

METODOS ESTADISTICOS PARA EL ANALISIS DE DATOS RETROSPECTIVOS SOBRE LA MIGRACION.

Ejemplos de aplicaciones en Quito y revisión metodológica

Olivier BARBARY
Estadístico, ORSTOM

Esta ponencia se dedica al examen metodológico de diferentes técnicas estadísticas y sus aplicaciones al análisis de los datos biográficos acerca de los itinerarios residenciales de los migrantes hacia ciudades. Más precisamente vamos a presentar heramientas estadísticas que pueden aplicarse al procesamiento de los datos recolectados en las encuestas retrospectivas que se realizan en el lugar de destino de los migrantes. El trabajo se desarrolló en el marco de un programa de investigación metodológica y comparativa a partir de dos encuestas de dicho tipo que realizó el ORSTOM en QUITO (ECUADOR) en 1987 y DAKAR (SENEGAL) en 1989. Obviamente estos ensayos de avances metodológicos tienen lugar en un marco problemático y temático de análisis del crecimiento urbano en los países en desarrollo, lo cual ya a sido objeto de muchas presentaciones y debates a lo largo del seminario. Solo queremos aquí hacer resaltar los dos objetivos temáticos principales que tenemos de vista en las aplicaciones; F. DUREAU les formula así :

- *"La medida y caracterización de los flujos de población tanto hacia las ciudades como al interior de ellas : número de inmigrantes, características generales de ellos (sexo, edad, nivel de instrucción ...) orígenes y trayectorias espaciales."* En este primer enfoque cuantitativo, que debe llevar al conocimiento de los grandes componentes de la migración hacia y dentro de la ciudad estudiada, una dimensión longitudinal en el análisis es obviamente muy deseable.

- *"La descripción de las prácticas residenciales y profesionales de los ciudadanos. Como ya lo han demostrado algunos trabajos desarrollados en los últimos años (CRDI-ORSTOM, 1988; DUPONT V. y DUREAU F., 1988), la cuidadosa observación de comportamientos de ciudadanos en lo relacionado con su residencia y actividad económica, constituye un instrumento esencial de análisis de la dinámica de las ciudades de países en desarrollo."* (DUREAU et BARBARY, 1992 : 2)

Frente a este interrogante general, cabe decir, antes de considerar problemas metodológicos más complejos, que ciertas técnicas simples de análisis demográfico (frecuencias simples y cruzadas, gráficas y mapas de evolución, etc...) quedan de muchos aportes en las fases de primera explotación de las encuestas para producir rápidamente resultados aplicables. No obstante el análisis estadístico de las biografías migratorias y profesionales plantea problemas particulares, dadas las características de la información recolectada, que requiere el empleo de métodos más específicos. Son estos problemas metodológicos a los que nos vamos a referir en nuestra presentación.

1. DATOS, METODOS Y RESULTADOS DE UN ESTUDIO HISTORICO Y GEOGRAFICO DE LA MIGRACION EN QUITO

Antes de dedicarse al desarrollo de técnicas estadísticas para el análisis de las biografías individuales completas, hemos querido, en una primera etapa de la explotación de las encuestas de Quito y Dakar, estudiar particularmente un momento decisivo de la vida de los migrantes, el de la primera llegada a la ciudad. El análisis de la información proporcionada por estas encuestas acerca de este evento, permite, con una técnica de implementación simple, evaluar y caracterizar los flujos de inmigrantes longitudinalmente a lo largo del periodo de la observación retrospectiva.

Con este fin se aplicó el análisis de correspondencia simple (AFC), llevando a cabo un doble análisis histórico :

- análisis de la evolución de los lugares de origen de esta migración

- análisis de la geografía de las zonas de primera residencia de los inmigrantes dentro de la urbe.

Para mostrar cual tipo de resultados permite alcanzar este enfoque, vamos a presentar rápidamente la técnica y los principales resultados obtenidos en el caso de Quito.

1. 1. Los datos y el método

En este análisis histórico de la migración, hemos utilizado dos tipos de informaciones recolectadas en las encuestas que se refieren a dos muestras diferentes de inmigrantes.

El análisis de la evolución de los lugares de origen de la migración se apoya en datos relativos al conjunto de los inmigrantes indentificados en el cuestionario general de hogares. En Quito, la encuesta de migración realizada por el Orstom en 1987 nos informa sobre el lugar de nacimiento y la fecha de la primera llegada de 5261 inmigrantes, muestra representativa de la población migrante "instalada" en Quito desde 1935 (población estimada de 399,000 personas). El AFC opera sobre el cuadro cruzado de las 42 modalidades de lugares de nacimiento con los 11 periodos quinquenales posibles para la fecha de primera llegada (variables principales)¹. Para revelar, dentro de poblaciones migrantes de origen geográfico distintos, eventuales diferencias de estructuras demográficas de edad o sexo, agregemos a las variables principales 18 modalidades suplementarias (2 para el sexo y 16 clases quinquenales de edad).

