



HAL
open science

Análisis estadístico de datos biográficos : métodos, ejemplos y perspectivas en el estudio de itinerarios migratorios

Olivier Barbary

► **To cite this version:**

Olivier Barbary. Análisis estadístico de datos biográficos : métodos, ejemplos y perspectivas en el estudio de itinerarios migratorios. Recolección y Análisis de Datos Longitudinales”, Bogotá, 9-13 décembre 1996, Olivier Barbary, Dec 1996, Bogota, Colombia. hal-02482583

HAL Id: hal-02482583

<https://hal.science/hal-02482583v1>

Submitted on 19 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ANÁLISIS ESTADÍSTICO DE DATOS BIOGRÁFICOS : METODOS, EJEMPLOS Y PERSPECTIVAS EN EL ESTUDIO DE ITINERARIOS MIGRATORIOS

BARBARY Olivier *

Resumen

El análisis armónico cualitativo, aunque reciente y perfectible, se consolida como una herramienta poderosa para la estadística descriptiva de los datos longitudinales. Partiendo de su aplicación a una encuesta sobre las trayectorias residenciales, profesionales y familiares de residentes en el área metropolitana de Bogotá, se muestra cómo este método, asociado con un proceso de clasificación automática, permite conseguir una tipología de la movilidad intra-urbana y relacionarla con los demás elementos de la biografía. Estos resultados descriptivos también son útiles cuando se quiere estimar modelos de duraciones de permanencia y sus factores determinantes.

*

* *

INTRODUCCIÓN

La reflexión teórica en ciencias sociales concede cada vez más importancia a los enfoques dinámicos, así como a la noción de estrategia de los individuos y de los grupos sociales. En este contexto se implementan nuevos métodos de recolección de la información y las encuestas retrospectivas sobre biografías individuales, basadas en muestreos representativos, toman importancia para alcanzar una observación más precisa del tiempo y de los espacios en los cuales transcurre la movilidad humana.

Pero mientras sabemos más precisamente por qué y cómo recolectar datos biográficos, subsisten muchas dificultades en cuanto a su análisis, particularmente en el campo de la estadística descriptiva. Cuando la prioridad es, en general, conseguir una tipología de las trayectorias individuales, el análisis se queda en muchos casos a un nivel transversal o en el manejo de indicadores longitudinales monovariados, necesariamente reductores de la información original. Lo que hace falta, entonces, es una herramienta de estadística exploratoria multivariada que nos permita respetar la riqueza y la continuidad de este tipo de datos.

* : Estadístico - ORSTOM, Departamento de matemáticas y estadística, Universidad Nacional de Colombia - Santafé de Bogotá D.C., Colombia. Email : obarbary@ciencias.ciencias.unal.edu.co

Los resultados de la investigación estadística sobre este problema son relativamente nuevos. Siguiendo los trabajos de Cox [1972] y apoyándose en la teoría de los procesos estocásticos, el análisis de los datos longitudinales ha conocido un desarrollo importante en el campo de la estadística inferencial y de la modelización (Courgeau y Lelievre [1989], Droesbeke, Fichet y Tassi [1989]). En cambio, la estadística descriptiva de este tipo de datos quedó durante mucho tiempo como “terra incógnita”. El análisis armónico cualitativo (AAC : Deville y Saporta [1980]), es una de las propuestas metodológicas recientes que, sin lugar a dudas, mejor se adapta a las necesidades de la fase descriptiva del análisis de los datos biográficos.

Es lo que pretendemos mostrar a partir de su aplicación a una encuesta sobre la movilidad espacial de las poblaciones de Bogotá, realizada en 1993 por un equipo de ORSTOM y de la Universidad de los Andes, y en la cual se intentó captar simultáneamente la biografía residencial, laboral y familiar de una muestra de 1031 individuos. El propósito de la primera fase de la explotación era identificar una tipología de la movilidad residencial dentro de la aglomeración urbana y relacionarla con los otros elementos biográficos observados. En la primera parte presentamos rápidamente el contexto de estudio, los datos de la encuesta y luego, con un poco más de detalles, los pasos de la implementación del AAC a estos datos. En la segunda parte, tratamos de ilustrar el interés de la metodología con algunos resultados : para mostrar cómo, gracias a las variables longitudinales ilustrativas, logramos evidenciar correspondencias entre los patrones de movilidad espacial y los demás componentes de la biografía; con este propósito, tomamos el ejemplo de la inserción residencial y laboral de los jóvenes migrantes.

1. DATOS Y MÉTODO

1.1. Los datos

Contexto y problemática de la encuesta

La encuesta y el trabajo estadístico presentado aquí se enmarcaron en una investigación emprendida en agosto de 1992 por un equipo de investigadores del Centro de Estudios sobre Desarrollo Económico (C.E. Florez) y de ORSTOM (F. Dureau y M.C. Hoyos), sobre la movilidad de las poblaciones de Bogotá y su impacto en la dinámica del área metropolitana. Paralelamente a este programa se desarrolló desde Febrero de 1994 una cooperación entre la Universidad Nacional de Colombia (J. Ramos, C.E. Pardo, C. Quintero y L.G. Díaz) y ORSTOM (O. Barbary) con el propósito de investigar los métodos de análisis estadístico de datos longitudinales.

Entre las capitales latinoamericana, Bogotá es la metrópoli que experimentó el más alto crecimiento demográfico durante las décadas de los cincuenta y sesenta (más de 6% anual). Entre 1951 y 1964 duplicó su población y alcanzó 2.5 millones de habitantes en 1970. Desde hace unos veinte años, el ritmo de crecimiento de la capital colombiana, al

igual que en las demás capitales del sub-continente, se ha venido desacelerando para llegar aproximadamente a 2.5% anual en el último censo de 1985. Al momento de la encuesta CEDE/ORSTOM (final de 1993), Bogotá tenía una población de cerca de 5.5 millones de habitantes y crecía a un ritmo ligeramente superior al 2% anual. Esta disminución de ritmo se explica por los efectos combinados de tres factores : la caída rápida de la fecundidad, un descenso en la intensidad de los flujos migratorios hacia la capital y la transformación del patrón de distribución geográfica de la población en beneficio de los municipios periféricos del área metropolitana. Dicho proceso de cambio demográfico se acompaña de nuevas estrategias de localización residencial, las cuales producen a su vez importantes modificaciones en el reparto de la población así como de las modalidades de segregación socio-espacial en la aglomeración urbana. Esta recomposición, que ha sido poco estudiada hasta el presente, constituye el centro de la problemática de la encuesta.

La muestra y el cuestionario

Por razones de costo, el diseño muestral no buscó garantizar la representatividad del conjunto del área metropolitana de Bogotá, sino solamente la observación confiable de 11 zonas de estudio (4 comunas o partes de comunas del área metropolitana y 7 barrios de Bogotá) escogidas a priori como típicas, y que tienen un valor heurístico para la problemática del programa (mapa 1). El cuestionario de la encuesta comprende, luego de un módulo de información socio-demográfica sobre el conjunto de los individuos de los hogares (1031 hogares en la muestra), una serie de calendarios cuyo propósito es registrar la biografía residencial, profesional y familiar de una submuestra. Dicha submuestra, llamada “muestra biográfica”, está compuesta con un individuo de más de 18 años por cada hogar encuestado (o sea 1031 personas). Su estructura es controlada por cuotas de sexo, edad, relación de parentesco con el jefe de hogar y estatus migratorio. Desde luego, las conclusiones de este análisis no son válidas sino al interior de este universo, para esa población particular y no para el conjunto de la población de Bogotá y de su área metropolitana.

