

DESCOMPOSICIÓN TENSORIAL TUCKER3 APLICADO A TABLAS DE CONTINGENCIAS DE TRES VÍAS

M. Sc. Lucy Gabriela Cuarita Ajno*

✉ lcuarita@fcpn.edu.bo

RESUMEN

El análisis de datos tensorial se encarga del estudio de datos obtenidos de la medición de más de una variable sobre un conjunto de individuos u objetos, los cuales son ordenados en un tensor de orden superior y donde interesa fundamentalmente la descomposición del tensor en componentes mucho más simples, de tal manera que faciliten la interpretación de los datos. En el campo del Análisis Multivariante, en particular, la técnica del Análisis de Correspondencias Múltiple permite identificar la interacción de los niveles correspondientes a un conjunto de variables de estudio, transformando la tabla de contingencias para luego aplicar la técnica del Análisis de Correspondencias Simple. Por otro lado, el modelo tensorial Tucker3 es un método de descomposición tensorial, que permite modelar la interacción entre las vías de un tensor de tercer orden y de sus componentes, preservando la estructura original de los datos. Hoy en día, los modelos tensoriales son una alternativa en el análisis de datos multivariantes, aunque la mayoría de los trabajos se encuentran en el campo del análisis de datos de tres vías, existen investigaciones que indican que la metodología continuará en ascenso mientras las estructuras de datos sean cada vez más complejas y los investigadores requieran un análisis integral de los datos.

PALABRAS CLAVE

Estadístico de Pearson, Inercia, Interacción, Descomposición Tensorial, Modelo Tucker3, Tensor.

ABSTRACT

Tensor data analysis is responsible for the study of data obtained from the measurement of more than one variable on a set of individuals or objects, which are arranged in a higher order tensor and where the decomposition of the tensor into much more components is of fundamental interest. simple, in such a way that they facilitate the interpretation of the data. In the field of Multivariate Analysis, in particular, the Multiple Correspondence Analysis technique allows identifying the interaction of the levels corresponding to a set of study variables, transforming the contingency table and then applying the Simple Correspondence Analysis technique. On the other hand, the Tucker3 tensor model is a tensor decomposition method that allows modeling the interaction between the pathways of a third order tensor and its components, preserving the original structure of the data. Today, tensor models are an alternative in multivariate data analysis, although most of the work is in the field of three-way data analysis, there is research that indicates that the methodology will continue to rise as long as the structures of data become increasingly complex and researchers require a comprehensive analysis of the data.

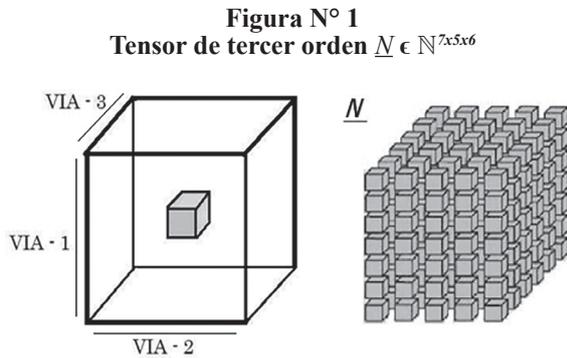
KEYWORDS

Pearson's statistic, Inertia, Interaction, Tensor Decomposition, Tucker3 Model, Tensor.

* Facultad de Ciencias Puras y Naturales de la Universidad Mayor de San Andrés – Bolivia

I. INTRODUCCIÓN

Una tabla de contingencias de tres vías, N , es un tensor de orden 3, que tiene una variable fila con I niveles (vía-1), una variable columna con J niveles (vía-2) y una variable tubo con K niveles (vía-3), es decir: $N \in \mathbb{N}^{I \times J \times K}$ (Amari S., Cichocki A., Huy A., Zdunek R., 2009).



El contenido de los cubos en el tensor son generalmente frecuencias absolutas n_{ijk} o frecuencias relativas p_{ijk} . Las diferencias entre las proporciones observadas pueden ser modeladas utilizando como fundamento teórico el modelo de independencia entre variables fila, columna y tubo (Kroonenberg P., 2008).