El análisis histórico de los lugares de llegada de los migrantes en la ciudad, recorre en cambio a la información sobre la localización de la primera residencia de los migrantes proporcionada por la parte biográfica de los cuestionarios. En el caso de Quito, esta parte del cuestionario ha sido administrada solamente a la submuestra de los jefes de hogares, lo que puede traer sesgos importantes en las estimaciones de los números de migrantes y errores en la interpretación de los resultados. El análisis debiera hacerse entonces estrictament en terminos relativos : consideraremos solo el porcentaje de la migración total que se dirige a tal lugar de la ciudad. La población analizada incluye 1587 individuos migrantes, jefes de hogares en el momento de la encuesta. Conocemos, para cada uno de ellos, la fecha de primera llegada a Quito y el barrio de esta primera residencia en la capital. Esta muestra es representativa de la población de los jefes de hogares residentes en Quito en 1987 (llegados desde 1935), la cual ha sido estimada por la encuesta a 130,000 personas. Para formar, a partir de esta información, un cuadro de correspondencias que cruce el lugar de primera residencia en Quito con la fecha de primera llegada, con efectivos suficientes en cada modalidad, clasifiquemos los barrios de residencia en 20 zonas y las fechas de llegada en 11 período quinquenales. Fue este cuadro 20 x 11 el sometido al análisis factorial².

1. 2. Resumen de los resultados

1.2.1 Evolución de la composición geográfica de la migración hacia Quito desde 1935 : un esquema general

Los resultados del análisis de los datos (AFC) permiten hacer un histórico bastante fino del origen de la migración hacia Quito durante este periodo. No entregaremos aquí el detalle de estos resultados. Solo presentaremos el fenómeno mayor de esta "historia", lo cual es la **modificación, en el curso del tiempo, de las contribuciones relativas a la migración total hacia la capital ecuatoriana de tres grandes espacios geográficos y culturales.**

¹ Con fin de comparar efectivos de inmigrantes durante períodos de iguales duraciones, los datos del período 1985-1987, conocidos en la fecha de encuesta, han sido proyectados al período 1985-1989 (multiplicación por 5/3).

² De nuevo los datos del periodo 1985-1989 son estimados a partir del periodo 1985-1987 observado.

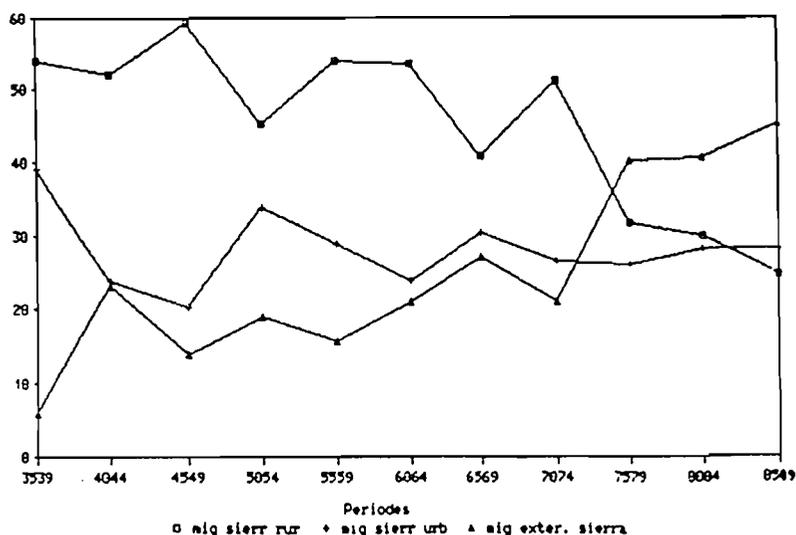
El primero es un medio con predominio rural, aunque incluye también pequeñas y medianas aglomeraciones urbanas. Corresponde aproximadamente al medio rural de la sierra ecuatoriana, del cual se debe sin embargo excluir, los campos de la sierra más lejanos de Quito, en el sur del país. En adelante haremos referencia a este espacio como el **espacio rural de la sierra centro y norte ecuatoriana**.

El segundo espacio agrupa, a parte de Quito, la mayoría de las ciudades importantes de la sierra (aglomeración de más de 20,000 habitantes, excepto Latacunga, ya incluida en el primer espacio); llamaremos este conjunto el **espacio urbano de la sierra ecuatoriana**.

En fin el tercer conjunto, de mayor heterogeneidad geográfica y cultural, tiene tres componentes: la costa rural y urbana (con dominación del urbano), el extranjero (con una mayoría de colombianos) y en fin, los sitios de la sierra que no fueron incluidos ya en los dos primeros espacios, es decir las ciudades medianas (10,000 a 20,000 habitantes) de la sierra norte y el medio rural de la sierra sur. Simplificando de nuevo, hablaremos ahí de **espacio exterior a la sierra ecuatoriana**.

