

Variabilidad en la clasificación de las trayectorias temporales
diferenciales en el análisis factorial dinámico, según diferentes
aproximaciones , jerárquicas y fuzzy

Jorge Blanco
Andrés Castrillejo
Laura Nalbarte

Octubre 2002

Serie documentos de trabajo
DT (02/01)

VARIABILIDAD EN LA CLASIFICACIÓN DE TRAYECTORIAS TEMPORALES
DIFERENCIALES EN ANÁLISIS FACTORIAL DINÁMICO, SEGÚN
DIFERENTES APROXIMACIONES JERÁRQUICAS Y FUZZY.

Jorge Blanco. Andrés Castrillejo. Laura Nalbarte.

Instituto de Estadística. F.C.E. y A. Universidad de la República. Uruguay.

RESUMEN

En el desarrollo del Modelo 1 (en sus enfoques directo y dual) del Análisis Factorial Dinámico (AFD) (Blanco J., Coppi R., Corazziari I. 1996-2002), se obtienen trayectorias temporales diferenciales de unidades y variables. Resulta en esa situación de gran importancia elaborar métodos que permitan detectar las similitudes entre estas trayectorias a los efectos de resumirlas, identificarlas, caracterizarlas y luego mediante adecuados procesos de suavizamiento utilizarlas como herramientas con fines predictivos.

En el presente trabajo se presentan los resultados obtenidos utilizando diversos tipos de enfoques para el cálculo de las distancias tales como los de multicomponent, estructural, cross time lagged, spline, etc. Luego se construyen mediante procedimientos de clasificación jerárquicos crisp y no jerárquicos fuzzy grupos de trayectorias y se analiza su variabilidad y robustez en el tratamiento de array temporales normalizados.

Se han considerado como punto de partida los resultados obtenidos con la estrategia base del AFD en el análisis del Mercado de Trabajo uruguayo en los últimos quince años.

PALABRAS CLAVES: Análisis Factorial Dinámico, Distancias entre Trayectorias diferenciales temporales multivariadas. Clasificación crisp y fuzzy, Robustez y variabilidad.

I. INTRODUCCIÓN

En los enfoques *Directo* y *Dual* del *Análisis Factorial Dinámico* (AFD) se descompone la variabilidad total del array $\mathbf{Z}(\text{IJT})$ expresada mediante matrices de varianzas y covarianzas \mathbf{S} , y de proximidades \mathbf{P} . Considerando el Modelo I, una de las fuentes de la variabilidad total es la dinámica diferencial de las unidades (*Directo*) y de las variables o grupos de ellas (*Dual*). Esto corresponde respectivamente a la variabilidad debida a la interacción *tiempos-unidades* y *tiempos-variables* y es modelizada mediante la proyección de unidades (variables) en espacios de baja dimensión con propiedades óptimas de inercia explicada.

Para el caso *Directo* (al que referiremos en adelante recordando que las consideraciones son válidas para el caso *Dual* sustituyendo oportunamente unidades por variables) la matriz a considerar es $\mathbf{S}_{\text{IT}} = \{s_{jj'}\}_{j,j'=1,\dots,J}$ matriz de varianzas y covarianzas de los valores $(z_{ijt} - \bar{z}_{ij} - \bar{z}_{.jt} + \bar{z}_{.j})$,

El elemento genérico es: $s_{jj'} = \frac{1}{\text{IT}} \cdot \sum_{t=1}^T \sum_{i=1}^I (z_{ijt} - \bar{z}_{ij} - \bar{z}_{.jt} + \bar{z}_{.j}) \cdot (z_{ij't} - \bar{z}_{ij'} - \bar{z}_{.jt} + \bar{z}_{.j'})$

El *primer modelo* considera grupos de observaciones definidos por el tiempo, descomponiendo \mathbf{S} en la siguiente forma $\mathbf{S} = {}^* \mathbf{S}_{\text{T}} + {}^* \mathbf{S}_{\text{I}} + \mathbf{S}_{\text{IT}} = {}^* \mathbf{S}_{\text{T}} + \bar{\mathbf{S}}_{\text{T}}$ donde:

$\bar{\mathbf{S}}_{\text{T}}$ es la media de la variabilidad “dentro” (*within*) cada grupo (los diferentes tiempos). Esta variabilidad es llamada *estructural* y está compuesta por la *estructura sintética* y la *dinámica diferencial*.

${}^* \mathbf{S}_{\text{T}}$ Expresa la variabilidad temporal de los centros medios $(\bar{z}_{.1t}, \bar{z}_{.2t}, \dots, \bar{z}_{.Jt})$, $t=1, \dots, T$. Esta es la *dinámica media*. Definido el modelo para el análisis de la dinámica media mediante regresión, las otras fuentes de variabilidad se analizan mediante análisis de componentes principales (ACP) con referencia al operador $\bar{\mathbf{S}}_{\text{T}}$.

La variabilidad aproximada por las componentes principales extraídas de $\bar{\mathbf{S}}_{\text{T}}$, puede ser descompuesta linealmente en la parte correspondiente a la traza de ${}^* \mathbf{S}_{\text{I}}$ y a la de \mathbf{S}_{IT} .

Las trayectorias de los individuos se obtienen considerando, para cada $i=1,\dots,I$, el conjunto de las coordenadas sobre las componentes principales con índices dados por $t=1,\dots,T$. Tales coordenadas resultan de la matriz $\mathbf{C} = \mathbf{Z}_c \times \mathbf{M} \times \mathbf{A}$ donde \mathbf{Z}_c es el array original $\mathbf{Z}(IT,J)$ considerando las variables centradas en cada ocasión, y \mathbf{A} es la matriz obtenida considerando los primeros H autovectores de $\bar{\mathbf{S}}_T$.

Tales trayectorias son el neto de la evolución media del sistema que caracteriza los centros $\bar{z}_{.jt}$ y sobre el plano considerado por los factores principales corresponde al baricentro θ .

Las trayectorias proyectadas pueden ser comparadas entre ellas definiendo oportunas matrices de distancias, y de esta manera identificar tipologías de individuos. Estamos entonces frente a la situación de clasificar procesos temporales multivariados y resultan de gran importancia los métodos elegidos para agrupar dichas trayectorias, lo que influirá directamente sobre su posterior caracterización.

En la literatura se han intentado diferentes aproximaciones a la solución de este problema. Presentamos aquí algunas de ellas y la variabilidad de los resultados que hemos obtenido analizando las trayectorias surgidas de la evolución del Mercado de Trabajo uruguayo en los últimos quince años. (ver Bellagamba, et al 2000).