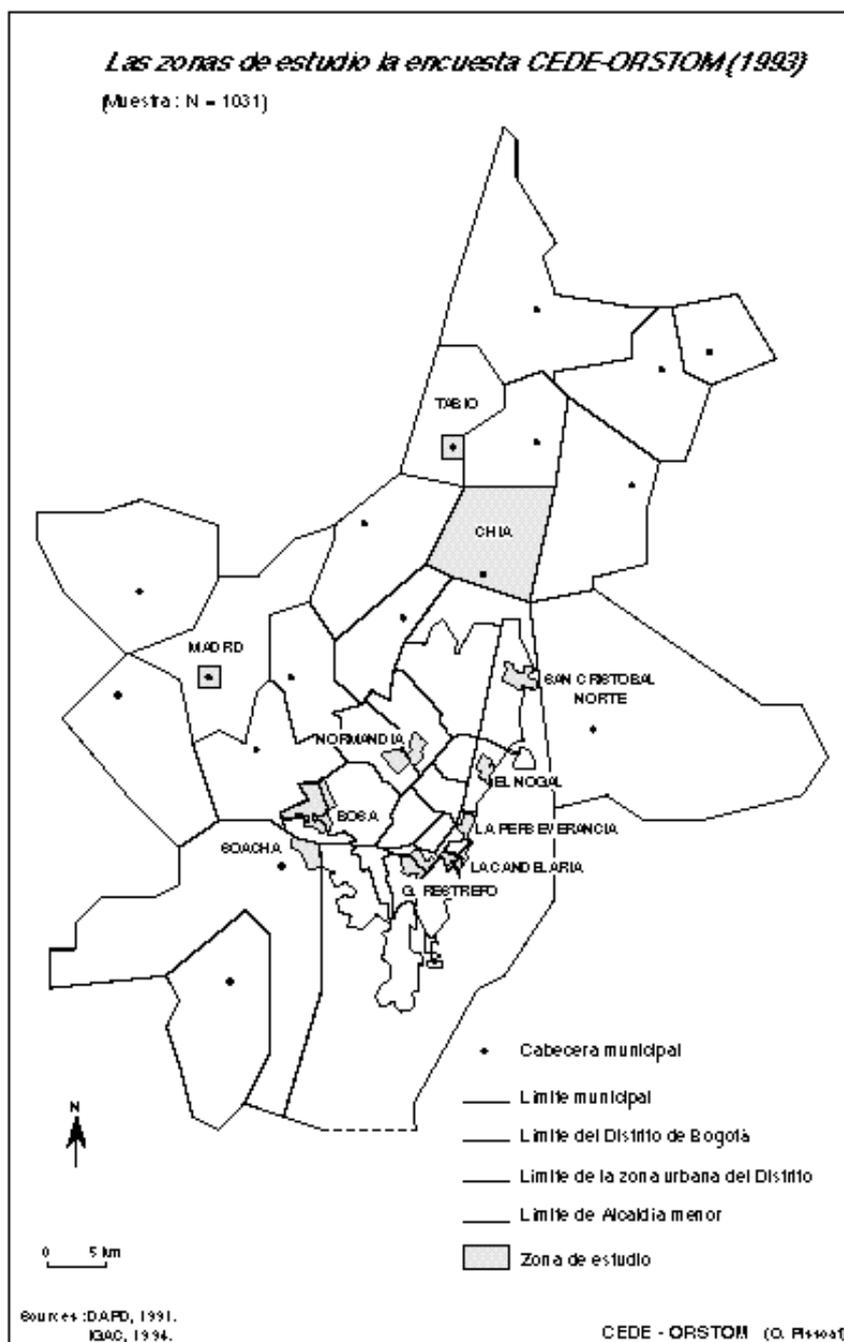
Las variables del análisis

Para lograr una tipología de los itinerarios residenciales en la ciudad, la variable longitudinal activa del análisis se basa en el conjunto de las etapas residenciales (con duración de un año o más) que nos han declarado los encuestados. El nivel de agregación geográfica de la variable debe suministrar la precisión máxima en el análisis de la movilidad espacial intraurbana, siempre y cuando conservemos un número suficiente de casos en cada modalidad. La nomenclatura que conviene mejor a este objetivo es aquella de las Alcaldías menores de Bogotá (19 unidades, ver mapa 1). A estas 19 modalidades se agrega una modalidad para las residencias dentro de la ciudad no especificadas, cuatro modalidades para las cuatro zonas de estudio de la periferia metropolitana (Chía, Tabio, Madrid y Soacha), una modalidad para los otros municipios del área metropolitana y, por último, una modalidad para las residencias situadas fuera del A.M. En total pues, la variable longitudinal activa cuenta con 26 modalidades.

Con fin de obtener una visión de conjunto de las relaciones existentes entre la trayectoria espacial y otros componentes de la biografía individual, introducimos 7 variables longitudinales ilustrativas. Ellas constituyen un resumen de cuatro capítulos de la biografía que podemos suponer determinados y/o determinantes con respecto a la trayectoria residencial: etapas del ciclo de vida (relación de parentesco con el jefe del hogar, estado matrimonial, coresidencia con los hijos, composición del hogar), acceso a la vivienda (condición de ocupación de la vivienda), carrera educativa (nivel educativo cursado) y movilidad socioprofesional (categoría socio-ocupacional). Cada una de estas variables longitudinales posee su propio “calendario”, independiente del número de etapas residenciales.

Finalmente, una caracterización socioeconómica de los individuos y de los hogares correspondiente a cada tipo de movilidad se obtiene con una serie de variables ilustrativas transversales. Se consideran primero las características socio-demográficas de los individuos : sexo, edad, status migratorio, nivel de instrucción, categoría socio-ocupacional, etc. Luego los hogares se caracterizan por su tamaño y composición, las características de la vivienda (número de cuartos, condición de ocupación, hacinamiento), y las del jefe del hogar (sexo, edad, status migratorio, edad promedio de los hijos).

MAPA 1:



1.2. El método de análisis tipológico

El análisis armónico cualitativo (AAC) fue presentado por primera vez en un artículo teórico por Deville y Saporta en 1979; luego más ampliamente por Deville, en un artículo de 1982. Con ésta proposición, el análisis factorial de correspondencias (AFC, Benzecri [1973]) encuentra un nuevo campo de aplicación: la estadística descriptiva y tipológica de los datos longitudinales. La presentación y justificación del método son objeto del texto de Gilbert Saporta en este volumen. Recordamos solamente de él una conclusión, fundamental en la práctica : se obtiene una aproximación bastante aceptable de la descomposición espectral del proceso original (análisis armónico de un proceso estocástico) aplicando el AFC a una codificación específica de los datos asincrónicos. Existen entonces varias soluciones; la más simple , adoptada en este trabajo, consiste en dividir el periodo de observación en un número razonable de intervalos (de duración constante o no), para luego calcular la proporción de tiempo que cada individuo pasó en cada uno de los estados de las variables longitudinales en el curso de cada intervalo. De esta manera se consigue una tabla de frecuencias caracterizando las trayectorias de los individuos, y podemos aplicarle el análisis de correspondencias.

Tal cual, según la expresión de Deville [1982], la técnica recibe sin lugar a dudas su “certificado de navegación”; en efecto su justificación teórica es sólida y su campo potencial de aplicación muy vasto. Sin embargo, en el caso del AAC la interpretación directa de los factores se vuelve a menudo compleja y a veces decepcionante cuando la información biográfica es detallada. Además, como lo veremos, una serie de parámetros se deben ajustar para que la codificación sea adaptada a cada problemática. Estas dificultades en la práctica del método, junto al hecho que no existe actualmente ningún paquete informático proponiendo su ejecución, explica probablemente el número limitado de aplicaciones... No obstante, se trata hoy en día del método de análisis descriptivo y tipológico de los procesos más poderoso y flexible que existe.

Con el fin de proponer una herramienta completa y de uso cómodo para el análisis de datos longitudinales basada en el AAC, hemos elaborado en SAS un programa de recodificación y de análisis de correspondencias, seguido por un proceso de clasificación automática, que permite una caracterización directa de los diferentes tipos de trayectorias. El programa efectúa, según la elección del usuario, diferentes tipos de recodificación de datos longitudinales para el análisis de correspondencias. El análisis factorial, la clasificación de los individuos en el espacio de factores y la caracterización de las clases puede hacerse también con SAS; pero para llevar a cabo una interpretación detallada nos pareció preferible utilizar el paquete especializado SPADN.

1.3. Los pasos de la recodificación y del análisis

El tiempo del análisis

El análisis se hizo según lo que se puede llamar el “tiempo biográfico individual”, es decir, siguiendo los individuos desde su nacimiento hasta la edad alcanzada en la fecha de la encuesta. Por supuesto, otras opciones son posibles, como el tiempo histórico (seguir a los individuos entre dos fechas) o un seguimiento a partir de un evento determinado (primera unión, primer nacimiento, etc.). Pero el tiempo biográfico es el más conveniente para nuestra problemática ya que permite relacionar, a nivel tanto individual como estadístico, los diferentes itinerarios (residenciales, familiares, educativos, profesionales), lo cual es indispensable para evidenciar la coherencia de las estrategias individuales.

El periodo de análisis

Una vez adoptada la opción del tiempo biográfico, conviene escoger un período de tiempo común a todos los individuos. Si se desea conservar la totalidad de las etapas, el individuo más viejo de la muestra (92 años) es quien fija el rango del período. Con el fin de evitar que el final de la tabla esté casi vacío, decidimos interrumpir el análisis a los 65 años de edad para la variable activa y a los 70 años para las variables ilustrativas, excepto la carrera educativa suspendida desde los 45 años.

Datos censurados a la derecha

La mayor parte de los individuos, ya que no han alcanzado ésta edad al momento de la encuesta, salen de observación antes del final del periodo de análisis (“censuramiento a la derecha” en la terminología del análisis longitudinal). Debemos anotar que el tipo de censuramiento depende del tipo de manejo de tiempo : con un tiempo histórico, por ejemplo, tendríamos censuras a la izquierda para los individuos que todavía no han nacido al principio del periodo. En este primer análisis, las censuras se trataron con una modalidad suplementaria para cada una de las variables en la que el individuo entra desde su edad de censuramiento. Esta opción, que permite conservar todos los individuos a lo largo de todo el análisis, puede criticarse en la medida en que las modalidades de censuramiento inciden en el resultado tipológico. Para limitar esta incidencia podríamos tratar estas modalidades como ilustrativas en el AFC, o también realizar análisis separados por cohortes de edades. No hemos adoptado ninguna de estas dos soluciones porque queremos una tipología global de la movilidad, es decir para el conjunto de la muestra y durante toda la duración de permanencia en el área metropolitana; en estas condiciones, nos pareció normal que un individuo de 20 años de edad no pueda pertenecer al mismo tipo que un individuo de 50 años.

Cortes en el tiempo y cálculo de frecuencias

Como se dijo en la presentación del método, la puesta en práctica del AAC reposa sobre el corte del período de análisis en un número “razonable” de intervalos de recodificación. Una vez realizada esta división, existen dos soluciones para construir la tabla de frecuencias :

- calcular las frecuencias de permanencia en cada estado en proporción a la duración total del análisis; en este caso la métrica sobre el tiempo es uniforme y la suma de cada fila es 1;
- calcular la frecuencia en proporción a la duración de cada intervalo de codificación : la métrica sobre el tiempo no es uniforme si los intervalos no tienen la misma duración, la suma de cada fila es igual al número de intervalos.

Para el cálculo algebraico, no se requiere que la métrica sobre el tiempo sea uniforme. De una parte unos cortes más detallados en los períodos donde se dan cambios de estado más numerosos proporcionan una mejor conservación de la información. De otra parte, para captar las estrategias residenciales, es natural interesarse más en los cambios de residencia que sobrevienen en la edad adulta, cuando la responsabilidad de la decisión pertenece en general al individuo, que en los cambios anteriores o posteriores, más frecuentemente decididos por terceros.