Los tres espacios de origen de la migración que venimos de definir agrupan, a todo lo largo de nuestro periodo de análisis, la casi totalidad de los migrantes: 98.4% de la migración total hacia Quito de todo el periodo.

Gráfico 1 - Distribución de la fechas quinquenales de llegada en Quito de los migrantes originarios de tres espacios geográficos desde 1935



El gráfico 1 indica que el papel de cada espacio ha variado desde 1935 y, sobre todo, que la composición geográfica de la migración hacia Quito se modificó radicalmente a partir de los años sesenta, siendo el periodo de esta transición situada entre 1960 y 1979.

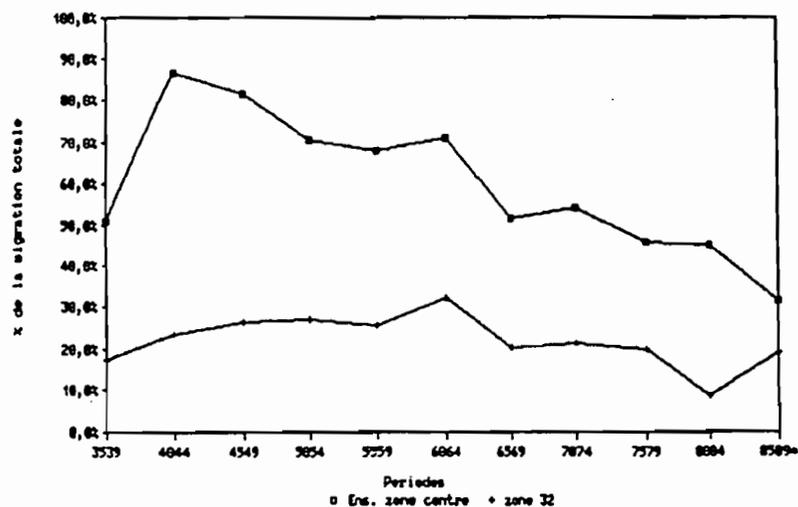
En efecto antes de 1960, la migración procedente de la sierra ejerce una fuerte dominación (más del 80% de los inmigrantes), siendo los papeles relativos de la sierra rural centro y norte y de la sierra urbana respectivamente alrededor del 55% (sierra rural) y del 30% (sierra urbana). La migración de origen exterior a la sierra constituye, durante la mayor parte del periodo, menos de 20% del total. Pero desde 1960, el papel de la migración exterior a la sierra aumentó casi constantemente, conociendo un muy fuerte crecimiento en los años 1975 a 1979. Su importancia sobrepasa los demás espacios desde 1975, para alcanzar 45% de la migración total en 1987. Dicho crecimiento fue realizado principalmente en detrimento de la migración rural de la sierra centro y norte, cuya contribución ha caído del 55% en 1960 hasta 25% en 1987. A lo contrario el papel de la migración urbana de la sierra queda estable desde 1960, siendo todavía en 1987 de 28%.

1. 2. 2. El desplazamiento de las zonas de primera residencia de los migrantes

Durante todo el periodo desde 1935 hasta 1987, pero sobre todo hasta 1965, los barrios céntricos de Quito - el centro histórico y el conjunto de sus "margenes" norte, sur, este y oeste, que llamaremos aquí la zona centro - juegan un papel muy importante en el recibimiento de la migración : esta zona agrupa más del 50% de las primeras residencias en Quito de los migrantes llegados durante todo el periodo. Este hecho es bien conocido de los analistas de la migración hacia Quito.

Pero constatamos en el gráfico 2 que un esquema presentado frecuentemente en la literatura sobre este tema, en el cual los barrios del centro histórico están identificados como principales lugares de llegada de los migrantes, no corresponde de hecho a la realidad. Desde 1935, para tener en cuenta la mayoría de las instalaciones de inmigrantes en Quito, tenemos que considerar una zona central mucho más amplia, y en cuanto al papel del centro histórico propiamente dicho, nunca sobrepasó, a lo largo de todo el periodo, 30% del total de las instalaciones. El AFC muestra, en efecto, que durante ciertos periodos, importantes flujos de inmigrantes se instalaron en otros barrios que los de la zona centro. En resumen desde 1935, dos desplazamientos principales de las zonas de primeras residencias de los migrantes se han producido (gráfico 2). El primero, durante el periodo 1935-1945, consiste en una reagrupación de dichas zonas en, y alrededor, del centro histórico. El segundo, a lo contrario, opera a partir de 1965 una diversificación hacia la periferia de los barrios de primeras residencias, la cual acompaña y participa al crecimiento espacial de la aglomeración quiteña. Estos dos traslados han sido separados por un periodo de estabilidad (1945-1965) durante el cual la mayor parte de la migración penetraba en la ciudad por los barrios de la zona centro.