La fuente de información para construir el array *unidades x variables x tiempos* provino de las Encuestas Continuas de Hogares (ECH). Se consideró para el análisis la población económicamente activa (PEA), clasificada según región, sexo, edad y nivel educativo. Como resultado de esta clasificación se conformaron 36 unidades estadísticas¹ que fueron luego caracterizadas por 19 variables medidas en los 15 tiempos. Las variables involucradas se refieren a la intensidad del trabajo (dedicación horaria, multiempleo), problemas en la inserción laboral (subempleo, precariedad, desempleo), características de la ocupación (categoría y código), características del establecimiento (rama y tamaño) y las remuneraciones provenientes del mercado laboral.

1

A los efectos de construir las unidades, la pea se organizó en clases definidas por: **Región:**(M) montevideo (I) interior urbano **Sexo:** (H) Hombres (M) Mujeres **Edad:** (1) 20-29 (2) 30-49 (3) 50 y más **Nivel educativo:** (1) Primaria (2) Secundaria o UTU (3) Universidad o Magisterio. de la clasificación resultaron 36 grupos, donde MH11 corresponde a montevideo, hombres, de entre 20 y 29 años y con educación primaria.

A continuación se presentan distintas formas de construcción de matrices de disimilaridad de las trayectorias determinadas por la dinámica diferencial de las unidades, estudiándose posteriormente la variabilidad de los resultados obtenidos mediante la consideración de diversos procedimientos de clasificación crisp y fuzzy.

II .CONSTRUCCIÓN DE LAS MATRICES DE DISIMILARIDAD

Sin considerar modelos probabilísticos en la generación de los datos las trayectorias multivariadas temporales pueden ser comparadas construyendo adecuadas medidas de disimilaridad (Para un desarrollo exhaustivo ver D'Urso 2000) . Entre ellas pueden destacarse diversas aproximaciones como las multicomponent, cross time lagged, estructural, etc. En este trabajo referiremos solamente a algunas de ellas.

II.1 Distancias Multicomponent

En el enfoque multicomponent las medidas de disimilaridad entre trayectorias temporales se definen teniendo en cuenta las características instantáneas (transversales) y/o evolutivas (longitudinales). Algunos de los aportes más significativos en este sentido son los siguientes:

a) Carlier (1986, 1991, 1999) ha definido tres tipos de medidas de disimilaridad entre trayectorias: 1) instantánea (transversal) 2) longitudinal (evolutiva) y 3) mixta.

$$1) \text{ Instantánea: } {}_{cs}d_{il}^2 = \sum_{t=1}^T \sum_{j=1}^J (x_{ijt} - x_{jlt})^2 = \sum_{t=1}^T \|x_{it} - x_{lt}\|^2 \quad {}_{cs}d_{il}^2 = 0 \Leftrightarrow x_{ijt} = x_{jlt}, \forall j, t$$

$$2) \text{ Longitudinal } {}_Ld_{il}^2 = \sum_{t=2}^T \sum_{j=1}^J [(x_{ijt} - x_{jlt-1})(x_{jlt} - x_{jlt-1})]^2 = \sum_{t=2}^T \|(x_{it} - x_{it-1})(x_{lt} - x_{lt-1})\|^2$$

$${}_Ld_{il}^2 = 0 \Leftrightarrow x_{ijt} = x_{jlt} + \mathbf{b}, \text{ para cualquier } \mathbf{b} \in R \forall j, t$$

3) Mixta: Medida de disimiliaridad que tiene en cuenta simultáneamente los aspectos instantáneos y evolutivos, considerando la combinación lineal:

$${}_c d_{il}^2 = \mathbf{a}_1 {}_{cs}d_{il}^2 + \mathbf{a}_2 {}_Ld_{il}^2. \quad \text{con } \mathbf{a}_1 + \mathbf{a}_2 = 1.$$

Las críticas fundamentales a las distancias de Carlier son las siguientes(D'Urso 2000): 1) la elección de los pesos es subjetiva 2) En el cálculo de la disimilaridad mixta se consideran T coordenadas en relación a la posición y T-1 coordenadas relativas a la parte longitudinal mientras cada trayectoria temporal es identificada por T coordenadas.

Esto implica que en el cálculo de la Mixta la influencia de las componentes transversal y longitudinal no sean equilibradas.

De estos dos tipos de críticas a nuestro entender la fundamental es la segunda pues en distintos trabajos realizados en el análisis de trayectorias multivariadas temporales la influencia de las disimilaridades instantáneas es ampliamente predominante. Esto ocurre fundamentalmente cuando el período considerado en el análisis de fenómenos reales no es demasiado extenso. Deben verificarse tipos de trayectorias muy particulares para que las distancias evolutivas adquieran una importancia comparable a las instantáneas.

En cuanto al carácter subjetivo de los pesos le restamos importancia porque como veremos más adelante otros autores pretenden dar un carácter “objetivo”, expresado a través de la optimización de ciertas funciones, y la subjetividad de los mismos está presente en la elección del tipo de modelo considerado.

b) Saporta y Lavellard(1996) proponen también una medida de disimilaridad mixta, compromiso entre la estructura y la evolución. A diferencia de lo que ocurría con las de Carlier buscan un procedimiento de optimización para asignar el peso a cada componente.

Para determinar la disimilaridad mixta se procede en dos pasos:

1. Se determina la configuración óptima $\{a_1, a_2, \dots, a_T\}$ donde $a_t = 0$ es coordenada de posición y para todo $t=2 \dots T$ si $a_t = 1$ se considera una coordenada de evolución y si $a_t = 0$ se considera una de posición.