El examen de la frecuencia de los cambios de estado según la edad (figura 1) permite definir unos cortes adaptados a los datos en 15 intervalos de amplitud variable que corresponden aproximadamente a los cuantiles de la distribución; la precisión es buena entre 15 y 25 años, menor antes y después. La misma estrategia se aplicó a las variables longitudinales ilustrativas. Para el conjunto de las variables, las frecuencias son calculadas en proporción al tiempo de cada intervalo de recodificación. La tabla final sometida al análisis de correspondencia comprende 1031 filas, 398 columnas activas y 625 columnas ilustrativas.

Figura 1 : Distribución de los cambios de residencia (RESIALC) según la edad

RESIALC	Midpoint	Freq	Perct	Cum.	
				Perct	
0	****	8	0.28	0.28	
1	*****	53	1.86	2.14	
2	*****	44	1.54	3.68	1 : 0-5 años
3	*****	39	1.37	5.05	
4	*****	33	1.16	6.21	
5	*****	50	1.75	7.96	
6	*****	44	1.54	9.50	
7	*****	49	1.72	11.22	2 : 6-9 años
8	*****	53	1.86	13.08	
9	*****	58	2.03	15.11	
10	*****	69	2.42	17.53	
11	*****	42	1.47	19.00	3 : 10-12 años
12	*****	68	2.38	21.39	
13	*****	62	2.17	23.56	4 : 13-14 años
14	*****	77	2.70	26.26	
15	*****	102	3.58	29.84	5 : 15-16 años
16	*****	96	3.37	33.20	
17	*****	117	4.10	37.31	6 : 17-18 años
18	*****	120	4.21	41.51	
19	*****	112	3.93	45.44	7 : 19-20 años
20	*****	117	4.10	49.54	
21	*****	118	4.14	53.68	8 : 21 años
22	*****	111	3.89	57.57	9 : 22-23 años
23	*****	88	3.09	60.66	
24	*****	93	3.26	63.92	10 : 24-25 años
25	*****	83	2.91	66.83	
26	*****	86	3.02	69.85	
27	*****	76	2.66	72.51	11 : 26-28 años
28	*****	52	1.82	74.33	
29	*****	54	1.89	76.23	
30	*****	79	2.77	79.00	12 : 29-31 años
31	*****	42	1.47	80.47	
32	*****	47	1.65	82.12	
33	*****	44	1.54	83.66	13 : 32-35 años
34	*****	45	1.58	85.24	
35	*****	52	1.82	87.06	
36	*****	39	1.37	88.43	
37	*****	27	0.95	89.38	
38	*****	30	1.05	90.43	
39	*****	25	0.88	91.30	14 : 36-42 años
40	*****	32	1.12	92.43	
41	*****	18	0.63	93.06	
42	*****	16	0.56	93.62	
43	*****	18	0.63	94.25	
44	*****	22	0.77	95.02	15 : 43-65 años
45	*****	15	0.53	95.55	
etc...					

El análisis factorial

El análisis de correspondencias suministra un histograma de valores propios muy liso (figura 2). Esto no debe sorprender dada la estructura de la tabla : gran número de columnas respecto al número de filas y abundancia de celdas nulas. Hay que recordar aquí que la codificación, si bien conserva las duraciones de permanencia de los individuos en los diferentes lugares, sacrifica en cambio la cronología de las etapas, dado que, para el resultado del AFC, cualquier permutación de las columnas es indiferente. En otros términos, dos individuos con los mismos itinerarios, simplemente rezagados en uno o dos años, tienen perfiles codificados muy diferentes. La tabla analizada es entonces muy "ruidosa" con respecto a la estructura de proximidades que nos interesa y el histograma de valores propios refleja este ruido.

Figura 2 : Histograma de los primeros valores propios del AFC de la Tabla

```

-----
SPAD.N Sistema Portable para el Analisis de Datos
Copyright (C) CISIA, 1987, 1991 - Version 2.52
-----

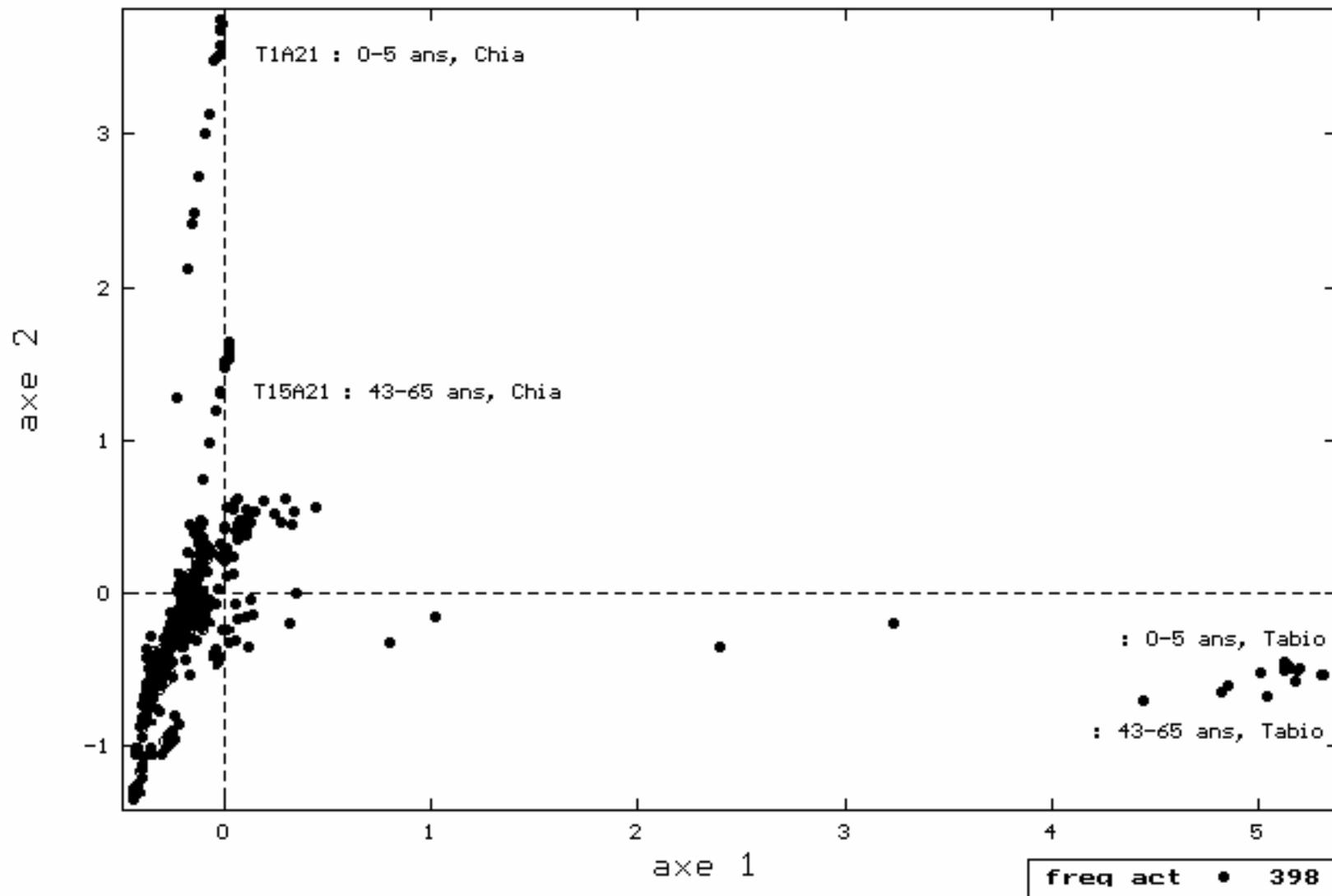
+---+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|NO |VALOR|PCENT|PCENT |
| |PROPI| |ACUMU.|
+---+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 1|.7099| 3.17| 3.17|*****
| 2|.6291| 2.81| 5.98|*****
| 3|.5973| 2.67| 8.65|*****
| 4|.5800| 2.59| 11.24|*****
| 5|.5562| 2.48| 13.72|*****
| 6|.5291| 2.36| 16.09|*****
| 7|.5046| 2.25| 18.34|*****
| 8|.4848| 2.17| 20.51|*****
| 9|.4834| 2.16| 22.67|*****
| 10|.4602| 2.06| 24.72|*****
| 11|.4562| 2.04| 26.76|*****
| 12|.4392| 1.96| 28.72|*****
| 13|.4266| 1.91| 30.63|*****
| 14|.4239| 1.89| 32.52|*****
| 15|.4197| 1.87| 34.40|*****
etc...

```

1031×398

A pesar de todo, la interpretación de los primeros factores es bastante fácil. De otra parte, si bien es cierto que con la codificación hemos perdido la cronología, esta meta-información estructura tan fuertemente los datos que resurge sobre todos los ejes factoriales útiles. Por ejemplo, los tres primeros factores son casi totalmente determinados por los municipios de Tabio, Chía y Madrid. El plano 1×2 (figura 3) muestra el tipo de estructura evidenciada por estos primeros ejes: las series de variables que representan un mismo lugar son agrupadas y ordenadas cronológicamente a lo largo de los ejes. Esto corresponde a unos subconjuntos de individuos minoritarios pero muy fuertemente caracterizados por cada grupo de variables. Se trata, naturalmente, de los subgrupos de la muestra estables durante toda la observación en Tabio, Chía y Madrid. A partir de los primeros factores del análisis, logramos sin dificultad la caracterización del conjunto de los grupos estables de la muestra. Estos individuos representan aproximadamente el 22% del conjunto de la muestra (ver más abajo el resultado de la clasificación) y alrededor de 52% de la inercia explicada por los diez primeros factores.