Gráfico 2 - Evolución del papel de la zona centro (zonas IMQ 22, 23, 32, 33, 34) y del centro histórico (zona IMQ 32) en el recibimiento de la migración desde 1935



Esta rápida presentación de algunos resultados obtenidos en Quito es por supuesto muy esquemática e incompleta. Su único propósito era de mostrar como, mediante técnicas de análisis de datos bastante simples, aplicadas a ciertos acontecimientos particulares de la biografía de los migrantes, se puede alcanzar conclusiones interesantes para el análisis de los flujos y la caracterización de algunas modalidades de la migración hacia las ciudades. El enfoque longitudinal, durante el periodo de la observación retrospectiva, que los datos y el método permiten, nos parece un aporte incontestable de

esta aproximación. Pero estos métodos simples son insuficientes para procesar las biografías individuales completas. Entremos pues a ver cuáles son, en el ámbito del análisis de datos, las propuestas metodológicas recientes que podrían cumplir este objetivo.

2. UNA REVISTA DE DIFERENTES METODOS ESTADISTICOS QUE PUEDEN APLICARSE AL ANALISIS DE LOS DATOS BIOGRAFICOS ¹

Antes de iniciar esta revista, quisieramos recordar los objetivos generales que fueron asignados al análisis estadístico de los datos biográficos después de la reflexión teórica que tuvimos con F. DUREAU al inicio de este programa de investigación. El interrogante metodológico puede resumirse de la siguiente manera : "*Cómo proceder a las clasificaciones una vez organizada la información (definición de las cohortes, caracterización de los espacios de salida y de llegada...)? Qué métodos emplear? Cómo poner en evidencia las interacciones entre fenómenos demográficos, los efectos de eventos contextuales?*" (DUREAU y BARBARY, 1992 : 9). Esperamos encontrar respuesta a estas preguntas por medio de dos tipos de enfoques teóricos.

2. 1. Modelos probabilísticos y análisis de datos. Dos aproximaciones complementarias para el análisis estadístico de las biografías

El enfoque de modelización probabilística, desarrollado por los estadísticos demógrafos a partir de modelos no paramétricos, paramétricos o semiparamétricos (COX, 1975 ; COURGEAU y LELIEVRE, 1989 ; FLOREZ, 1990) tiene como punto de partida una formalización matemática que plantea a las biografías individuales como **procesos aleatorios**: la biografía de un individuo está representada por una serie de variables aleatorias que son el tiempo de permanencia en los diferentes estados que la componen, tomados en orden cronológico. El trabajo consiste entonces en estimar a partir de los datos de la encuesta, un modelo de distribución de cada una de estas variables, para intentar luego la modelización de la distribución más compleja del conjunto de la trayectoria (distribución conjunta). Los métodos estadísticos y los programas informáticos generados por este enfoque, permiten un sofisticado análisis de interacciones entre fenómenos demográficos y apreciar los efectos de las características individuales (sociales, económicas, culturales, etc.) y de los eventos "exteriores", (contexto económico, sociopolítico, catástrofes naturales, etc.) sobre el tiempo de permanencia en diferentes estados.

Una limitación importante de este método tiene que ver con el carácter necesariamente limitado del sistema de eventos o estados que pueda considerarse; además de que el sistema debe definirse *a priori*. Finalmente el enfoque de la modelización no brinda ninguna herramienta descriptiva o tipológica de las trayectorias individuales, no satisface por lo tanto más que uno de los dos objetivos del análisis estadístico de las biografías identificadas anteriormente. Siendo la bibliografía sobre estos métodos considerable, no hemos incluido ejemplos de este enfoque en nuestra revista.

Las técnicas de análisis exploratorio de grandes tablas de datos constituyen, por su parte, herramientas competitivas de descripción y clasificación: análisis factoriales, clasificaciones automáticas. Cómo pueden aplicarse en nuestro caso?

Los métodos de análisis de datos han sido aplicados a menudo, especialmente en economía, a series cronológicas sincrónicas, con un fin descriptivo o incluso de previsión. Pero el problema aquí es totalmente diferente: opuestos a las series cronológicas habituales, los datos cronológicos de las

¹ Esta parte proviene de un trabajo presentado en el seminario sobre el uso de historias de vida en las ciencias sociales organizado en Villa de Leyva del 17 al 20 de marzo de 1992. para una exposición mas completa, vease DUREAU y BARBARY (1992 : 10-17).

biografías individuales son por un lado asincrónicas -no se trata de una serie de "medidas" efectuadas en fechas fijas para todos los individuos sino por el contrario, de un conjunto de secuencias de eventos fechados, cada uno con "reloj" propio-, y por otro lado, el número de eventos y el orden en el que ocurren son diferentes *a priori* de un individuo a otro.

Debido estas características particulares de los datos, sólo recientemente teóricos y usuarios del análisis exploratorio de datos se han interesado en la elaboración, la evaluación y el uso de herramientas específicas para analizar el tiempo pasado en situaciones sucesivas. Una búsqueda bibliográfica, aunque incompleta, muestra que existe desde entonces un legado en materia de análisis tipológico de los datos biográficos; estos métodos abren perspectivas interesantes para una primera fase de procesamiento de las biografías, complementaria al enfoque probabilístico.

