

**UNIVERSIDAD DE COSTA RICA
FACULTAD DE CIENCIAS SOCIALES
INSTITUTO DE INVESTIGACIONES SOCIALES**

16

CONTRIBUCIONES

**"UN ESTUDIO MULTIVARIADO SOBRE
LA MORBILIDAD DE LA FUERZA DE
TRABAJO EN COSTA RICA"**

**Jaime Lobo S.
William Castillo W.**

**Consejo Editorial: Licda. Dina Krauskopf, M.Sc. Sui Moy Li,
Dr. Héctor Pérez, Dr. Tomás Guerra.**

**Apoyo Administrativo: Coordinación, Lisbeth Vega;
Secretaria, Jenny Sánchez; Auxiliar de Imprenta, Jorge
Oconitrillo.**

Diseño de Portada: Sonia Calvo.

Setiembre 1993



INDICE

PRESENTACION	ii
I. INTRODUCCION	1
II. ESTUDIOS PRELIMINARES SOBRE EGRESOS HOSPITALARIOS	3
III. PLANTEAMIENTO DEL PROBLEMA Y DEFINICION DE LAS TABLAS DE DATOS	5
IV. EL ANALISIS DE DATOS	9
4.1 Clasificación de diagnósticos y ocupaciones, caso hombres	9
4.2 Clasificación de diagnósticos y ocupaciones, caso mujeres	10
V. INTERPRETACION DE LAS CLASES. PRIMEROS ANALISIS DE CORRESPONDENCIAS	11
5.1 Caso de los hombres	11
5.2 Caso de las mujeres	17
5.3 Comparación entre los sexos	21
VI. ANALISIS DE LAS TENDENCIAS	23
6.1 Tendencias en los hombres	23
6.2 Tendencias en las mujeres	26
6.3 Comparación de las tendencias según sexo	29
VII. CONCLUSIONES	31
APENDICE	33
BIBLIOGRAFIA	39

PRESENTACION

Es para nosotros motivo de satisfacción publicar en la serie de Contribuciones, el presente trabajo titulado "Un análisis multivariado sobre la morbilidad de la fuerza de trabajo en Costa Rica", de los profesores Jaime Lobo y William Castillo, de la Escuela de Matemática. El trabajo es producto del proyecto de investigación "Fuerza de trabajo y morbilidad hospitalaria en Costa Rica, 1980-1987", llevado a cabo en el Instituto de Investigaciones Sociales, y en el cual participaron los autores.

Hemos considerado importante, dar a conocer una forma de aplicación de modelos matemáticos al estudio de un fenómeno concreto, como la relación entre enfermedad y ocupación, con la esperanza de que un tema como éste, promueva el desarrollo de un espacio de trabajo con otras disciplinas, que sin duda vendría a enriquecer nuestro quehacer.

Con este nuevo número, el Instituto de Investigaciones Sociales reitera su interés en promover el trabajo interdisciplinario y la discusión académica.

Dr. Manuel Solís A.
Director
Instituto de Investigaciones Sociales

UN ESTUDIO MULTIVARIADO SOBRE LA MORBILIDAD Y FUERZA DE TRABAJO EN COSTA RICA

I. INTRODUCCION

A pesar de que la morbilidad es un indicador indirecto y por tanto insuficiente sobre la salud de la población, ya que muestra únicamente una faceta parcial de este fenómeno, su estudio es de gran utilidad dada la ausencia de estos análisis a nivel general y particularmente en Costa Rica. En este país las estadísticas de salud se siguen presentando fundamentalmente en base a datos de mortalidad, lo que impide un mayor conocimiento de las condiciones de salud de la sociedad.

Realizar un estudio sobre la morbilidad en Costa Rica resulta una tarea posible gracias a la existencia de fuentes de información bastante completas y confiables. El acopio de datos que centralizadamente realiza la Caja Costarricense de Seguro Social (CCSS), sobre los egresos hospitalarios a nivel nacional, desde el año 1980, y considerando la amplia cobertura de esta institución (cerca del 80% de la población total), constituyen ventajas importantes que deben aprovecharse para mejorar y ampliar el conocimiento acerca de la salud de la población costarricense.

En particular, es de gran interés actual, estudiar las características de la morbilidad en la población entre los 20-59 años, que conforma el grueso de la población laboral del país y de la cual se conoce muy poco, sobre todo en lo que a salud se refiere. En esta población de edades intermedias, resulta especialmente pertinente analizar la morbilidad, ya que en ella se concentra gran parte de las enfermedades y no así las defunciones, las cuales se ubican principalmente en las edades extremas, como son los primeros años de la vida y la vejez.

Para estudiar la morbilidad de este sector de la población, a partir de los egresos hospitalarios, la CCSS facilitó varias cintas magnéticas con toda la información obtenida de las boletas de egresos hospitalarios de todos los centros, durante el periodo 1980-1987. De estos datos se tomaron para el presente análisis dos variables fundamentales: ocupación y diagnóstico principal, que en los datos aparecen clasificadas según las clasificaciones COTA y la Clasificación Internacional de Enfermedades (CIE). La clasificación COTA es una clasificación de las ocupaciones adaptada de la Clasificación Internacional de Ocupaciones, que ha sido la empleada también en los últimos censos de población de Costa Rica.

En este estudio se analiza la morbilidad hospitalaria de los trabajadores mediante las técnicas del llamado "análisis multivariado", que forman parte de la estadística descriptiva. Se completan así los estudios sobre este mismo tema reportados en (1) y (2), realizados en el marco del proyecto de investigación "Fuerza de trabajo y morbilidad hospitalaria en Costa Rica, 1980-1987".

II ESTUDIOS PRELIMINARES SOBRE EGRESOS HOSPITALARIOS

Nos referimos a ciertos resultados importantes derivados del estudio de los datos proporcionados por la CCSS, resumidos en (1) y (2), y que pueden servir de preliminar al estudio que expondremos posteriormente. Es oportuno señalar que las conclusiones obtenidas en estos estudios están basadas en un análisis de las estadísticas de morbilidad que no incluyeron los métodos multivariados, Por lo tanto se deducen del examen de las tablas de efectivos totales que se obtuvieron de los datos sin entrar en problemas de cruce de variables. Nuestro trabajo fue de tipo ante todo exploratorio, tratando de reafirmar por medio de la inspección de las tablas de datos la importancia de aspectos como el sexo y la edad, supuestos determinantes de las diferencias en los cuadros de morbilidad en las personas.

Se determinó que estas diferencias eran tanto cualitativas como cuantitativas. Así, durante el período 1980-1987, el número de egresos hospitalarios femeninos llegó a triplicar los egresos masculinos, concentrándose en las mujeres jóvenes (20-44 años). Las causas de egresos en los hombres revelaron ser más variadas, aunque se destacan algunas como los problemas mentales, fracturas, intoxicaciones y traumatismos, ligadas a factores socio-laborales. Las principales causas de egresos hospitalarios en mujeres están relacionadas con enfermedades obstétricas (más del 60% de los egresos), y en el grupo de 45-59 años estas causas pierden importancia con respecto a otras como diabetes e hipertensión.

