

ANÁLISIS FACTORIAL DINÁMICO MEDIANTE EL MÉTODO TUCKER3

Jeannette L. Amaya L. *

Pedro N. Pacheco D. **

Resumen

El método Tucker3, propuesto por Leyard Tucker en 1966, identificado dentro de las técnicas diseñadas para análisis de datos a tres vías, se considera como una generalización del Análisis de Componentes Principales (PCA) y la Descomposición en Valores Singulares (SVD), hace uso de cálculos matriciales y procedimientos de descomposición para la estimación de los parámetros del modelo y su correspondiente representación gráfica en espacios de dimensión inferior a la del arreglo de datos original. Mediante este método es posible explicar la información de un conjunto de datos tres-modos resumiendo las entidades asociadas por medio de unos pocos componentes, y describir las interacciones entre las tres fuentes de variación presentes en los datos mediante el uso de un arreglo central. En este artículo se describen los principios teóricos y los soportes analíticos del método Tucker3, los cuales se ejemplifican con base en datos hipotéticos.

Palabras Clave: Análisis factorial dinámico, datos cúbicos, modelo Tucker3, descomposición en valores singulares, métodos factoriales, trayectorias

* Estadística. Profesora Asociada, Pontificia Universidad Javeriana, Centro de Investigaciones CENDEX

** Profesor Asociado. Departamento de Estadística, Universidad Nacional de Colombia ; email: pacheco@matematicas.edu.co

Abstract

The Tucker3 method, proposed by Leyard Tucker en 1966, is identified as a technique designed to three-mode data analysis and is considered a generalization of the Principal Component Analysis (PCA) and the Singular Value Decomposition (SVD). This method uses matrix operations and decomposition procedures to estimate model parameters and its corresponding graphic representation in spaces of lower dimension to the original data array. By this method it is possible to explain information of a three-mode data set summarizing the associate entities through a few components, and to describe the possible interactions between the three sources of variation in the data using one central array. In this paper theoretical principles and the analytic supports for this method are presented and exemplified using hypotheticalal data.

Key Words: *dynamic factorial analysis, cubic data, Tucker3 model, singular value decomposition, factorial planes, trajectories.*

1. Introducción

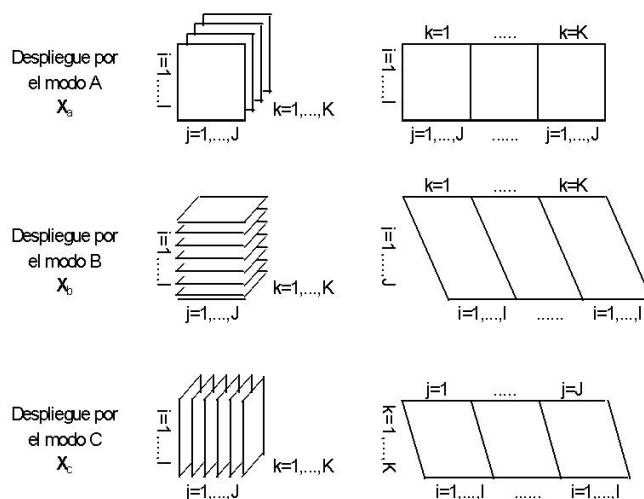
El análisis factorial dinámico surge del estudio de factores clásico, cuyo propósito primordial es definir la estructura fundamental de una matriz de datos resumiendo la información en un conjunto de nuevas dimensiones conocidas como factores. Cuando a la estructura dos-modos se le adiciona un conjunto independiente de entidades, el número de modos se incrementa y se da paso al análisis factorial dinámico. La estructura tres-modos contempla tres conjuntos de entidades contenidas en cada uno de los modos del arreglo, los cuales se refieren regularmente a sujetos, variables y ocasiones. En general, los análisis factoriales buscan reducir la dimensionalidad de un conjunto de datos original, y en este sentido, el método Tucker3 permite resumir mediante unos pocos componentes cada uno de los modos del arreglo, y condensarlos en un arreglo central que define la estructura fundamental del conjunto de datos. Luego de explicar algunos términos y procedimientos básicos, se presentará el ejemplo ilustrativo con base en el cual se verificará la capacidad que tiene el método para capturar el comportamiento general del arreglo de datos mediante el uso de componentes de resumen (LAW H., SNYDER C., HATTIE J., MCDONALD R., 1984).

2. Consideraciones preliminares

Antes de dar inicio al estudio del modelo Tucker3, se presentarán algunas definiciones y procedimientos requeridos para el desarrollo del método. Con base en estos, y como una etapa inicial del análisis de los datos, se mostrará la manera de aplicar procedimientos de pre-procesamiento sobre los datos y determinar el número de componentes para resumir cada uno de los modos del arreglo.