2. 2. El análisis exploratorio de los datos biográficos

J. PICARD (1987), en un artículo consagrado a un método de clasificación de perfiles individuales, introduce su exposición con las siguientes referencias:

"El análisis de la evolución multivariada de individuos conlleva el estudio de las estructuras de un cuadro tridimensional, en el que los tres índices corresponden respectivamente a los individuos, las variables y el tiempo. Los métodos utilizados para este estudio se basan esencialmente en dos tipos de enfoques factoriales:

*El primer enfoque ha sido propuesto por LEBART (1966-1969) bajo el nombre de **análisis factorial local (AFL)**, que aplicó en **contiguidades espaciales**.*

El segundo enfoque es el del método STATIS desarrollado por L'HERMIER DES PLANTES y ESCOUFIER (1978)." (PICARD, 1987 : 27)

De estas dos familias de métodos de análisis factorial, presentaremos aquí las ideas principales del artículo de LEBART (1969) y una aplicación de los métodos STATIS. Es necesario igualmente, agregar otras dos aproximaciones factoriales de estos problemas al inventario de PICARD :

-La primera, que presentaremos brevemente también, es el **análisis armónico cualitativo**, una aplicación particular del análisis de las correspondencias múltiples desarrollado y justificado por DEVILLE y SAPORTA, en dos artículos (1980-1982).

-La segunda, por ser la más simple de todas, no puede por ello omitirse. Consiste en aplicar el **análisis de las correspondencias múltiples clásico** a un cuadro en el que se han resumido las trayectorias individuales a estudiar, en un determinado número de variables cuantitativas o cualitativas que las describirán en forma resumida. Todo depende entonces del escogimiento de estos resúmenes. Dada su naturaleza clásica, no cabe presentar esta metodología aquí. Un excelente ejemplo de resultados interesantes obtenidos mediante estos métodos se encuentra en MARPSAT (1984).

2. 2.1. El análisis estadístico de la contiguidad (o AFL) de LEBART¹

"Los estadísticos tienen que ver a menudo con conjuntos de medidas u observaciones que (...) se realizan sobre un grafo (por ejemplo el caso en que se privilegia un conjunto de parejas de observaciones: parejas de observación sucesivas para las series temporales, parejas de observaciones contiguas para las variables regionales, etc.)." (LEBART, 1969 : 81)

Es así como LEBART introduce el problema al que aporta elementos de solución en su artículo publicado en 1969. La idea de partida es que en muchos casos, la evolución temporal y/o espacial de los datos observados, no es independiente de la estructura del grafo que arroja los datos (grafo de contiguidades espaciales o temporales). Como se trata precisamente de la hipótesis que se puede formular a propósito de los datos biográficos individuales, los desarrollos de LEBART pueden abastecernos certeramente de pistas metodológicas interesantes. En efecto, uno de los objetivos del

¹ Véase LEBART (1969).

artículo es el de desarrollar una nueva técnica de análisis factorial que permita evidenciar la escala y la localización de los enlaces espaciales o temporales entre los datos.

Al iniciar la empresa, LEBART retoma la noción de coeficiente de contigüidad definido por GEARY en 1933. Este coeficiente puede comprenderse como la razón obtenida con dos estimadores de la varianza de la variable: en el numerador, el estimador toma en cuenta la contigüidad o no de las observaciones sobre el grafo, mientras que en el denominador encontramos la varianza empírica clásica. Pero LEBART generaliza el coeficiente de GEARY a la noción de coeficiente de contigüidad en el nivel beta(β), tomando las parejas de vértices distantes de beta(β) sobre el grafo. Esta generalización le permite probar, no solamente la dependencia con respecto a la estricta vecindad en el grafo, sino además las dependencias con respecto a vecindades de nivel beta(β). Así se logra identificar la escala espacial o temporal en la que se ejerce la autocorrelación de los datos.

LEBART pasa luego a la aplicación de estas nociones en el análisis factorial de un conjunto de observaciones contiguas. Considera algunas variables en un mismo grafo e introduce la matriz de contigüidad de las observaciones: matriz de covarianza de variables, calculada sobre las observaciones que poseen determinada distancia sobre el grafo. La idea es que si no hay independencia de las evoluciones espaciales o temporales en relación con la estructura del grafo, se podrá interpretar las diferencias entre los planos factoriales del análisis en componentes principales (ACP) de la matriz de contigüidad y aquellos de la matriz de covarianza.