II. MODELO DE INDEPENDENCIA

El modelo de independencia postula que p_{ijk} puede ser expresada como el producto de las proporciones marginales:

$$p_{ijk} = p_{i..} p_{.j.} p_{..k}$$

donde:

$$p_{i..} = \sum_{j=1}^J \sum_{k=1}^K p_{ijk}, p_{.j.} = \sum_{i=1}^I \sum_{k=1}^K p_{ijk}$$

y

$$p_{..k} = \sum_{i=1}^I \sum_{j=1}^J p_{ijk}.$$

Luego, la contribución del modelo de independencia debe ser sustraído a cada proporción almacenada en los cubos del tensor, es decir: $(p_{ijk} - p_{i..} p_{.j.} p_{..k})$, quedando definidos los cubos con la dependencia entre los niveles de las tres vías.

El análisis de la dependencia implica estandarizar los valores de la dependencia utilizando la raíz cuadrada de los valores esperados, es decir:

$$(p_{ijk} - p_{i..} p_{.j.} p_{..k}) / \sqrt{p_{i..} p_{.j.} p_{..k}}$$

A los que se denominan *residuales estandarizados* del modelo de independencia. Luego, la suma de los residuales estandarizados al cuadrado es el *coeficiente de contingencia del promedio al cuadrado de Pearson* más conocida como **Inercia** y denotada por Φ^2 . Además, si n es el total de observaciones entonces el estadístico χ^2 de Pearson se define como: $\chi^2 = \Phi^2 n$.

III. MEDICIÓN DE LA DEPENDENCIA

Mientras que en tablas de dos vías existe un solo tipo de dependencia, en tablas de tres vías se pueden distinguir: la dependencia total que es la desviación del modelo de independencia de tres vías, la dependencia marginal que es el resultado de la interacción de dos vías y la dependencia de las tres vías la cual es debido a la interacción de las tres vías (Kroonenberg P., 2008).

Medición de la dependencia total.

En tablas de contingencias de tres vías con dimensiones I, J y K, la dependencia total es medida por la inercia Φ^2 , definida como:

$$\begin{aligned} \Phi^2 &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left[\frac{p_{ijk} - p_{i..} p_{.j.} p_{..k}}{\sqrt{p_{i..} p_{.j.} p_{..k}}} \right]^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K p_{i..} p_{.j.} p_{..k} \left[\frac{p_{ijk} - p_{i..} p_{.j.} p_{..k}}{p_{i..} p_{.j.} p_{..k}} \right]^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K p_{i..} p_{.j.} p_{..k} [\Pi_{ijk}]^2 \end{aligned} \quad (1)$$

donde Π_{ijk} es la medida de la dependencia total en la celda (i,j,k) en la tabla de contingencias de tres vías.

Descomposición tensorial Tucker3 aplicado a tablas de contingencias de tres vías

Medición de la dependencia marginal y la dependencia de tres vías.

La dependencia de la celda Π_{ijk} puede ser dividida en contribuciones separadas de las interacciones de dos y tres vías, (Carrier A. y Kroonenberg P., 1996). Luego, la descomposición de Π_{ijk} es:

$$\Pi_{ijk} = \frac{p_{ij.} - p_{i.} p_{.j.}}{p_{i.} p_{.j.}} + \frac{p_{i.k} - p_{i.} p_{..k}}{p_{i.} p_{..k}} + \frac{p_{.jk} - p_{.j.} p_{..k}}{p_{.j.} p_{..k}} + \frac{p_{ijk} - \alpha p_{ijk}}{p_{i.} p_{.j.} p_{..k}} \quad (2)$$

Donde el término que mide el tamaño de la interacción de las tres vías para la celda (i, j, k) es: $\alpha p_{ijk} = p_{ij.} p_{..k} + p_{i.k} p_{.j.} + p_{.jk} p_{i.} - 2p_{i.} p_{.j.} p_{..k}$. Aplicando la definición de los totales marginales de dos vías, y la última definición para Π_{ijk} , la inercia Φ^2 puede ser particionada como:

$$\begin{aligned} \Phi^2 &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K p_{i.} p_{.j.} p_{..k} [\Pi_{ijk}]^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J p_{i.} p_{.j.} [\Pi_{ij.}]^2 + \sum_{i=1}^I \sum_{k=1}^K p_{i.} p_{..k} [\Pi_{i.k}]^2 + \\ &\quad + \sum_{j=1}^J \sum_{k=1}^K p_{.j.} p_{..k} [\Pi_{.jk}]^2 + \\ &\quad + \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K p_{i.} p_{.j.} p_{..k} \left[\frac{p_{ijk} - \alpha p_{ijk}}{p_{i.} p_{.j.} p_{..k}} \right]^2 \\ &= \Phi_{IJ}^2 + \Phi_{JK}^2 + \Phi_{IK}^2 + \Phi_{IJK}^2 \quad (3) \end{aligned}$$

La última relación es una medida de los ajustes para cada interacción, además proporciona las contribuciones de estas interacciones a la dependencia total.