Para la configuración $\{a_1, a_2, \dots, a_T\}$, I_T representa la inercia total que se puede descomponer en $I_T = I_{T_1} + I_{T_2}$

$$\text{donde: } I_{T_1} = \sum_{i=1}^I \sum_{\{t:a_t=1\}} \|x_{it} - \bar{x}_t\|^2 \quad I_{T_2} = \sum_{i=1}^I \sum_{\{t:a_t=0\}} \|(x_{it} - x_{it-1}) - (\bar{x}_t - \bar{x}_{t-1})\|^2$$

La configuración óptima es aquella que se obtiene minimizando: $|I_{T_1} - I_{T_2}|$

2. Una vez obtenida la configuración óptima se calcula la disimilaridad compromiso de la siguiente forma:

$${}_{SL}d_{it}^2 = \sum_{\{t:a_t=0\}} \|x_{it} - x_{it}\|^2 + \sum_{\{t:a_t=0\}} \|(x_{it} - x_{it-1}) - (x_{it} - x_{it-1})\|^2$$

En este caso se superan algunos de los problemas que presentaban las distancias de Carlier. Sin embargo el método de Saporta-Lavellard al pretender minimizar la

diferencia entre los componentes de la inercia privilegia en la mayoría de los casos la evolutiva. Se podría mejorar para casos particulares cambiando la relación final minimizando más veces la posicional respecto a la evolutiva. De todas maneras constituye un aporte interesante y ha sido la base para nuevas búsquedas de soluciones. Entre ellas surge la propuesta de:

c) **D'Urso- Vichi** (1998) proponen una medida de disimilaridad que combina una transversal y dos evolutivas

la transversal está definida por:
$${}_1d_{il}^2 = \sum_{t=1}^T \sum_{j=1}^J (x_{ijt_s} - x_{ijt_s})^2 = \sum_{t=1}^T \|x_{it_s} - x_{lt_s}\|^2$$

y las de disimilaridad longitudinal se basan en los conceptos de velocidad y aceleración:

$${}_2d_{il}^2 = \sum_{s=2}^T \sum_{j=1}^J (v_{ijt_s} - v_{ijt_s})^2 = \sum_{s=2}^T \|v_{it_s} - v_{lt_s}\|^2$$
 mide la disimilaridad evolutiva-velocidad,

mientras que

$${}_3d_{il}^2 = \sum_{s=3}^T \sum_{j=1}^J (a_{ijt_s} - a_{ijt_s})^2 = \sum_{a=2}^T \|a_{it_s} - a_{lt_s}\|^2$$
 mide la disimilaridad evolutiva-aceleración.

La velocidad de la ***i*-ésima trayectoria temporal** relativa a la ***j*-ésima** variable, en el intervalo $[t_{s-1}, t_s]$, mide la variación de ***i*** de t_{s-1} a t_s con:
$$v_{ijt_s} = \frac{x_{ijt_s} - x_{ijt_{s-1}}}{t_s - t_{s-1}} \quad s=2, \dots, T.$$
 e

indica la dirección de la trayectoria temporal en cada intervalo $[t_{s-1}, t_s]$

La aceleración es la variación de ***i*** en cierto intervalo de tiempo. Con
$$a_{ijt_s} = \frac{v_{ijt_s} - v_{ijt_{s-1}}}{t_s - t_{s-2}}$$
 $s=3, \dots, T$

Medida de disimilaridad mixta:
$${}_{DV}d_{il}^2 = \sum_{r=1}^3 \mathbf{a}_r d_{il}^2, \quad r=1,2,3$$
 . En este caso los ***a*** son

elegidos haciendo máxima la varianza de la disimilaridad compromiso.

El aporte mayor de los autores es la incorporación del concepto de aceleración. Se mantienen sin embargo los problemas ya vistos para la velocidad en Saporta – Lavellard. A su vez, se plantea una forma diferente de ponderar cada una de las componentes relacionándolas con la descomposición espectral de una matriz formada con vectores que representan cada una de las distancias.

II.2 Distancias estructurales.

En este caso se suponen válidas las hipótesis clásicas de que la trayectoria temporal puede ser descompuesta en ciclo, tendencia, estacionalidad y componente irregular. De este modo cada pareja de trayectorias puede ser comparada considerando oportunos parámetros que caracterizan su estructura. Una de las medidas de disimilaridad consiste en comparar los coeficientes del modelo trend elegido.

a) Medidas de disimilaridad entre trend.

Dada la i -ésima variable y la j -ésima trayectoria temporal se define el trend polinomial de orden g de la siguiente forma: $\mathbf{x}_{ij} = \mathbf{T}\mathbf{b}_{ij} + \mathbf{e}_{ij}$ donde \mathbf{x}_{ij} es el vector de dimensión $T \times 1$ de las observaciones relativas a la i -ésima unidad y la j -ésima variable y \mathbf{T} es la matriz de dimensión $T \times (g+1)$, \mathbf{b}_{ij} es el vector $(g+1) \times 1$ de los parámetros de regresión del modelo y \mathbf{e}_{ij} el vector de residuos.

Considerando que respecto a la variable j ($j=1, J$) todas las I trayectorias temporales tengan un trend del mismo orden, la disimilaridad considerada entre parejas de trayectorias es: $d_{il}^2 = \sum_{j=1}^J \|\widehat{\mathbf{b}}_{ij} - \widehat{\mathbf{b}}_{lj}\|^2 = \|\widehat{\mathbf{B}}_i - \widehat{\mathbf{B}}_l\|$ donde $\widehat{\mathbf{B}}_i$ y $\widehat{\mathbf{B}}_l$ son las matrices $(g+1) \times T$ de los parámetros de los modelos trend de orden g , de la i -ésima y l -ésima trayectoria temporal.

Considerando esta medida de disimilaridad, las trayectorias temporales son comparadas en base a sus trend polinomiales mediante la comparación de los respectivos coeficientes de regresión.

III. ANÁLISIS DE RESULTADOS

De la aplicación del Modelo I del AFD al array de datos del mercado de trabajo surgen cuatro nuevas variables (ejes factoriales) que se describen a continuación:

El *primer eje* discrimina según el nivel de remuneraciones y estabilidad en el mercado laboral. En efecto, se correlacionan positivamente con el eje los profesionales y técnicos, los ingresos, el multiempleo, los empleados públicos y los servicios financieros y a empresas. En el extremo opuesto del eje inciden la precariedad, los servicios personales, los empleados privados y la rama industrial. Se verifica un

ordenamiento ascendente de las unidades según su nivel educativo. A su vez, dentro de las unidades con la misma educación, se produce un ordenamiento por tramo etario. En cuanto al *segundo eje* la lectura que se realiza del mismo tiene que ver con la dedicación horaria en el mercado de trabajo así, se correlacionan positivamente el tiempo parcial, el desempleo y el subempleo y, en sentido contrario, se encuentra fundamentalmente la condición de jefatura de hogar (“indicadora de alta dedicación laboral”) y en menor medida los cuenta propia sin local y trabajadores manuales. El *tercer eje* factorial brinda información acerca del tipo de relación respecto al puesto de trabajo, en tanto contrapone a los empleados de oficina que son los que desempeñan tareas bajo las órdenes de un superior (correlación positiva) de los cuenta propia con local que corresponden a trabajadores independientes (correlación negativa). En la formación del *cuarto eje* factorial intervienen fundamentalmente los empleados públicos (correlación negativa) por oposición a los trabajadores de la rama servicios financieros y a empresas (correlación positiva).