2.1. Procedimiento de despliegue

Se parte de un arreglo de datos tres-modos, donde el índice I indica el número de sujetos, el índice J representa las variables y el índice K representa las ocasiones en las cuales fueron tomadas las mediciones. La alta dimensionalidad de los modos hace necesario el uso de un procedimiento, denominado despliegue o matriciación, que se basa en la transformación o reordenamiento del arreglo tres-modos en una matriz. En el caso de datos cúbicos, se presentan tres posibles formas de despliegue: el despliegue por el modo A, por el modo B y por el modo C, los cuales son indicados respectivamente mediante los subíndices a, b y c. El despliegue se realiza manteniendo en las columnas las entidades del modo por el cual se desea desplegar el arreglo de datos, y confundiendo en las columnas las entidades de los otros dos modos (KIERS H, 2000).



2.2. Procedimientos de pre-procesamiento

Los procedimientos de pre-procesamiento se aplican cuando se requiere eliminar diferencias de nivel y escala en los datos. Centrar a través de los sujetos y escalar entre las variables son los procedimientos más comunes, los cuales al ser aplicados conjuntamente proporcionan la estandarización de los datos. El centrado se realiza a través de las columnas de la matriz, mientras que el escalamiento se aplica sobre tajadas completas o submatrices del arreglo. Por ejemplo, centrar a través de los sujetos significa obtener X_a (despliegue por el modo A del arreglo X) y restar de cada elemento su correspondiente promedio columna, mientras que escalar entre las variables sugiere tomar la matriz X_b (despliegue por el modo B del arreglo X)

y dividir cada elemento por la correspondiente desviación S_j , como se observa en las siguientes ecuaciones (BRO R., 1998).

$$x_{ijk}^{centrado} = x_{ijk} - \frac{\sum_{i=1}^I x_{ijk}}{I} = x_{ijk} - \bar{x}_{.jk} \text{ y } x_{ijk}^{escalado} = \frac{x_{ijk}}{S_j} \text{ con } S_j = \sqrt{\sum_{i=1}^I \sum_{k=1}^K x_{ijk}^2}$$

2.3. Número de componentes

Resulta esencial antes del análisis, determinar el número de componentes por medio de los cuales se resumirán las entidades de cada uno de los modos del arreglo. Una manera de determinar la dimensionalidad óptima es establecer los porcentajes de ajuste del modelo para cada posible combinación de número de componentes en los tres modos del arreglo. Entonces, se establece como la mejor combinación de número de componentes aquella que presenta poca variación en el ajuste del modelo al considerar un componente adicional en alguno de los modos del arreglo (BRO, 1998).

2.4. Porcentaje de ajuste

El porcentaje de ajuste del modelo ($AM \%$) se calcula como el cociente entre las sumas de cuadrados de los elementos del arreglo central y del arreglo tres-modos original, como se presenta a continuación.

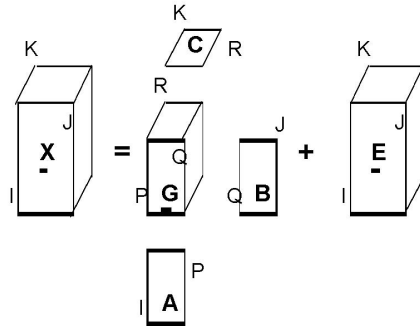
$$AM \% = \left(\frac{\sum \hat{g}_{pqr}^2}{\sum_{ijk} x_{ijk}^2} \right) * 100 \%$$

Similarmente, es posible determinar la contribución al ajuste del modelo de cada uno de los componentes, generando la suma de cuadrados de los elementos correspondientes contenidos dentro del arreglo central (KIERS H., MECHELEN I., 2001).

3. Modelo tucker3

El modelo Tucker3 descompone un arreglo $\underline{X}^{(IxJxK)}$ en cuatro conjuntos de parámetros, teniendo en cuenta la extracción de diferente número de componentes en cada uno de los modos. Los parámetros corresponden a una matriz $\mathbf{A}^{(IxP)}$ que resume mediante P componentes las I entidades del modo A (sujetos), una matriz $\mathbf{B}^{(JxQ)}$ que resume mediante Q componentes las J entidades del modo B (variables), una matriz $\mathbf{C}^{(KxR)}$ que resume mediante R componentes las K entidades

del modo C (ocasiones) y un arreglo central tres-modos $\mathbf{G}^{(PxQxR)}$, generado con base en las tres matrices componentes, que constituye el aporte más importante del método Tucker3 (TUCKER R., 1966).



