07
2005	-0.75	-0.12
2006	-0.63	-0.07
2007	-1.19	0.16

En la situación planteada los dos métodos difieren en la interpretación obtenida para los componentes relacionados con las condiciones. El AFM extrae tres factores, identificando solo el primero como relevante, mientras el Tucker 3 extrae y justifica la extracción de los dos factores sugeridos por el modelo.

Las representaciones de los años sobre el primer plano del AFM presentado en la figura 4.8 destaca la alta relación entre todos los años y el primer factor, mencionada anteriormente.

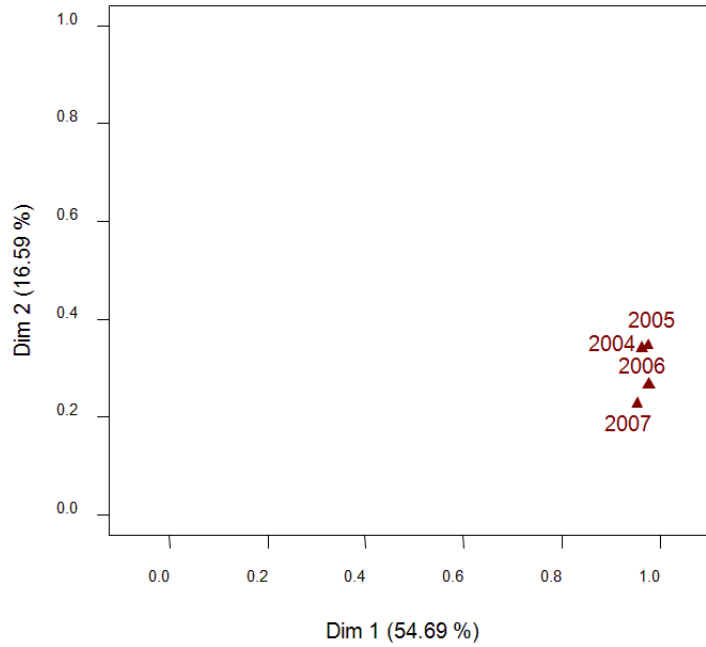


Figura 4.8: Condiciones (años 2004-2007) sobre el plano (1,2) en AFM

En Tucker 3, la distribución de los años evaluados se presenta sobre el primer plano en la figura 4.9. En la representación, los diferentes años se ubican en la misma dirección sobre el primer eje. Respecto al segundo eje, el año 2007 se ubica opuesto a los años 2004-2006.

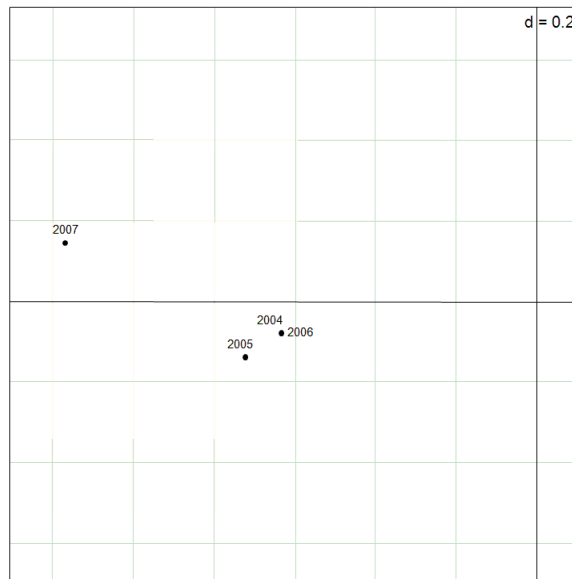


Figura 4.9: Condiciones (años 2004-2007) sobre el plano (1,2) en Tucker 3

A diferencia del AFM, el Tucker 3 permite una mejor separación de los años, distanciando el año 2007 de 2004-2006 sobre el primer y segundo eje, permitiendo justificar la

inclusión de un segundo componente por el modo C.

En general, el AFM se destaca por aportar medidas específicas de calidad de representación y de contribución muy importantes para la interpretación de los planos factoriales, de las cuales carece el modelo de Tucker 3, siendo esta una ventaja del AFM sobre el Tucker 3.

## 4.6. Trayectorias

En el AFM, la justificación de estudiar la representación simultánea de los departamentos se encuentra al analizar la razón de inercia presentada en la tabla 4.9, donde se observan valores cercanos a 1 para el primer y segundo factor. Esto indica que estos factores mantienen durante los cuatro años de estudio una estructura muy similar y que la representación simultánea obtenida para todos los departamentos conserva características comunes a través de los años, que justifica un estudio de las diferencias. Esta estabilidad se observa en la representación del primer plano (figura 4.10). Aunque el valor de razón de inercia obtenido para el eje 3 es un poco inferior que los anteriores, refleja una estabilidad importante de considerar y la utilidad de su estudio. La estabilidad menor en este eje se puede identificar en la representación del plano (2,3) (ver figura 4.12).

Tabla 4.9: Razón de inercia = Inercia inter/Inercia total

Dim.1	0.97
Dim.2	0.92
Dim.3	0.77

En el primer plano (figura 4.10), la disposición de los cuatro años ante el centroide de cada departamento es muy similar, es decir que el nivel de calidad de vida y el comportamiento frente a la mortalidad durante los cuatro años estudiados en cada departamento parece ser muy similar. Al respecto, se destaca la estructura del departamento de Chocó que en los últimos años (2006-2007) presenta tendencia a alejarse aún más de su centroide, mostrando un comportamiento heterogéneo en los años analizados, relacionado posiblemente con incremento en los niveles de pobreza. Por el contrario, Bogotá presenta un comportamiento más homogéneo, siendo uno de los más estables.

Se identifican algunos departamentos con estructura similar. Esto es, sus centros de gravedad están muy próximos. Un ejemplo de tal estructura lo conforma el grupo de departamentos de la costa atlántica (Guajira, Magdalena, Córdoba, Bolívar y Cesar), caracterizados por presentar valores elevados de natalidad.

En Tucker 3, la figura 4.11 muestra la representación conjunta de cada departamento en los cuatro años sobre el primer plano factorial. Respecto al primer eje, se destaca el comportamiento inestable en las trayectorias generadas para los departamentos de Bogotá y Chocó, marcado principalmente por el alejamiento en el año 2007 y en relación con pobreza, que en el Chocó obedece a incremento de las variables que la evalúan (PNBI, PBLI y PBLP) y en Bogotá a disminución de las mismas. En relación al segundo eje, se destaca el comportamiento inestable en la trayectoria obtenida para la Guajira, con alejamiento destacado hacia el año 2007 y en relación con incremento en natalidad y

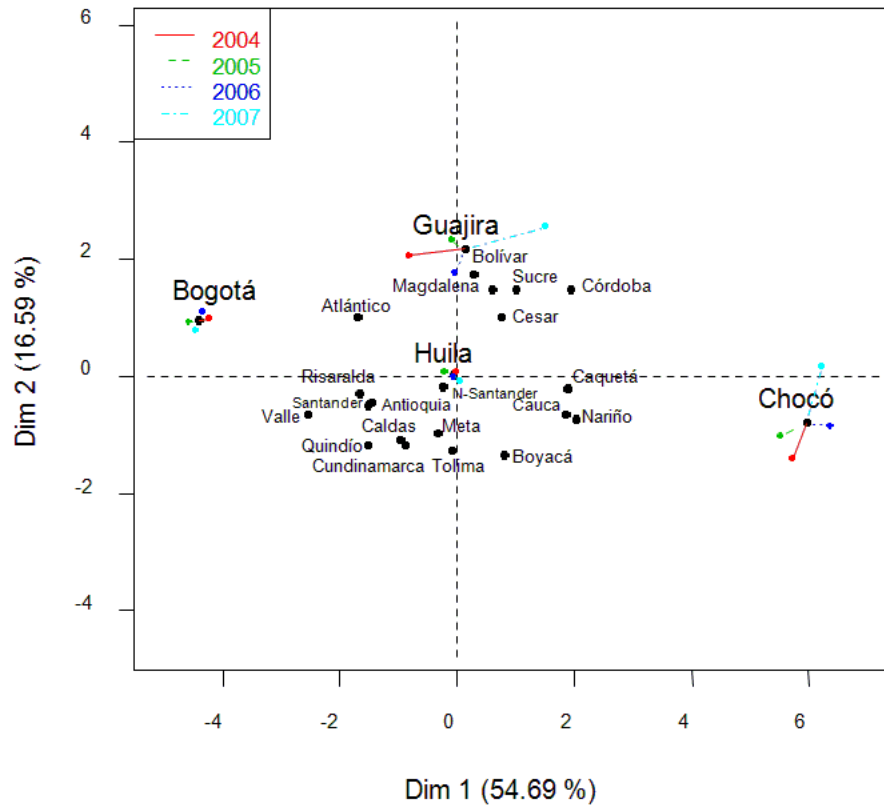


Figura 4.10: Trayectorias de los departamentos en el periodo 2004-2007 en AFM, plano (1,2)

necesidades básicas insatisfechas.

En el plano (2,3) del AFM (figura 4.12), se destaca la ubicación y estructura inestable del departamento de Chocó que junto a la moderada calidad de su representación destacada previamente (4.3) sugiere una cuidadosa interpretación del comportamiento observado para este departamento en esta representación, a pesar de ser el más contributivo al tercer eje. En oposición a este, se destaca Sucre y Córdoba (departamentos contributivos y bien representados), por ubicarse entre los departamentos que hacia el año 2007 mostraron las tasas de mortalidad materna y neonatal más bajas.