Teniendo en cuenta las 4 variables, 36 unidades y 15 tiempos se construyen matrices de disimilaridades según las distintas aproximaciones antes mencionadas², a partir de las mismas se clasifican las unidades en forma jerárquica a través del algoritmo Ward y en forma no jerárquica a través de la clasificación fuzzy.

Se presentan a continuación los resultados obtenidos en la clasificación jerárquica según las distintas definiciones de disimilaridad seguidos de los resultados obtenidos en la clasificación fuzzy.

III.1. Clasificación crisp

En este tipo de procedimientos de clasificación cada objeto considerado (en este caso cada trayectoria temporal multivariada) es asignado a un grupo. Cuando se producen transformaciones importantes en el tiempo, este tipo de métodos basados en algoritmos en general de tipo jerárquicos carecen de la flexibilidad necesaria para describir los procesos de cambio que se han producido en el período, lo que tendrá repercusiones en los mecanismos de previsión.

² En el cálculo de disimilaridades se utilizó el proc IML del SAS. En los procedimientos de clasificación se usó tanto el SAS como el SPLUS.

En el cuadro siguiente se presentan las distintas unidades y los grupos a los que fueron asignadas de acuerdo a las distintas definiciones de disimilaridades; se toma como referencia los grupos definidos según la de Carlier (En el anexo se puede apreciar la descripción de cada uno de los grupos formados).

Comparación de la conformación de los clusters según tipo de distancia. Cluster Crisp.

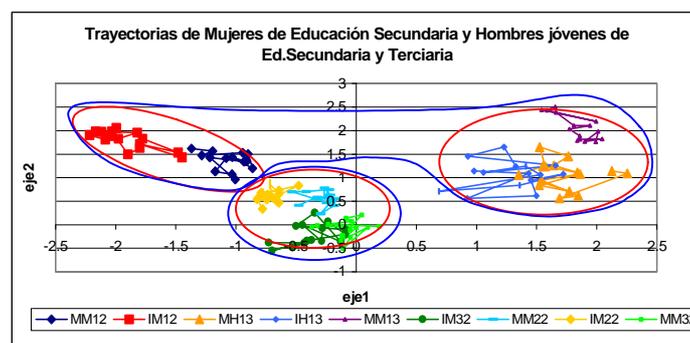
	Carlier	D'Urso-Vichi	Saporta-Lavellard	Trend
MH23	1	1	1	1
MH33	1	1	1	1
IH23	1	1	1	1
IH33	1	1	1	1
MM23	2	2	2	1
MM33	2	2	2	1
IM23	2	2	2	1
IM33	2	2	2	1
IM13	2	2	2	2
MM13	3	3	3	2
MH13	3	3	3	2
IH13	3	3	3	2
MM11	4	4	4	3
MM21	4	4	4	4
MM31	4	4	4	4
IM11	4	4	4	3
IM21	4	4	4	4
IM31	4	4	4	4
MM12	5	3	3	4
IM12	5	3	3	4
IH12	6	5	5	5
MH12	6	5	5	5
MM22	6	3	5	4
MM32	6	3	5	4
IM22	6	3	5	4
IM32	6	3	5	4
MH31	6	5	5	5
MH32	6	5	5	5
IH11	6	5	5	5
IH21	6	5	5	5
IH31	6	5	5	5
MH11	6	5	5	5
MH21	6	5	5	5
MH22	6	5	5	5
IH22	6	5	5	5
IH32	6	5	5	5

Existe una diferencia apreciable entre las estructuras de grupos encontradas bajo distancias de tipo multicomponent (Carlier, D'Urso-Vichi y Saporta-Lavellard) y las de tipo estructural (Trend).

Las de tipo *multicomponent* separan en los tres casos a las mujeres de Educación primaria de todas las edades, a los hombres mayores de 30 años de Educación terciaria y a las mujeres de Educación terciaria (con excepción de las mujeres jóvenes de Montevideo). Existen diferencias en la agrupación de los jóvenes de educación terciaria y de las mujeres jóvenes de Educación secundaria

En el gráfico siguiente se presentan las trayectorias en el plano principal de tres grupos de unidades (según Carlier) que equivalen a dos grupos bajo Saporta y se transforman en uno según D'Urso.

Clasificación realizada con las disimilaridades de Carlier, Saporta-Lavellard y D'Urso -Vichi.



Las unidades que tienen un comportamiento diferencial según Carlier son las *mujeres jóvenes de educación secundaria*. Estas presentan una inserción distinta en el mercado de trabajo con menor dedicación horaria que las mayores y en condiciones de mayor "inestabilidad" respecto a las mujeres y hombres de su misma edad pero de mayor nivel educativo (por un análisis más detallado de las trayectorias ver Bellagamba et al, 2000).

El seguimiento de las unidades y sus trayectorias en relación a la inserción en el mercado laboral (posición y movimiento en los distintos planos factoriales) refleja que las diferencias sustanciales entre unidades se verifican en los ejes 1 y 2 (Ver anexo).

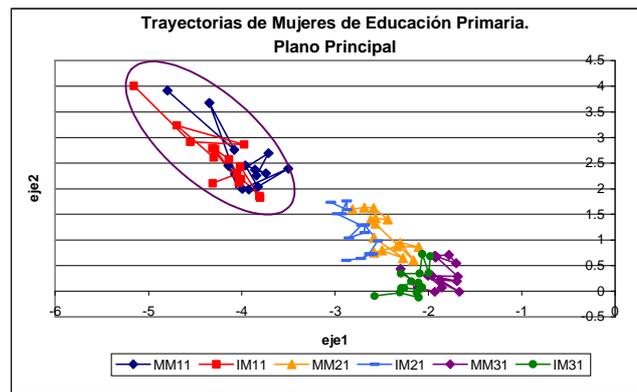
En cuanto a la comparación de las disimilaridades *estructurales* y *multicomponent* las diferencias más importantes se observan por un lado, en los grupos de mujeres de nivel educativo mas bajo y por otro en los mayores de 30 años de Educación superior.

Las disimilaridades *multicomponent* agrupan a las mujeres de Educación Primaria de todas las edades mientras que la del Trend separa a las más jóvenes.