El modelo Tucker3 se define en términos de una triple sumatoria entre los elementos contenidos en cada una de las matrices componentes y del arreglo central, mas un término de error correspondiente a cada estimación de los valores del arreglo original.

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

El modelo puede ser escrito en términos matriciales teniendo en cuenta alguna de las formas de despliegue de los arreglos tres-modos \mathbf{X} , \mathbf{G} y \mathbf{E} , que representan respectivamente al arreglo de datos original, el arreglo central y el arreglo que contiene los errores del modelo.

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

El despliegue por el modo A del arreglo tres-modos original, por ejemplo, corresponde al producto entre la matriz componente \mathbf{A} , el arreglo central desplegado por el modo A y el producto Kronecker entre las transpuestas de las matrices componentes \mathbf{B} y \mathbf{C} (TUCKER R., 1966).

4. Algoritmo Tucker3

El algoritmo Tucker3, utilizado para la estimación de los parámetros del modelo, se basa en la descomposición en valores singulares de diferentes arreglos matriciales, procedimiento que se desarrolla iterativamente hasta obtener el criterio

de convergencia establecido para el algoritmo. El procedimiento algorítmico, junto con los procedimientos de despliegue y pre-procesamiento de los datos, fue programado por Andersson y Bro (2000) en lenguaje MATLAB, Daniel Kappler en lenguaje GAUSS y Pieter M. Kroonenberg mediante el programa TUCKALS3, y recientemente por Amaya L. (2002) en lenguaje SAS, mediante el procedimiento IML abundante en estructuras de control, operadores matriciales y funciones. Cada una de las programaciones mencionadas difiere en el criterio de convergencia del algoritmo, el cual, en el caso de la programación realizada en SAS, consistió en una diferencia no superior a $1E-10$ entre la suma de cuadrados de los elementos del arreglo central en dos iteraciones consecutivas.

El algoritmo es el siguiente:

Determinar el criterio de convergencia del algoritmo

Inicializar \mathbf{B} y \mathbf{C}

1. $[\mathbf{A}, \mathbf{S}, \mathbf{V}] = svd(\mathbf{X}_a(\mathbf{C} \otimes \mathbf{B}), P)$
2. $[\mathbf{B}, \mathbf{S}, \mathbf{V}] = svd(\mathbf{X}_b(\mathbf{C} \otimes \mathbf{A}), Q)$
3. $[\mathbf{C}, \mathbf{S}, \mathbf{V}] = svd(\mathbf{X}_c(\mathbf{B} \otimes \mathbf{A}), R)$
4. Ir al paso 1 hasta llegar a la convergencia
5. $\mathbf{G}_a = \mathbf{A}^T \mathbf{X}_a(\mathbf{C} \otimes \mathbf{B})$

En general, la expresión $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = svd(\mathbf{X}, P)$ significa que se realiza una SVD reducida de \mathbf{X} reteniendo los primeros P componentes. La matriz \mathbf{X} es descompuesta como $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, donde la matriz \mathbf{U} contiene los primeros P vectores singulares izquierdos, \mathbf{V} los primeros P vectores singulares derechos, y donde \mathbf{S} es una matriz diagonal cuyos elementos diagonales corresponden a la raíz cuadrada de los primeros P valores singulares (JOHNSON A., WICHERN D., 1998).

Luego de haber determinado el criterio de convergencia del algoritmo, se procede a generar las estimaciones iniciales de las matrices componentes \mathbf{B} y \mathbf{C} . La estimación de la matriz componente \mathbf{B} se obtiene al seleccionar los Q primeros vectores singulares izquierdos de la SVD de \mathbf{X}_b , y la estimación inicial de la matriz componente \mathbf{C} corresponde a los R primeros vectores singulares izquierdos de la SVD de \mathbf{X}_c . Con base en estas estimaciones iniciales, se da paso al proceso iterativo aplicando SVD sobre los arreglos matriciales $\mathbf{X}_a(\mathbf{C} \otimes \mathbf{B})$, $\mathbf{X}_b(\mathbf{C} \otimes \mathbf{A})$ y $\mathbf{X}_c(\mathbf{B} \otimes \mathbf{A})$ para obtener las estimaciones de las tres matrices componentes al seleccionar respectivamente los P , Q y R vectores singulares izquierdos de la descomposición. Adicionalmente, con base en las tres matrices componentes estimadas, se obtiene el arreglo central en alguna de sus formas de despliegue, por ejemplo, el despliegue por el modo A como se muestra en el paso 5 del algoritmo (BRO, 1998).