2. 2. 2. El método STATIS de L'HERMIER DES PLANTES y ESCOUFIER ¹

El instrumento estadístico que utilizan los métodos multicuadros STATIS es el ACP de los diferentes cuadros de distancias, contruídos a partir de las dimensiones de las tablas de datos originales: distancias interindividuos, intervariables, interfechas. Daremos una ojeada esquemática a esta aproximación, tomando como ejemplo el artículo de UBERTALLI y PERNIN (1990), donde estudian las trayectorias profesionales de las primeras promociones de la escuela de enfermería de Roane.

Tratándose de ACP de cuadros de distancias, es necesario en primer lugar representar individuos, variables y fechas en espacios euclidianos. Para situarse en este marco, los autores deben discretizar y sincronizar el tiempo de observación de los individuos: a partir de la fecha de obtención del diploma, se escogen "instantes clave" de observación de individuos (diploma más un año, diploma más tres años, etc.). Luego se resume la carrera individual para cada una de estas fechas por medio de un número determinado de variables métricas: tiempo de ejercicio como enfermero, tiempo de estudios complementarios, número de períodos de inactividad, etc.

Luego la aplicación de los métodos STATIS se desarrolla en cuatro etapas principales.

a. Análisis de la "interestructura"

Para cada fecha se calcula la matriz de distancias euclidianas entre individuos, en el espacio de las variables. Esta matriz es la de un operador lineal que traduce la "estructura individuos" en cada fecha. El producto escalar de HILBERT-SMITH entre operadores lineales, permite construir una matriz de distancias interfechas que es sometida al ACP (de ahí el nombre de "interestructura"). La interpretación de los planos factoriales conduce al análisis de la **evolución temporal promedia de los individuos**.

b. Análisis de las "distancias-compromiso".

A partir de las matrices de las distancias interindividuales en cada fecha, se construye ahora una matriz de "distancias-compromiso", es decir la suma de las distancias en cada fecha ponderada por las componentes del primer vector propio de la "interestructura" (es en síntesis, una distancia interindividual promedia a lo largo del conjunto de fechas). El ACP de este cuadro proporciona un **resumen temporal de las trayectorias individuales**.

¹ Vease ESCOUFIER y L'HERMIER DES PLANTES (1978) y UBERTALLI y PERNIN (1990).

c. Análisis de la "intraestructura" de las variables.

Consideremos ahora el gran cuadro rectangular obtenido "fusionando" los diferentes cuadros individuos por variables en cada fecha, nuevamente ponderados por las raíces cuadradas de las componentes del primer vector propio de la "interestructura", para permitir la superposición de planos factoriales con aquellos obtenidos en el ACP de las "distancias-compromisos". El análisis factorial de este cuadro y la proyección de la nube de "variables x fechas", permite entonces medir la importancia relativa de cada variable en la discriminación de las diferentes formas de trayectorias.

d. Análisis de la "intraestructura" de los individuos.

El análisis de la "intraestructura" de los individuos se hará por el medio de la adición, como elementos suplementarios, de los cuadros individuos x variables en cada fecha en el acp del cuadro de "distancias-compromisos". Sobre los planos factoriales del "compromiso", existe un punto por individuo y por fecha que se puede relacionar en orden cronológico para obtener trayectorias individuales interpretables en función de su posición con respecto a los ejes del "compromiso".

2. 2. 3. El análisis armónico cualitativo de DEVILLE y SAPORTA¹

En el análisis armónico cualitativo, presentado por DEVILLE y SAPORTA (1980-1982), la estructura de los datos considerados es realmente idéntica a la de los datos que consideramos aquí. En efecto, el problema que plantean los autores es el siguiente : *"Diversos datos de encuestas permiten trazar la historia de los individuos durante un período de tiempo determinado: la evolución de la actividad profesional, de la situación matrimonial, de la residencia. El objetivo de este artículo es mostrar que datos de este tipo pueden analizarse de manera análoga a lo que haríamos con un proceso escalar (análisis armónico) y para un conjunto finito de variables cualitativas no temporales (análisis canónico generalizado o análisis de correspondencias múltiples)".* (DEVILLE y SAPORTA, 1980 : 376)

El alcance de este objetivo arrastra a DEVILLE y SAPORTA hacia desarrollos teóricos complejos que aquí tan sólo resumiremos.

El punto de partida es considerar este tipo de datos biográficos como la realización de un proceso cualitativo en un tiempo continuo $X_t(i)$ de m modalidades ($X_t(i)$ es el estado, entre los m estados posibles, en el que se encuentra el individuo i en un instante t). Puesto que se trata de un proceso cualitativo que no se sabe analizar, se considera un código escalar de dicho proceso : $Y_t = f_t(X_t)$. El análisis armónico del proceso Y_t es entonces posible, pero no es su interpretación que nos interesa, porque DEVILLE y SAPORTA hacen luego una observación de gran importancia práctica : *"El análisis armónico cualitativo es una generalización de métodos familiares de análisis de datos y es reductible a estos métodos, al plano numérico. Todas las técnicas de interpretaciones habituales en estos métodos tienen validez en nuestro caso"* (op. cit. : 380).