Las mujeres jóvenes de Educación primaria son el grupo con peor inserción en el mercado laboral, mayores niveles de desempleo y mayor precariedad. En términos dinámicos su evolución es prácticamente nula en el período, en oposición, las mujeres de otras edades se desplazan hacia la derecha y hacia abajo en el plano principal, lo que estaría indicando una mejora en las condiciones laborales y un incremento en la dedicación horaria.

En el gráfico se presenta la trayectoria de las mujeres con educación primaria en el plano principal.

Clasificación realizada según las disimilaridades de Carlier, Saporta-Lavellard y D’Urso –Vichi versus Trend.



III.2. Clasificación Fuzzy

En los métodos fuzzy se asigna a cada observación un coeficiente u_{iv} que indica el grado de pertenencia de la observación i al grupo v . Dicho coeficiente debe satisfacer las siguientes condiciones:

1. $u_{iv} > 0 \quad \forall i = 1, \dots, n \quad v = 1, \dots, k$
2. $\sum_{v=1}^k u_{iv} = 1 \quad \forall i = 1, \dots, n$

La función objetivo a minimizar en este caso es:

$$\sum_{v=1}^k \frac{\sum_{i,j=1}^n u_{iv}^2 u_{jv}^2 d(i,j)}{2 \sum_{j=1}^n u_{jv}^2}$$

En caso que la clasificación sea totalmente difusa todos los $u_{iv} = 1/k$ mientras que para el caso de clasificación “crisp” se verificará $u_{iv} = 0$ ó 1 . (Por una descripción detallada del método ver Kaufman and Rousseeuw, 1990).

En el caso de trayectorias temporales multivariadas los grados de pertenencia indican diferente evolución de las unidades en el tiempo.

Se presentan en forma breve algunos de los resultados de la aplicación de la clasificación fuzzy en D’Urso-Vichi y en el Trend. Se observó la existencia de unidades con coeficientes de pertenencia difusos. Esto es, la unidad puede pertenecer a uno u otro grupo, reflejando un comportamiento temporal no homogéneo. Por ejemplo, si se considera la clasificación fuzzy según la disimilaridad de D’Urso- Vichi los *hombres jóvenes de Ed. secundaria* (IH12) tienen coeficiente de pertenencia casi idéntico para los grupos 1 y 2 (0.40 y 0.41 respectivamente) lo que los hace “ubicables” en cualquiera de ellos.

Clusters Fuzzy . Unidades cuyo coeficiente de pertenencia es “difuso”.

Disimilaridad D’Urso -Vichi						Disimilaridad Trend					
Coef.de pertenencia						Coef.de pertenencia					
	1	2	3	4	5		1	2	3	4	5
IH12	0.40	0.41	0.04	0.10	0.05	IH11	0.34	0.21	0.06	0.09	0.29
IH13	0.15	0.29	0.19	0.10	0.26	IM12	0.10	0.31	0.06	0.16	0.37
MH13	0.14	0.29	0.24	0.09	0.23	IM21	0.12	0.37	0.34	0.05	0.11
MM13	0.12	0.27	0.20	0.10	0.31	IM32	0.27	0.37	0.11	0.03	0.21

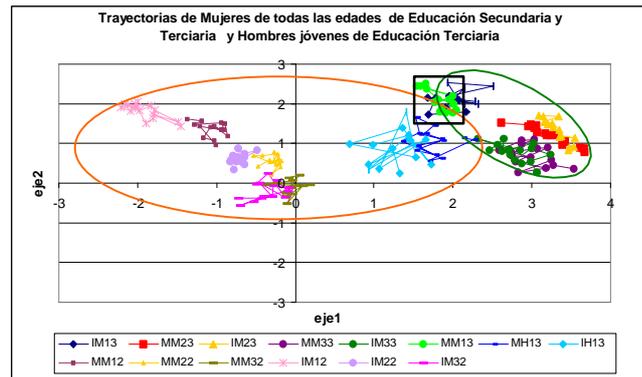
Según se aprecia en el cuadro anterior las *mujeres jóvenes de Educación Terciaria de Montevideo* (MM13) pueden ser clasificadas tanto en el grupo 5 (mujeres de nivel educativo alto) como en el grupo 2(mujeres de Educación Secundaria y hombres jóvenes de Educación Terciaria). En efecto, al observar la trayectoria de esta unidad podemos verificar que la misma presenta por períodos características similares a las de sus mayores (iguales en género y educación) y en otros períodos se asemeja a los hombres jóvenes de igual nivel educativo.

En términos generales comparando la trayectoria de los individuos de educación terciaria de ambos sexos, se constata que la mujer ha logrado una mejora relativa en el período, alcanzando una posición muy próxima a la situación de los hombres en términos de remuneraciones y estabilidad laboral, aunque manteniendo una dedicación horaria inferior.

A continuación se presenta en forma gráfica la comparación de clasificaciones crisp y fuzzy. Vemos en el plano principal dos agrupaciones que son prácticamente coincidentes en los dos tipos de clasificación utilizados bajo una misma definición de

disimilaridad, la diferencia radica en la unidad de pertenencia difusa antes mencionada (los grupos son delineados con un ovalo de distinto color mientras que la unidad “difusa” es marcada con un rectángulo). Es importante remarcar que el agrupamiento no se ve tan claramente si se siguen las unidades en los distintos planos factoriales (ver anexo).

Comparación de la clasificación crisp y fuzzy en las disimilaridades definidas por D'Urso -Vichi



IV. CONCLUSIONES

La importancia de agrupar y caracterizar trayectorias temporales multivariadas presenta dificultades importantes y es además de singular importancia dentro del ámbito de técnicas multiway como el Análisis Factorial Dinámico. En ella se construyen trayectorias optimizadas que tienen en cuenta múltiples variables todas ellas influyentes sobre un fenómeno que se desarrolla en el tiempo. La variabilidad de los resultados según el tipo de procedimientos que se utilice puede poner en evidencia diferentes aspectos de las trayectorias y a su vez tratar con mayor flexibilidad o rigidez estos procesos.

La utilización de técnicas fuzzy se demuestra sumamente útil para este tipo de análisis por la mayor flexibilidad frente a las transformaciones en el tiempo de las unidades o las variables consideradas.

En este artículo se han expuesto en modo sintético algunos de los primeros resultados obtenidos combinando la técnica de AFD con diversos procedimientos de clasificación. En la actualidad se están analizando el comportamiento de estos y otros métodos frente a la influencia de los datos mediante mecanismos de simulación y remuestreo.