5. Planos factoriales

En el caso de arreglos tres-modos, se encuentran siete posibles representaciones gráficas; tres de ellas corresponden a los planos que exhiben la configuración de las entidades de un modo particular en relación con las entidades de los dos modos restantes, las otras tres se identifican como la configuración de las entidades de dos modos particulares mediante trayectorias en relación con las entidades del modo restante, y la última muestra la versión más reducida de los datos, generada con base en los valores del arreglo central. Por ejemplo, para obtener la configuración de las entidades del modo B y proyectar sobre el mismo plano las entidades de los modos A y C, se obtiene $\tilde{\mathbf{X}}_b = \mathbf{B}((\mathbf{C} \otimes \mathbf{A})\mathbf{G}_b^T)^T$, se encuentra una base ortonormal para la matriz columna ortogonal $\mathbf{F} \equiv (\mathbf{C} \otimes \mathbf{A})\mathbf{G}_b^T$ al dividir cada una de sus columnas por las raíces cuadradas de los valores singulares obtenidos de la descomposición, y este escalamiento se compensa en \mathbf{B} al multiplicar sus columnas por los mismos valores. Las matrices así calculadas, $\tilde{\mathbf{B}}$ y $\tilde{\mathbf{F}}$, son usadas respectivamente para obtener el plano de las J variables y las proyecciones de los IK ejes originales. Por otra parte, para la configuración de las entidades de los modos A y C con proyección de las entidades del modo B, se hace uso respectivo de las filas de la matriz ortogonal \mathbf{F} conectando los puntos pertenecientes al mismo sujeto en ocasiones consecutivas, y de la matriz ortonormal \mathbf{B} . Finalmente, los valores del arreglo central \mathbf{G} se grafican trazando trayectorias a través de las ocasiones para cada combinación sujeto - variable (KIERS H., 2000).

6. Ejemplo ilustrativo

Los datos presentados en la Tabla 1 corresponden a señales de $I = 6$ sujetos sobre $J = 4$ variables respuesta que indican la manera en que cada sujeto manifiesta un comportamiento emocional, sensitivo, cuidadoso y esmerado en $K = 4$ situaciones diferentes referidas a hacer un examen, dar una opinión, un paseo familiar y una nueva cita (KIERS H., MECHELEN I., 2001).

Los datos cuentan con una estructura preestablecida de tal manera que es posible representarlos de forma resumida mediante la selección de dos componentes en cada uno de los modos del arreglo de datos. Se observa que Ana, Diana y Francia presentan respuestas diferentes respecto a las proporcionadas por Bernardo y Camilo, encontrándose Edna con valores intermedios entre los dos comportamientos. En el caso de las variables se diferencian claramente lo emocional y sensitivo de lo cuidadoso y esmerado, y finalmente, los valores obtenidos en las primeras ocasiones son distintos a los recogidos en las dos últimas.

	<i>Hacer un examen</i>				<i>Dar una opinión</i>				<i>Paseo Familiar</i>				<i>Nueva Cita</i>			
	<i>Emocional</i>	<i>Sensitivo</i>	<i>Cuidadoso</i>	<i>Esmerado</i>	<i>Emocional</i>	<i>Sensitivo</i>	<i>Cuidadoso</i>	<i>Esmerado</i>	<i>Emocional</i>	<i>Sensitivo</i>	<i>Cuidadoso</i>	<i>Esmerado</i>	<i>Emocional</i>	<i>Sensitivo</i>	<i>Cuidadoso</i>	<i>Esmerado</i>
<i>Ana</i>	0,0	0,0	4,0	4,0	0,6	0,6	2,4	2,4	4,0	4,0	0,0	0,0	4,6	4,6	0,9	0,9
<i>Bernardo</i>	0,0	0,0	2,0	2,0	0,2	0,2	1,8	1,8	1,0	1,0	1,0	1,0	1,2	1,2	1,8	1,8
<i>Camilo</i>	0,0	0,0	2,0	2,0	0,2	0,2	1,8	1,8	1,0	1,0	1,0	1,0	1,2	1,2	1,8	1,8
<i>Diana</i>	0,0	0,0	4,0	4,0	0,6	0,6	2,4	2,4	4,0	4,0	0,0	0,0	4,6	4,6	0,9	0,9
<i>Edna</i>	0,0	0,0	2,5	2,5	0,4	0,4	2,1	2,1	2,0	2,0	0,5	0,5	2,4	2,4	1,4	1,4
<i>Francia</i>	0,0	0,0	4,0	4,0	0,6	0,6	2,4	2,4	4,0	4,0	0,0	0,0	4,6	4,6	0,9	0,9