Figura 3 : Proyección de las variables longitudinales activas sobre el plano factorial 1x2



Los patrones de movilidad que caracterizan el resto de los individuos (el 78% de la muestra) se identifican menos fácilmente a partir de la interpretación de los factores. Un cierto número de ejes muestran asociaciones entre modalidades que significan transiciones frecuentes a una edad dada, entre diferentes lugares. No tenemos, como en el caso de la estabilidad, grupos de individuos con trayectorias globalmente idénticas, sino un grupo de individuos teniendo solamente en común una cierta transición a una cierta edad. Este tipo de resultado, aunque clásico en la aplicación del análisis armónico cualitativo a los datos de calendarios (J.C. Deville [1982], P. Béret [1988] [1995]), es insuficiente con respecto a nuestro objetivo tipológico.

Para conseguir una tipología completa de las trayectorias, la idea es proceder a la clasificación de los individuos en el espacio vectorial de los primeros factores del análisis¹. Veremos a continuación que se pueden considerar los diez primeros ejes como los factores útiles, totalizando así el 25% de la inercia total. Despreciar el 75% de la información original puede sorprender, pero recordemos que con 26 modalidades activas tenemos una tabla bastante dispersa donde una parte de la variabilidad (los pequeños desfases entre trayectorias equivalentes), aún si representa la mayor parte de la inercia, carece de interés para nuestro propósito. Finalmente, el análisis de correspondencias tiene como mérito principal -pero ya es mucho- eliminar la mayoría del ruido de la información bruta y proporcionarnos un espacio euclidiano para la clasificación de los individuos.

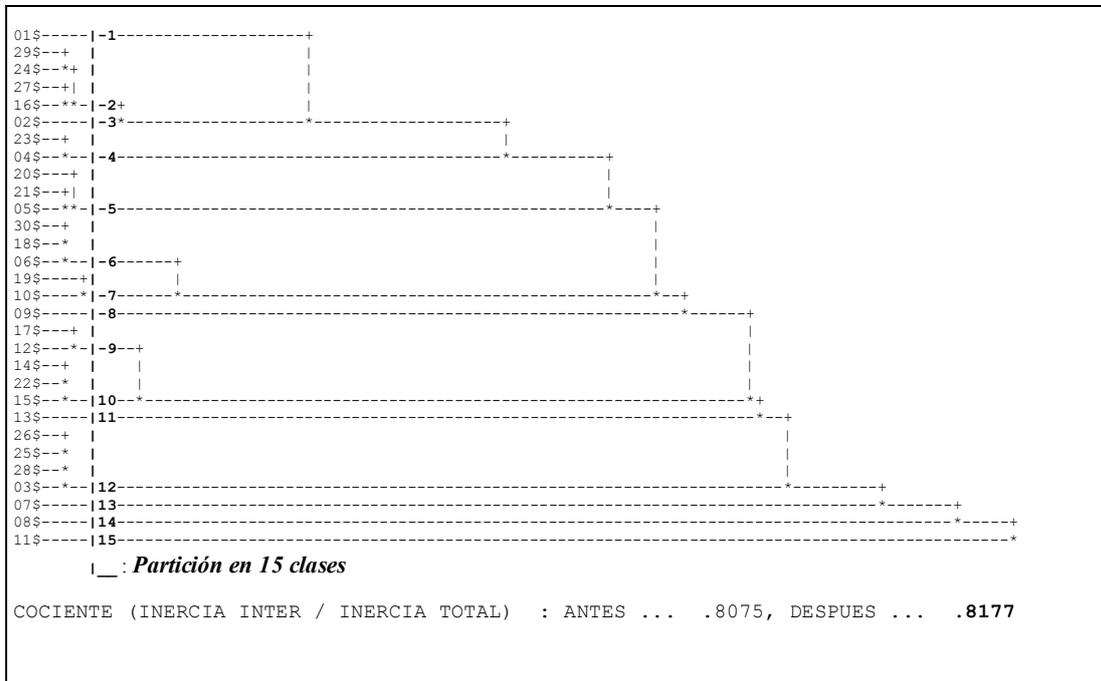
La clasificación automática

En este espacio la nube de individuos puede someterse a algoritmos de clasificación ascendente jerárquica o semijerárquica (CLUSTER - criterio de Ward - en SAS, SEMIS en SPADN). Ambos procedimientos dan resultados muy próximos y conservamos los de SPADN, ya que tienen la ventaja de optimizar la partición una vez escogido el número de clases. Después de varios ensayos, haciendo variar el número de factores, constatamos que, hasta 10 factores, la tipología gana en precisión, luego el tamaño de la clase más importante aumenta. Con esta base empírica hemos conservado el árbol de clasificación en el espacio de los diez primeros factores (figura 4).

La partición en 15 clases constituye la tipología que vamos a describir. Se obtiene cortando el árbol de clasificación y optimizando la clasificación por asignación de los individuos a la clase de centro más próximo, maximizando así la inercia interclase (PROC PARTI de SPADN). Dicha clasificación conserva el 82% de la inercia total de la nube en el espacio de los diez primeros factores. Por lo tanto, podemos concluir que la metodología permite captar la mayor parte de la información significativa (es decir excluyendo el ruido) que nos proporciona la encuesta acerca de la heterogeneidad de las trayectorias residenciales.

¹ : Son los mismos pasos que aplican A. Degenne, M. O. Lebeaux, L. Mounier [1995] a los datos sobre inserción profesional de los jóvenes en Francia.

Figura 4 : Parte superior del árbol de clasificación en el espacio de los diez primeros factores



Caracterización de las clases y otros resultados

Hasta ahora hemos seguido los pasos clásicos del análisis tipológico. Pero debido al carácter longitudinal de los datos, el trabajo de caracterización de las clases se vuelve un tanto particular. En efecto, aunque siempre es posible describir la tipología con una interpretación “indirecta” de las clases a partir de los factores, es sin lugar a dudas más directo, objetivo y enriquecedor, volver a la tabla de frecuencias codificada. Asignando a cada individuo de la muestra su clase de pertenencia y calculando para cada clase las frecuencias promedias de permanencia de los individuos en los estados durante los intervalos de codificación, obtenemos los perfiles de movilidad de las clases y podemos traducirlos gráficamente (gráficas 1 a 15 del anexo). Con la misma técnica, calculamos los perfiles de clase correspondientes a las variables longitudinales ilustrativas; estos perfiles son comparables entre ellos y al perfil del conjunto de la muestra ² (gráficas 16 a 18). Por último, se puede producir la serie de tablas cruzadas que relaciona la tipología con las características transversales de los individuos y de los hogares. El grado de “asociación” de las clases con cada característica se mide con la prueba de ###² asociada a la tabla, correlaciones de rangos o regresiones logísticas.

² : Para permitir estas comparaciones basta con eliminar el efecto de las estructuras por edades diferentes en cada clase, directamente traducido por la importancia de la modalidad de censuramiento a medida que se progresa en la edad. Para lograr esto se calculan las frecuencias para cada edad entre los individuos de la clase que alcanzaron esta edad (es decir excluyendo los individuos censurados). Así el total de cada columna (periodos de edad) queda normalizado al 100%.

Este material sirve de base para la descripción del patrón de comportamiento residencial de cada clase y del itinerario espacial de los encuestados. A partir de las variables longitudinal ilustrativas, podemos relacionar esta trayectoria con la sucesión de los eventos del ciclo de vida. Apuntamos también las características demográficas y socioeconómicas de los individuos y de los hogares, completando así la “cédula de identidad” de la clase. Se desprenden, entonces, hipótesis acerca de las interacciones que estructuran la biografía de los individuos, que podemos someter a verificación.

Por último, algunos análisis adicionales pueden complementar estos resultados descriptivos :

- análisis armónicos cualitativos en subconjuntos de la muestra para detallar la tipología;
- estimaciones no paramétricas de las duraciones de permanencia en diferentes estados;
- modelos paramétricos y semiparamétricos que evidenciarán los determinantes de estas duraciones.

2. DOS EJEMPLOS DE RESULTADOS

2.1 Un esquema global de la movilidad espacial

Vista en su conjunto, la tipología en 15 clases refleja dos fenómenos principales: la movilidad intra-urbana entre alcaldías menores y la migración procedente del exterior del área metropolitana - más precisamente, la edad y el lugar de llegada de los migrantes al área metropolitana. Si se considera los dos indicadores sintéticos de la tabla 1 (Imobalc - col. 8- y Pctmig -col. 5-), la correlación con el rango de las clases es evidente; podemos entonces distinguir 4 grupos:

1. Muy estables (8% de los individuos, gráficas 1 a 3 del anexo)

Este grupo conformado de las clases con la movilidad intra-urbana más baja se ubica en su totalidad fuera de Bogotá: estables en Chía (clase 14), estables en Madrid (clase 13) y estables en Tabio (clase 15). Encontramos lógicamente una gran mayoría de nativos del área metropolitana en las tres clases, pero, sin embargo, los migrantes no están totalmente ausentes (35% en la clase 14).

2. Estables (18% de los individuos, gráficas 4 a 8)

Un poco más móviles que las del grupo precedente, las clases 9, 10, 11 y 12 están constituidas por núcleos de población estable en diversos barrios de Bogotá : Usaqué (clase 10), Rafael Uribe y Antonio Nariño (clase 11), La Candelaria y Santa Fé (clase 12), a los cuales se agrega la clase 2 compuesta por migrantes jóvenes estabilizados en Madrid y la clase 9, igualmente joven, estabilizada en Usaqué.