REFERENCIAS

Bellagamba. A, Furtado, M, Nalbarte.L, (2001), *El mercado de trabajo uruguayo (1986-2000) descripción y análisis desde la perspectiva del análisis factorial dinámico*. Actualización de artículo publicado en Revista de Economía del BCU noviembre 2000.

Blanco J., Coppi R., Corazziari I. Con la collaborazione di Nalbarte L., *L'analisi fattoriale dinamica. Una tecnica per lo studio delle serie temporali multivariate*. In edizione. 2002.

Blanco J., Coppi R., Camaño G., Corazziari I. (1996), *Descomposición Factorial y regresiva de la variabilidad de un array a tres vías*. Actas. Seminario sobre datos temporales multivariados. Bogotá . Colombia . pp 355-392. Revista Quantum. Número 10. pp.81- 107. Uruguay.

Blanco J. Coppi I, Corazziari I.(1999), *Su alcuni sviluppi metodologici dell' Analisi Fattoriale Dinamica*. Dipartimento di Statistica, Probabilità e Statistiche Applicate. La Sapienza. Roma Serie ricerche. N° 13.

Coppi . R Zanella F. (1978), *L'analisi fattoriale di una serie temporale multipla relativo allo stesso insieme di unità statistiche* . Atti riunione della SIS pp. 61-67.

Coppi R. (1994), *An introduction to Multiway Data and their Analysis*. Computational Statistics and Data Analysis. Col 18 , n4 pp 3-13.

Corazziari I (1999), *Dynamic Factor Analysis, in Classification and Data Analysis. Theory and Application*. Springer – Verlag

D'Urso Pier Paolo (2000), *Classificazione fuzzy per matrici a tre vie temporali*. Tesi dottorato. Roma.

Saporta, G., Lavellard F (1996), *L'Analyse des Données Temporelles. Methodes et Applications*. Editions Technip, Paris..

Splus 4 (1997). *Guide to Statistics*, Data Analysis Products Division. MathSoft, Seattle

ANEXO

Clasificación jerárquica según las distintas distancias utilizadas.

Clusters Crisp formados según distancia Carlier.		
Cluster	unidades	descripción
1	MH23, MH33, IH23, IH33	Hombres de Educación terciaria mayores de 30 años
2	MM23, MM33, IM13, IM23, IM33	Mujeres de Educación terciaria
3	MH13, MM13, IH13	Jóvenes de Educación terciaria
4	MM11, MM21, MM31, IM11, IM21, IM31	Mujeres de Educación primaria
5	MM12, IM12	Mujeres jóvenes de Educación media
6	Resto	Hombres de Educación primaria y media y mujeres de educación media mayores de 30 años

Clusters Crisp formados según distancia D'Urso Vichi.		
Cluster	unidades	descripción
1	MH23, MH33, IH23, IH33	Hombres de Educación terciaria mayores de 30 años
2	MM23, MM33, IM13, IM23, IM33	Mujeres de Ed. terciaria (excepción jóvenes Montevideo).
3	MH13, IH13, MM12, IM12, MM13, MM22, IM22, MM32, IM32,	Hombres (Mont. e Int.) y Mujeres (mont) jóvenes de Ed. Terciaria. Mujeres de Ed. Secundaria
4	MM11, MM21, MM31, IM11, IM21, IM31	Mujeres de Educación Primaria
5	MH11, IH11, MH12, IH12, MH21, IH21, MH22, IH22, MH31, IH31, MH32, IH32	Hombres de Educación Primaria y Secundaria

Clusters Fuzzy formados según distancia D'Urso Vichi .		
Grupos	Unidades	descripción
1	MH11, IH11, MH21, IH21, MH31, IH31, MH22, IH22, MH32, IH32	Hombres de Ed. Primaria y Hombres mayores de 30 de Ed. Secundaria
2	MH12, IH12(1) , MH13(3) , IH13(3) , MM12, IM12, MM22, IM22, MM32, IM32	
3	MH23, IH23, MH33, IH33	Hombres mayores de 30 años de Ed. Terciaria
4	MM11, IM11, MM21, IM21, MM31, IM31	Mujeres de educación primaria
5	MM13(3) , IM13, MM23, IM23, MM33, IM33	Mujeres de educación terciaria

Clusters formados según distancia Saporta Lavellard.		
Cluster	unidades	descripción
1	MH23, IH23, MH33, IH33	Hombres de educación terciaria mayores de 30 años
2	IM13, MM23, IM23, MM33, IM33	Mujeres de Ed terciaria (excepción jóvenes Montevideo).
3	MM13, MH13, IH13, MM12, IM12	Jóvenes de Ed. Terciaria y Mujeres jóvenes de Ed. Secundaria.
4	MM11, MM21, MM31, IM11, IM21, IM31	Mujeres de Educación Primaria
5	Resto	Hombres de Ed. Primaria y Secundaria y Mujeres de educación media mayores de 30 años

Clusters Fuzzy formados según distancia Saporta Lavellard		
Grupos	Unidades	descripción
1	MH11, IH11, MH21, IH21, MH31, IH31, MH22, IH22, MH32, IH32	Hombres de ed. primaria y Hombres mayores de 30 de ed. secundaria
2	MH12, IH12, MH13, IH13, MM12, IM12, MM22, IM22, MM32, IM32	
3	MH23, IH23, MH33, IH33	Hombres mayores de 30 años de Ed. Terciaria
4	MM11, IM11, MM21, IM21, MM31, IM31	Mujeres de educación primaria
5	MM13, IM13, MM23, IM23, MM33, IM33	Mujeres de educación terciaria