En Tucker 3 la figura 4.13 respecto al eje 2, destaca el comportamiento de la Guajira mostrando una tendencia de alejamiento en el último año, posiblemente relacionado con incremento en la natalidad. Respecto al eje 3, se destaca Chocó por presentar en el año 2007 una trayectoria de acercamiento, relacionada con disminución de la mortalidad materna, indicando posibles mejoras en la calidad de la atención en salud que podrían llevar a la disminución de los valores en esta variable. Este mismo comportamiento se observa en la trayectoria de Sucre y Córdoba, pero con ubicación opuesta a la de Chocó debido a que estos departamentos se registran dentro de aquellos con menor mortalidad materna y neonatal.

El Tucker 3 permite generar adicionalmente las trayectorias o comportamiento de cada departamento a través de las variables y de las variables a través de los cuatro años. Esta

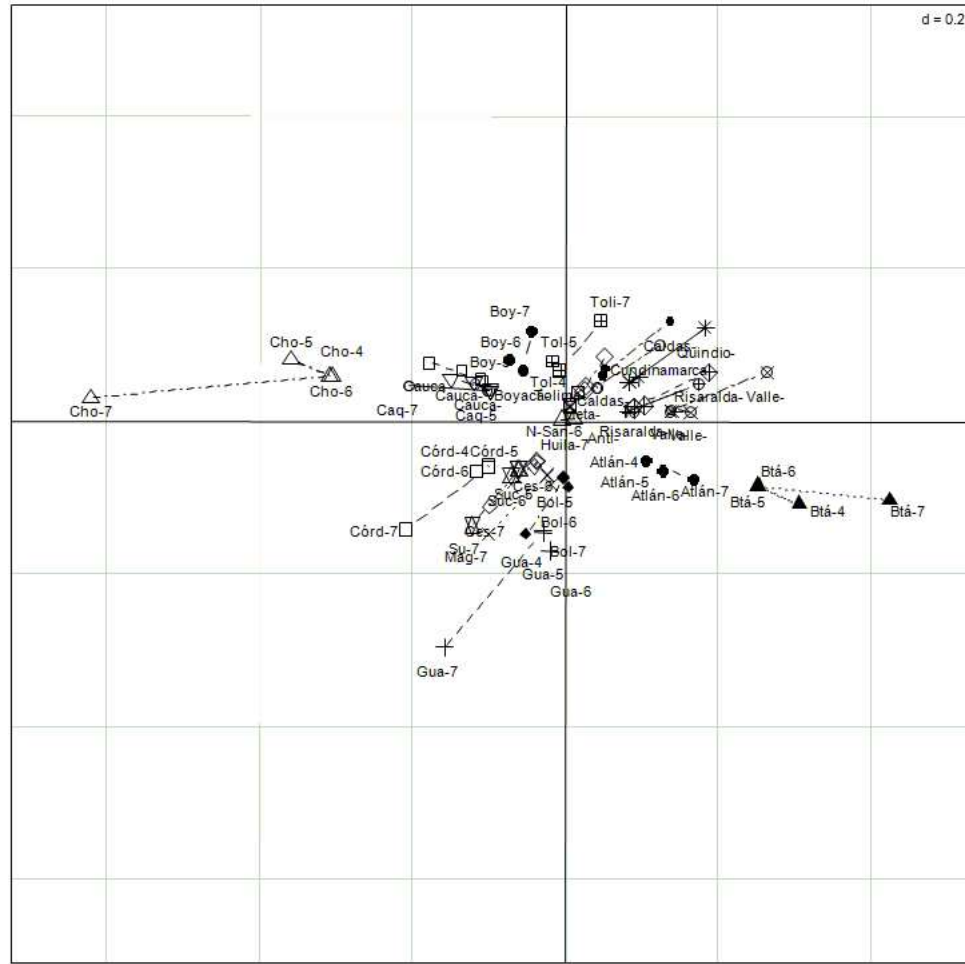
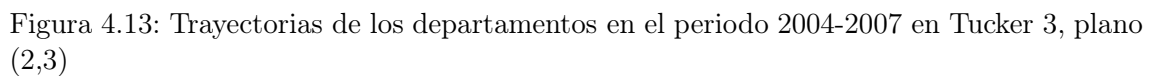
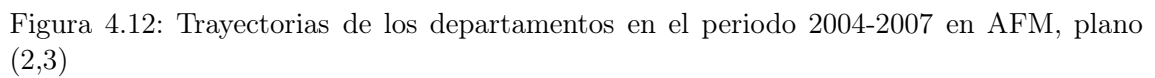


Figura 4.11: Trayectorias de los departamentos en el periodo 2004-2007 en Tucker 3, plano (1,2)

última representación se muestra en la figura 4.14, en la cual se destaca la inestabilidad presentada en la mayoría de variables, las cuales en el año 2007 tomaron valores más alejados que los reportados en años anteriores.

Las trayectorias de los departamentos son similares en ambos métodos, sin embargo Tucker 3 destaca más los puntos correspondientes al año 2007.

El método de Tucker 3 genera trayectorias de variables a través de condiciones e individuos a través de variables, las cuales son adicionales a la única trayectoria generada por el AFM, aportando con estas representaciones información adicional al análisis.



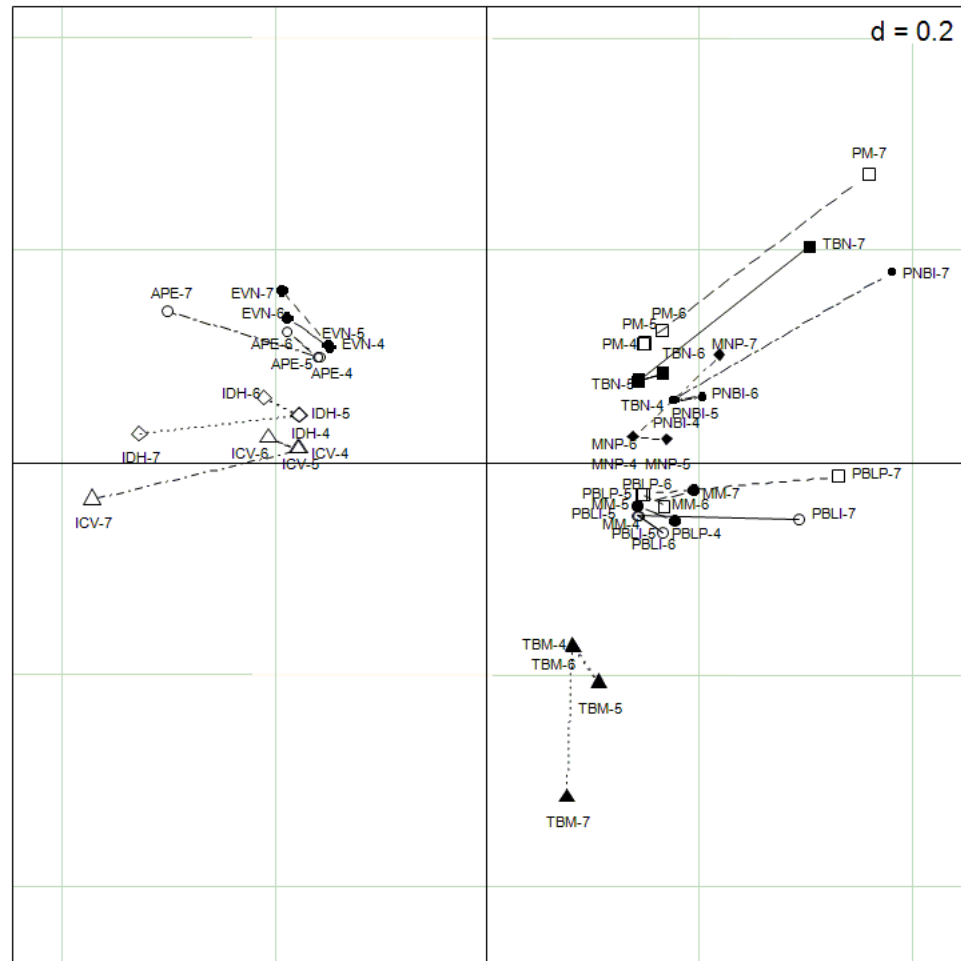


Figura 4.14: Trayectorias de las variables en el periodo 2004-2007 en Tucker 3, plano (1,2)



## 4.7. Medidas de relación

Los dos métodos permiten obtener información concerniente al establecimiento de relaciones, sin embargo, difieren respecto al tipo de relación establecida. El AFM establece tres medidas de relación: -correlación ejes parciales-eje general, -correlación entre ejes parciales y -correlación entre condiciones mediante el coeficiente RV. El Tucker 3 determina la relación entre los componentes del arreglo central.

La tabla 4.10 presenta la contribución de los ejes parciales a la conformación de los 3 factores generales extraídos. Los ejes parciales 1,2 y 3 contribuyen de manera importante a la conformación del primer, segundo y tercer factor general, respectivamente, reflejando la existencia de tres factores comunes importantes, durante los cuatro años evaluados. Para los ejes 1 y 2, esta contribución corrobora la alta correlación observada en la figura 4.15.