Tabla 1. Datos de niveles de comportamiento

El porcentaje de ajuste del modelo fue determinado para cada una de las posibles combinaciones de número de componentes en cada uno de los modos del arreglo, como se muestra en la Tabla 2. Las negrillas cursivas indican el mayor porcentaje de ajuste obtenido por cada aumento en la suma de los componentes de los tres modos, con base en las cuales se determina la dimensionalidad óptima para el ajuste del modelo. Al seleccionar un componente en el modo A y dos en el modo B y el modo C, se obtiene un aumento del 32,5% en el ajuste del modelo respecto al ajuste obtenido al considerar tan solo un componente en cada modo. Ahora, al seleccionar un componente adicional en el modo A, se obtiene un incremento del 3,91% en el ajuste, el cual es muy superior al obtenido al adicionar un componente en el modo C (incremento del 0,14%), por lo cual se considera como la mejor opción la selección de dos componentes en cada modo como se evidenciaba en el conjunto de datos original.

<i>P</i>	<i>Q</i>	<i>R</i>	<i>P+Q+R</i>	<i>Ajuste (%)</i>	<i>Incremento (%)</i>
<i>1</i>	<i>1</i>	<i>1</i>	<i>3</i>	<i>63,43</i>	
<i>2</i>	<i>1</i>	<i>2</i>	<i>5</i>	<i>63,57</i>	
<i>2</i>	<i>2</i>	<i>1</i>	<i>5</i>	<i>67,07</i>	
<i>1</i>	<i>2</i>	<i>2</i>	<i>5</i>	<i>95,93</i>	<i>32,5</i>
<i>2</i>	<i>2</i>	<i>2</i>	<i>6</i>	<i>99,84</i>	<i>3,91</i>
<i>3</i>	<i>1</i>	<i>3</i>	<i>7</i>	<i>63,57</i>	
<i>3</i>	<i>3</i>	<i>1</i>	<i>7</i>	<i>67,07</i>	
<i>1</i>	<i>3</i>	<i>3</i>	<i>7</i>	<i>95,93</i>	
<i>2</i>	<i>3</i>	<i>2</i>	<i>7</i>	<i>99,84</i>	
<i>3</i>	<i>2</i>	<i>2</i>	<i>7</i>	<i>99,84</i>	
<i>2</i>	<i>2</i>	<i>3</i>	<i>7</i>	<i>99,98</i>	<i>0,14</i>

Tabla 2. Porcentaje de ajuste del modelo

Las estimaciones finales de los parámetros del modelo obtenidas al aplicar el proceso iterativo del algoritmo Tucker3, con un ajuste del 99,84%, son presentadas a continuación. La denominación dada a cada componente de las matrices se realiza teniendo en cuenta los mayores valores absolutos de cada entidad del modo dentro de los componentes. Así, en la Tabla 3 se observa que los sujetos del género femenino se encuentran con altos puntajes sobre el primer componente al que se denominó "Feminidad", mientras aquellos del género masculino se identifican más claramente en el segundo componente denominado "Masculinidad".

	Feminidad	Masculinidad
Ana	0,519	-0,228
Bernardo	0,217	0,620
Camilo	0,217	0,620
Diana	0,519	-0,228
Edna	0,315	0,274
Francia	0,519	-0,228

Tabla 3. Matriz Componente A

En la Tabla 4, que presenta los dos componentes obtenidos para el modo variables, se identifican las variables emocional y sensitivo sobre el primer componente denominado "Emocionalidad", y las variables cuidadoso y esmerado en el segundo componente denominado "Conciencia".

	Emocionalidad	Conciencia
Emocional	0,589	-0,392
Sensitivo	0,589	-0,392
Cuidadoso	0,392	0,589
Esmerado	0,392	0,589

Tabla 4. Matriz Componente B

Con respecto a los componentes del modo situaciones, mostrados en la Tabla 5, se verifica el hecho que hacer un examen y dar una opinión son "Situaciones de Desempeño", mientras que los paseos familiares y la asistencia a nuevas citas se relacionan con "Situaciones Sociales".

	Situaciones Sociales	Situaciones de Desempeño
Hacer un examen	0,333	0,771
Dar una opinión	0,310	0,435
Paseo familiar	0,538	-0,387
Nueva cita	0,709	-0,259

Tabla 5. Matriz Componente C

Finalmente, el cuarto parámetro del modelo corresponde al arreglo central $\underline{\mathbf{G}}$, que resume toda la información teniendo en cuenta los componentes obtenidos para cada modo. En la Tabla 6 se identifica con negrillas la combinación de componentes más significantes dentro del modelo, que corresponden a los valores más sobresalientes del arreglo central.