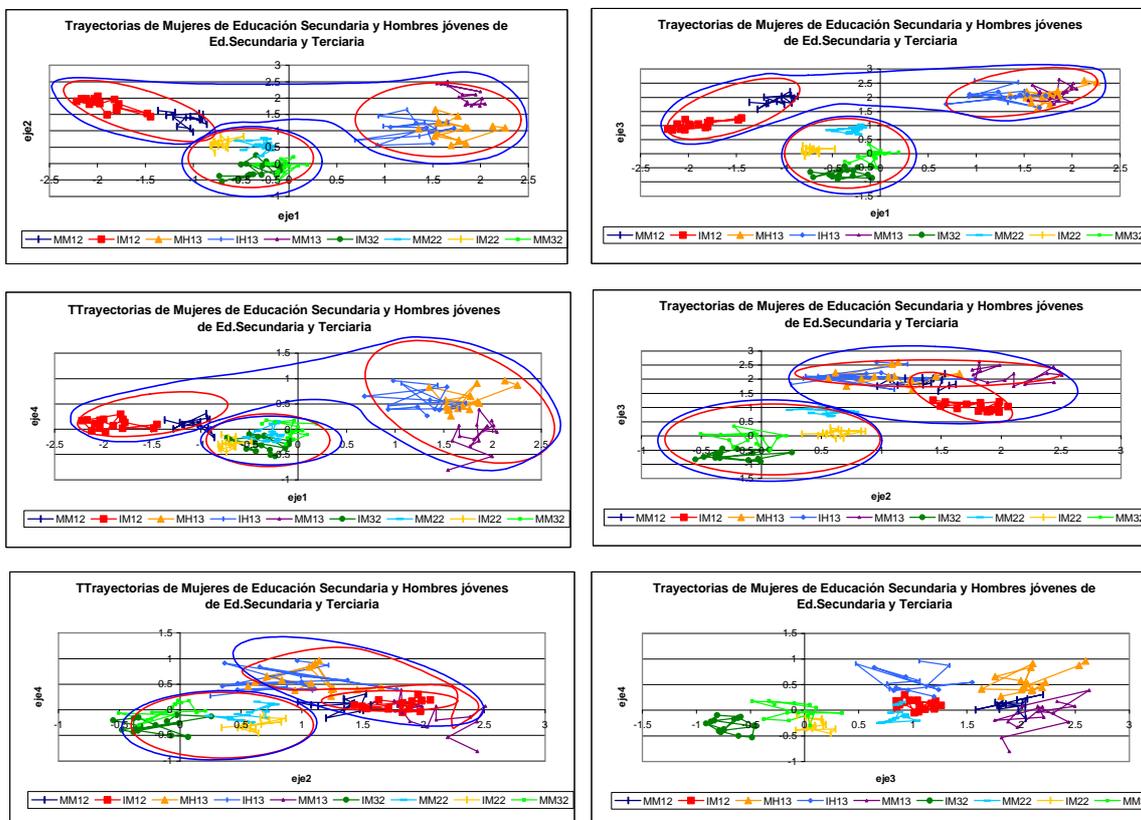
Clusters Crisp formados según distancia trend.

Cluster	UNIDADES	DESCRIPCIÓN
1	MH23, IH23, MM23, IM23, MH33, IH33, MM33, IM33	Hombres y Mujeres mayores de 30 años de Ed.Terciaria
2	MH13, MM13, IH13, IM13	Jóvenes de educación terciaria
3	MM11, IM11	Mujeres jóvenes de Educación primaria
4	MM21, IM21, MM31, IM31, MM12, IM12, MM22, IM22, MM32, IM32	Mujeres de Educación primaria (excepción de los jóvenes) y secundaria
5	MH11, IH11, MH21, IH21, MH31, IH31, MH12, IH12, MH22, IH22, MH32, IH32	Hombres de Educación Primaria y Secundaria.

Clusters Fuzzy formados según distancia Trend.

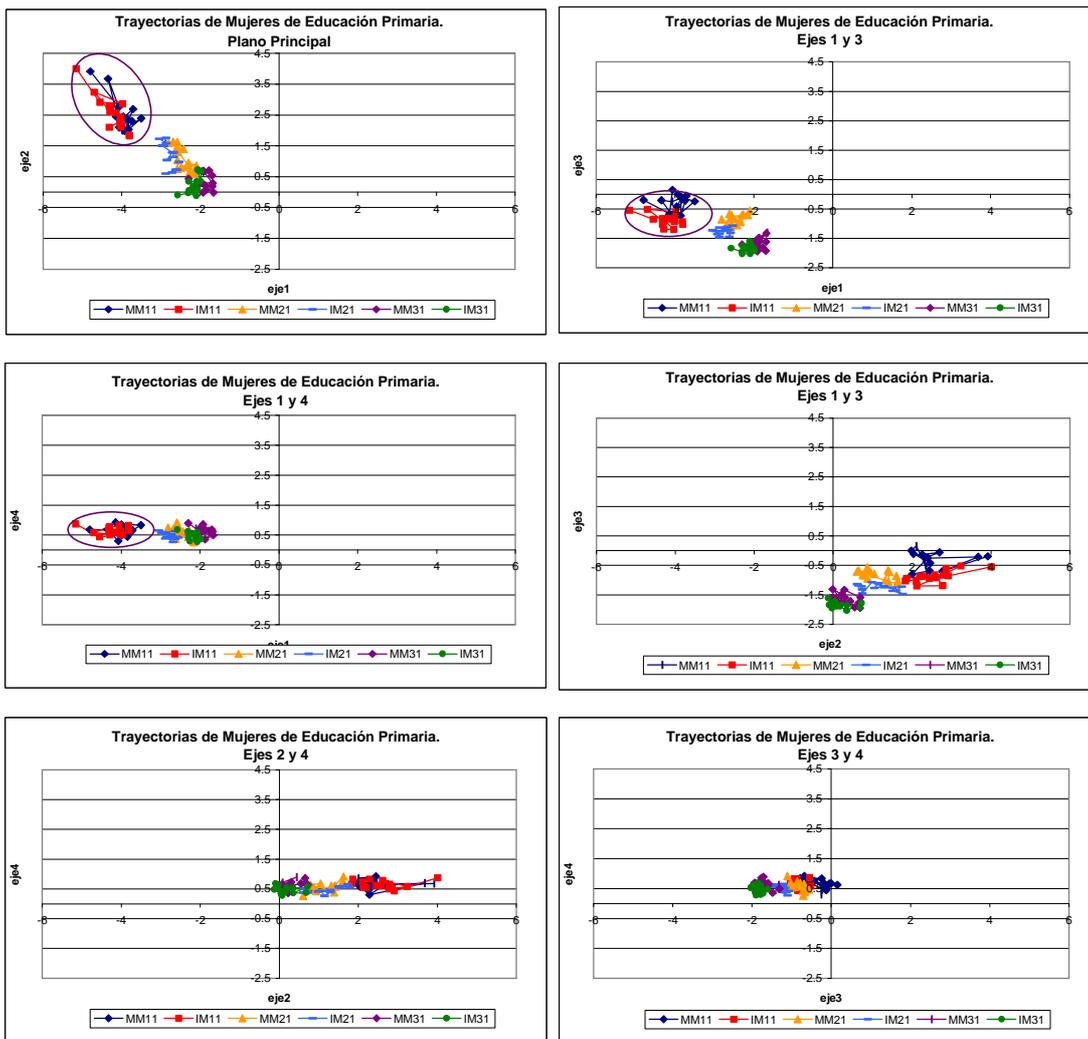
Grupo	unidades	descripción
1	MH11, IH11 (5) , IH12(2) , MH21, IH21, MH22, IH22, MH31, IH31, MH32, IH32	Hombres de Educación primaria y secundaria (excepción de los jóvenes de Montevideo)
2	MH12 (1) , MM12, MH13, IH13, MM13, IM13, MM22, IM22, MM32, IM32(1)	Mujeres mayores de 30 años y HyM jóvenes de Mont.de Ed. Secundaria y jóvenes de Ed. Terciaria
3	MH23, IH23, MH33, IH33, MM23, IM23, MM33, IM33	Hombres y Mujeres mayores de 30 años de Educación Terciaria
4	MM11, IM11	Mujeres jóvenes de Educación Primaria
5	IM12 (2) , MM21, IM21, MM31, IM31	Mujeres mayores de 30 años de Educación Primaria y Mujeres Jóvenes del Interior de Ed. Secundaria