Tabla 4.10: Contribución de los ejes parciales en AFM

	Dim.1	Dim.2	Dim.3
Dim1.2004	24.87	0.03	0.17
Dim2.2004	0.00	81.22	3.49
Dim3.2004	0.00	1.21	75.85
Dim4.2004	0.03	0.08	43.40
Dim5.2004	0.29	0.52	20.67
Dim1.2005	25.14	0.00	0.96
Dim2.2005	0.00	83.70	0.30
Dim3.2005	0.23	0.09	108.88
Dim4.2005	0.04	0.16	24.33
Dim5.2005	0.00	0.17	4.60
Dim1.2006	25.22	0.13	0.01
Dim2.2006	0.03	77.39	8.48
Dim3.2006	0.01	3.84	108.47
Dim4.2006	0.13	0.41	15.08
Dim5.2006	0.00	0.31	0.00
Dim1.2007	24.56	0.30	1.29
Dim2.2007	0.14	69.46	2.95
Dim3.2007	0.17	3.05	120.42
Dim4.2007	0.00	0.71	3.47
Dim5.2007	0.01	0.02	0.12

La correlación entre ejes parciales del análisis, se presenta en la tabla 4.11. Las primeras y segundas componentes muestran una elevada correlación durante los cuatro años estudiados. Este hecho puede observarse claramente en la figura 4.15 para el plano (1,2).

Tabla 4.11: Correlación entre ejes parciales en AFM

	1.2004	2.2004	3.2004	1.2005	2.2005	3.2005	1.2006	2.2006	3.2006	1.2007	2.2007	3.2007
1.2004	1.00	-0.00	-0.00	0.98	-0.01	-0.02	0.95	0.02	-0.02	0.93	-0.11	-0.09
2.2004	-0.00	1.00	-0.00	0.03	0.98	0.11	-0.03	0.90	-0.32	0.15	0.84	-0.31
3.2004	-0.00	-0.00	1.00	-0.05	0.13	-0.68	-0.02	0.31	0.33	0.07	0.08	0.49
1.2005	0.98	0.03	-0.05	1.00	-0.00	0.00	0.97	0.01	-0.06	0.94	-0.07	-0.1
2.2005	-0.01	0.98	0.13	-0.00	1.00	-0.00	-0.05	0.95	-0.27	0.05	0.86	-0.23
3.2005	-0.02	0.11	-0.68	0.00	-0.00	1.00	-0.13	-0.23	-0.55	-0.20	-0.10	-0.60
1.2006	0.95	-0.03	-0.02	0.97	-0.05	-0.13	1.00	0.00	0.00	0.97	-0.10	-0.09
2.2006	0.02	0.90	0.31	0.01	0.95	-0.23	0.00	1.00	-0.00	-0.00	0.84	0.01
3.2006	-0.02	-0.32	0.33	-0.06	-0.27	-0.55	0.00	-0.00	1.00	0.01	-0.01	0.87
1.2007	0.93	0.04	0.07	0.94	0.05	-0.20	0.97	0.11	0.01	1.00	-0.00	-0.00
2.2007	-0.11	0.84	0.08	-0.07	0.86	-0.10	-0.10	0.84	-0.01	-0.00	1.00	-0.00
3.2007	-0.09	-0.31	0.49	-0.14	-0.23	-0.60	-0.09	0.01	0.87	-0.00	-0.00	1.00

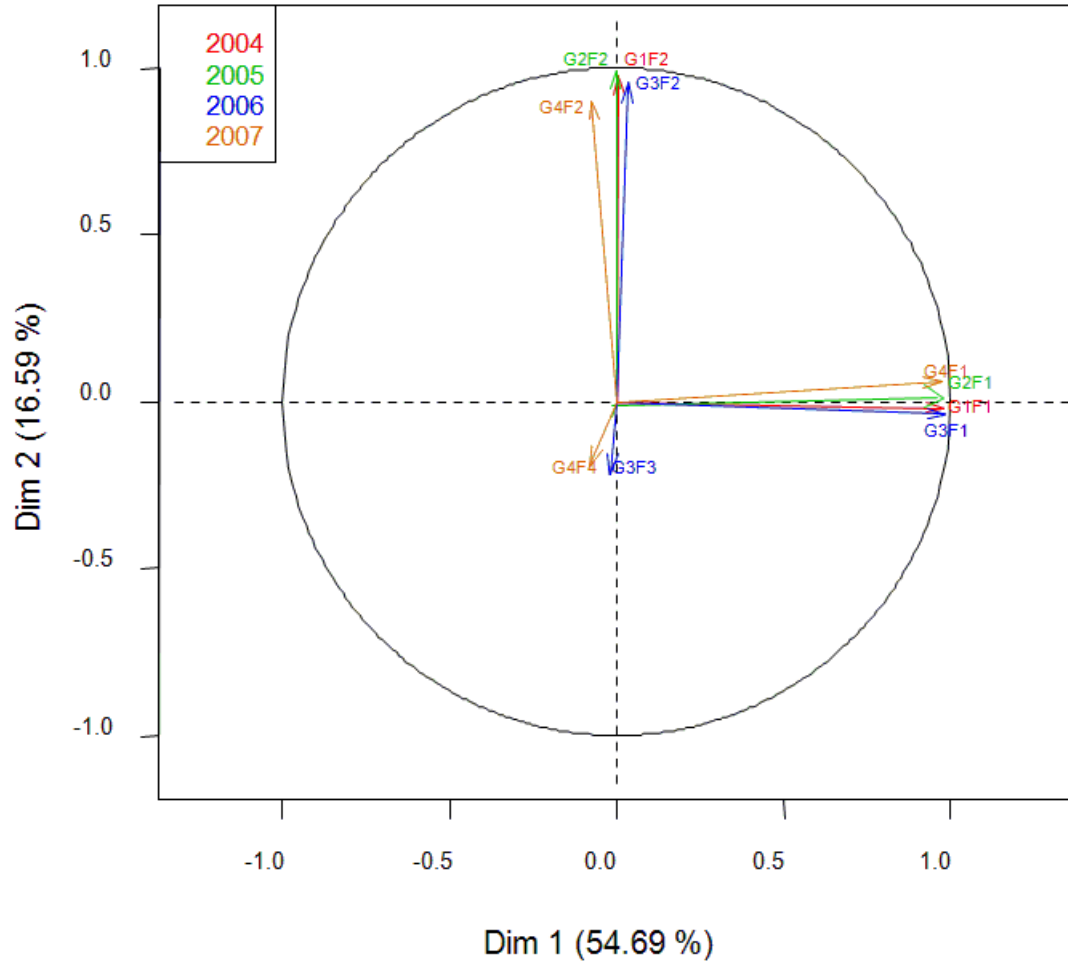


Figura 4.15: Correlación ejes parciales eje general

Los valores para el coeficiente RV determinados por el AFM se presentan en la tabla 4.12. Se generaron valores superiores a 0.85, indicando una relación alta del conjunto de variables evaluadas entre los diferentes años, tendencia esperada en este caso por tratarse de una situación que está evaluando comportamiento de las variables en el tiempo.

El modelo de Tucker 3 estimó el arreglo central  $\mathbf{G}_a$ , de dimensión  $3 \times 6$ , que contiene los valores del arreglo central (tabla 4.13), en la cual se destaca la combinación de los primeros componentes extraídos por cada modo por capturar de manera importante la interacción tres vías presente en los datos, que en términos de la situación estudiada corresponde a la interacción entre departamentos que difieren respecto a calidad de vida en el promedio de años estudiados.

La representación de estos valores corresponde a las trayectorias presentadas en la

Tabla 4.12: Coeficientes RV en AFM

	2004	2005	2006	2007	MFA
2004	1.00	0.95	0.90	0.85	0.96
2005	0.95	1.00	0.92	0.86	0.97
2006	0.90	0.92	1.00	0.93	0.97
2007	0.85	0.86	0.93	1.00	0.94
MFA	0.96	0.97	0.97	0.94	1.00

Tabla 4.13: Arreglo central  $\mathbf{G}_a$  en Tucker 3

	B1C1	B2C1	B3C1	B1C2	B2C2	B3C2
A1	1.39	-0.00	0.01	-0.03	-0.10	0.10
A2	-0.00	0.76	0.04	-0.14	-0.01	-0.02
A3	0.01	0.03	-0.51	0.07	0.00	-0.08

figura 4.16, las cuales no son paralelas, indicando diferencias en el comportamiento de los departamentos respecto a los tres componentes identificados como: calidad de vida, mortalidad y calidad de los servicios de salud durante el periodo estudiado. La primera representación de la figura corresponde a la triada de componentes que explica el 55 % de la variabilidad, la cual se interpreta diciendo que el puntaje de departamentos en los que se registra alta calidad de vida y alta pobreza difiere si se considera el periodo estudiado como un promedio de los cuatro años, a si se consideran por separado 2004-2006 vs 2007.

Las relaciones establecidas por cada método son diferentes, siendo estas: la relación entre componentes en el caso de Tucker 3 y la relación entre ejes parciales y globales y entre condiciones en el caso del AFM. Esta diferencia le aporta información adicional y complementaria al análisis de datos tres vías.