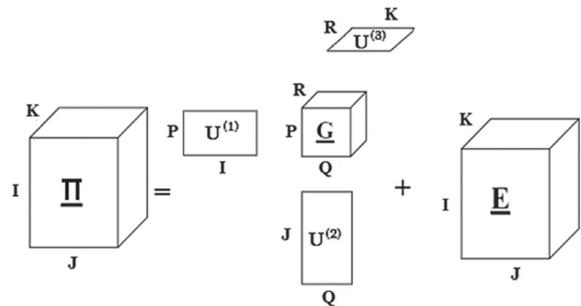
IV. MODELACIÓN DE LA DEPENDENCIA TOTAL

En el caso de tensores de tres vías la modelación implica aplicar una Descomposición en Valores Singulares Generalizada. Al respecto, existen varios candidatos, (Kroonenberg P. M.; 2008). En particular, se elige la *Descomposición en Valores Singulares de tres Modos*, más

conocida con el nombre de Descomposición Tucker3 y por consiguiente Modelo Tucker3.

La aplicación del modelo Tucker3 en la medida de la dependencia total, implica expresar el tensor Π de orden tres como:

Figura N° 2
Modelo Tensorial Tucker3 para el tensor Π



Alternativamente, los valores Π_{ijk} se denotan por:

$$\Pi_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} u_{ip}^{(1)} u_{jq}^{(2)} u_{kr}^{(3)} + \epsilon_{ijk} \quad (4)$$

V. DIMENSIONALIDAD DEL MODELO TUCKER3

Una característica sobresaliente en la estimación de los parámetros del modelo Tucker3 es la identificación de la dimensionalidad del arreglo central, G , donde no se consideran todas las combinaciones (P, Q, R) , pues no siempre son factibles. Por tal motivo surgen procedimientos algorítmicos como ser: el *criterio st* (Ceulemans E. y Kiers H., 2006) y el *diffit* (Timmerman M. y Kiers H., 2000) que permiten determinar las ternas con los mejores porcentajes de ajuste o las menores sumas de residuales al cuadrado como es el caso del *scree plot multivía*, (Timmerman M. y Kiers H., 2000).

VI. ANÁLISIS DE LOS RESIDUALES

El modelo tensorial tucker3 permite detectar características particulares en los datos, como ser puntos atípicos o datos inusuales en los residuales, que pueden revelar características

especiales de algunos datos que no pueden ser modelados y tienen un efecto directo en la salida o estimación del modelo, (Kroonenberg P. M., 2008).

Análisis de los residuales estructurados al Cuadrado

La identificación de puntos anómalos en el ajuste, es el resultado de analizar los residuales dentro de la suma de cuadrados para los elementos de cada modo separadamente, utilizando la suma de los residuales al cuadrado y empleando los gráficos de la suma de cuadrados de los residuales relativos-

Análisis de los residuales no estructurados al cuadrado

Los residuales multivía son mucho más complejos que los residuales de dos vías, sin embargo, un estudio no estructurado sería esencialmente el mismo, es decir que se pueden utilizar gráficos bidimensionales de los residuales y de los valores ajustados.

VII. APLICACIÓN

La Encuesta de Hogares del año 2011 cuenta con 33821 personas encuestadas en sus hogares, de donde son de interés personas mayores de edad que dieron respuesta a la pregunta en relación a su *identificación con algún pueblo originario o indígena*, además de su *idioma materno* y el *departamento* de procedencia; realizando las depuraciones correspondientes sólo 28644 personas cumplieron las características especificadas. Posteriormente, se construye una tabla de contingencias donde la vía-1 corresponde a la variable *Identidad* con cinco niveles (Quechua, Aymara, Otros Nativos, Ninguno y NS\NR), la vía-2 tiene asociada a la variable *Idioma Materno* con cuatro niveles (Aymara, Castellano, Otros y Quechua) y a la vía-3 le corresponde la variable *Departamento* con nueve niveles

(Chuquisaca, La Paz, Cochabamba, Oruro, Potosí, Tarija, Santa Cruz, Beni y Pando). En general, las especificaciones dadas describen la estructura de un tensor de orden 3, $N \in \mathbb{N}^{5 \times 4 \times 9}$.