Se estudió, asimismo, la relación entre actividad económica y egresos, que busca identificar determinantes socioeconómicos del proceso salud-enfermedad. Al respecto puede mencionarse que el comportamiento de esta variable no presenta diferencias significativas entre hombres y mujeres; además es congruente con la distribución de la PEA en la producción del país. La mayor cantidad de egresos corresponde a personas dentro de la categoría "sin actividad" (más del 30% del total de egresos). Esta categoría incluye a la población indigente y asegurados voluntarios, que constituyen grupos socioeconómicos deprimidos. Es relevante mencionar que el 71% de los egresos por psicosis y trastornos neuróticos se producen en hombres sin actividad económica y el 54% en mujeres de la misma condición. En hombres la principal actividad económica es la rama de agricultura, caza, silvicultura y pesca, con el 19% del total de egresos. Aquí predominan las enfermedades infecciosas, traumas y heridas, intoxicaciones por insecticidas, cáncer de piel, alcoholismo y anemia. En caso de las mujeres, la rama más importante es la de servicios comunales, sociales y personales, siendo las enfermedades más frecuentes las del sistema osteomuscular, lumbago, asma, várices, enfermedades nerviosas y tumores benignos. Esto sin incluir las obstétricas que constituyen la principal causa en mujeres de 20-44 años.

Otro aspecto que se trató en el estudio, se refiere a la

relación entre las enfermedades y el lugar de residencia de los pacientes, en el entendido de que el espacio vital de los individuos afecta su salud. Este aspecto no fue considerado en el estudio multivariado.

Parte de las conclusiones de nuestro presente trabajo consistirá en comparar estos resultados con los obtenidos de los análisis multivariados que describiremos a continuación.

III PLANTEAMIENTO DEL PROBLEMA Y DEFINICION DE LAS TABLAS DE DATOS

Los datos de egresos hospitalarios ofrecidos por la CCSS nos posibilita una amplia gama de análisis cuantitativo. Para efectos de un estudio multivariado debemos, sin embargo, centrarnos en aquellos que respondan a las siguientes condiciones.

- Las variables diagnóstico y ocupación son las que en primera instancia explican mejor el fenómeno trabajo-enfermedad en el contexto que lo situamos.
- Un análisis apropiado de estas dos variables debe contribuir a identificar grupos bien diferenciados de ocupaciones y diagnósticos y sus relaciones, lo que llamaremos una "estructura" de la población en estudio.
- Se busca detectar cambios o tendencias de dichas estructuras en ciertos periodos determinados.

Se determinó que la mejor manera de tratar estos datos y que se prestaran a estos objetivos, es bajo la forma de tablas de contingencia, es decir, tablas que cruzan las variables diagnóstico y ocupación, cada casilla de dicha tabla conteniendo el efectivo total para un diagnóstico u una ocupación dados. De los archivos de datos ofrecidos por la CCSS hemos procedido entonces, a la formación de cuatro matrices o tablas de contingencia, cruzando la variable diagnóstico con la de ocupación según sexo y periodo, de la siguiente manera: mujeres 1980-1983, mujeres 1984-1987, hombre 1980-1983, hombres 1984-1987. La codificación de los diagnósticos corresponde a los dos primeros dígitos de la clasificación CIE; los de ocupaciones a los dos primeros dígitos de la clasificación COTA. Las casillas o entradas de cada una de estas matrices contienen, para un sexo y un periodo determinado, el efectivo total de egresos hospitalarios para cada tipo de diagnóstico y ocupación, siendo por lo tanto matrices de contingencia.

Además de las razones mencionadas sobre la escogencia de las variables diagnóstico y ocupación, existen otras que justifican la conformación de las matrices anteriores:

- la primera, y la más obvia, es la necesidad de distinguir un estudio separado para cada sexo, por ser un hecho reconocido que existen diferencias sustanciales en sus cuadros de morbilidad;
- la distinción de dos periodos: 1980-1983 y 1984-1987 responde al interés de realizar un estudio comparativo en el tiempo tanto como nos lo permiten los datos disponibles. El corte efectuado en el año 1983 tiene la ventaja que sitúa el censo de 1984 a mitad de nuestro periodo de estudio, permitiendo un uso razonable de los datos de este censo para ambos periodos. Por otra parte consideramos que para el estudio de tendencias era innecesario efectuar una disgregación del periodo total en

más de dos periodos, pues periodos muy cortos y consecutivos no arrojarían mayor información de los cambios que deseamos detectar. Podemos agregar finalmente que los métodos analíticos de comparación de tablas de contingencia no siendo todavía suficientemente desarrollados a nivel computacional, no hemos querido introducir demasiadas complicaciones en este primer estudio preliminar;

- número de dígitos seleccionados para la lectura de los códigos COTA y CIE: la boleta de egresos hospitalarios permite escoger a tres dígitos el código COTA en el caso de las ocupaciones y a tres dígitos el código CIE en el caso de los diagnósticos. De haber escogido este número de dígitos para la formación de las tablas hubiéramos obtenido tablas de dimensión 1000X1000 (un millón de casillas), con las evidentes dificultades técnicas que implica su manejo. Por otra parte, de haber reducido el número de dígitos a solo uno, mucha información valiosa de los datos se hubiera perdido para cualquiera de las variables. De estas consideraciones hemos seguido un principio de parcimonia decidiendo tomar dos dígitos para la codificación de ambas variables. Las matrices obtenidas no pueden así sobrepasar la dimensión 100X100, y por otro lado la agregación entre ellas no es demasiado grande como para que la información más valiosa no se pierda.

A partir de los datos organizados en las tablas descritas nos preguntamos ahora cuál es el tipo de análisis que permite cumplir los objetivos: explicación del fenómeno trabajo-enfermedad, establecer una estructura de la población en estudio, detectar cambios de las estructuras. Para responder a esto debemos interpretar en términos más cuantitativos a conceptos vagos como "estructura del fenómeno" y que sean operacionables. La manera más idónea es recurriendo a los conceptos mismos del análisis multivariado. Entre ellos es fundamental el concepto de perfil, que en nuestro caso se define cuantitativamente como frecuencias relativas de columnas o de filas derivadas de las tablas de contingencia. Cada tabla define así perfiles ocupacionales para los diagnósticos y perfiles de morbilidad para las ocupaciones, para cada sexo y periodo de estudio. Estos perfiles son en realidad vectores en espacios de gran dimensión (menor que 100 en nuestro caso) y pueden ser analizados gráficamente en los llamados planos de correspondencias de los ejes principales. En este estudio se hará mención de varias nociones y técnicas del análisis multivariado, y el lector hallará un resumen muy sucinto de la materia en el apéndice al final de este trabajo.