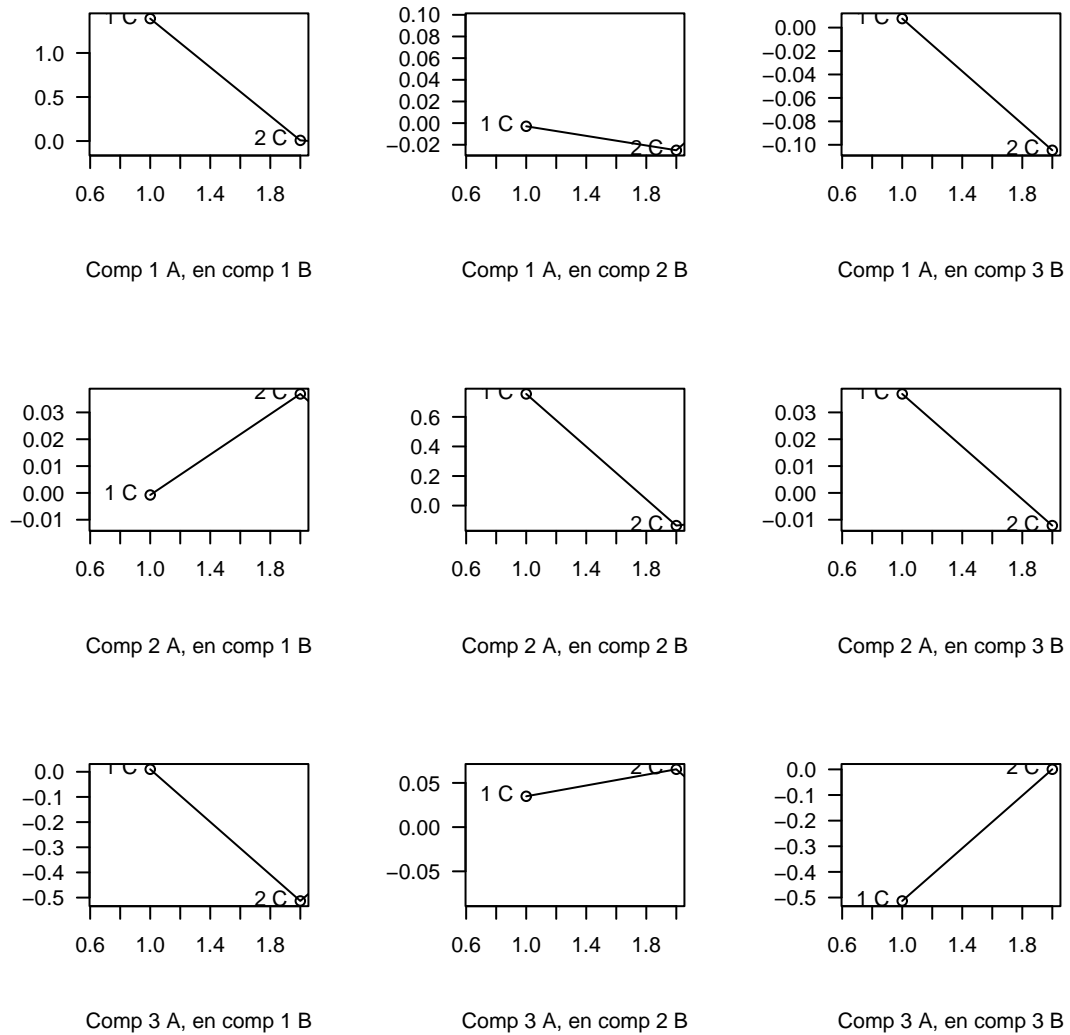


Figura 4.16: Trayectorias de los elementos del arreglo central en Tucker 3

---

## Conclusiones y Recomendaciones

---

- A pesar de las diferencias metodológicas presentes entre los dos métodos, estos muestran elementos en común relacionados con: el preprocesamiento adecuado de los datos, las representaciones de individuos, variables, condiciones y trayectorias de individuos a través de condiciones.

- La ventaja del AFM respecto al Tucker 3 está definida por la generación de medidas específicas de contribuciones y calidad de representación de individuos y variables a la inercia de cada factor. El Tucker 3 por su parte presenta una ventaja respecto al AFM en relación con las trayectorias, las cuales favorecen la interpretación y son más diversas. La identificación de ventajas por parte de los dos métodos son los elementos que permiten sugerir considerarlos como complementarios (capítulo 2).

- Las funciones programadas permitieron la implementación de los aspectos relacionados con matriciación, preprocesamiento de datos, estimación de parámetros y elección del mejor modelo, determinación de ajuste y generación de representaciones de los diferentes modos y combinación de estos para el modelo Tucker 3 (capítulo 3).

- El análisis de la situación sociodemográfica y de salud de los departamentos de Colombia a través del modelo de Tucker 3 y del AFM puede resumirse en ambos casos a través de tres factores. El primero, relacionado con desarrollo y pobreza, sobre el cual Bogotá presenta las mejores condiciones de vida en oposición a la situación de Chocó, que muestra niveles extremos de pobreza. El segundo, asociado con natalidad y mortalidad, que presenta a los departamentos de la costa atlántica con elevada natalidad, mientras que los departamentos ubicados en la zona central del país tienen niveles de natalidad bajos. El tercero relacionado con mortalidad materna y neonatal, presenta a Chocó con niveles altos sobre este factor, en contraste con los departamentos de Sucre y Córdoba que presentan los niveles más bajos del país, posiblemente asociado con mejoras en la calidad de los servicios de salud. El comportamiento de los departamentos en los cuatro años son similares pero en el 2007 se nota, sobre todo, una acentuación de la pobreza en el Chocó en oposición a una mejoría en Bogotá (capítulo 4).

- Se recomienda continuar la programación del método de Tucker 3 en ambiente R, para permitir un conocimiento del método a profundidad y facilitar su accesibilidad, enfocándose en los aspectos relacionados con posprocesamiento, validación y análisis de residuales sugeridos por Kroonenberg (2008).

---

## Apéndices

---

### A.1. Programación del modelo de Tucker 3 en ambiente R

#### A.1.1. Matriciación del arreglo de datos original (función: *matriciacion*)

```
matriciacion=function(lista){
  N=length(lista)
  Xa=as.matrix(lista[[1]])
  for (i in 2:N){
    Xa=cbind(Xa,lista[[i]])
  }
  matricesT=lapply(lista,t)
  Xb=matricesT[[1]]
  for (i in 2:N){
    Xb=cbind(Xb,matricesT[[i]])
  }
  Xc=as.vector(as.matrix(lista[[1]]))
  for (i in 2:N){
    Xc=rbind(Xc,as.vector(as.matrix(lista[[i]])))
  }
  Xa=as.data.frame(Xa)
  Xb=as.data.frame(Xb)
  row.names(Xc)=NULL
  Xc=as.data.frame(Xc)
  matriciaciones=list(Xa,Xb,Xc)
  names(matriciaciones)=c("Xa","Xb","Xc")
  k=length(names(lista))
  j=ncol(lista[[1]])
  indcond=paste(row.names(lista[[1]]),names(lista)[1],sep="-")
  varcond=paste(names(lista[[1]]),names(lista)[1],sep="-")
  indvar=paste(row.names(lista[[1]]),names(lista[[1]])[1],sep="-")
  for(i in 2:j){
    indvar=c(indvar,paste(row.names(lista[[1]]),names(lista[[1]])[i],sep="-"))
  }
  for(i in 2:k){
    indcond=c(indcond,paste(row.names(lista[[1]]),names(lista)[i],sep="-"))
    varcond=c(varcond,paste(names(lista[[1]]),names(lista)[i],sep="-"))
  }
}
```

```

}
names(matriciaciones$Xa)=varcond
names(matriciaciones$Xb)=indcond
names(matriciaciones$Xc)=indvar
row.names(matriciaciones$Xc)=names(lista)
class(matriciaciones)="matriciacion"
matriciaciones

```

### A.1.2. Preprocesamiento (función: *preproc*)

```

preproc=function(X,scale=TRUE){
  ind=row.names(X$Xa);var=row.names(X$Xb);cond=row.names(X$Xc)
  indcond=names(X$Xb);varcond=names(X$Xa);indvar=names(X$Xc)
  Xa=as.matrix(X$Xa);Xb=as.matrix(X$Xb);Xc=as.matrix(X$Xc)
  na=nrow(Xa);nb=nrow(Xb);nc=nrow(Xc)
  if(class(X) != "matriciacion")
    stop("Objetct must be of class 'matriciacion'")
  Xacol=ncol(Xa)
  I=na
  Port=diag(c(rep(1,I)))-(1/I)*c(rep(1,I))%*%t(c(rep(1,I)))
  Xac=Port%*%Xa
  Xbac=t(Xac[,1:nb])
  for (i in seq((nb+1),nb*nc,nb)){
    Xbac=cbind(Xbac,t(Xac[,i:(i+(nb-1))]))
  }
  for(j in 0:(nb-1)){
    D=Xac[, (j+1)]
    for(i in (j + seq((nb+1),nb*nc,nb))){
      D=cbind(D,Xac[,i])
    }
    if(j==0) {
      Xcac=t(D)
      next
    }
    Xcac=cbind(Xcac,t(D))
  }
  if(scale==TRUE){
    Wj=diag(1/sqrt(diag(Xbac%*%t(Xbac))/(na*nc)))
    Wk=diag(1/sqrt(diag(Xcac%*%t(Xcac))/(na*nb)))
    Ya=Xac%*%kronecker(Wk,Wj)
    Yb=t(Ya[,1:(1+(nb-1))])
    for (i in seq((nb+1),nb*nc,nb)){
      Yb=cbind(Yb,t(Ya[,i:(i+(nb-1))]))
    }
    for(j in 0:(nb-1)){
      D=Ya[, (j+1)]
      for(i in (j + seq((nb+1),nb*nc,nb))){
        D=cbind(D,Ya[,i])
      }
      if(j==0) {
        Yc=t(D)
        next
      }
      Yc=cbind(Yc,t(D))
    }
  }
  Ya=as.data.frame(Ya)