	Situaciones Sociales	Situaciones de Desempeño
	Emocionalidad	Conciencia
Feminidad	17,098	-0,489
Masculinidad	0,624	4,106

Tabla 6. Despliegue por el modo A del arreglo central $\underline{\mathbf{G}}$

Frente a situaciones de desempeño, las mujeres se caracterizan por comportamientos regidos por la emocionalidad, mientras los hombres se identifican por su conciencia. Ante situaciones sociales, se evidencia el comportamiento consciente de las mujeres, sin encontrar grandes diferencias entre el comportamiento emocional de los hombres y las mujeres cuando se enfrentan a esta clase de situaciones. Con base en los valores cuadrados de la matriz presentada en la Tabla 6, se obtiene el porcentaje de ajuste del modelo mostrado en la Tabla 2 al considerar dos componentes en cada uno de los modos del arreglo. A continuación se presentan las sumas de cuadrados de los valores del arreglo central y del arreglo de datos original.

$$AM \% = \left(\frac{\sum_{p=1}^2 \sum_{q=1}^2 \sum_{r=1}^2}{\sum_{i=1}^6 \sum_{j=1}^4 \sum_{k=1}^4} \right) * 100 \% = \left(\frac{461,222}{461,960} \right) * 100 \% = 99,84 \%$$

Se presentan tres de las posibles representaciones gráficas, haciendo uso de las matrices $\underline{\mathbf{B}}$ y $\underline{\mathbf{F}}$ obtenidas a partir de la matriz componente $\underline{\mathbf{B}}$ y la matriz columna ortogonal $\underline{\mathbf{F}} \equiv (\underline{\mathbf{C}} \otimes \underline{\mathbf{A}}) \underline{\mathbf{G}}_b^T$, dadas por

$$F = \begin{matrix} & & \text{Emocio.} & \text{Consc.} \\ \text{Hacer un exámen} & \text{Ana} & 0,176 & 0,270 \\ \text{Hacer un exámen} & \text{Bernardo} & 0,100 & 0,147 \\ \text{Hacer un exámen} & \text{Camilo} & 0,100 & 0,147 \\ \text{Hacer un exámen} & \text{Diana} & 0,176 & 0,270 \\ \text{Hacer un exámen} & \text{Edna} & 0,122 & 0,184 \\ \text{Hacer un exámen} & \text{Francia} & 0,176 & 0,270 \\ \text{Dar una opinión} & \text{Ana} & 0,162 & 0,144 \\ \text{Dar una opinión} & \text{Bernardo} & 0,086 & 0,100 \\ \text{Dar una opinión} & \text{Camilo} & 0,086 & 0,100 \\ \text{Dar una opinión} & \text{Diana} & 0,162 & 0,144 \\ \text{Dar una opinión} & \text{Edna} & 0,109 & 0,111 \\ \text{Dar una opinión} & \text{Francia} & 0,162 & 0,144 \\ \text{Paseofamiliar} & \text{Ana} & 0,271 & -0,185 \\ \text{Paseofamiliar} & \text{Bernardo} & 0,119 & 0,027 \\ \text{Paseofamiliar} & \text{Camilo} & 0,119 & 0,027 \\ \text{Paseofamiliar} & \text{Diana} & 0,271 & -0,185 \\ \text{Paseofamiliar} & \text{Edna} & 0,168 & -0,052 \\ \text{Paseofamiliar} & \text{Francia} & 0,271 & -0,185 \\ \text{Nuevacita} & \text{Ana} & 0,359 & -0,148 \\ \text{Nuevacita} & \text{Bernardo} & 0,163 & 0,068 \\ \text{Nuevacita} & \text{Camilo} & 0,163 & 0,068 \\ \text{Nuevacita} & \text{Diana} & 0,359 & -0,148 \\ \text{Nuevacita} & \text{Edna} & 0,226 & -0,015 \\ \text{Nuevacita} & \text{Francia} & 0,359 & -0,148 \end{matrix}$$

$$y \quad \tilde{E} = \begin{matrix} & \text{Emocio.} & \text{Consc.} \\ \text{Emocional} & 10,082 & -5,076 \\ \text{Sensitivo} & 10,082 & -5,076 \\ \text{Cuidadoso} & 6,708 & 7,629 \\ \text{Esmerado} & 6,708 & 7,629 \end{matrix}$$

La Figura 1 muestra la configuración de las entidades del modo variables y la proyección, sobre el mismo plano, de los vectores que representan todas las posibles combinaciones entre las entidades del modo sujetos y el modo ocasiones. Se identifican claramente dos grupos, correspondientes a las variables emocional y sensitivo del componente "Emocionalidad" y a las variables cuidadoso y esmerado del componente "Conciencia".