En resumen, con la interpretación cuantitativa del análisis multivariado nos proponemos un análisis que reduzca al mínimo las hipótesis a priori sobre las características de la población en estudio. Se harán los siguientes análisis sobre las tablas de contingencia:

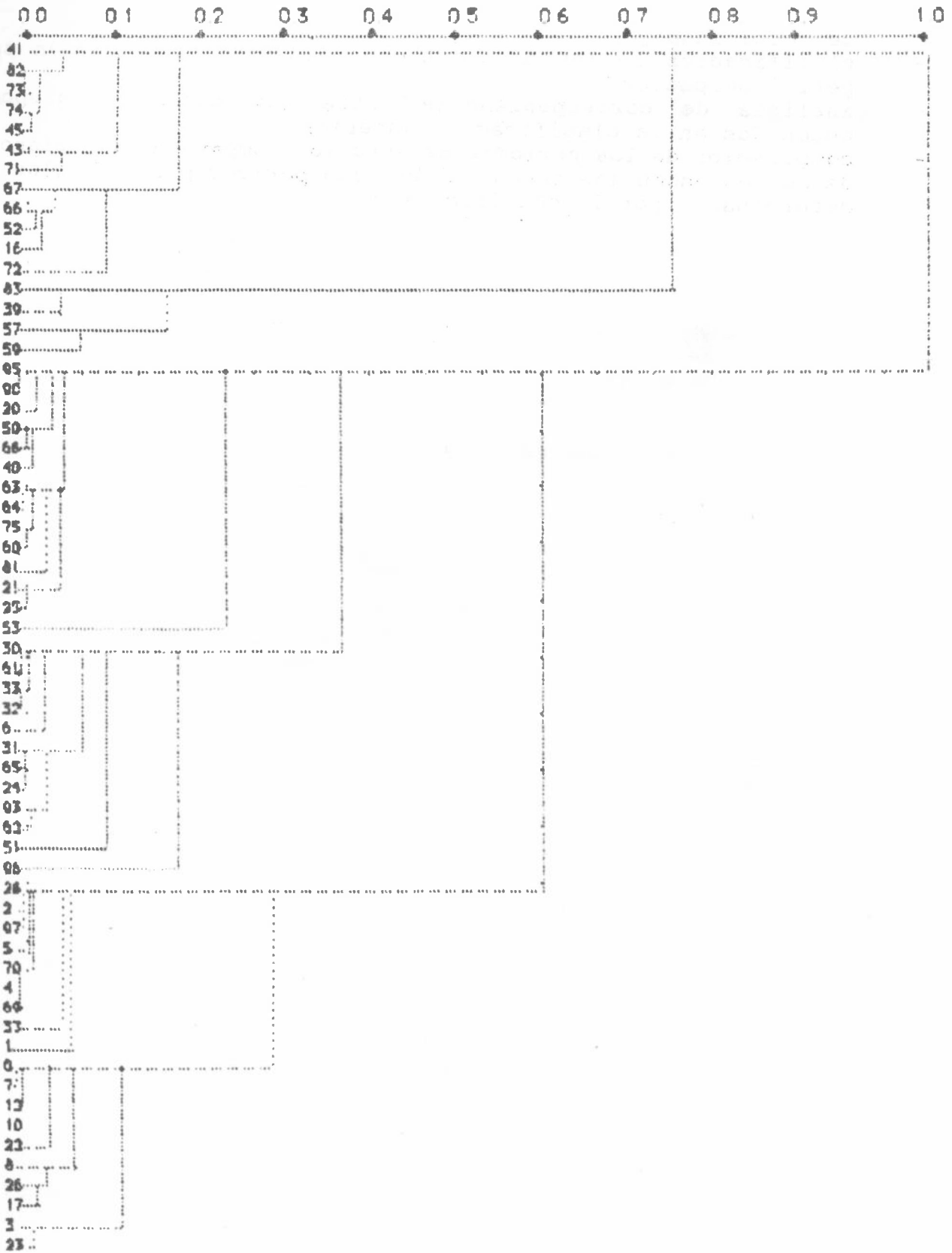
- **clasificación** de ocupaciones para cada sexo, según su perfil

de morbilidad

- **clasificación** de los diagnósticos para cada sexo, según su perfil ocupacional
- **análisis de correspondencias** entre las clases o grupos obtenidos en la clasificación anterior
- **comparación de los períodos de estudio** comparando los planos de correspondencias obtenidos en cada período para las clases determinadas por la clasificación.

GRAFICO 1

Arbol jerárquico de ocupaciones, hombres



Nota : corte del arbol al nivel de 0,3

IV EL ANALISIS DE DATOS

Para efectos de obtener una clasificación de los diagnósticos y las ocupaciones, por sexo, se consideró únicamente el período 1980-1983. Las matrices resultantes son dos tablas de contingencia, una por cada sexo, obtenidas del cruce diagnóstico -x- ocupación. Una vez removidas ciertas modalidades, como se explica en 4.1 y 4.2, se procedió a aplicar la siguiente metodología de análisis de datos.

1. Reducción de la dimensión: sobre cada una de las dos matrices de diagnóstico x ocupación se hizo un análisis factorial de correspondencias (AFC). Esto permitió transformar las tablas originales (tablas de contingencia) en tablas de tipo individuo X variables, donde las variables son en todos los casos 10 componentes principales extraídas del AFC. Los individuos son los diagnósticos o las ocupaciones representadas en los nuevos "ejes perpendiculares", resultando del AFC. Las tablas transformadas son 4, dos de ocupaciones por sexo y dos de diagnósticos, también por sexo.
2. Utilizando las tablas transformadas como se describió en el punto anterior, se aplica el método de Ward para construir un árbol de clasificaciones sobre las filas de cada una de las cuatro tablas transformadas. "Cortando" horizontalmente cada uno de estos árboles en un nivel adecuadamente escogido, se obtienen 4 clasificaciones que son precisamente, las entradas en la etapa siguiente.
3. Para mejorar el valor del criterio de calidad (inercia interclase) se usa el método de nubes dinámicas inicializado con las clasificaciones obtenidas en el punto 2, anterior y, como datos de base, las tablas transformadas de acuerdo con el punto 1.
4. Se construyen las tablas (clases de diagnóstico) X (clases de ocupación) por sexo, según las clasificaciones finalmente obtenidas en el punto 3. Se opera un análisis de correspondencias sobre esas tablas para obtener una representación gráfica de clases de diagnóstico y ocupación, superpuestas (diagramas 1 y 2).

4.1 Clasificación de diagnósticos y ocupaciones, caso hombres

La matriz Hombres 1980-1983 constaba originalmente de 96 columnas (ocupaciones), y de 98 filas (diagnósticos). Se comenzó por reducirla eliminando varias columnas de ocupación, en total 21, de código no especificado en la clasificación COTA, cuya existencia en la matriz solo se explican por errores de codificación o de digitación. Su efectivo era muy pequeño.

Luego de aplicar los puntos 1 y 2 la primera clasificación

sobre la matriz transformada fue poco satisfactoria. Analizando la tabla de datos detectamos la posible perturbación en el análisis ocasionada por grupos de efectivos muy pequeños o por el contrario, muy grandes. Se eliminaron entonces los diagnósticos del código CIE a dos dígitos siguientes: 2, 26, 50, 62, 63, 64, 65, 66, 79, 6, 67. Cabe observar que los códigos 60-70 corresponden a egresos ginecológicos, lo que explica su exclusión. Se eliminaron también dos grandes grupos de ocupaciones: trabajadores agropecuarios y otros en agricultura (código 42 COTA), y sin ocupación (código 42). Estos dos grupos se retomarán en calidad de información suplementaria.

Con esta reducción se aplicó de nuevo el algoritmo de Ward, obteniéndose para ambas variables un árbol jerárquico más satisfactorio (más grupos y mejor separados). Se cortó el árbol jerárquico de diagnósticos al nivel de 0,27, obteniéndose una primera partición de 8 clases, y de ocupaciones al nivel de 0,3, obteniéndose 6 clases. A las clases de ocupaciones se incorporaron después los dos grandes grupos de ocupaciones antes mencionadas.