```

```

Yb=as.data.frame(Yb)
row.names(Yc)=NULL
Yc=as.data.frame(Yc)
row.names(Ya)=ind;row.names(Yb)=var;row.names(Yc)=cond
names(Ya)=varcond;names(Yb)=indcond;names(Yc)=indvar
Y=list(Ya,Yb,Yc)
names(Y)=c("Ya","Yb","Yc")
class(Y)="matriciacion"
}
Xac=as.data.frame(Xac)
Xbac=as.data.frame(Xbac)
row.names(Xcac)=NULL
Xcac=as.data.frame(Xcac)
row.names(Xac)=ind
row.names(Xbac)=var
row.names(Xcac)=cond
names(Xac)=varcond
names(Xbac)=indcond
names(Xcac)=indvar
X1=list(Xac,Xbac,Xcac)
names(X1)=c("Xac","Xbac","Xcac")
class(X1)="matriciacion"
if(scale==TRUE) return(Y)
if(scale==FALSE) return(X1)
}

```

### A.1.3. Estimación de parámetros del modelo de Tucker 3 (función: *tucker3*)

```

tucker3=function(lista,N=100){
  if(class(lista)!="matriciacion") stop("Object must be of class 'matriciacion'")
  varcond=names(lista[[1]]);indcond=names(lista[[2]]);indvar=names(lista[[3]])
  cond=row.names(lista[[3]]);var=row.names(lista[[2]]);ind=row.names(lista[[1]])
  X1=as.matrix(lista[[1]]);X2=as.matrix(lista[[2]]);X3=as.matrix(lista[[3]])
  P=nrow(X1)-1;Q=nrow(X2)-1;R=nrow(X3)-1
  c1=c(1:R);c2=c(1:Q);c3=c(1:P)
  a1=c(rep(R,Q))
  a2=c(rep(Q*R,P))
  s1=c(rep(c2,a1))
  w1=c(rep(c1,P*Q))
  w2=c(rep(s1,P))
  w3=c(rep(c3,a2))
  Z=cbind(w3,w2,w1)
  pq=w3*w2;qr=w2*w1;pr=w3*w1
  L=pq>=w1 & qr>=w3 & pr>=w2
  Z=Z[L,]
  SCE1=c(rep(0,N))
  SSX1=sum(X1^2)
  A=list();B=list();C=list();G1=list();G2=list();G3=list()
  X1est=c();G1t=c();G2t=c();G3t=c();At=c();Bt=c();Ct=c()
  Ncomp=as.vector(tapply(Z,row(Z),sum))
  Z=cbind(Z,Ncomp)
  maxi=min((P+1),((Q+1)*(R+1)))+min((Q+1),
    ((P+1)*(R+1)))+min((R+1),((P+1)*(Q+1)))-3
  if (maxi==P+Q+R){
    Z=Z[Z[,4]!=max(Z[,4]),]
  }
}

```



```

    Z=Z[Z[,4]!=max(Z[4,]),]
  }
  else{
    Z=Z[Z[,4]<maxi,]
  }
  nZ=nrow(Z)
  SCE=c(rep(0,nZ))
  o=order(Z[,4])
  for(i in 1:nZ){
    Bt=svd(X2)
    Ct=svd(X3)
    for(j in 2:(N+1)){
      Aant=At;Bant=Bt;Cant=Ct
      G1ant=G1t;G2ant=G2t;G3ant=G3t
      SSG1ant=sum(G1ant^2)
      X1estant=X1est
      At=svd(X1%*%kronecker(Ct$u[,1:Z[i,2:3][2]],Bt$u[,1:Z[i,2:3][1]]))
      Bt=svd(X2%*%kronecker(Ct$u[,1:Z[i,2:3][2]],At$u[,1:Z[i,1]]))
      Ct=svd(X3%*%kronecker(Bt$u[,1:Z[i,2:3][1]],At$u[,1:Z[i,1]]))
      G1t=t(At$u[,1:Z[i,1]])%*%X1%*%kronecker(Ct$u[,1:Z[i,2:3][2]],
      Bt$u[,1:Z[i,2:3][1]])
      G2t=t(Bt$u[,1:Z[i,2:3][1]])%*%X2%*%kronecker(Ct$u[,1:Z[i,2:3][2]],
      At$u[,1:Z[i,1]])
      G3t=t(Ct$u[,1:Z[i,2:3][2]])%*%X3%*%kronecker(Bt$u[,1:Z[i,2:3][1]],
      At$u[,1:Z[i,1]])
      X1est<-(At$u[,1:Z[i,1]]%*%G1t%*%kronecker(t(Ct$u[,1:Z[i,2:3][2]]),
      t(Bt$u[,1:Z[i,2:3][1]])))
      E1=X1-X1est
      SCE1[j]=sum(E1^2)
      dif=abs(SCE1[j] - SCE1[(j-1)])
      if(dif<=1e-10) {
        A[[i]]=Aant$u[,1:Z[i,1]];B[[i]]=Bant$u[,1:Z[i,2]];C[[i]]
        =Cant$u[,1:Z[i,3]]
        G1[[i]]=G1ant;G2[[i]]=G2ant;G3[[i]]=G3ant;SCE[i]=SCE1[j-1]
        break
      }
      if(j==(N+1)) {
        SSG1=sum(G1t^2)
        A[[i]]=At$u[,1:Z[i,1]];B[[i]]=Bt$u[,1:Z[i,2]];C[[i]]=Ct$u[,1:Z[i,3]]
        G1[[i]]=G1t;G2[[i]]=G2t;G3[[i]]=G3t
        SCE[i]=SCE1[j]
        break
      }
    }
  }
}
SCE=round(SSE,2)
SPM=cbind(Z,SSE)
SPM=SPM[o,]
id=1:nrow(SPM)
SPM=as.data.frame(SPM)
names(SPM)=c("P","Q","R","No de comp.,"SCE")
SSEmin=tapply(SPM[,5], SPM[,4],min)
ncomp=as.numeric(names(SCEmin))
plot(ncomp,SCEmin,type="b",xlab="Suma del número de componentes S=(P+Q+R)",
ylab="SCE",cex=.5)
print(SPM)
length(SCEmin)->n
SSE2=SCEmin[2:n]

```

```

SSE1=SCEmin[1:(n-1)]
diffS=SCE1-SSE2
b=c()
nd=length(diffS)
for(i in 1:(nd-1)){
  if(diffS[(i+1)]!=0) b[i]=diffS[i]/max(c(diffS[(i+1):nd]))
  if(diffS[(i+1)]==0) b[i]=0
}
nb=names(diffS)[2:(n-1)]
mod=nb[b==max(b)]
id=row.names(SPM[SPM[,4]==mod & SPM[,5]==SCEmin[names(SCEmin)==mod],])
cat("Selected model with Dif-fit criterion is:",id,"\n")
cat("Select the combination of components (number between 1 and", nZ,")","\n")
j=as.integer(readLines(n = 1))
jo=o[j]
def=as.integer(SPM[jo,1:3])
p=def[1];q=def[2];r=def[3]
A=A[[jo]];B=B[[jo]];C=C[[jo]]
G1=G1[[jo]];G2=G2[[jo]];G3=G3[[jo]]
res=list(A,B,C,G1,G2,G3,X1,varcond,indcond,indvar,ind,var,cond)
names(res)=c("A", "B", "C", "Ga", "Gb", "Gc", "Xa", "varcond", "indcond",
"indvar", "ind", "var", "cond")
return(res)
}

```

#### A.1.4. Cálculo de medidas de ajuste del modelo de Tucker 3 (función: *ajuste*)

```

ajuste=function(lista){
  AJtotal=(sum(lista$Ga^2))/(sum(lista$Xa^2))*100
  contrib.compsA=(diag(as.matrix(lista$Ga)%*%t(as.matrix(lista$Ga)))/
/sum(lista$Xa^2))*100
  contrib.compsB=(diag(as.matrix(lista$Gb)%*%t(as.matrix(lista$Gb)))/
/sum(lista$Xa^2))*100
  contrib.compsC=(diag(as.matrix(lista$Gc)%*%t(as.matrix(lista$Gc)))/
/sum(lista$Xa^2))*100
  contrib.triadacomps=(as.matrix(lista$Ga)*as.matrix(lista$Ga))*
(100/sum(lista$Xa^2))
  res=list(AJtotal,contrib.compsA,contrib.compsB,contrib.compsC,
contrib.triadacomps)
  names(res)=c("AJtotal", "contrib.compsA", "contrib.compsB", "contrib.compsC",
"contrib.triadacomps")
  return(res)
}

```

#### A.1.5. Representaciones de las entidades de cada modo (función: *conf.mode*)

```

conf.mode=function(lista,x=1,y=2,s=1,mode="A",xlim=NULL,ylim=NULL){
  library(ade4)
  library(MASS)
  A=lista$A;B=lista$B;C=lista$C
  nind=nrow(A);nvar=nrow(B);ncond=nrow(C)
  Ga=lista$Ga;Gb=lista$Gb;Gc=lista$Gc
  ind=lista$ind;var=lista$var;cond=lista$cond
  varcond=lista$varcond;indcond=lista$indcond;indvar=lista$indvar