Al inspeccionar las proporciones marginales (masas de filas, columnas y tubos) de la *Tabla N° 1*, se advierte que el 70% de la población no se identifica como perteneciente a ningún pueblo originario o indígena, es posible que sea debido a que la lengua materna del 73% de la población es el Castellano, sin embargo, un porcentaje significativo de la población (27%) se identifica con el pueblo Quechua o Aymara.

Tabla N° 1. Datos sobre la Identidad: Proporciones Marginales de fila, columna y tubo

Identidad		Idioma	Materno	Departamento	
Quechua	0.16	Aymara	0.10	Chuquisaca	0.07
Aymara	0.11	Castellano	0.73	La Paz	0.23
Otros Nativos	0.02	Otros	0.01	Cochabamba	0.18
Ninguno	0.70	Quechua	0.17	Oruro	0.06
NS-NR	0.01			Potosí	0.09
				Tarija	0.07
				Santa Cruz	0.21
				Beni	0.06
				Pando	0.04

Cálculo de la Inercia y el estadístico de Pearson

Utilizando los datos del tensor definido en el apartado anterior y aplicando las relaciones de la Inercia (1) y (3), además de la definición del estadístico χ^2 de Pearson, los resultados de la *Tabla N° 2*, muestran que existe una correspondencia significativa de la identidad de la población si se toma en cuenta la región de nacimiento y el idioma materno, con un 38% de la inercia, a pesar de que la proporción marginal, *Tabla N° 1*, mostró que el 70% de la población no tiene una identificación con un pueblo originario.

Además, es bueno señalar que, la correspondencia entre la Identidad y el Idioma Materno es significativa con un 28.77% del total de la inercia. Por lo expuesto, se advierte que las correspondencias identificadas muestran que en nuestro País aún existen grupos humanos

Descomposición tensorial Tucker3 aplicado a tablas de contingencias de tres vías

que conservan o se identifican con las raíces culturales de su región o territorio.

Tabla 2. Particionamiento de la dependencia total

Interdependencia	χ^2	Φ^2	Porcentaje
Identidad x IdiomaMat	24891.64	0.869	28.77
Identidad x Departamento	16498.94	0.576	19.06
IdiomaMat x Departamento	12259.63	0.428	14.17
Identidad x IdiomaMat x Departamento	32883.31	1.148	38.00
Total	86533.52	3.021	100.00

Identificación del Modelo

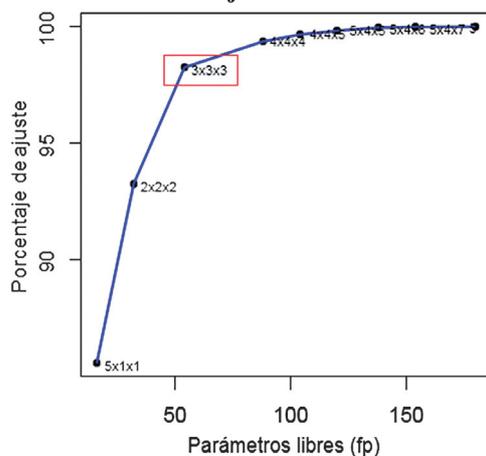
Sea el tensor $\underline{N} \in \mathbb{N}^{5 \times 4 \times 9}$, correspondiente a los datos sobre la Identidad, se construye el tensor de proporciones $\underline{P} \in \mathbb{R}^{5 \times 4 \times 9}$, luego con los valores p_{ijk} de \underline{P} es necesario estimar el tensor $\underline{\Pi} \in \mathbb{R}^{5 \times 4 \times 9}$, cuyas celdas representan las dependencias entre los diferentes niveles de la tres vías. La identificación inicia con la determinación de la dimensión del arreglo central, de la descomposición de Tucker, \underline{G} , para tal efecto se utilizan los criterios de dimensionalidad.