Sigue en el artículo la exposición de un método de análisis en tiempo discreto operacional que es una aplicación particular del análisis de correspondencias múltiples (ACM). El principio es sencillo. Se comienza por discretizar el período de observación del proceso, considerando todos los intervalos de tiempo en los que el proceso se mantiene estable, es decir sin que individuo alguno cambie de estado (no se pierde, pues, información sobre la cronología de los eventos recolectados). El análisis de este proceso en tiempo discreto se hace mediante el ACM de un cuadro disyuntivo particular con n líneas y mp columnas (n es el número de individuos, m el número de estados posibles y p el número de períodos de estabilidad). La casilla elemental del cuadro vale uno si el individuo está en el estado considerado durante el período y cero si no lo está. Pero inmediatamente surge un problema práctico: en los casos concretos de aplicación, los números n , m y p generan un cuadro de tamaño asombroso.

¹ Vease DEVILLE y SAPORTA (1980) y DEVILLE (1982).

DEVILLE y SAPORTA proponen entonces una solución de aproximación : se dividirá el intervalo de observación en un número razonable de períodos sin tener en cuenta los cambios de estado individuales. El cuadro sometido al ACM se construye con la proporción de tiempo que a permanecido cada individuo en cada uno de los estados posibles a lo largo de cada período.

2. 3. Las especificidades y la complementariedad de los métodos exploratorios para el análisis de los datos biográficos

Evidentemente no podemos prejuzgar ahora los resultados, que en el plano temático, arroje el procesamiento estadístico de las biografías completas recolectadas en Quito y Dakar. Desde el punto de vista de la metodología estadística, en cambio, quisiéramos intentar una síntesis inicial, examinando globalmente la adecuación de los métodos de análisis exploratorio frente a un objetivo esencial del procesamiento de los datos biográficos : **describir y clasificar**, con el fin de medir la importancia relativa de cada tipo, las diferentes trayectorias individuales. La rápida revista metodológica que nos precede, ha mostrado que esta adecuación debe medirse con **dos criterios principales** relacionados con los problemas de recodificación de los datos:

- El grado en que se conserva, en la recodificación discreta y sincrónica de los tiempos de observación individuales, la información continua y asincrónica de la variable original?
- La conservación más o menos adecuada, mediante la recodificación de los estados residenciales y profesionales, de la riqueza de la información recolectada.

Arriesgándonos en un primer intento de síntesis, pero desde un punto de vista estrictamente teórico y sin ninguna pretensión definitiva, podemos proponer una "clasificación" inicial de los cuatro métodos arriba considerados.

El análisis de las correspondencias múltiples sobre el cuadro que resume las trayectorias individuales permite desembocar en una tipología de las trayectorias, probablemente muy reduccionista desde el punto de vista de los dos criterios que nos interesan: la discretización/sincronización de los tiempos es total; la selección y codificación de las variables para describir sintéticamente el itinerario residencial y profesional determina en gran medida la pertinencia de los resultados, de donde nace la necesidad de conservar las nomenclaturas más refinadas. Pero su ejecución es simple y rápida; bien aplicada, puede prestar apreciables servicios en una primera fase de descripción general de los datos biográficos.

El análisis de la contiguidad, aunque no sea propiamente un método completo de análisis tipológico, puede proveer herramientas valiosas para la preparación de los datos para tal análisis. Permite en efecto la identificación de escalas y la localización de las autocorrelaciones temporales, espaciales o económicas que caracterizan los itinerarios biográficos, y por consiguiente optimizar sus recodificaciones . La discretización/sincronización de los tiempos practicada en este análisis es maleable, pero, basado sobre el ACP, implica la constitución, para cada intervalo de tiempo, de variables métricas que resuman los eventos residenciales y profesionales acaecidos.

El método STATIS, constituye una muy completa "caja de herramientas" para el análisis tipológico de datos biográficos individuales que permite :

- evidenciar a partir de la "interestructura," las fechas claves en la diferenciación de los itinerarios;
- explicar con la ayuda de la "intraestructura" de las variables, el papel de los eventos residenciales y profesionales en esta diferenciación;
- clasificar las trayectorias individuales, ("intraestructura" de los individuos) en el espacio factorial del "compromiso", interpretar la tipología e ilustrarla por medio de trayectorias-tipos.

El tiempo se encuentra discretizado y sincronizado en el seno de un número necesariamente reducido de intervalos, ya que cada fecha introduce un nuevo cuadro de análisis; reducir la información se convierte en un imperativo. Además, también basada en el ACP, presentará la misma debilidad que el método precedente, desde el punto de vista del segundo criterio. Sin embargo, en un marco de discusión bien definido, la riqueza de la interpretación puede convertirle en un precioso método.