```

```

Fa=kronecker(C,B)%*%t(Ga);Fb=kronecker(C,A)%*%t(Gb);Fc=kronecker(B,A)%*%t(Gc)
Ta=diag(svd(Fa)$d);Tb=diag(svd(Fb)$d);Tc=diag(svd(Fc)$d)
Amod=A%*%Ta;Bmod=B%*%Tb;Cmod=C%*%Tc
Famod=Fa%*%ginv(Ta);Fbmod=Fb%*%ginv(Tb);Fcm=Fc%*%ginv(Tc)
sub="Configuracion de las entidades del modo";sub=paste(sub,mode);
ejes="(ejes ";ejes=paste(ejes,x,y,"");sub=paste(sub,ejes)
if(mode=="A"){
  Ad=A[,c(x,y)]
  Amodd=Amod[,c(x,y)]
  Fad=Fa[,c(x,y)]
  Famodd=Famod[,c(x,y)]
  if(ncol(A)==1) stop("Selected mode must have for than one column")
  s.label(Amodd,clabel=0.5,label=ind,xlim=xlim,ylim=ylim,sub=sub,csup=1.25)
  s.arrow(s*Amodd,xax=1,yax=2,label=varcond,clabel=0.5,boxes=FALSE,
  add.plot=TRUE)
}
if(mode=="B"){
  Bd=B[,c(x,y)]
  Bmodd=Bmod[,c(x,y)]
  Fbd=Fb[,c(x,y)]
  Fbmodd=Fbmod[,c(x,y)]
  if(ncol(B)==1) stop("Selected mode must have for than one column")
  s.label(Bmodd,clabel=0.5,label=var,xlim=xlim,ylim=ylim,sub=sub,csup=1.25)
  s.arrow(s*Bmodd,xax=1,yax=2,label=indcond,clabel=0.5,boxes=FALSE,
  add.plot=TRUE)
}
if(mode=="C"){
  Cd=C[,c(x,y)]
  Cmodd=Cmod[,c(x,y)]
  Fcd=Fc[,c(x,y)]
  Fcmodd=Fcm[,c(x,y)]
  if(ncol(C)==1) stop("Selected mode must have for than one column")
  s.label(Cmodd,clabel=0.5,label=cond,xlim=xlim,ylim=ylim,sub=sub,csup=1.25)
  s.arrow(s*Cmodd,xax=1,yax=2,label=indvar,clabel=0.5,boxes=FALSE,
  add.plot=TRUE)
}
}
}

```

### A.1.6. Representaciones de las trayectorias de dos modos (función: *traj.modes*)

```

traj.modes=function(lista,comb="AC",x=1,y=2,s=1,xlim=NULL,ylim=NULL){
  library(ade4)
  A=lista$A;B=lista$B;C=lista$C
  nind=nrow(A);nvar=nrow(B);ncond=nrow(C)
  Ga=lista$Ga;Gb=lista$Gb;Gc=lista$Gc
  ind=lista$ind;var=lista$var;cond=lista$cond
  varcond=lista$varcond;indcond=lista$indcond;indvar=lista$indvar
  Fa=kronecker(C,B)%*%t(Ga);Fb=kronecker(C,A)%*%t(Gb);Fc=kronecker(B,A)%*%t(Gc)
  fac1=factor(rep(1:nind,ncond))
  sub="Configuracion de la combinacion";sub=paste(sub,comb);
  ejes="(ejes ";ejes=paste(ejes,x,y,"");sub=paste(sub,ejes)
  if(comb=="AC"){
    Bd=B[,c(x,y)]
    Fbd=Fb[,c(x,y)]
    if(ncol(B)==1) stop("Selected modes combination cannot be computed")
    s.trajet(Fbd,fac=fac1,ord = (1:length(fac1)), xax = 1, yax = 2,
    label = NULL,

```

```

        clabel = 0.8, cpoint = 1, pch = 1, grid = TRUE, addaxes = TRUE,
        edge = FALSE, origin = c(0,0),
        include.origin = TRUE, sub = sub, csub = 1.25, possub = "bottomleft",
        cgrid = 0.5, pixmap = NULL, contour = NULL, xlim = xlim, ylim = ylim,
        area = NULL)
        s.arrow(s*Bd,xax=1,yax=2,label=var,clabel=0.5,boxes=TRUE,add.plot=TRUE)
        text(Fbd[,1],Fbd[,2],indcond,pos=1,cex=0.5)
    }
    fac2=factor(rep(1:nvar,ncond))
    if(comb=="BC"){
        Ad=A[,c(x,y)]
        Fad=Fa[,c(x,y)]
        if(ncol(A)==1) stop("Selected modes combination cannot be computed")
        s.traject(Fad,fac=fac2,ord = (1:length(fac2)), xax = 1, yax = 2,
        label = NULL, clabel = 0.5, cpoint = 1, pch = 1, xlim = xlim, ylim = ylim,
        grid = TRUE, addaxes = TRUE, edge = FALSE, origin = c(0,0),
        include.origin = TRUE, sub = sub, csub = 1.25, possub = "bottomleft",
        cgrid = 1, pixmap = NULL, contour = NULL, area = NULL)
        s.arrow(s*Ad,xax=1,yax=2,label=ind,clabel=0.5,boxes=TRUE,add.plot=TRUE)
        text(Fad[,1],Fad[,2],varcond,pos=1,cex=0.5)
    }
    fac3=factor(rep(1:nind,nvar))
    if(comb=="AB"){
        Cd=C[,c(x,y)]
        Fcd=Fc[,c(x,y)]
        if(ncol(C)==1) stop("Selected modes combination cannot be computed")
        s.traject(Fcd,fac=fac3,ord = (1:length(fac3)), xax = 1, yax = 2,
        label = NULL, clabel = 0.8, cpoint = 1, pch = 1, xlim = xlim, ylim = ylim,
        grid = TRUE, addaxes = TRUE, edge = FALSE, origin = c(0,0),
        include.origin = TRUE, sub = sub, csub = 1.25, possub = "bottomleft",
        cgrid = 1, pixmap = NULL, contour = NULL, area = NULL)
        s.arrow(s*Cd,xax=1,yax=2,label=cond,clabel=0.5,boxes=TRUE,add.plot=TRUE)
        text(Fcd[,1],Fcd[,2],indvar,pos=1,cex=0.5)
    }
}

```

#### A.1.7. Representación de los elementos del arreglo central (función: *rep-core*)

```

function(lista,compA,compB){
    ini=lista$Ga[compA,]
    n=ncol(lista$Ga)
    nc=nrow(lista$Gc);nb=nrow(lista$Gb)
    ind=seq(compB,n-nc+compB,by=nc)
    ind2=1:nc
    texto1=""
    texto2="C"
    texto1=rep(texto1,nc)
    texto2=rep(texto2,nc)
    t=paste(texto1,ind2)
    t=paste(t,texto2)
    sub=paste("Comp",compA,"A, en comp",compB,"B")
    plot(ini[ind],sub=sub,xlim=c(0.65,nc),xlab="",ylab="",las=1)
    lines(ini[ind])
    text(ini[ind],labels=t,pos=2)
}

```

## A.2. Datos de las variables demográficas y socioeconómicas evaluadas en 24 departamentos de Colombia durante el periodo 2004-2007

	Año	TBN	EVN	TBM	MNP	MM	PNBI	PBLI	PBLP	IDH	ICV	PM	APE
Antioquia	2004	21.07	70.56	6.29	8.00	67.70	18.30	20.00	59.40	0.75	78.70	5.20	7.20
Atlántico	2004	22.59	73.15	4.97	11.40	87.80	23.00	28.10	66.70	0.76	79.90	7.10	8.10
Bogotá	2004	20.10	73.23	4.79	6.30	77.10	12.40	14.90	49.60	0.81	86.90	1.30	9.60
Bolívar	2004	25.54	73.76	4.88	16.70	98.90	35.20	31.90	65.80	0.74	68.40	13.00	7.00
Boyacá	2004	22.82	70.77	6.87	7.40	111.70	24.50	29.20	65.60	0.72	65.60	4.70	5.90
Caldas	2004	19.58	71.43	6.17	9.60	120.60	15.40	20.70	61.30	0.73	74.90	3.60	6.40
Caquetá	2004	30.60	69.14	6.18	17.00	262.20	22.20	28.70	69.20	0.71	68.30	6.60	5.90
Cauca	2004	26.06	69.87	6.42	10.60	202.40	28.90	39.70	77.20	0.68	64.90	11.30	5.50
Cesar	2004	26.63	71.08	5.51	10.80	76.00	36.30	16.40	61.40	0.72	67.90	13.20	6.20
Chocó	2004	23.74	66.78	7.32	16.10	396.60	60.60	47.40	78.00	0.66	56.60	17.60	5.50
Córdoba	2004	22.73	73.52	5.05	11.90	109.50	44.10	36.70	69.40	0.71	62.90	18.10	6.40
Cundinamarca	2004	19.23	71.55	6.49	6.10	100.20	23.40	28.10	59.00	0.79	73.80	5.10	6.50
Huila	2004	23.54	71.57	5.64	9.10	86.40	28.60	26.80	62.40	0.74	69.00	6.60	6.30
Guajira	2004	26.08	73.16	5.05	11.20	39.80	32.60	23.60	56.60	0.78	71.40	11.70	7.70
Magdalena	2004	24.75	72.73	5.09	11.80	156.00	34.00	18.40	59.50	0.72	69.90	16.30	7.10
Meta	2004	25.87	67.57	6.22	7.00	147.80	26.20	17.50	55.50	0.75	74.30	7.50	6.60
Nariño	2004	24.79	70.12	6.38	7.90	169.60	32.60	40.60	74.70	0.69	64.00	9.80	5.80
N-Santander	2004	25.75	70.93	5.74	11.40	65.50	19.20	20.20	53.00	0.72	76.50	3.20	6.70
Quindío	2004	19.68	70.57	6.52	7.90	94.00	18.10	11.50	50.60	0.75	78.10	3.40	6.80
Risaralda	2004	20.23	72.01	5.74	8.20	80.10	16.70	23.20	58.20	0.76	77.60	4.00	6.80
Santander	2004	22.45	71.80	6.00	5.70	86.50	20.80	23.10	57.70	0.75	74.80	4.10	6.70
Sucre	2004	25.10	74.13	4.91	7.00	80.50	39.80	26.60	67.40	0.70	63.30	15.00	6.50
Tolima	2004	20.51	70.94	6.77	7.30	64.10	27.30	31.00	66.40	0.76	71.20	8.70	6.20
Valle	2004	19.38	71.54	6.00	7.40	96.90	15.70	15.80	52.30	0.78	81.40	2.60	7.80