Criterio st de Ceulemans y Kiers

La aplicación del criterio st indica que la dimensión $(P, Q, R) = (3, 3, 3)$ es adecuada para representar el arreglo central de la Descomposición de Tucker.

Figura N° 3

Criterio st Parámetros libres y el porcentaje de ajuste



Scree plot multivía

En la figura 6.14 (gráfico de la derecha), se observa que a partir del orden $(3, 3, 3)$ el descenso de la Suma de Errores al cuadrado

es prácticamente constante, luego se considera que este parámetro es un referente significativo de la dimensionalidad del arreglo central en el modelo Tucker3.

Criterio diffit

En este criterio es relevante analizar la razón del incremento del porcentaje de ajuste resultante del m -ésimo componente en relación al ajuste porcentual del $m+1$ ésimo componente, donde el número total de componentes es $S=P+Q+R$. La Tabla N° 3, muestra que la aplicación del criterio diffit determina que la dimensión, del arreglo central, a ser elegida es $(P, Q, R) = (3, 3, 3)$.

Figura N° 4

Scree plot multi vía

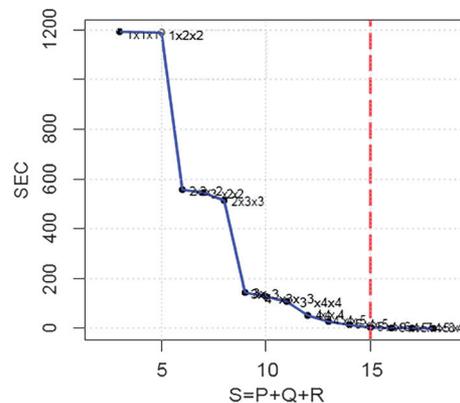


Tabla N° 3. Resultados de la aplicación del Criterio diffit

P	Q	R	S	sec	fit(%)	difs	btm
1	1	1	3	1193.19	85.58	85.58	11.20
1	2	2	5	1189.99	85.62	0.04	-
2	2	2	6	557.92	93.26	7.64	1.71
3	2	2	7	544.98	93.41	0.15	-
2	3	3	8	514.87	93.78	0.37	-
3	3	3	9	144.21	98.26	4.48	6.59
4	3	3	10	127.30	98.46	0.20	-
3	4	4	11	108.10	98.69	0.23	-
4	4	4	12	52.08	99.37	0.68	2.27
4	4	5	13	27.70	99.67	0.30	1.88
5	4	5	14	14.37	99.83	0.16	1.23
5	4	6	15	3.31	99.96	0.13	4.33
5	4	7	16	0.90	99.99	0.03	3.00
5	4	8	17	0.42	99.99	0.00	-
5	4	9	18	0.00	100.00	0.01	∞

Contribución de las componentes en cada vía

Los espacios originales de las vías en estudio, debido a la descomposición de Tucker3 empleada, se han reducido a un espacio tridimensional, en donde interesa conocer el porcentaje de la dependencia retenida en cada una de ellas, luego se presentan los siguientes resultados:

Tabla N° 4. Contribución porcentual en cada vía

Vía	Dim.1	Dim.2	Dim.3	Total
Identidad	85.738	8.045	4.474	98.26
Idioma Materno	85.730	8.038	4.490	98.26
Departamento	85.657	7.910	4.691	98.26

Donde 98.26 corresponde al porcentaje total de contribución o dependencia explicada por el modelo Tucker3, en cada una de sus vías.

Tabla N° 5. Porcentaje de ajuste de la inercia con el Modelo Tucker3 seleccionado

Interdependencia	$\hat{\chi}^2$	$\hat{\Phi}^2$	Porcentaje Ajuste
Identidad x IdiomaMat	20451.82	0.714	82.16
Identidad x Departamento	10942.01	0.382	66.32
IdiomaMat x Departamento	8936.93	0.312	72.90
Identidad x IdiomaMat x Departamento	35403.98	1.236	123.03

La Tabla N° 5 muestra en su última columna el porcentaje de ajuste de la inercia con el modelo seleccionado, donde la interacción entre el *Idioma Materno* y la *Identidad* tiene el porcentaje más significativo con 82 %. Por otro lado, llama la atención el sobre ajuste identificado para la triple interacción sin embargo no es alarmante si se comparan con las inercias y los valores del estadístico χ^2 de los datos originales y estimados, puesto que se preserva la interpretación de la dependencia dada para cada una de las interacciones.