Al aplicar el punto 3 la nueva participación obtenida resultó ser muy parecida a la inicial, lo que demuestra la buena estabilidad de ambas participaciones respecto al criterio que optimiza el método de nubes dinámicas (ver descripción de las clases finales en cuadros 1 y 2).

4.2 Clasificación de diagnósticos y ocupaciones, caso mujeres

La clasificación usando la matriz Mujeres 1980-1983 siguió, a grandes rasgos, los mismos pasos que la de los hombres. Esta matriz constaba inicialmente de 87 columnas (ocupaciones) y 100 líneas (diagnósticos), se comenzó por eliminar las columnas correspondientes a códigos inexistentes en la clasificación COTA.

Al aplicar los puntos 1 y 2 resultó una clasificación jerárquica poco satisfactoria debido a la presencia de muchas clases de efectivo casi insignificante y de otras muy masivas. Considerando la posible perturbación de casillas muy masivas de la matriz, se eliminaron las columnas y filas de efectivo superior a 5 dígitos, corresponden a la columna de ocupación 39 (oficios domésticos), y las líneas de diagnósticos 62, 63, 64, 66 (diagnósticos ginecológicos). Estos grupos se considerarán como información suplementaria en el análisis.

Se aplica 2, dando por resultado dos árboles de clasificación jerárquica. Al cortar el de ocupaciones al nivel 0,22 dio 5 clases de ocupaciones mientras que el de diagnósticos al nivel 0,15 dio 4 clases. Lo que se pudo observar esta vez fue la relativa estabilidad de las clases obtenidas luego de aplicar 3. En efecto, la clasificación por nubes dinámicas, arrojó participaciones casi idénticas a las variables de inicialización (45% de la inercia total en diagnósticos, 55% en las ocupaciones). Las modificaciones obtenidas se muestran en los cuadros 3 y 4. Este resultado se mantiene al utilizar variables de inicialización aleatorias.

V INTERPRETACION DE LAS CLASES. PRIMEROS ANALISIS DE CORRESPONDENCIAS

Una vez obtenidas las clasificaciones de ocupaciones y diagnósticos, una tarea importante a realizar es la de interpretar dichas clases, de manera que nos revelen fenómenos sociales de algún significado. No hay que olvidar que para la conformación de cada una de estas clases se han seguido criterios muy bien definidos (similitud de perfiles de morbilidad en el caso de las ocupaciones, de ocupación en el caso de los diagnósticos), criterios que no son los usualmente utilizados, por ejemplo en las clasificaciones COTA o CIE. Es de esperar pues, que la interpretación de dichas clases plantee ciertas dificultades.

La manera más directa de caracterizar una clase es, tratando de distinguir algunos rasgos comunes entre sus componentes que expliquen la similitud de sus perfiles. Es claro que en esta distinción de rasgos comunes nos basamos esencialmente en los conocimientos a priori que tenemos de las ocupaciones y diagnósticos, provenientes de otros estudios de tipo sociológico o clínico. Más como estos pueden ser insuficientes para nuestro objetivo, recurriremos luego a los resultados de los análisis de correspondencias que podrían aclarar ciertos aspectos ambiguos de la primera caracterización. En este estudio recurrimos más que todo al primer plano de correspondencias (ejes 1 y 2), que aparecen en los diagramas 1 y 2. Distinguiremos, como anteriormente, entre hombres y mujeres para tratar luego de establecer una comparación entre los sexos.

5.1 Caso de los hombres

Examinemos primero los diagnósticos (cuadro 1). Los aspectos más relevantes que podemos detectar en estas clases son:

- las enfermedades asociadas a ocupaciones de alto riesgo son claramente agrupadas en tres clases principales: I, II y V. Estas están dominadas por los diagnósticos: traumatismos y fracturas (clase I), quemaduras (el total de la clase II), y heridas de los miembros (clase V). Debemos deducir entonces que por alguna razón, aunque siendo diagnósticos de perfil ocupacional de alto riesgo, estos riesgos no son del mismo tipo en cada una de las clases. En el caso de las quemaduras, que es un caso interesante por estar netamente separado de las otras clases, no es difícil suponer que está ligado a ocupaciones relacionadas con la manipulación de líquidos inflamables o máquinas que producen altas temperaturas;
- el grueso de las enfermedades mentales de origen orgánico están agrupadas en la clase VI;
- las enfermedades de diagnóstico impreciso ("otras de..") se agrupan en la clase VII, además de aquellas relacionadas con el estilo de vida (stress);



CUADRO NO. 1

CLASIFICACION DE DIAGNOSTICOS HOMBRES 1980-1983
--

CLASE I DE DIAGNOSTICOS			
NOMBRE RESUMIDO DE LA CIE	EFFECTIVO	PORC CLASE	PORC ACUM
Traumatismo intracraneal (salvo cráneo)	7054	0.16	0.16
Dosopatias	6113	0.14	0.29
Pulmonar obstructiva crónica	3506	0.08	0.37
Fractura miembro inferior	3492	0.08	0.45
Fractura miembro superior	3084	0.07	0.52
Fractura cráneo, cuello, tronco, huesos	2577	0.06	0.58
Traumatismo nervio, médula espinal	2156	0.05	0.62
Infecciones intestinales	2108	0.05	0.67
Infecciones piel, tejido celular subcutáneo	1883	0.04	0.71
CLASE II DE DIAGNOSTICOS			
Quemaduras	827		
CLASE III DE DIAGNOSTICOS			
De la sangre y órganos hematopoyéticos	855	0.35	0.35
Tumor maligno genitourinario	454	0.19	0.54
Tumores evolución incierta	392	0.16	0.71
CLASE IV DE DIAGNOSTICOS			
Otras del sistema nervioso central	1879	0.69	0.69
Envenenamientos drogas, medicamentos, biológicos	139	0.05	0.75
CLASE V DE DIAGNOSTICOS			
Herida miembro superior	3007	0.45	0.45
Herida miembro inferior	1685	0.25	0.70
CLASE VI DE DIAGNOSTICOS			
Psicosis orgánicas	7390	0.85	0.85
Retraso mental	1344	0.15	1.00

CLASE VII DE DIAGNOSTICOS

NOMBRE RESUMIDO DE LA CIE	EFFECTIVO	PORC CLASE	PORC ACUM
Otras del aparato urinario	5422	0.16	0.16
Transtornos neuróticos, de la personalidad, y otros no psicóticos	5379	0.16	0.32
Otras del aparato digestivo	3455	0.10	0.42
Signos, síntomas mal definidos	3031	0.09	0.51
Hipertensiva	3031	0.09	0.60
Otras de las glándulas endocrinas	2551	0.08	0.67
Isquémica del corazón y circulación pulmonar	1960	0.06	0.73

CLASE VIII DE DIAGNOSTICOS

De los organos genitales masculinos	3871	0.17	0.17
De las venas, vasos linfáticos, y otras	3200	0.14	0.31
Luxación	2059	0.09	0.40
Artropatías y afines	2056	0.09	0.49
Otras vías respiratorias superiores	1863	0.08	0.57
Osteopatías, candelopatías y deformaciones adquiridas del sistema osteomuscular	1458	0.06	0.63
Otras debidas a virus y clamidias	1386	0.06	0.69