3. Movilidad intra-urbana media (44% de los individuos gráficas 9 a 13)

El grupo más importante de la tipología está formado por individuos que han tenido alrededor de dos cambios de alcaldía desde su entrada al área metropolitana y se caracterizan, entonces, por estadías largas en ciertos barrios, sin que por lo tanto las residencias en la fecha de la encuesta estén muy concentradas. El grupo contiene, en orden creciente de movilidad, las siguientes clases:

- clase 8 : estadías largas en otros municipios del área metropolitana
- clase 7 : estadías largas en la Candelaria y Teusaquillo
- clase 5 : estadías largas en Rafael Uribe
- clase 4 : estadías largas en la periferia Oeste
- clase 1 : viejos migrantes en todos los barrios

4. Movilidad intra-urbana fuerte (30% de los individuos, gráficas 14 y 15)

El grupo más móvil comprende la clase 6, compuesta de viejos migrantes llegados jóvenes y nativos que han tenido una fuerte movilidad en los barrios del peri-centro norte y de la periferia Norte y Oeste; y la clase 3 que reagrupa a los jóvenes migrantes llegados entre los 10 y 20 años, cuyos itinerarios se ubican frecuentemente en la periferia Oeste y Sur. En la clase 3 las residencias en la fecha de la encuesta están bastante concentradas: Bosa, Soacha y Gustavo Restrepo.

Tabla 1 : Indicadores de movilidad promedia para las quince clases de la tipología

CLASES (1)	NINDIV (2)	PESO (3)	PCTPESO (4)	PCTMIG (5)	DURMED (6)	IMOBRES (7)	IMOBALC (8)
1	217	36156	22,32	95,9	21,3	12,7	7,86
2	25	2326	1,44	69,0	21,8	9,32	4,02
3	196	36557	22,57	83,3	15,8	18,5	14,23
4	72	13890	8,58	39,9	27,6	9,57	7,08
5	65	14706	9,08	30,1	30,0	9,33	6,67
6	89	11889	7,34	54,7	38,6	14,0	11,05
7	32	3039	1,88	12,2	47,4	8,69	6,13
8	34	4244	2,62	24,8	40,8	9,16	6,00
9	37	6515	4,02	40,3	23,6	10,2	5,56
10	25	4294	2,65	11,3	34,8	5,91	2,83
11	26	6157	3,80	25,8	35,6	4,98	3,15
12	90	9486	5,86	19,2	39,3	7,23	3,60
13	39	3529	2,18	13,6	29,7	5,20	1,56
14	48	7956	4,91	35,2	29,2	7,44	1,16
15	36	1234	0,76	6,6	43,2	5,41	2,17
Total	1031	161978	100	58,4	26,4	10,9	7,23

(1) : Rango de la clase en la tipología

(2) : Número de individuos de la muestra biográfica

(3) : Población expandida de la clase

(4) : Porcentaje de la población total expandida

(5) : Porcentaje de individuos nacidos fuera del a.m.

(6) : Duración promedia de residencia en el a.m.

(7) : Frecuencia de cambios de viv. en el a.m. (100 x número de cambios de viv. / número de años de residencia en el a.m.)

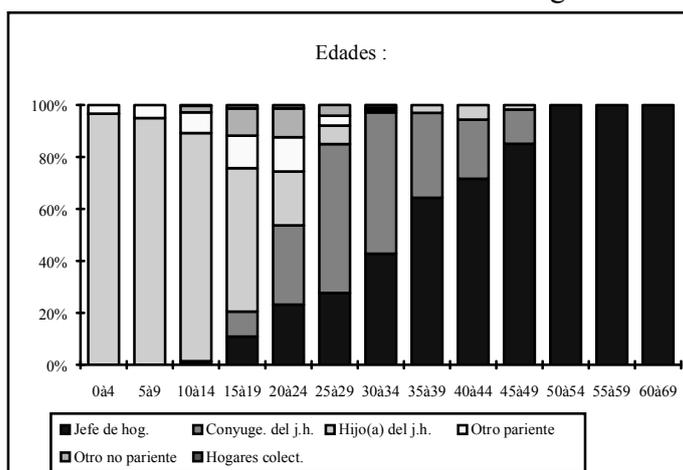
(8) : Frecuencia de los cambios de alcald. en el a.m. (100 x número de cambios de alc. / número de años de resid. en el a.m.)

2.2 Algunos elementos sobre la inserción de los jóvenes migrantes en Bogotá

2.2.1. Modalidades de inserción residencial al llegar a la ciudad

En la clase 3 (jóvenes migrantes llegados en Bogotá entre los 10 y 25 años de edad), la frecuencia de los status “otros parientes” y “otros no parientes” del jefe del hogar entre 15 y 24 años, demuestra la importancia de varias formas de hospedaje en el momento de la llegada a la aglomeración urbana.

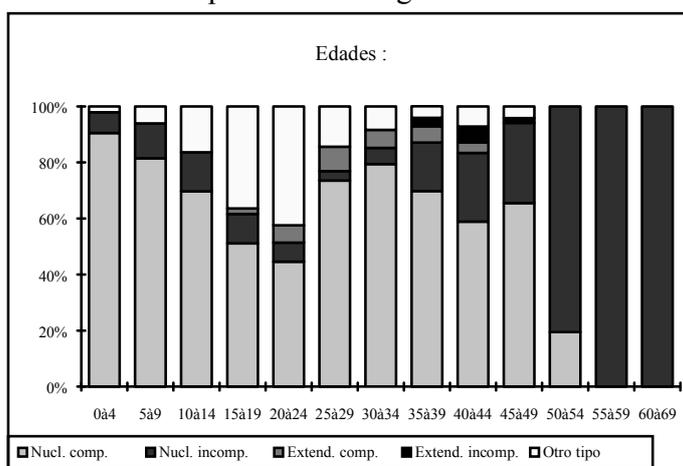
Gráfica 16 : condición del individuo en el hogar



El hecho destacable en el perfil longitudinal de la relación de parentesco con el jefe del hogar (gráfica 16) es la salida muy rápida del hogar de los padres. Pero, sobre todo, estas salidas no siempre corresponden a entradas directas a la condición de jefe del hogar o de cónyuge : entre los 15 y 24 años, las frecuencias de los otros parientes y otros no parientes del jefe del hogar alcanzan el doble de los valores observados en toda la muestra. Además, la

“circulación” es rápida en estas situaciones de transición : así, el 54% de los individuos de la clase experimentó un episodio de este tipo en algún momento, siendo el hecho mucho más frecuente en el caso de las mujeres (65%) que en el de los hombres (37%). Con tal nivel de frecuencia, estas formas de hospedaje aparecen como una de las modalidades principales de inserción residencial de los jóvenes migrantes a su llegada a Bogotá.

Gráfica 17 : composición del hogar



El perfil de la composición de los hogares en los cuales viven los jóvenes migrantes (gráfica 17) también muestra particularidades al momento de la llegada al área metropolitana : entre 10 y 25 años es impactante la frecuencia de los “otros tipos” de hogar. Esta modalidad reagrupa los hogares unipersonales y los hogares sin núcleo familiar al nivel del jefe del hogar. Allí de nuevo, a

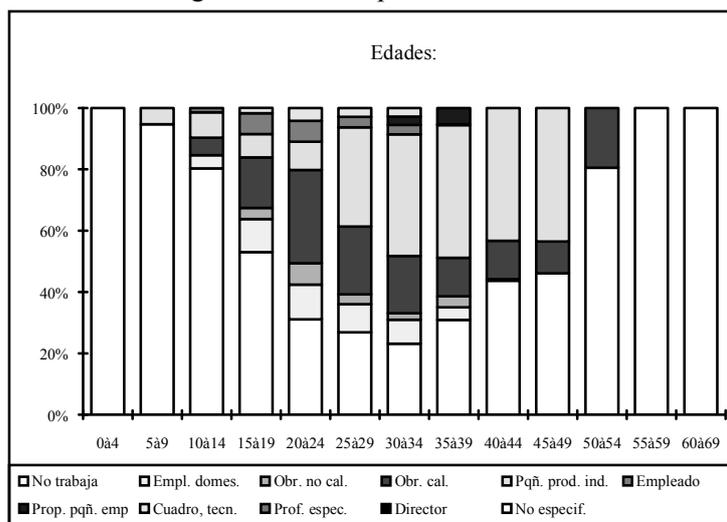
causa de la fuerte “circulación”, la frecuencia real de estos episodios es superior a lo que hace aparecer la gráfica : conciernen el 72% de los individuos de la clase. Esta vez la frecuencia es mayor en el caso de los hombres (84%, contra 64% de las mujeres). Si volvemos a los datos individuales, aparece que para las mujeres el 93% de los episodios de vida en hogares sin núcleo familiar tienen lugar bajo las condiciones de “otro pariente” o “otro no pariente” del jefe de hogar (hospedajes), mientras que, para los hombres, una fracción importante (30.4%) corresponde a una condición de jefe de hogar en un hogar “no familiar” (unipersonal o de individuos emparentados colaterales o no emparentados). La misma situación no se encuentra sino para el 7% de las mujeres.

Así, la inserción residencial de los jóvenes migrantes entre los 15 y 30 años de edad se caracteriza por el hospedaje y la vida en hogares no familiares, aunque las duraciones de permanencia en estos estados sean generalmente cortas. Por una parte este patrón aparece bastante distinto de aquel que se da, a las mismas edades, para los otros individuos de la muestra. Por otra parte existe un fuerte diferencial según el género, con hospedajes muy frecuentes para las mujeres mientras los hombre experimentan casi sistemáticamente episodios en hogares unipersonales o no familiares. Ambos perfiles nos confirman que estas situaciones son eminentemente transitorias, ya que a partir de los 25 años se observa muy rápidamente el acceso a las condiciones de jefes de hogares y cónyugues (gráfica 16), así como la formación de hogares nucleares completos (gráfica 17).

2.2 Una movilidad profesional fuerte y particular

Los jóvenes migrantes se distinguen también por sus formas de inserción laboral (gráfica 18), en particular a causa de una movilidad específica entre empleos asalariados e independientes.