Representación gráfica e interpretación

La representación gráfica de la dependencia estimada consiste en expresar las coordenadas de las componentes de una vía en un espacio determinado y proyectar sobre el las coordenadas de las componentes de las otras dos vías. Por tal motivo, las coordenadas

Estimación de la Inercia y el estadístico de Pearson

El modelo Tucker3, relación (4), permite descomponer el tensor $\underline{\Pi}$ como sigue:

$$\Pi_{ijk} = \sum_{p=1}^3 \sum_{q=1}^3 \sum_{r=1}^3 g_{pqr} u_{ip}^{(1)} u_{jq}^{(2)} u_{kr}^{(3)} + \epsilon_{ijk}$$

de donde $\hat{\underline{\Pi}}$ se define como:

$$\hat{\Pi}_{ijk} = \sum_{p=1}^3 \sum_{q=1}^3 \sum_{r=1}^3 \hat{g}_{pqr} \hat{u}_{ip}^{(1)} \hat{u}_{jq}^{(2)} \hat{u}_{kr}^{(3)}$$

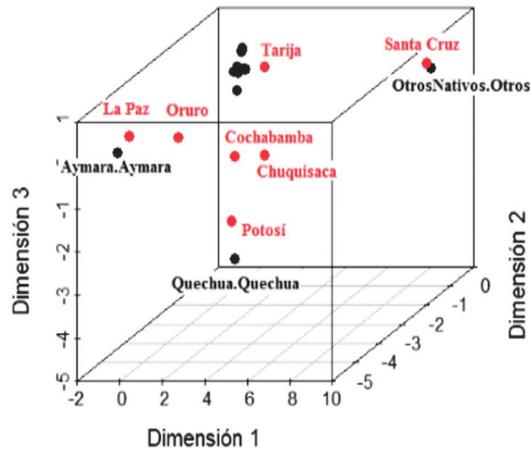
Estas estimaciones permiten aproximar los valores calculados de las inercias, expuestos en el Tabla N° 2, de acuerdo a las relaciones (1) y (3).

de los niveles de la vía *Departamento* son transformadas a un espacio en particular, donde luego las coordenadas de los niveles combinados de las vías *Identidad e Idioma materno* son proyectadas en este espacio, como se muestra en la Figura N° 5.

En la Figura N° 5, se pone de manifiesto la dependencia de las tres vías, puesto que el *idioma materno* determina en la mayoría de los casos la identificación de la persona con el pueblo originario o indígena donde se hable ese idioma, por ejemplo si el idioma materno es el Aymara en general su identificación será con el pueblo Aymara que geográficamente e históricamente abarca los departamentos de La Paz y Oruro; un fenómeno similar sucede con las personas Potosinas, Cochabambinas y Chuquisaqueñas cuyo idioma materno es el Quechua su identificación generalmente es con el pueblo Quechua.

Descomposición tensorial Tucker3 aplicado a tablas de contingencias de tres vías

Figura N° 5:
Representación de la dependencia estimada de la vía Departamento y las vías combinadas Identidad e Idioma Materno.



En Santa Cruz es donde se tiene una mayor diversidad de pueblos indígenas, luego la identificación con un pueblo indígena tiene una correspondencia directa con su lengua materna. Hay que señalar, que los departamentos de Tarija, Beni y Pando son aquellos donde el idioma materno que impera es el castellano, luego es evidente que su identificación se halla relacionada principalmente con los niveles “Ninguno” y “NS-NR”.

Análisis de los Residuales Estructurados

Sea el tensor de residuales, $\underline{E} \in \mathbb{R}^{5 \times 4 \times 9}$ definido por:

$$E = \Pi - \hat{\Pi}$$

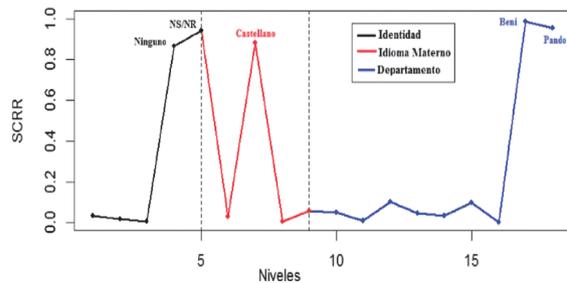
donde Π es el tensor con las dependencias y $\hat{\Pi}$ la estimación de las dependencias hallada a través del Modelo Tucker3 de dimensión $3 \times 3 \times 3$. El análisis de los residuales estructurados, inicia con el cálculo y representación gráfica de las sumas de cuadrados.