En esta clasificación se suprimieron los siguientes diagnósticos de código:

3, 27, 51, 63, 64, 65, 66, 67, 80, 7, 68

CUADRO NO. 2

CLASIFICACION DE OCUPACIONES, HOMBRES, 1980-1983

NOMBRE RESUMIDO COTA	CLASE I DE OCUPACIONES		
	EFFECTIVO	PORC CLASE	PORC ACUM
Electricistas, electrónica, operario radio	1590	0.32	0.32
Fontaneros, tuberías, soldadura, metales	1396	0.28	0.60
Pescadores y afines	506	0.10	0.70
CLASE II DE OCUPACIONES			
Otros obreros, jornaleros NCEOG	2997	0.62	0.68
Ocupaciones mal definidas	908	0.21	0.89
Oficios domésticos	487	0.11	1.00
CLASE III DE OCUPACIONES			
Agricultores propietarios en general, especializados	7896	0.18	0.18
Porteros, conserjes, limpieza, afines	6572	0.15	0.33
Choferes, carreteros, cocheros, conductores	6440	0.15	0.48
Mecánic., ajustadores, operadores maquin. no elect.	4119	0.10	0.58
Carpinteros, ebanistas, toneleros	3949	0.09	0.67
Albañiles, ladrillos, vidrios	3122	0.07	0.74
CLASE IV DE OCUPACIONES			
Comerciantes - prop., comercio por mayor y menor	4610	0.26	0.26
Dependientes de tienda, vendedores ambulantes	3919	0.22	0.48
Profesores y maestros	2566	0.14	0.62
Cocineros, cantineros, camareros	1430	0.08	0.70
CLASE V DE OCUPACIONES			
Otros oficinistas, afines	5599	0.51	0.51
Estudiante	2367	0.21	0.72
CLASE VI DE OCUPACIONES			
Arquitectos, ingenieros, afines	2123	0.26	0.26
Otros directores, gerentes, prop. administración	2080	0.26	0.52
Matemáticos, estadísticos, sociólogos, economistas	2034	0.25	0.77

En esta clasificación se suprimieron las ocupaciones agricultores y desocupados y las de código: 94, 11, 79, 77, 92, 76, 27, 80, 78, 91, 44

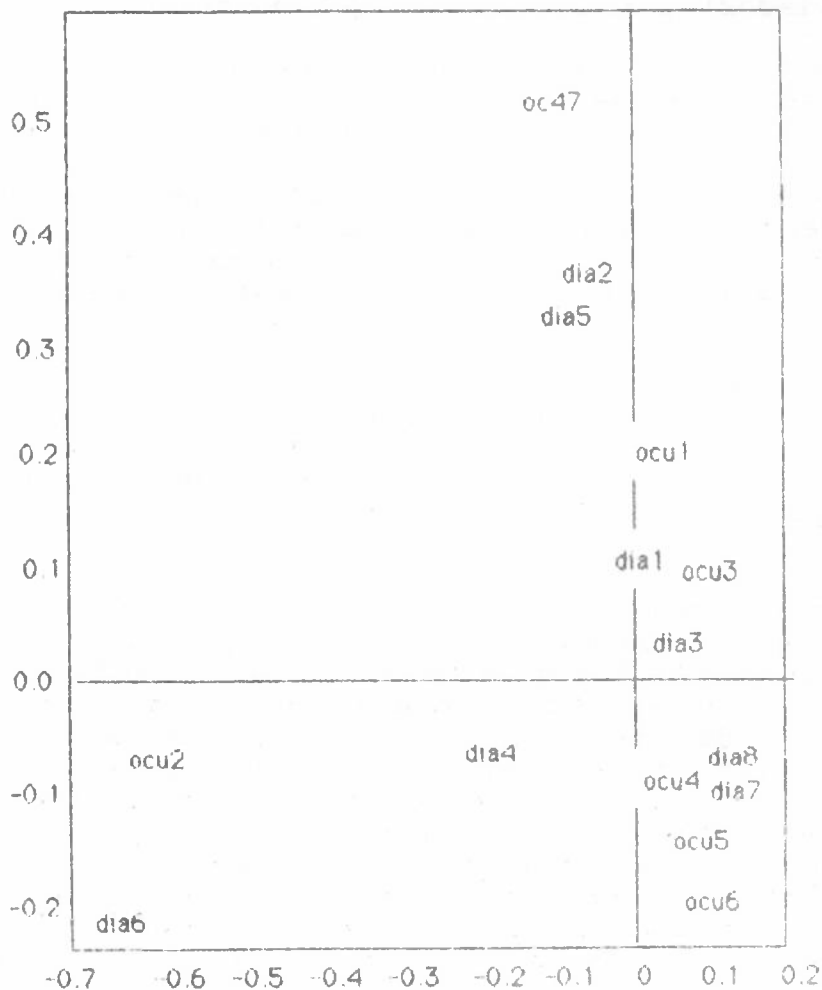
DIAGRAMA 1

ANALISIS DE CORRESPONDENCIAS:
COMPONENTES PRINCIPALES 1 Y 2

CENTROS DE GRAVEDAD DE CLASES DE
DIAGNOSTICO Y OCUPACION, HOMBRES 80-83

Eje 1: Componente principal 1
Eje 2: Componente principal 2

EJE 2



EJE 1

- las otras clases son de interpretación más difícil.

Consideremos ahora las clases de ocupaciones (cuadro 2) en sus aspectos más relevantes:

- las ocupaciones de tipo manual y de mayor trabajo físico se agrupan en las clases I, II, III. La clase I reúne una serie de labores manuales especializadas y que utilizan instrumentos y equipo pesado. Se podría considerar como mano de obra calificada. La clase II por el contrario agrupa ocupaciones poco calificadas, inestables y mal remuneradas, donde se podrían encontrar trabajadores subempleados o del sector informal. La clase III es menos definida;
- la clase IV reúne a artesanos y trabajadores independientes (no asalariados) y a asalariados del comercio;
- las clases V y VI reúnen a la mayoría de las ocupaciones de alta formación académica. La V está dominada por los burócratas y la VI por los profesionales.

Completamos las interpretaciones anteriores por medio del examen del primer plano de correspondencias según los dos primeros ejes principales (diagrama 1). La ocupación cód. 42 (agricultores en general) y el diagnóstico V6 (desocupados) no intervinieron en la determinación de los ejes y para efectos de representación fueron proyectados sobre los ejes y para efectos de representación fueron proyectados sobre los mismos. Los puntos más distantes del origen en este plano, siguiendo un cierto eje, son los que mejor explican la dispersión de los perfiles sobre este eje. De esta manera podemos establecer las clases más importantes de ocupaciones y diagnósticos que explican los dos primeros ejes:

En la primera componente la clase II de ocupaciones explica la primera dispersión de los perfiles de las ocupaciones, y la clase VI de diagnóstico, la de los diagnósticos. Analizando la cercanía entre estas dos clases notamos que están siempre asociadas y relativamente alejadas del origen, lo que permite afirmar que la primera explicación de dispersión de los perfiles de las ocupaciones es debido a las enfermedades del tipo VI, es decir mentales de origen orgánico. Y viceversa, la ocupación del tipo II explica mayormente la dispersión de los perfiles ocupacionales de los diagnósticos, la importancia capital de estas clases queda de manifiesto al constatar que el resto de las clases en el primer plano de correspondencia tienden a aglomerarse a lo largo del segundo eje.