Gráfica 18 : categorías socio-ocupacionales



La distribución de los empleos no se aleja mucho del patrón global hasta los 24 años, apenas se nota una proporción un poco superior de empleados domésticos (11% entre 15 y 24 años vs. 8% global). Después de los 25 años, en cambio, el perfil se destaca por la proporción muy superior al promedio de pequeños productores independientes (sin equivalentes en la tipología) :

44% de los empleos ocupados entre 25 y 29 años (vs. 19% global) y hasta 77% entre 40 y 44 años (vs. 33%). Dado que se trata de una clase que contiene a cerca del 23% de la muestra, estamos en presencia de un hecho de importancia mayor. Volviendo a los datos individuales constatamos que su explicación reside en una movilidad entre el empleo asalariado y el independiente particularmente fuerte y concentrada en el tiempo, ameritando que se detalle su mecanismo.

En el conjunto de los individuos que han ejercido un empleo en la fecha de la encuesta (84% de la clase), el 35.3% ha ocupado al menos un empleo independiente, pero sus itinerarios profesionales son de dos tipos distintos. Para el 22% de ellos, dicha actividad ha sido el primer empleo y, en este caso, todos la iniciaron antes de los 28 años y 90% antes de los 25 años. Los otros (13%) acceden a un empleo independiente después de otros empleos, y esta transición se da para 21% de ellos antes de los 22 años y para 74% entre los 22 y los 25 años (solamente 5% después de los 25 años), lo que causa la modificación marcada del perfil de actividad de la clase a esta edad. Notemos que la transición inversa (pequeños productores hacia otros empleos) no es menos frecuente : 15% de los individuos que han tenido un empleo en la fecha de la encuesta la han experimentado, pero todos antes de los 22 años. Eso contribuye también a explicar el perfil longitudinal, limitando las densidades de presencia en los empleos independientes antes de los 24 años. Así, no es propiamente el saldo de la movilidad entre los empleos de pequeños productores y los otros lo que explica la recomposición observada (en la clase este saldo es más bien favorable a los otros empleos). El fenómeno se debe al desfase en el tiempo de los dos tipos de movilidad : las salidas de los empleos de pequeños productores hacia otros empleos ocurren entre los 15 y los 22 años mientras que, en la mayoría de los casos, las entradas en estos empleos viniendo de otros empleos se dan entre los 22 y los 25 años. Queda por saber si el tipo de actividad y el ingreso procurado son equivalentes en ambos casos.

CONCLUSIÓN

Cuando se plantea su aplicación a datos biográficos complejos, el análisis armónico cualitativo presenta, tanto a nivel teórico como práctico, una serie de ventajas importantes.

Primero el manejo del tiempo biográfico individual es flexible y permite varias estrategias con el fin de conservar de la mejor forma la información continua y asincrónica original, y de ajustar la estrategia de codificación a la problemática del análisis. Se pueden, en efecto, adoptar otras lógicas de división del periodo de análisis distintas de la que se empleó aquí. Por ejemplo, la del tiempo histórico, donde los límites de los intervalos de recodificación son escogidos en fechas que se consideran determinantes en el contexto conyuntural, político o social del periodo. El análisis toma entonces un giro más histórico, lo que puede proporcionar otro punto de vista interesante sobre los datos (véase por ejemplo O. Barbary [1993], pp 30-64). Una

tercera posibilidad consiste en un tiempo biográfico más colectivo, donde la sincronización no se hace según la edad de los individuos, sino según el tiempo transcurrido después (o antes) de un acontecimiento clave del ciclo de vida : primera llegada de los migrantes al lugar de la encuesta, primera salida del hogar de los padres, primera unión, primer acceso a la propiedad de la vivienda etc... Varios análisis adicionales de este tipo serían seguramente de gran interés para la problemática de la encuesta CEDE/ORSTOM.

Otro punto es la cuestión de la división óptima del periodo de análisis que corresponde a cada una de estas lógicas de manejo del tiempo. Esta pregunta se puede plantear en un sentido teórico : optimización, con un cierto criterio estadístico (conservación de la cantidad de información), del número y de los límites de los intervalos de codificación del proceso, lo que abre espacios para investigaciones futuras. Quedándonos en el terreno empírico, hemos visto que no hay obligación de que los intervalos de codificación sean de duración constante o la métrica sea uniforme en el tiempo. Estos dos parámetros son importantes para ajustar el método a la estructura de los datos y la problemática de la encuesta.

Para realizar el análisis multivariado de diferentes tipos de eventos biográficos, el análisis armónico cualitativo ofrece dos posibilidades.

La primera, que no hemos adoptado aquí, consiste en analizar la variable compleja resultante del cruce de las diferentes situaciones residenciales, profesionales, familiares etc. Parece atractiva ya que agrupa los distintos temas en la variable activa y, desde luego, les da el mismo peso en el resultado tipológico. Sin embargo, no se debe creer que un buen uso del método es el que consiste en mezclar todas las informaciones biográficas disponibles y esperar del cálculo algebraico que evidencie el conjunto de relaciones que estructuran los datos. Existen limitantes de varios órdenes a esta práctica. Una condición "epistemológica" primero. Aún cuando el enfoque del análisis de datos tiene vocación exploratoria, su empleo nunca dispensa de una problemática bien definida cuando se analiza una encuesta. Por otra parte, desde el punto de vista de la fiabilidad estadística, el tamaño de la muestra no permite generalmente aumentar, más allá de un cierto número, las modalidades de los estados biográficos (este límite existe igualmente para la división del periodo de análisis). Finalmente, siendo la metodología muy reciente en su desarrollo, hace falta todavía la acumulación de experiencia en la interpretación de este tipo de resultados.

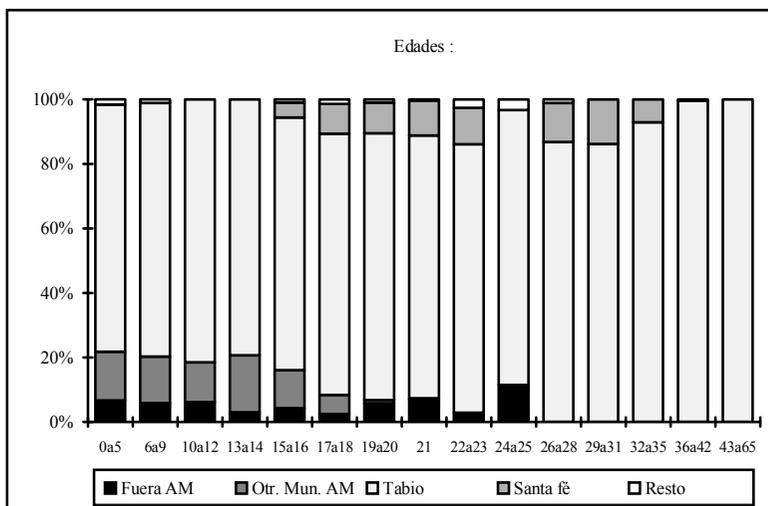
La segunda, la técnica de las variables longitudinales ilustrativas, nos parece la mejor alternativa para el análisis descriptivo multivariado de diferentes itinerarios. En primer lugar la prioridad que se da a la variable activa es afirmación del propósito experimental asignado al análisis. En este marco, la interpretación de los resultados encuentra un principio director que lleva al estadístico, si se esfuerza un poco, a formular hipótesis y/o conclusiones precisas e interesantes para sus interlocutores de ciencias sociales. También permite mantener normas de representatividad aceptables con

muestras que no pueden ser muy grandes (la recolección de biografías resulta compleja y costosa!).

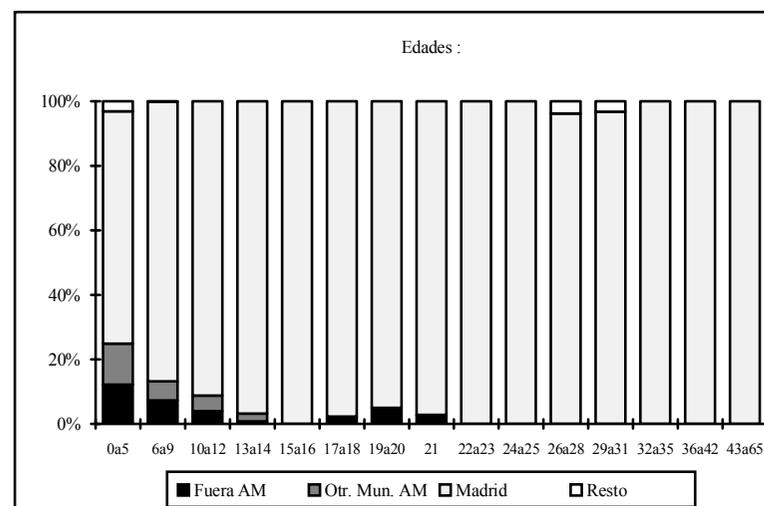
Para finalizar, cabe resaltar que, además de los resultados descriptivos que logra evidenciar, el método presenta también un interés evidente en una perspectiva de modelización. Por una parte, el resultado tipológico en si es fundamental para reducir la heterogeneidad interna de la muestra cuando se busca estimar directamente el efecto de variables independientes. Por otra parte, cuando el número de casos en la clase es suficiente, los perfiles de movilidad sugieren a menudo las formas y los parámetros de las distribuciones de los tiempos de permanencia en los estados biográficos; estas hipótesis pueden ser confirmadas luego, eliminando el efecto de los datos censurados, con la estimación no paramétrica de las “funciones de sobrevida” (Kaplan Meier). Este material empírico, cuya ausencia es una crítica frecuentemente dirigida a los enfoques de estadística inferencial en ciencias sociales, será de gran utilidad en la construcción de modelos paramétricos o semi-paramétricos.