Suma de Cuadrados de los residuales relativos (SCRR) de componentes por vía.

El gráfico de la Figura N° 6 presenta a las tres vías del tensor de datos: *Identidad, Idioma Mat*

y *Departamento* en sus diferentes niveles con las respectivas SCRR. Las representaciones gráficas muestran que los niveles *Ninguno* y *NS –NR de Identidad; Castellano de Idioma Materno; Beni y Pando de Departamento* no presentan estimaciones razonables (valores cercanos a la unidad) en relación a las restantes componentes en sus vías respectivas.

Figura N° 6.
De izquierda a derecha; Suma de Cuadrados de los Residuales Relativos de cada una de las componentes correspondientes a las vías Identidad, Idioma Materno y Departamento.



La Figura N°6 presenta de manera conjunta la SCRR en donde es evidente la pobre estimación de los niveles *Ninguno, NS.NR, Castellano, Beni y Pando* con el modelo Tucker3 de dimensión $3 \times 3 \times 3$. Las causas pueden ser debido a que:

- La pregunta con respecto a la identificación con un pueblo originario o indígena, causa que la población en su gran mayoría opte por las opciones *Ninguno, NS.NR*; aspecto que puede asociarse al efecto de la globalización, confusión o carencia de conocimiento sobre las implicaciones del término identidad o identificación con pueblos originarios, entre otros.
- La población en un número significativo asocia su identidad con el idioma materno, es decir que si su idioma materno es el *Castellano* no se identifica con ningún pueblo originario o indígena, sin tomar en cuenta sus raíces y las expresiones

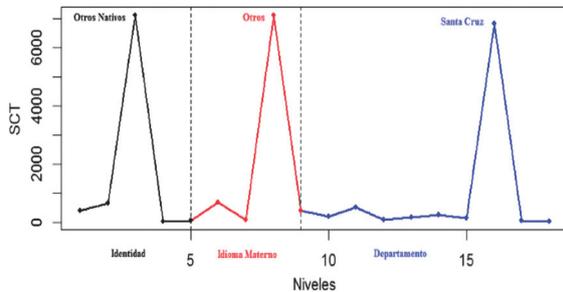
culturales de su región.

- Los departamentos de *Beni* y *Pando* son regiones afectadas por la migración, en el caso de la migración interna se observa que, si bien el idioma materno es el Castellano, la identificación con los pueblos Aymara, Quechua y otros nativos es significativa, lo que marca una diferencia sobresaliente con relación a los otros departamentos.

Suma de Cuadrados de los totales de componentes por vía (SCT)

La Figura N° 7 muestra la SCT por modo, en donde se identifican tres valores extremadamente altos correspondientes a los niveles *Otrosnativos*, *Otros* y *SantaCruz* de las vías *Identidad*, *Idioma Materno* y *Departamento*, respectivamente.

Figura N° 7. De izquierda a derecha; Suma de Cuadrados de los Totales de las componentes correspondientes a las vías Identidad, Idioma Materno y Departamento.



Considerando, que la SCT es calculada del tensor Π implica que en los componentes identificados existe una interacción significativa, en especial en la interacción de tres vías. Sin embargo, en la Tabla N° 1 se muestra claramente que las proporciones marginales de *Otrosnativos* en *Identidad* y *Otros* en *Idioma Materno* corresponden solamente al 2% y 1%, respectivamente, de la población encuestada. Por otro lado, la población encuestada en Santa Cruz es 6105 y si tomamos en cuenta a las personas con los dos niveles anteriores se tiene que es

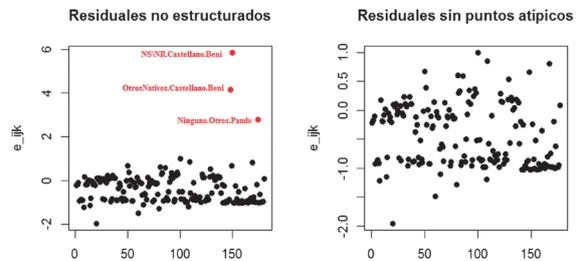
solamente el 1.04 %; luego es posible que en el departamento de Santa Cruz sea significativa esta relación de niveles, sin embargo, puede estar afectando la interacción de las vías en los otros departamentos.