Finalmente, el análisis armónico cualitativo, evaluado en el terreno teórico de acuerdo a los dos criterios que nos interesan, parece el más prometedor: conserva óptimamente el orden y la cronología de los eventos residenciales y profesionales, y el empleo del análisis de correspondencias múltiples

permite la mezcla de variables cualitativas y cuantitativas en la descripción de los estados. De la mano con las técnicas de clasificación automática, desemboca en tipologías de los itinerarios individuales susceptibles de interpretación con el auxilio de reglas sencillas.

Subrayemos finalmente la óptima **complementariedad** potencial de los cuatro métodos. El primer método, utilizado en la etapa de descripción general de los itinerarios biográficos, permitirá la definición de tipos globales en los que luego operara el análisis tipológico detallado. El análisis de contiguidades, aportando conclusiones sobre las escalas en que interviene la autocorrelación temporal, geográfica y económica de los datos biográficos, contribuirá a definir intervalos de tiempo y nomenclaturas para recodificar variables de estadios residenciales y profesionales. Podremos luego alternar o llevar conjuntamente los análisis tipológicos refinados con la ayuda de métodos STATIS y con el análisis armónico cualitativo, mediante la selección que mejor se adapte al problema, o sacando provecho, en un campo común de aplicación, de las dos iluminaciones complementarias que permiten técnicas de cálculo y de interpretación diferentes.

CONCLUSION

Como lo ha dicho F. DUREAU *"la innovación metodológica en la etapa de recolección impone el desarrollo de nuevas técnicas de análisis estadístico si no se quiere perder la ventaja de una observación mejorada"* (DUREAU y BARBARY, 1992 : 17). El ajuste de herramientas de análisis exploratorio de datos biográficos, que hemos intentado presentarles, nos parece constituir una etapa fundamental en esta fase necesaria de innovación metodológica. Así mismo, podrá proponerse un proceso coherente de procesamiento de encuestas biográficas que alterne la fase de identificación y caracterización de los itinerarios por medio del enfoque exploratorio del análisis de datos con las fases de modelización o explicación por medio de un enfoque probabilístico, cuando se estudia un sistema particular de acontecimientos, como en el caso de los eventos migratorios y profesionales claves.

BIBLIOGRAFIA

- BERET P., 1988. Analyse de données chronologiques relatives à l'insertion professionnelle. en : Les Cahiers de l'Analyse des Données, vol. XIII, n° 2, pp 159-174.
- COURGEAU D., LELIEVRE E., 1989. Analyse démographique des biographies. INED, Paris, 268 p.
- COX D.B., 1972. Regression models and life tables (with discussion). en : Journal of Royal Statistical Society", B 34, pp. 187-220.
- DEVILLE J.C., SAPORTA G., 1980. Analyse harmonique qualitative. en : 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 ? en : Annales de l'INSEE, n° 45, pp. 45-104.
- DUREAU F., BARBARY O., 1992. De la recolección al análisis estadístico de biografías migratorias y profesionales. Ejemplo de una encuesta realizada en Quito. Ponencia presentada en el seminario "El uso de las historias ed vida en las ciencias sociales - Villa de Leyva, Colombia, 17-20 de marzo de 1992, 21 p.
- ESCOUFIER Y., L'HERMIER DES PLANTES, 1978. A propos de la comparaison graphique des matrices de variance. en : Biometrical Journal, Vol 20, 5, pp. 491-497.

- FOUCART T., 1983. Une nouvelle approche de la méthode STATIS. en : *Revue de Statistique Appliquée*, Vol. 31, 2, pp. 61-95.
- LEBART L., 1969. Analyse statistique de la contiguité. Publications de l'Institut de Statistique de l'Université de Paris, Vol. XVIII, pp. 81-112.
- LE FOLL Y., 1982. Pondération des distances en analyse factorielle. en : *Statistique et analyse des données*, 1982, 1, pp. 13-31.
- MARPSAT M., 1984. Chômage et profession dans les années trente. en : *Economie et Statistique*, n° 170, pp. 53-69.
- PICARD J., 1987. Classification des profils évolutifs incomplets et asynchrones. en : *Revue de Statistique Appliquée*, vol. XXXV (2), pp. 27-38.
- UBERTALLI B., PERNIN M.O., 1990. Utilisation d'une méthode multitableaux en sciences sociales. Une étude longitudinale de carrières : les 12 premières promotions de l'école d'infirmières de Roanne. en : *Population*, n°6, pp. 1092-1100.

Barbary Olivier (1995)

Metodos estadisticos para el analisis de datos retrospectivos sobre la migracion : ejemplos de aplicaciones en Quito y revision metodologica

In : Dureau Françoise (ed.), Florez C.E. (collab.), Hoyos Maria Cristina (collab.), Villa M. (collab.) Las nuevas formas de movilidad de las poblaciones urbanas en America Latina

Bogota : CEDE, (97), 163-173 multigr.. (Documento CEDE ; 97)

Taller Metodologico sobre las Nuevas Formas de Movilidad de las Poblaciones Urbanas en Americana Latina, Bogota (COL), 1992/12/7-11