Anexo : Los grupos estables y estabilizados en los municipios periféricos (Tabio, Chía, Madrid)

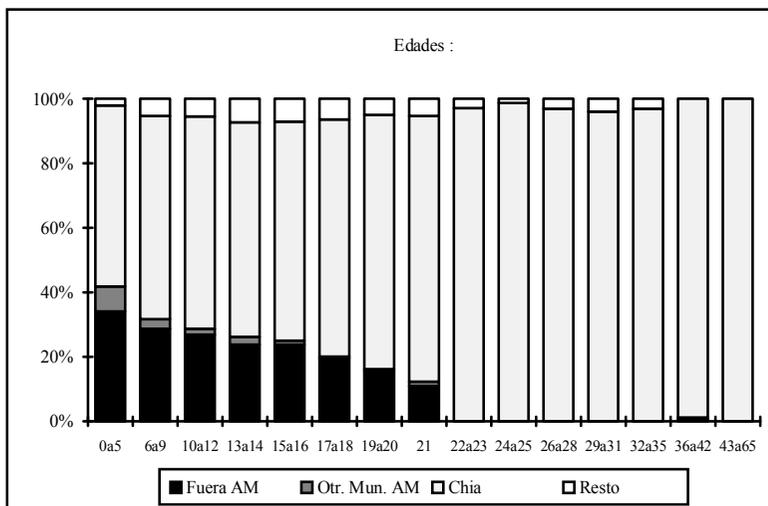
Gráfica 1 : Clase 15, estable en Tabio



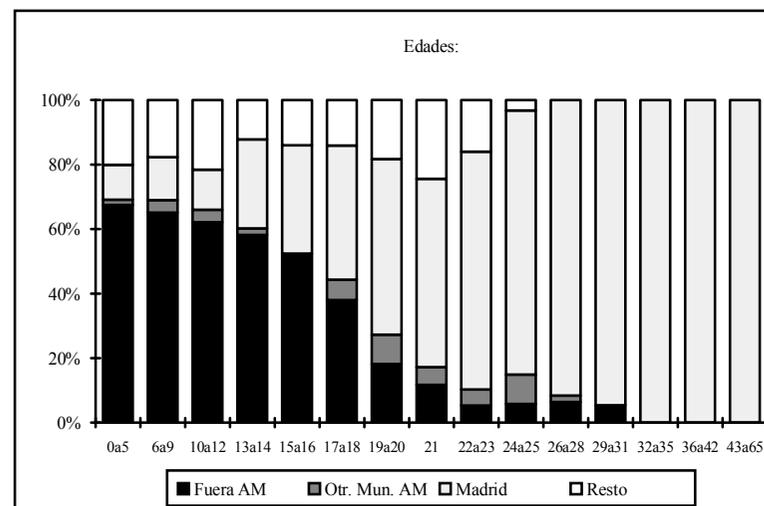
Gráfica 3 : Clase 13, estable en Madrid



Gráfica 2 : Clase 14, estable en Chía

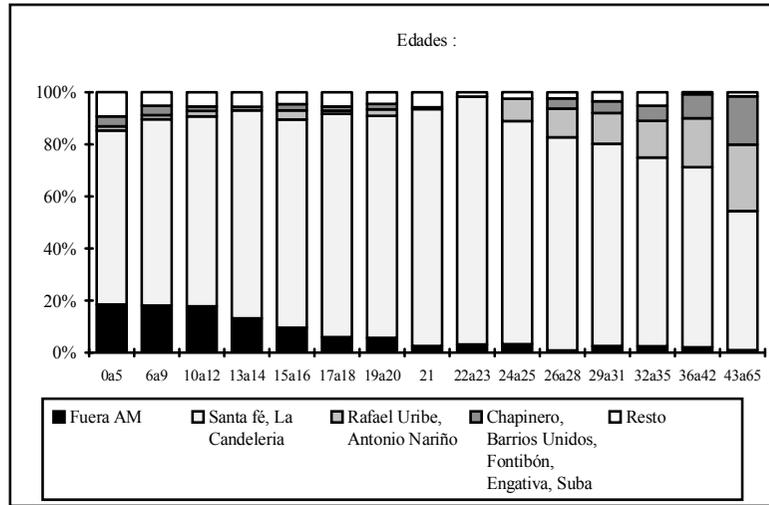


Gráfica 4 : Clase 2, estabilizada en Madrid

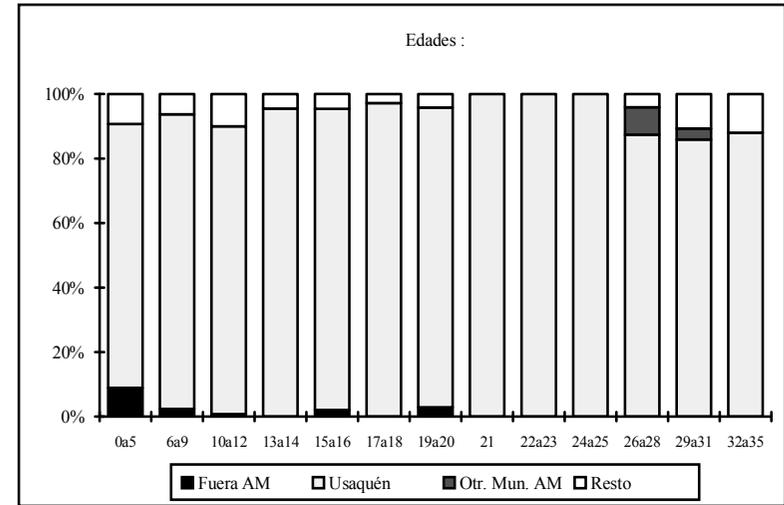


Anexo : Los grupos estables y estabilizados en Bogotá

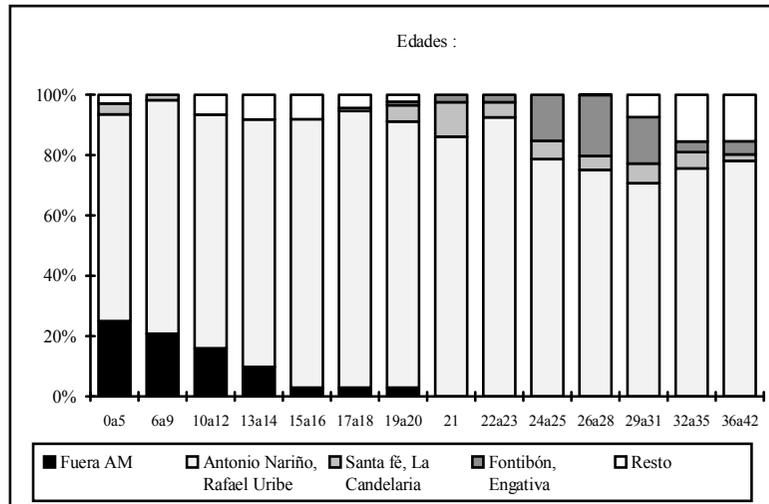
Gráfica 5 : Clase 12, estable en Santa Fé y La Candelaria