	Año	TBN	EVN	TBM	MNP	MM	PNBI	PBLI	PBLP	IDH	ICV	PM	APE
Antioquia	2005	19.28	73.19	5.68	7.10	58.70	18.30	20.00	59.40	0.79	81.74	5.20	7.20
Atlántico	2005	20.90	73.96	4.94	9.40	81.10	23.00	28.10	66.70	0.78	83.55	7.10	8.10
Bogotá	2005	18.69	74.72	4.70	6.10	64.20	12.40	14.90	49.60	0.83	89.57	1.30	9.60
Bolívar	2005	23.34	74.11	4.86	12.60	83.90	35.20	31.90	65.80	0.76	76.94	13.00	7.00
Boyacá	2005	20.92	71.51	6.84	8.60	98.10	24.50	29.20	65.60	0.76	70.66	4.70	5.90
Caldas	2005	18.31	72.75	6.12	8.30	91.90	15.40	20.70	61.30	0.77	78.35	3.60	6.40
Caquetá	2005	22.75	71.24	5.53	15.10	190.70	22.20	28.70	69.20	0.74	72.53	6.60	5.90
Cauca	2005	24.08	70.53	6.36	9.10	133.70	28.90	39.70	77.20	0.72	71.02	11.30	5.50
Cesar	2005	24.48	72.55	5.19	9.00	65.80	36.30	16.40	61.40	0.73	71.79	13.20	6.20
Chocó	2005	21.73	68.08	6.97	16.90	315.70	60.60	28.10	59.00	0.68	60.47	17.60	5.50
Córdoba	2005	21.03	73.91	5.20	11.20	114.00	44.10	47.40	78.00	0.75	65.68	18.10	6.40
Cundinamarca	2005	17.90	72.17	6.84	6.40	70.00	23.40	36.70	69.40	0.78	75.38	5.10	6.50
Huila	2005	22.22	72.83	5.49	6.60	78.40	28.60	26.80	62.40	0.76	74.28	6.60	6.30
Guajira	2005	24.29	74.21	4.86	11.90	48.00	32.60	23.60	56.60	0.76	71.67	11.70	7.70
Magdalena	2005	23.33	73.32	5.08	11.00	132.90	34.00	18.40	59.50	0.74	79.52	16.30	7.10
Meta	2005	23.80	68.35	6.19	6.50	104.80	26.20	17.50	55.50	0.76	75.58	7.50	6.60
Nariño	2005	22.17	70.72	6.27	9.60	158.20	32.60	40.60	74.70	0.71	69.44	9.80	5.80
N-Santander	2005	23.59	71.95	5.56	9.40	77.60	19.20	20.20	53.00	0.73	76.15	3.20	6.70
Quindío	2005	18.63	72.30	6.34	8.50	80.10	18.10	11.50	50.60	0.76	78.33	3.40	6.80
Risaralda	2005	18.86	74.05	5.41	8.60	35.00	16.70	23.20	58.20	0.77	79.06	4.00	6.80
Santander	2005	20.68	72.87	5.92	5.50	42.70	28.00	23.10	57.70	0.79	79.12	4.10	6.70
Sucre	2005	23.07	74.41	4.98	7.10	82.60	39.80	26.60	67.40	0.73	69.45	15.00	6.50
Tolima	2005	18.80	72.31	6.81	6.10	103.60	27.30	31.00	66.40	0.77	75.33	8.70	6.20
Valle	2005	18.13	72.93	5.88	7.30	88.00	15.70	15.80	52.30	0.80	83.69	2.60	7.80

	Año	TBN	EVN	TBM	MNP	MM	PNBI	PBLI	PBLP	IDH	ICV	PM	APE
Antioquia	2006	19.28	73.19	5.68	7.80	68.30	18.20	18.40	54.10	0.79	81.74	5.30	7.60
Atlántico	2006	20.90	73.96	4.94	7.70	72.50	17.50	11.50	48.20	0.79	83.55	4.40	8.90
Bogotá	2006	18.69	74.62	4.70	6.40	64.20	7.70	5.90	29.50	0.83	89.57	0.70	10.20
Bolívar	2006	23.34	74.11	4.86	9.40	92.40	32.70	16.70	54.60	0.78	73.94	10.40	7.40
Boyacá	2006	20.92	71.51	6.84	6.50	66.60	26.70	40.00	71.50	0.76	70.66	6.30	5.90
Caldas	2006	18.31	72.75	6.12	10.20	81.10	16.00	17.10	57.70	0.78	78.35	2.40	6.90
Caquetá	2006	27.74	71.24	5.53	10.10	198.60	26.90	19.40	56.80	0.74	72.53	6.70	6.10
Cauca	2006	24.08	70.53	6.36	7.40	120.70	24.90	28.20	63.00	0.73	71.02	4.50	6.30
Cesar	2006	24.48	72.55	5.19	7.90	95.30	34.00	18.80	59.30	0.75	71.79	10.60	6.60
Chocó	2006	21.73	68.08	6.97	17.40	429.20	71.30	39.30	71.60	0.68	60.47	19.40	6.00
Córdoba	2006	21.03	73.91	5.20	10.40	112.00	43.00	33.60	70.80	0.75	65.68	16.10	6.60
Cundinamarca	2006	17.90	72.17	6.84	6.80	84.80	20.10	18.10	53.60	0.79	75.38	4.10	6.70
Huila	2006	22.22	72.83	5.49	5.30	64.60	22.20	29.40	66.30	0.78	74.28	5.60	7.00
Guajira	2006	24.29	74.21	4.86	7.70	125.80	32.90	13.10	52.80	0.78	71.67	11.90	7.60
Magdalena	2006	23.33	73.32	5.08	9.20	90.60	36.90	14.30	55.00	0.75	69.52	12.50	7.30
Meta	2006	23.80	68.35	6.19	5.60	69.60	22.80	11.20	42.50	0.77	75.58	4.70	7.00
Nariño	2006	22.17	70.72	6.27	8.10	144.10	27.40	28.20	67.30	0.73	69.44	6.70	6.40
N-Santander	2006	23.59	71.95	5.56	8.10	65.80	23.20	20.40	57.90	0.75	76.15	6.00	6.70
Quindio	2006	18.63	72.30	6.34	7.30	50.40	17.00	15.50	47.30	0.77	78.33	2.40	7.70
Risaralda	2006	18.86	74.05	5.41	6.30	65.50	17.80	8.90	44.70	0.77	79.06	2.50	7.40
Santander	2006	20.68	72.87	5.92	5.60	46.10	12.70	13.50	48.60	0.80	79.12	2.00	7.20
Sucre	2006	23.07	74.41	4.98	5.60	63.30	37.60	23.40	65.70	0.73	69.45	13.40	6.70
Tolima	2006	18.80	72.31	6.81	4.70	41.70	22.50	25.20	60.10	0.77	75.33	5.50	6.70
Valle	2006	18.13	72.93	5.88	6.80	49.10	12.70	10.40	38.90	0.80	83.69	1.90	8.10

	Año	TBN	EVN	TBM	MNP	MM	PNBI	PBLI	PBLP	IDH	ICV	PM	APE
Antioquia	2007	17.75	73.41	6.17	7.10	65.50	22.59	18.82	54.91	0.79	81.40	8.04	7.52
Atlántico	2007	19.65	74.81	5.35	7.20	71.40	24.68	11.37	48.82	0.79	83.00	8.14	8.89
Bogotá	2007	16.74	77.08	4.39	6.50	59.60	9.16	4.54	28.41	0.83	89.70	1.38	10.05
Bolívar	2007	22.45	73.85	5.53	8.20	91.10	46.56	13.26	52.94	0.78	73.70	23.24	7.46
Boyacá	2007	19.60	74.25	6.95	7.40	78.00	30.70	34.42	67.61	0.76	70.20	10.38	6.47
Caldas	2007	17.56	73.24	7.22	6.80	42.60	17.69	15.56	57.34	0.78	78.70	3.15	7.03
Caquetá	2007	26.28	68.32	6.88	9.90	110.40	41.60	26.01	62.18	0.74	70.10	15.56	6.07
Cauca	2007	21.25	70.84	6.85	7.90	125.90	46.41	27.48	61.06	0.73	72.00	20.44	6.48
Cesar	2007	24.79	72.71	5.26	7.90	39.20	44.53	14.78	58.24	0.75	69.40	21.90	6.64
Chocó	2007	31.28	67.80	7.35	17.40	250.90	79.05	48.70	78.54	0.68	60.47	31.80	5.64
Córdoba	2007	24.14	73.17	5.81	9.70	25.90	59.02	28.47	66.27	0.75	68.40	30.22	6.93
Cundinamarca	2007	19.44	73.88	6.19	7.00	69.60	21.16	16.60	49.67	0.79	76.30	5.20	7.30
Huila	2007	22.29	72.53	6.16	7.20	38.30	32.56	24.26	59.72	0.78	73.10	11.54	6.89
Guajira	2007	31.19	73.79	5.20	11.10	131.00	65.18	16.02	55.90	0.78	69.80	47.07	7.57
Magdalena	2007	24.71	73.45	5.48	10.10	109.60	47.63	15.19	57.45	0.75	71.90	22.97	7.00
Meta	2007	22.07	70.63	6.32	5.60	81.60	24.84	12.21	44.83	0.77	78.30	8.03	7.36
Nariño	2007	21.59	72.88	6.20	6.00	76.60	43.75	23.74	64.04	0.73	69.30	17.02	6.36
N-Santander	2007	20.68	71.82	6.46	6.60	90.80	30.25	18.49	57.55	0.75	75.40	10.90	6.76
Quindio	2007	17.26	73.10	7.13	7.60	80.00	16.04	11.80	45.63	0.77	81.20	2.99	7.83
Risaralda	2007	17.21	73.33	6.67	7.90	69.50	16.64	10.54	45.72	0.77	80.70	3.67	7.49
Santander	2007	17.74	74.30	6.18	5.70	56.30	21.85	14.57	48.99	0.80	78.80	6.63	7.45
Sucre	2007	21.97	74.21	5.66	4.20	51.50	54.93	22.60	69.46	0.73	66.60	26.86	6.63
Tolima	2007	19.52	72.23	7.56	5.10	85.90	29.79	19.64	56.85	0.77	74.30	10.44	6.66
Valle	2007	17.04	73.70	6.33	6.20	89.90	15.57	9.56	37.56	0.80	83.00	3.49	8.14

### A.3. Definición de las variables demográficas y socioeconómicas evaluadas en 24 departamentos de Colombia durante el periodo 2004-2007

**Tasa bruta de natalidad (TBN)(por 1000 habitantes):** mide el efecto de la fecundidad sobre el crecimiento de la población relacionando el número total de nacimientos con la población total en un año determinado.