La segunda componente aporta una precisión más. La oposición de las ocupaciones tipo VI y I parece mostrar una oposición entre ocupaciones de tipo manual especializado y de esfuerzo físico a las

profesionales. La oposición entre los diagnósticos V, II y VI es más difícil de interpretar. Sin embargo, la clara oposición entre las ocupaciones tipo I y diagnósticos II, y las ocupaciones tipo II con la de diagnósticos VI parece indicar una oposición entre enfermedades debido al riesgo.

Es importante notar que las ocupaciones tipo III y los diagnósticos tipo I juegan el papel de promedio entre ocupaciones y diagnósticos respectivamente, pues al examinar otros planes de correspondencias en el que se incluyen la tercera componente principal siempre estas aparecen muy cerca del origen.

5.2 Caso de las mujeres

La clasificación de las ocupaciones para las mujeres se muestra en el cuadro 4. Los resultados significativos son:

- Notamos la presencia de una clase masiva de ocupaciones muy heterogéneas que van desde profesionales hasta artesanas, operarias y oficinistas. Está encabezada por ocupaciones relacionadas con el servicio doméstico, la enseñanza de labores de oficina, las cuales resultan comunes en el espacio laboral en el que se desenvuelven la mayoría de las mujeres. Además son todas ocupaciones más bien típicas del medio urbano.
- Las ocupaciones típicas del medio rural se agrupan en una sola clase, la III.
- Los puestos de gran responsabilidad y de alta formación académica se agrupan en una sola clase la II.
- El grupo más vulnerable desde el punto de vista socioeconómico es el de la clase V (sin ocupación u ocupación mal definida).
- Existe una clase muy pequeña y singular compuesta de religiosas, (clase IV). Su estilo de vida podría explicar el hecho de que se les agrupe en una clase distinta.

El cuadro 3 muestra la clasificación de diagnósticos. Observamos lo siguiente:

- La clase más claramente definida es la que agrupa a los trastornos mentales (clase V).
- Las enfermedades asociadas claramente a características del medio ambiente y sociales se reúnen en dos clases (III y IV). Es de notar que la IV constan de un sólo diagnóstico (traumatismo superficial), lo que deja suponer que dicho

CLASIFICACION DE DIAGNOSTICOS MUJERES, PERIODO 1980-1983

NOMBRE RESUMIDO CIE	EFECTIVO	CLASE I	
		PORC	PORC ACUM
Otros del aparato genital femenino	21978	14%	14%
Transtornos mama, inflam. órganos pelv. endometriosis	11521	7%	21%
De las venas, vasos linfáticos y otras	9082	6%	27%
V2	8736	6%	33%
Pulmonar obstructiva crónica	7053	5%	37%
Otras aparato digestivo	6398	4%	42%
Tumores benignos	5670	4%	45%
Dorsopatías	4892	3%	48%
Otras glándulas endocrinas	4683	3%	51%
Otras aparato urinario	4430	3%	54%
Transtornos neuróticos, personalidad	4367	3%	57%
Hernia cav. abdom., enteritis y colitis no infec.	3816	2%	59%
Carcinoma in situ, evolución incierta	3780	2%	62%
Hipertensión	2518	2%	63%
Del esófago, estómago, duodeno	2474	2%	65%
De las glándulas tiroides	2467	2%	67%
Causas externas no espec., complicac. atenc. médica	2439	2%	68%
Artropatías y afines	2392	2%	70%
CLASE II			
Tumor maligno tejido linfático	666	79%	79%
CLASE III			
Efectos tóxicos sustancias no medic.	448	27%	27
Herida miembro superior	445	27%	53
Herida del miembro inferior	273	16%	70
CLASE IV			
Traumatismo superficial	60		
CLASE V			
Psicosis orgánica	5792	64%	64%
Otras sistema nervioso central	1960	22%	86%

En esta clasificación se suprimieron los diagnósticos de código: 65, 64, 66, 67, 63

CUADRO NO. 4

CLASIFICACION DE OCUPACIONES MUJERES, PERIODO 1980-1983
--

			CLASE I
Cocineras, sirvientas y criadas	17294	16%	16%
Profesoras y maestras	14097	13%	30%
Otras oficinistas	10433	10%	39%
Mecanógrafas, taquígrafas, operadoras máquinas de oficina	8835	8%	48%
Porteras, conserjes, ascensoristas, limpieza	8669	8%	56%
Cocineras, meseras, Camareras	6335	6%	62%
Modistas, costureras, confección	5586	5%	67%
Dependientes de tienda y vend. ambulantes	4830	5%	72%
<hr/>			
			CLASE II
Empleadas directivas administ. pública	134	100%	100%
<hr/>			
			CLASE III
Trabajadoras agropecuarias y otras agric.	741	79%	79%
Agricultoras-propietarias, en general, espec.	202	21%	100%
<hr/>			
			CLASE IV
Religiosas	140	100%	100%
<hr/>			
			CLASE V
Sin ocupación	2765	92%	92%
Ocupación mal definida	246	8%	100%

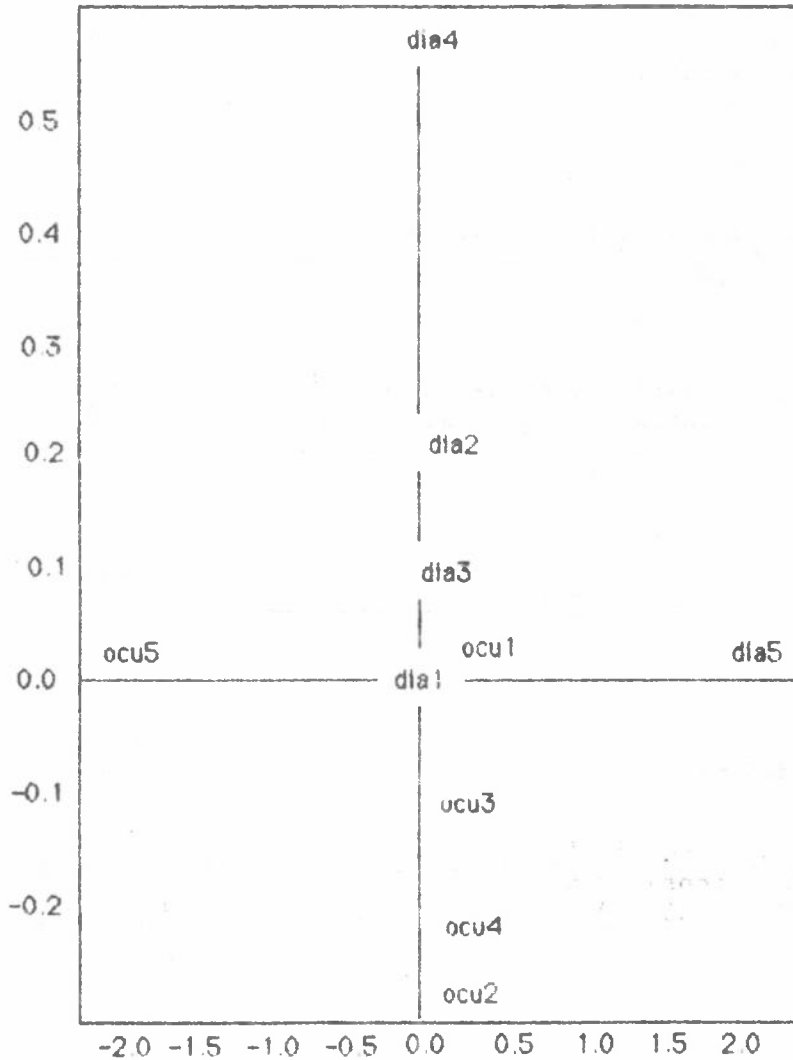
Se suprimieron de esta clasificación las ocupaciones de código siguientes:
Y9 (oficios domésticos), y 67, 2, 25, 90, 91, 78, 72, 69, 32, 63, 66, 76,
24, 65, 44, 83, 73, 45, 68, 50, 40, 11, 77, 80, 64, 43, 71, 81