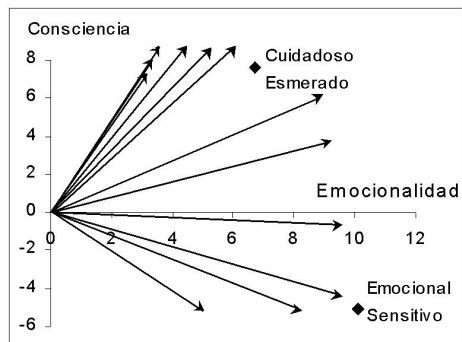


Figura 1. Configuración de las entidades del modo B

Al proyectar sobre el mismo plano los vectores que identifican las 24 combinaciones de sujetos y ocasiones, se verifica el hecho que los comportamientos regidos

por la emocionalidad son característicos de las mujeres frente a situaciones sociales (líneas punteadas hacia abajo), y aquellos regidos por la conciencia se identifican más claramente en las mujeres cuando se enfrentan a situaciones de desempeño (líneas punteadas hacia arriba). Por otra parte, se observa que los vectores que identifican a los hombres ante situaciones sociales (líneas continuas), tienden hacia los comportamientos conscientes más que hacia los emocionales.

Pasando ahora a la configuración de las entidades de dos modos del arreglo de datos, en la Figura 2 se muestran las trayectorias de los sujetos a través de las ocasiones, y las proyecciones de los vectores que identifican las entidades del modo variables. Se observa que las trayectorias con iguales para Ana, Diana y Francia, por una parte, y para Bernardo y Camilo por otra, lo cual verifica los puntajes iguales obtenidos por ellos para todas las variables medidas en cada una de las ocasiones. De otro lado, Edna se encuentra identificada por una trayectoria intermedia debido a que sus puntajes corresponden a valores intermedios entre los dos grupos de sujetos.

Finalmente, al proyectar sobre el mismo plano los vectores correspondientes a las variables, se observa que aquellas variables que identifican la emocionalidad (líneas punteadas) tienden hacia los comportamientos de las mujeres frente a situaciones sociales, mientras que las variables que identifican la conciencia (líneas continuas) se inclinan hacia las mujeres en situaciones de desempeño. Por otra parte, la trayectoria de los hombres ante situaciones sociales se inclinan más hacia los comportamientos regidos por la conciencia que hacia los regidos por la emocionalidad.

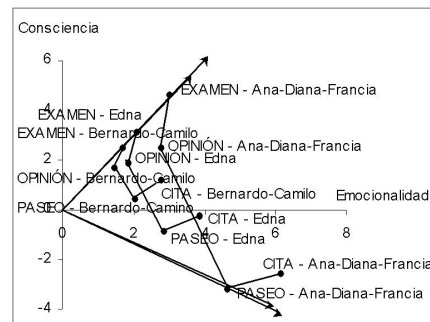


Figura 2. Configuración de entidades del modo A y C

Similar al procedimiento gráfico presentado anteriormente, es posible generar el plano factorial de las entidades del modo A y proyectar sobre él todas las posibles combinaciones entre las entidades del modo B y C, y el plano factorial de las entidades del modo C y proyectar las combinaciones de las entidades del modo A

y B. Por otra parte, para visualizar el comportamiento general de los datos con base en todos los componentes obtenidos, se obtiene el gráfico de trayectorias de los valores del arreglo central, como se muestra en la Figura 3.

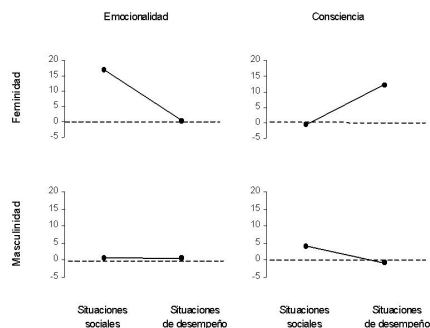


Figura 3. Trayectorias de los valores del arreglo central

Las trayectorias identifican concisamente los niveles de comportamiento de los sujetos medidas para las variables en cada una de las ocasiones, y permiten evidenciar el comportamiento de las mujeres que es emocional cuando se enfrentan a situaciones sociales, y consciente al encontrarse ante situaciones de desempeño, en contraste con los hombres, que tienden hacia comportamientos conscientes al enfrentar las situaciones de desempeño consideradas en el estudio.