Análisis de los Residuales No Estructurados

El tensor de residuales $\underline{E} \in \mathbb{R}^{5 \times 4 \times 9}$ cuenta con 180 residuales de las diferentes vías, en este apartado interesa solamente el conjunto de datos sin tomar en cuenta las vías y sus componentes.

En la Figura N° 8 se muestra el comportamiento de los 180 residuales, en donde claramente se observa la presencia de tres puntos atípicos que corresponden a las ternas: (NS-NR, Castellano, Beni), (Otros-Nativos, Castellano, Beni) y (Ninguno, Otros, Pando). En general, se conoce que, en Beni o Pando, no existe una identificación directa con algún pueblo originario, luego una alternativa razonable puede ser modelar los datos de la dependencia eliminando la influencia de estos departamentos.

Figura N° 8. De izquierda a derecha; a) Residuales no estructurados donde se identifican tres puntos atípicos; b) residuales no estructurados sin considerar los atípicos.



Análisis de Correspondencias Múltiple

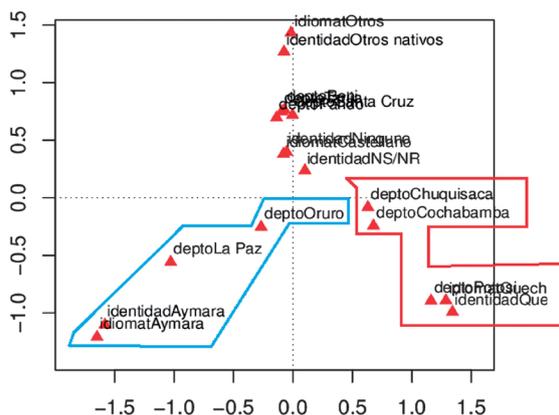
El análisis de tablas de contingencias de tres vías se realiza aplicando la Técnica del Análisis de Correspondencias Múltiple, que es la aplicación de un Análisis de Correspondencias Simple en tablas de dos vías donde la tabla de tres vías es transformada en una tabla de dos vías, (Greenacre M., 2007).

Descomposición tensorial Tucker3 aplicado a tablas de contingencias de tres vías

Transformando los datos en la presente aplicación y utilizando la matriz de Burt, se obtiene los datos de la inercia total y el valor del estadístico de Pearson, 2.082948 y 59663,97, respectivamente. Note que el valor de la inercia total está por debajo de la inercia hallada con el modelo Tucker3, además que no es posible calcular las inercias de la interacción de dos y tres vías. Sin embargo, si se observa lo expuesto en la Figura N° 9, se evidencia similitud con las conclusiones emitidas en relación a la Figura N° 5.

Figura N° 9.

Análisis de Correspondencias Múltiple aplicado a la tabla de contingencias de 3 vías, donde se considera las variables Identidad, Idioma Materno y Departamento



VIII. CONCLUSIÓN

La potencialidad del Modelo Tucker3 en la identificación de la interacción o correspondencia entre vías y sus componentes en una tabla de contingencias de 3 vías, se pone de manifiesto si se comparan los resultados obtenidos con métodos convencionales como ser el Análisis de Correspondencias Múltiple. Actualmente, los modelos tensoriales se constituyen en nuevas herramientas en el campo multivariante, los cuales proponen alternativas que pueden robustecer o complementar las conclusiones emitidas con las técnicas convencionales multivariantes.

BIBLIOGRAFÍA

1. Amari S., Cichocki A., Huy A. y Zdunek R. (2009). Non negative matrix and tensor factorizations: Applications to exploratory Multi-Way data Analysis and blind source separation. Editado por John Wiley & Sons Ltd, Reino Unido
2. Carlier A. y Kroonenberg P. (1996). Decompositions and blots in three-way correspondence analysis. *Psychometrika*, vol. 61, No 2, 355-373.
3. Ceulemans E. y Kiers H. (2006). Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. *British Journal of Mathematical and Statistical Psychology*, No 59, 133-150
4. Kroonenberg P. M. (2008). *Applied Multiway Data Analysis*. Wiley Series in Probability and Statistics. Estados Unidos de Norte América.
5. Timmerman M. y Kiers H. (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.