ANALISIS DE CORRESPONDENCIAS: COMPONENTES
PRINCIPALES 1 Y 2

CENTROS DE GRAVEDAD DE CLASES DE
DIAGNOSTICO Y OCUPACION, MUJERES 80-83

Eje 1: Componente principal 1
Eje 2: Componente principal 2

EJE 2



EJE 1

diagnóstico está asociado a ocupaciones muy específicas que los distinguen de otras enfermedades del medio ambiente.

- Como en las ocupaciones, notamos la presencia de una clase muy masiva de enfermedades, muy heterogéneas, que hace difícil su interpretación.

Examinemos ahora el análisis de correspondencias para mujeres que aparecen en el diagrama 2, que es el primer plano de correspondencias de las dos primeras componentes principales. La explicación de estos ejes se resumen en el cuadro:

# de eje	Clase de ocupación	Clase de diagnóstico
1	5	5
2	2	4

Las explicaciones de los ejes resultan en este caso más sencillas que en la de los hombres, pues las clases mostradas en el cuadro son fáciles de interpretar, según se ha visto. Sin embargo los planos de correspondencias tienen ahora la particularidad de que es difícil establecer asociaciones entre diagnósticos y ocupaciones (notar la aglomeración a lo largo de cada eje, y la figura en forma de L de las nubes de ocupaciones y diagnósticos en este plano). Podemos notar el rol importante jugado por las clases de efectivo pequeño (religiosas, empleadas directivas de la administración pública). En el caso de la clase II podría explicarse la singularidad de su perfil por tratarse de ocupaciones muy emparentadas con las de los varones.

Al realizar otros planos de correspondencias notamos que en las clases masivas de ocupaciones y diagnósticos (I y I en ambos casos), se encuentran siempre muy cerca del origen de coordenadas. Podemos entonces considerarlas como las clases de perfil promedio.

5.3 Comparación entre los sexos

La analogía más importante entre los sexos es la de la explicación del primer eje principal. En ambos casos se hallan presentes las enfermedades de tipo mental y las ocupaciones mal definidas (con ciertas variaciones), lo que muestra la importancia de estas clases para explicar el primer factor de desviación a la media en la población total.

A partir del segundo eje comenzamos a notar diferencias interesantes. Mientras que en los hombres la explicación del segundo eje es atribuirle a varias clases de ocupaciones y diagnósticos, en las mujeres parece ser determinada por un tipo muy específico de clase. Las enfermedades mentales siguen siendo

importantes para el segundo eje en el caso de los varones, mientras que para las mujeres es una clase muy pequeña la que juega ese papel.

Lo más notable es que el tipo de asociaciones entre diagnósticos y ocupaciones que aparecen en los hombres no aparezcan en las mujeres. Tal vez se pueda explicar a la reciente y parcial inserción de la mujer en el campo laboral del país, lo que explica también la presencia de grandes clases residuales en este sexo.

VI ANALISIS DE LAS TENDENCIAS

El estudio de las secciones anteriores se ha centrado en el periodo 1980-1983 de las respectivas matrices de contingencia. Con el estudio de las matrices 1984-1987 buscamos detectar cambios de alguna significancia en el segundo periodo, lo que supone entonces un estudio comparativo. Debido a que el concepto de perfil es el que nos ha permitido establecer una estructura de la población en el primer periodo, la atención para efectos comparativos se centrará entonces en los cambios ocurridos en dichos perfiles.

Para que dicha comparación sea factible se decidió mantener en el periodo 1984-1987 las clases de ocupaciones y de diagnósticos obtenidas en el primer periodo. La justificación de esto es la necesidad de identificar los mismos individuos (clases) en ambos periodos, lo que no se hubiera logrado con clasificaciones generadas por cada periodo por separado. La comparación de estructuras se traduce ahora en la comparación de los planos de correspondencias principales entre periodos. Esto tiene sentido por varias razones; la primera, siendo que las componentes principales de cada periodo están determinadas por sus matrices correspondientes. Los movimientos de las clases con respecto a las componentes principales reflejarían de algún modo cambios en la estructura de la población. La hipótesis asumida aquí es que las clases de ocupaciones-diagnósticos se conservan, grosso modo, a lo largo del tiempo, mas no necesariamente su contribución a los planos principales.

6.1 Tendencias en los hombres

El diagrama 3 muestra los representantes de las clases de ocupaciones y diagnósticos en el primer plano de correspondencias generado por la matriz de contingencia Hombres 1984-1987. Se ha procedido de la misma manera que en el periodo anterior, es decir reducción de la matriz original a una cruzando clases y análisis de correspondencias de esta última matriz. Las ocupaciones masivas (códigos 42 y Y6) son clases separadas y se muestran proyectadas en el plano (sin contribuir al análisis propiamente dicho).

La inspección de este plano permite resumir para cada eje, las clases que mejor lo explican:

# de eje	Clase de ocupación	Clase de diagnóstico
1	2, (47)	6
2	2, (42, 47)	2, 5, 6 <--->3

(la flecha indica oposición entre grupos de clases en los ejes correspondientes).