7. Conclusiones

El artículo presenta las bases conceptuales, teóricas, analíticas y prácticas que se requieren para el análisis de un arreglo de datos cúbicos mediante el método Tucker3, considerado dentro de las técnicas de análisis factorial dinámico, mostrando conjuntamente una situación aplicada sobre datos hipotéticos con estructura preestablecida. Los conceptos básicos y los principios teóricos tienen su base en el análisis de factores clásico, el análisis de componentes principales, la descomposición en valores singulares, y diferentes manejos matriciales y mecanismos de despliegue usados con el fin de reducir tiempo de cálculo en la estimación de los parámetros del modelo.

El método Tucker3 permite el análisis de arreglos tres-modos referidos regularmente a mediciones de variables sobre sujetos en diferentes ocasiones, mediante procedimientos de reducción que permiten la fácil interpretación y representación en espacios de dimensiones menores a las del arreglo de datos original. El modelo Tucker3 se basa en un mecanismo de descomposición que permite resumir las

entidades de cada uno de los modos del arreglo por medio de unos pocos componentes, y condensar la información total en un arreglo central que especifica las interacciones entre las tres fuentes de variación.

El uso de la descomposición en valores singulares, que proporciona matrices componentes ortonormales, hace del método Tucker3 una buena alternativa para la descripción de datos cúbicos, y permite la estimación de un arreglo central que se constituye en uno de los principales aportes del método, puesto que resume toda la información original en una sola tabla construida con base en las reducciones de cada modo del arreglo de datos. Los parámetros del modelo son estimados mediante el proceso iterativo del algoritmo Tucker3, el cual considera conjuntamente los modos reducidos del arreglo mediante el uso del producto Kronecker.

Las posibilidades interpretativas del método abarcan el análisis de los parámetros del modelo con base en datos tabulados, el porcentaje de ajuste del modelo y de componentes particulares, y todas las posibles representaciones gráficas por medio de planos factoriales. Como un aporte importante del trabajo realizado por Amaya L. (2002), se encuentra la programación que incluye los procedimientos de pre-procesamiento y despliegues, el algoritmo Tucker3 para la estimación de los parámetros, y los cálculos de error y ajuste del modelo.

8. Recomendaciones

Adicional al criterio de ajuste aplicado para determinar el número de componentes a seleccionar en cada modo, es posible establecer la dimensionalidad óptima haciendo uso de otras herramientas tales como validación cruzada, función indicadora de Malinowski, análisis split-half, estudio de residuales y diagnóstico de consistencia central.

Otras de las implementaciones utilizadas para estimar los parámetros del modelo, adicional a la aquí presentada SVD, corresponden a la ortogonalización de Gram-Schmidt, la ortonormalización de Bauer-Rutishauser, la ortonormalización QR y NIPALS.

En caso de aplicar transformaciones sobre alguno de los parámetros del modelo, la manera de seleccionar bases ortonormales para las representaciones gráficas no es tan obvia, y se requiere la aplicación de procedimientos de ortonormalización tales como la ortonormalización de Gram-Schmidt.

Referencias

- [1] AMAYA L. (2002). Análisis Factorial Dinámico mediante el método Tucker3. Universidad Nacional de Colombia. Bogotá, Colombia.

- [2] BRO R. (1998). Multi-way Analysis in the Food Industry. Models, Algorithms, and Applications. Academish proefschrift. Dinamarca.
- [3] JOHNSON A., WICHERN D. (1998). Applied Multivariate Statistical Analysis. Cuarta edición. Prentice-Hall. New Jersey.
- [4] KIERS H. (2000). Towards A Standardized Notation And Terminology In Multiway Analysis. *Journal of Chemometrics*, 14, p.105-122.
- [5] KIERS H. (2000). Some Procedures For Displaying Results From Three-Way Methods. *Journal of Chemometrics*, 14, p.151-170.
- [6] KIERS H., MECHELEN I. (2001). Three-Way Component Analysis: Principles And Illustrative Application. Research Fund of the University of Leuven, Netherlands.
- [7] LAW H., SNYDER C., HATTIE J., MCDONALD R. (1984). Research Methods For Multimode Data Analysis. New York: Praeger.
- [8] TUCKER R. (1966). Some Mathematical Notes On Three-Mode Factor Analysis. *Psychometrika*, 31, p.279-311.