Comparación de las distintas distancias en clasificación "crisp". Grupos según distancias de Carlier Saporta-Lavellard y D'Urso-Vichi



 grupos formados según distancia de Carlier
 grupos formados según distancia de Saporta-Lavellard

Grupos según distancias de Carlier Saporta-Lavellard y D'Urso-Vichi versus grupos según distancia del Trend.

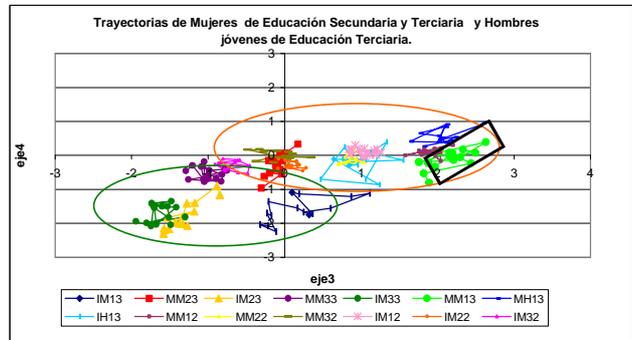
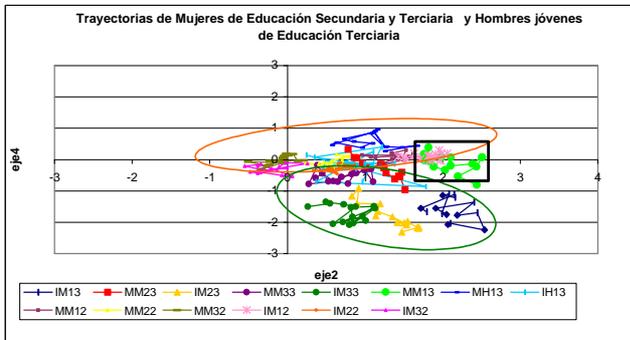
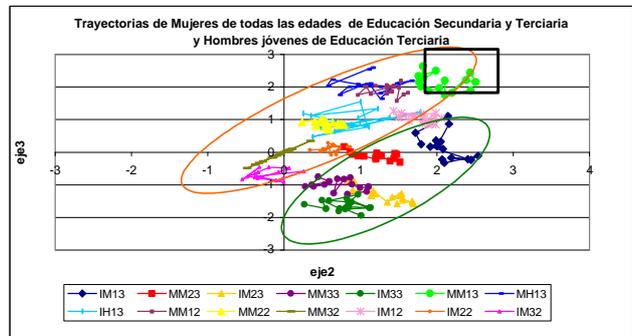
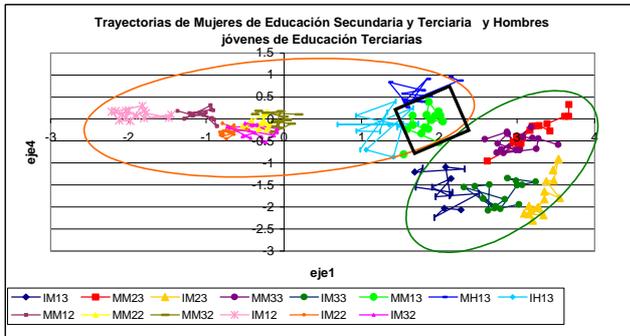
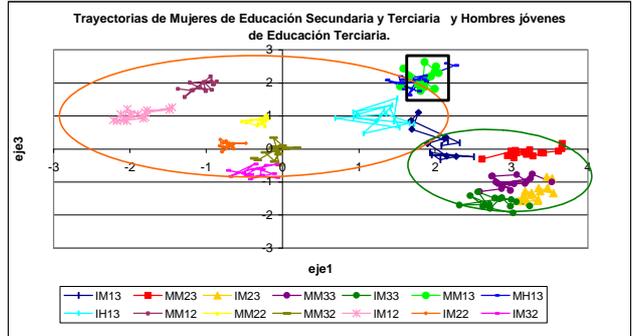
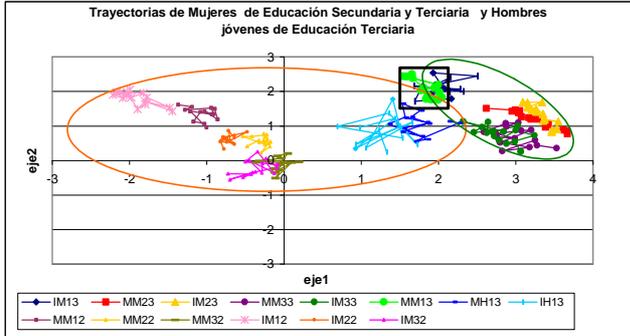


GRUPO CONFORMADO SEGÚN DISTANCIA DEL TREND

NOTA: Las mujeres de Educación Primaria conforman un único grupo según las distancias de Carlier, Saporta y D'Urso mientras que las mujeres jóvenes son separadas en un grupo aparte según la distancia del trend.

Clasificación crisp y clasificación fuzzy.

Algunos de los grupos formados según distancia D'Urso Vichi.



 CLUSTER "CRISP"

 CLUSTER "CRISP"

 UNIDAD QUE POSEE COEFICIENTE DE PERTENENCIA "DIFUSO" EN LA CLASIFICACIÓN FUZZY EL MAYOR COEFICIENTE DE PERTENENCIA LO POSEE EL GRUPO DE "EDUCACION TERCIARIA" (El de menor cantidad de integrantes)