$TBN = [\text{Número de nacidos vivos de mujeres de todas las edades durante el año} / \text{Población total calculada a mediados de año}] * 1000$

**Esperanza de vida al nacer (EVN):** señala el número de años que se espera que viva una persona nacida en el año que se estudia. Se estima a partir de las tasas específicas de mortalidad por grupos quinquenales de edad y género para un año determinado con base en las cuales se construye una cohorte teórica.

**Tasa bruta de mortalidad (TBM) (por 1000 habitantes):** mide la frecuencia de defunciones en una población por cada mil habitantes en un determinado año.

$TBM = [\text{Número de defunciones de todas las edades ocurridas durante un periodo determinado} / \text{Población total calculada a mediados de año}] * 1000$

**Mortalidad neonatal precoz (MNP) (por 1000 nacidos vivos):** indica la probabilidad de muerte de recién nacidos vivos que sobreviven brevemente (días o semanas) en relación con el número de nacidos vivos en un periodo determinado.

$MNP = [\text{Número de defunciones neonatales tempranas durante un periodo especificado} / \text{Número de nacidos vivos}] * 1000$

**Razón de mortalidad materna (MM) (por 100000 nacidos vivos):** mide la probabilidad de que una mujer fallezca por complicaciones el embarazo, parto o puerperio en un periodo dado en relación con el número de nacidos vivos.

$MM = [\text{Número de defunciones maternas durante un periodo especificado} / \text{Número de nacidos vivos}] * 100000$

**Población en necesidades básicas insatisfechas (PNBI):** indica la incidencia de la pobreza en el total de la población en hogares particulares, que no alcanzan a cubrir las necesidades básicas establecidas en el país.

$PNBI = [\text{Población en hogares con NBI} / \text{total de la población}] * 100$

**Población bajo línea de indigencia (PBLI):** revela la proporción de personas cuyos ingresos no alcanzan para satisfacer un conjunto de necesidades básicas alimentarias capaz de satisfacer un umbral mínimo de necesidades energéticas y protéicas.

$PBLI = [\text{Población de personas con ingreso per capita familiar inferior al valor de la línea de indigencia por adulto equivalente} / \text{Total de población en hogares}] * 100$

**Población bajo línea de pobreza (PBLP):** describe el número de personas cuyo ingreso no alcanza para adquirir una canasta de bienes y servicios básicos para la vida. Para Colombia el nivel que define este nivel de ingresos equivale a un salario mínimo.

**Índice de desarrollo humano (IDH):** mide el logro medio en una región en tres dimensiones básicas de desarrollo humano: vida larga y saludable, conocimientos y nivel de vida decente. Contiene tres variables: la esperanza de vida al nacer, el logro educacional (alfabetización de adultos y la tasa bruta de matriculación primaria, secundaria y terciaria combinada) y el Producto Interno Bruto (PIB) real per cápita ajustado en dólares.

**Índice de condiciones de vida (ICV):** refleja el estado de la calidad de vida de la población que habita en los municipios que tienen territorio en un área de especial interés para el análisis de los ecosistemas y de la biodiversidad. Obtiene información sobre condiciones de vida de los hogares como educación del jefe de hogar, educación del resto de personas mayores de 12 años, asistencia de las personas entre 12 y 18 años a la educación secundaria y universitaria, asistencia de los niños entre 5 y 11 años a la educación primaria, material de construcción de las paredes de la vivienda, material de construcción de los pisos de la vivienda, tipo de servicio sanitario disponible en la vivienda, tipo de abastecimiento de agua para el hogar, tipo de energía que emplea el hogar para la cocina, sistema de disposición o recolección de basuras que emplean, número de niños de 6 o menos años de edad en el hogar y número promedio de personas en el hogar por cada habitación de la vivienda.

**Población en miseria (PM):** mide la privación en cuanto al desarrollo humano básico en las mismas dimensiones que el IDH.

$$PM = \left[ \frac{\text{Población de personas con privación en cuanto al desarrollo humano básico}}{\text{Total de población en hogares}} \right] * 100$$

**Años promedio de educación para la población de 15 años y más (APE):** determina el número promedio de años escolares cursados por personas con edad igual o superior a 15 años.



---

## Bibliografía

---

- Amaya, J. & Pacheco, P. (2002), ‘Análisis factorial dinámico mediante el método tucker3’, *Revista Colombiana de Estadística* **25**, 43–57.
- Bro, R. (1998), Multi-way Analysis in The Food Industry. Models, Algorithms and Applications, PhD thesis, University of Amsterdam.
- Bro, R. & Smilde, A. (2003), ‘Centering and Scaling in Component Analysis’, *Journal of Chemometrics* **17**, 16–33.
- Carroll, J. & Chang, J. (1970), ‘Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition.’, *Psychometrika* **35**, 283–319.
- Dray, S., Dufour, A. & Chessel, D. (2007), ‘The ade4 package-II: Two-table and K-table methods. ’, *R News* **7**(2), 47–52.
- Escofier, B. & Pagés, J. (1982), Comparaison de groupes de variables définies sur le même ensembles définies sur le même ensemble d’individus, Technical report, Institut National de Recherche en Informatique, Paris.
- Escofier, B. & Pagés, J. (1992), *Análisis factoriales simples y múltiples. Objetivos, métodos e interpretación.*, Universidad del País Vasco., Bilbao.
- Escofier, B. & Pagés, J. (1994), ‘Multiple Factor Analysis (AFMULT package)’, *Computational Statistics and Data Analysis* **18**, 121–140.
- Harshman, R. A. (1970), ‘Foundations of the PARAFAC procedure: models and conditions for an “explanatory” multi-mode factor analysis’, *UCLA Working Papers in Phonetics* (16), 1–84. Citado por Kiers (1991).
- Husson, F., Josse, J., Le, S. & Mazet, J. (2008), *FactoMineR: Factor Analysis and Data Mining with R*. R package version 1.10.  
\*<http://factominer.free.fr>, <http://www.agrocampus-rennes.fr/math/>
- Husson, F. & Pagés, J. (2006), ‘INDSCAL Model: Geometrical Interpretation and Methodology.’, *Computational Statistics and Data Analysis* **50**, 357–378.
- Kiers, H. (1991), ‘Hierarchical Relations Among Three-way Methods.’, *Psychometrika* **56**, 449–470.

- Kiers, H. (2000a), 'Some Procedures for Displaying Results from Three-way Methods.', *Journal of Chemometrics* **14**, 151–170.
- Kiers, H. (2000b), 'Towards a Standardized Notation and Terminology in Multiway Analysis.', *Journal of Chemometrics* **14**, 105–122.
- Kiers, H. A. L. & Mechelen, I. (2001), 'Three-way Component Analysis: Principles and Illustrative Application.', *Psychological Methods* **6**, 84–110.
- Kroonenberg, P. (2008), *Applied Multiway Data Analysis.*, John Wiley and Sons., New Jersey.
- Kroonenberg, P. & De Leeuw, J. (1980), 'Principal Component Analysis of Three-mode Data by Means of Alternating Least Squares Algorithms.', *Psychometrika* **45**, 69–97.
- Lavit, C., Escoufier, Y., Sabatier, R. & Traissac, P. (1994), 'The ACT (STATIS method)', *Computational Statistics and Data Analysis* **18**, 97–119.
- Lebart, L., Morineau, A. & Piron, M. (1995), *Statistique exploratoire multidimensionnelle.*, Dunod, París.
- Mures, Q., Vallejo, P. & García, G. (2006), 'Comparación empírica de técnicas estadísticas para tablas de tres entradas: la construcción en Castilla y León en el periodo 2002–2004.', *Pecunia* pp. 95–140.
- Pagés, J. (1996), 'Eléments de comparaison entre l'analyse factorielle multiple et la méthode statis.', *Revue de Statistique Appliquée* **44**, 81–95.
- R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.  
\*<http://www.R-project.org>
- Timmerman, M. E. & Kiers, H. A. L. (2000), 'Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima', *British Journal of Mathematical and Statistical Psychology* **53**, 1–16. Citado por Kroonenberg (2008).
- Tucker, L. (1966), 'Some Mathematical Notes on Three-mode Factor Analysis.', *Psychometrika* **31**, 279–311.