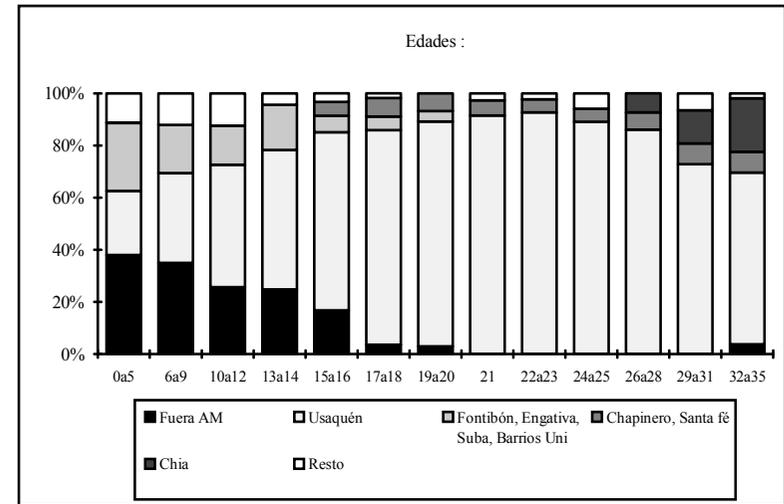
Gráfica 7 : Clase 10, estable en Usaqué



Gráfica 6 : Clase 11, estable en Antonio Nariño y Rafael Uribe

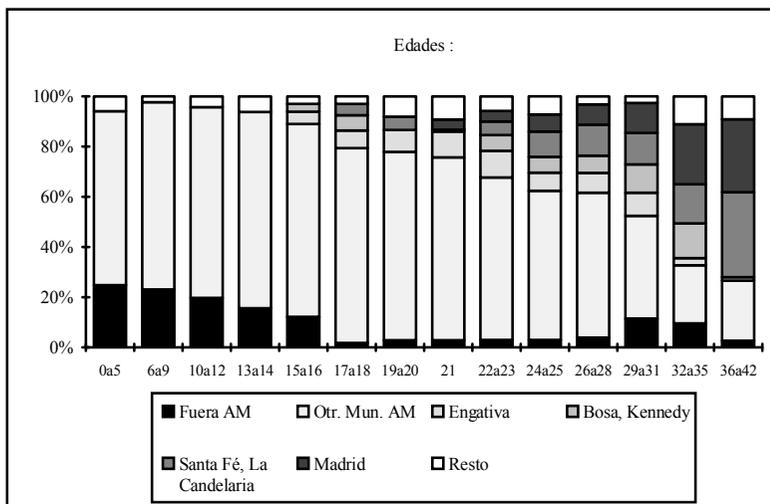


Gráfica 8 : Clase 9, estabilizada en Usaqué

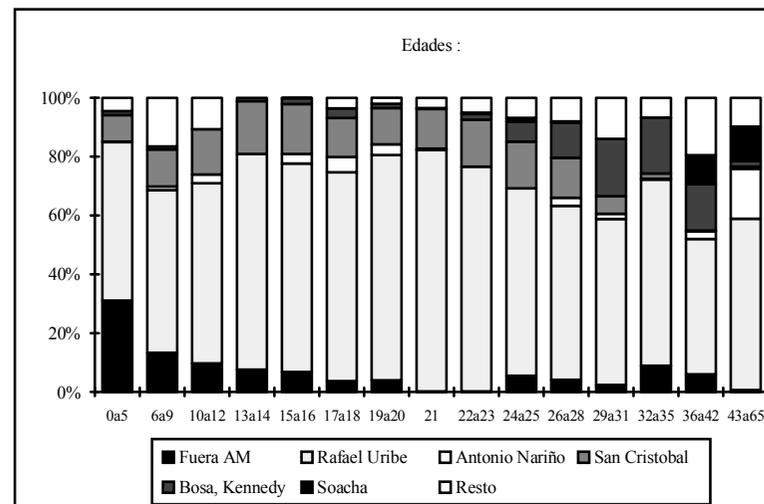


Anexo : Los grupos con movilidad intra-urbana media

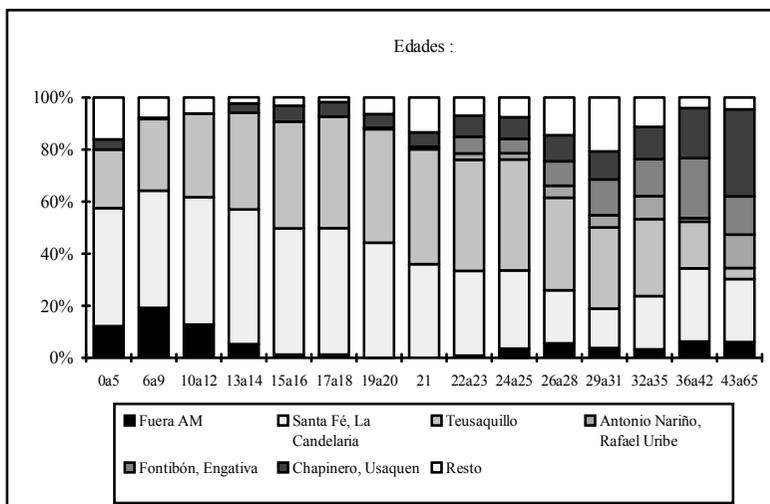
Gráfica 9 : Clase 8, estadías en otros municipios del a.m.



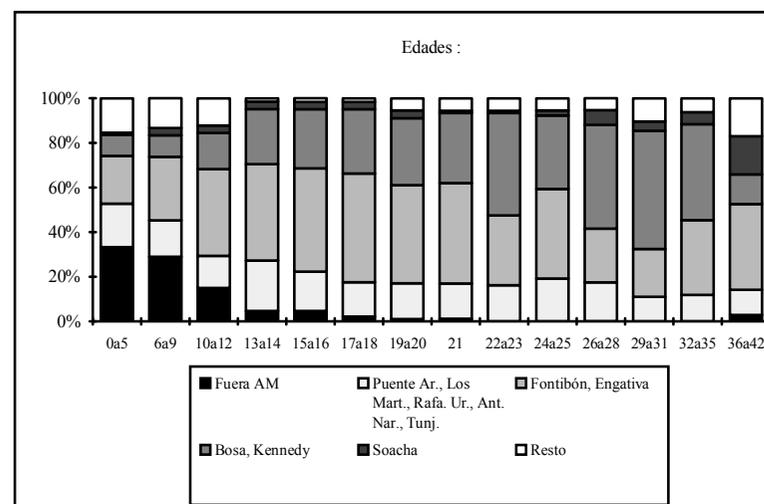
Gráfica 11 : Clase 5, estadías en Rafael Uribe y San Cristóbal



Gráfica 10 : Clase 7, estadías en La Candelaria y Teusaquillo

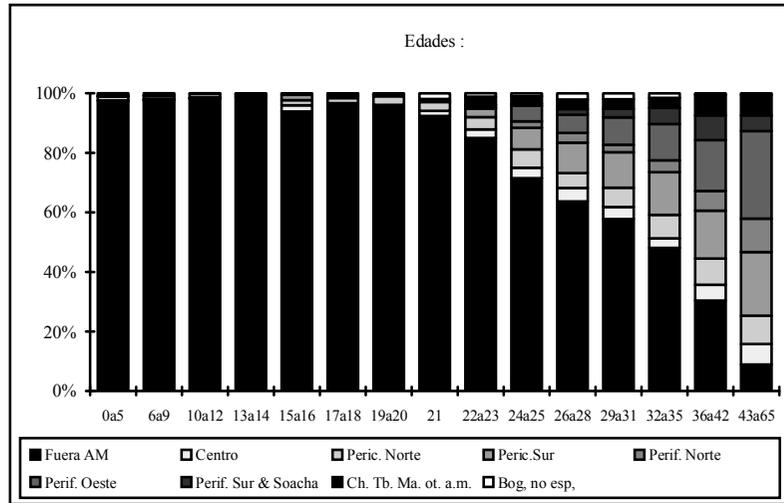


Gráfica 12 : Clase 4, estadías en la periferia Oeste

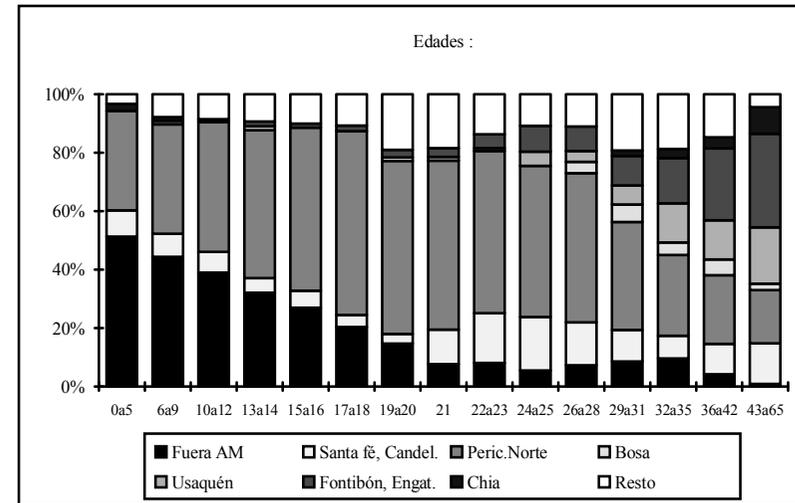


Anexo : Los migrantes y los grupos con movilidad intra-urbana fuerte

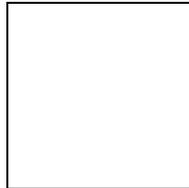
Gráfica 13 : Clase 1, migrantes mayores llegados entre los 20 y 40 años de edad



Gráfica 14 : Clase 6, migrantes mayores llegados antes de los 20 años de edad



Gráfica 15 : Clase 3, migrantes jóvenes llegados entre los 10 y 25 años de edad



NOTAS BIBLIOGRÁFICAS :

- ARIÑO M., ROSAS M.E., TORRADO S., [1987] : Metodología para el análisis de la estructura socio-ocupacional argentina (1960-1980). Consejo Federal de Inversiones, Documentos de la Secretaria General, Buenos Aires, 31p.
- BARBARY O. (1993) avec la collaboration de E.H.A. DIOP, "L'insertion urbaine : le cas de Dakar, rapport d'exploitation d'enquête", Dakar, 213 p. multigr.
- BENZECRI J.P. et al. [1973], "L'analyse des données. Tome 2 : L'analyse des correspondances" DUNOD, Paris, 632 p.
- BERET P. [1988] : Analyse de données chronologiques relatives à l'insertion professionnelle, in "Les Cahiers de l'Analyse des Données" vol. XIII, n° 2, pp. 159-174.
- COURGEAU D., LELIEVRE E. [1989] : Analyse démographique des biographies, éditions de l'INED, Paris, 268 p.
- COX D.B. [1972] : Regression models and life tables (with discussion), in "Journal of Royal Statistical Society", B 34, pp. 187-220.
- DEGENNE A., LEBEAUX M.O., MOUNIER L. [1995] : Construction d'une typologie de trajectoires à partir de l'enquête de suivi des jeunes des niveaux V, Vbis et VI. Communication aux Journées CEREQ-LASMAS-IDL sur les données longitudinales dans l'analyse du marché du travail, Caen, 28-29 juin 1995, 14 p.
- DEVILLE J.C., SAPORTA G. [1979] : Analyse harmonique qualitative, in "Data Analysis and Informatics", E. DIDAY et al. éditeurs, North Holland Publishing Compagny, pp. 375-389.
- DEVILLE J.C. [1982] : Analyse des données chronologiques qualitatives, comment analyser les calendriers ? "Annales de l'INSEE", n° 45, pp 45-104.
- DROESBEKE J.J., FICHET B., TASSI Ph. éditeurs [1989] : Analyse statistique des durées de vie, modélisation des données censurées, ECONOMICA, Paris, 282 p.
- DUREAU F., BARBARY O., FLOREZ C.E. [1994] : "La movilidad de las poblaciones y su impacto sobre la dinámica del área metropolitana de Bogotá", resultados preliminares de la encuesta cuantitativa. Documento de trabajo N° 3, CEDE/ORSTOM, Bogotá, 309 p.
- FLOREZ C.E, MENDEZ R., [1994] : Boletín de estadísticas sobre hogar y familia en Colombia, 1972-1992. Informe final, Universidad de los Andes, CEDE, Bogotá, 43p.
- VAN DER HEIJDEN P. G. M. [1987] : Correspondence analysis of longitudinal categorical data, DSWO PRESS, Leiden.