DIAGRAMA 3

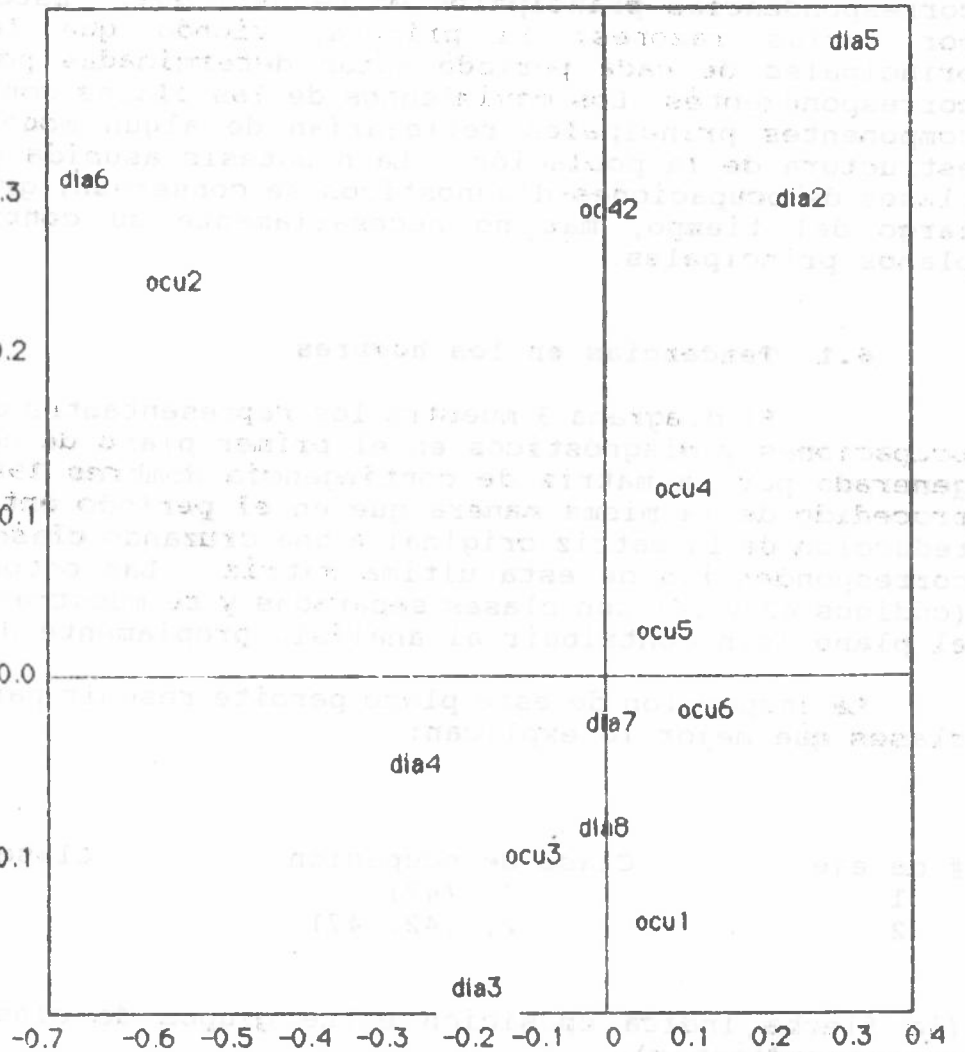
Análisis de correspondencias : componentes principales 1 y 2

Centros de gravedad de clases de diagnóstico y ocupación , hombres 84-87

Eje 1 : componente principal 1

Eje 2 : componente principal 2

EJE 2



EJE 1

puntos superpuestos : día 1 : eje 1 = 0.06, eje 2 = 0.02

Un primer hecho que llama la atención es la situación de la clase de los desocupados (código V6) proyectada en los planos. Esta tiende a diferenciarse marcadamente de las demás en los primeros ejes tal como lo muestra el valor de sus coordenadas. Su posición con respecto al primer eje contrasta con la del primer período en la que incluso no aparecía como una clase de particular contribución. Media aquí sin embargo, la dificultad de explicar la interpretación de una clase que ha sido proyectada, por lo que el resultado debe ser tomado con precaución. Retomaremos este caso después de discutir el aporte de las demás clases.

En este segundo período constatamos que la explicación de la primera componente principal sigue siendo determinada por la clase II de ocupaciones (jornaleros NCEOG, ocupaciones más definidas) y la clase VI de diagnósticos (enfermedades mentales de origen orgánico), así como sus cercanías en todos los planos de correspondencias. Esta constancia del primer eje permite sin ambigüedad concluir que estas clases son la que mejor explican las dispersiones de los perfiles en ambos períodos.

Estas clases aparecen también como explicativas del segundo eje. Es de notar aquí la preponderancia de la ocupación II para la explicación del segundo eje que parece no haberla tenido en el primer período, donde la explicación la ofrecían otras clases de ocupaciones. Se podría deducir una importancia cada vez más grande de la clase II desplazando con esto a otras clases. En cuanto a los diagnósticos la explicación del segundo eje es similar a la del primer período.'

Estos cambios podrían sugerir la aparición de nuevas correspondencias entre ocupaciones y enfermedades, así como la desaparición de otras del primer período. Fuera del mantenimiento de la asociación clase II- clase VI antes mencionada hallamos pocos casos de asociaciones que se mantengan en el segundo período. Este fenómeno puede explicarse geoméricamente al constatar que las nubes de clases de ocupaciones en diferentes períodos son simétricas entre sí con respecto al primer eje, mientras que las nubes de clases de diagnósticos parecen estables.

Finalmente es interesante referirse a las clases que juegan el papel de "promedio". En el primer período habíamos encontrado cerca del origen a las ocupaciones III y diagnósticos I. En este segundo período se les suman los diagnósticos VIII y las ocupaciones V y VI. Esta tendencia parece indicar una mayor concentración de las clases alrededor de un promedio hecho que es correlativo al mayor peso explicativo que adquieren otras según se observara anteriormente. ¿Sería esto acaso un signo de polarización progresiva de los perfiles?

6.2 Tendencias en las mujeres

El diagrama 4 muestra el primer plano de correspondencia de la matriz Mujeres 1984-1987, representando las clases de ocupaciones y diagnósticos. Las clases son representadas por sus centroides. Como en el primer período la interpretación de este plano de correspondencias reveló ser más difícil que el de los hombres, agregamos en este caso los cuadros de los valores de las inercias explicadas por los ejes, así como los valores de los índices de bondad de representación de las clases con respecto a estos (cuadros 5 y 6). El índice de bondad de representación, denotado COS, mide para cada variable y para cada eje principal la calidad de la representación de esta variable con respecto a este eje. En el cuadro 6, QL 15 denota la suma de los $COS^2 \cdot 100$ en proyección sobre los 15 primeros ejes.

La explicación de los primeros ejes queda resumida así:

# de eje	Clase de ocupación	Clase de diagnóstico
1	5	4
2	4	4 <---> 5

Notamos en primer lugar que el papel de clase promedio jugado por las clases I de diagnósticos y I de ocupaciones se mantiene de nuevo en este período. Esto se deduce por sus cercanías al origen.

En las ocupaciones es notable constatar que la distribución en forma de L de la nube de ocupaciones se mantiene en este período, a condición de efectuar una rotación adecuada de los ejes. Lo mismo pasa con la posición relativa entre las clases. Esto explica que la clase V siga determinando la desviación más importante en el primer eje que además muestra un COS muy alto según se observa en el cuadro 6. La mayor diferencia se manifiesta en el segundo eje dada la importancia que adquiere ahora la clase IV para su explicación. La inercia explicada acumulada siendo del 75% para los dos primeros ejes, es un signo de que estos ejes ofrecen una buena imagen de la matriz original y por ende la interpretación dada puede ser considerada satisfactoria.

La estructura de las clases de diagnósticos ha cambiado por la relevancia que ha adquirido la clase IV, desplazando en importancia a la clase V. Es posible incluso notar ahora una oposición entre estas dos clases en el segundo eje. Globalmente existe por lo tanto un mayor dominio de las clases IV y V. El análisis de los diagnósticos debe sin embargo ser tomado con cautela por la aparente débil representación de las clases en los ejes (48% de inercia acumulada explicada). La explicación es particularmente deficiente para la clase IV (73% de QL 15).

En contraste el primer período es posible, sin embargo efectuar asociaciones entre clases de ocupaciones y diagnósticos. Así por ejemplo, la clase IV de ocupaciones parecer asociarse con

DIAGRAMA 4

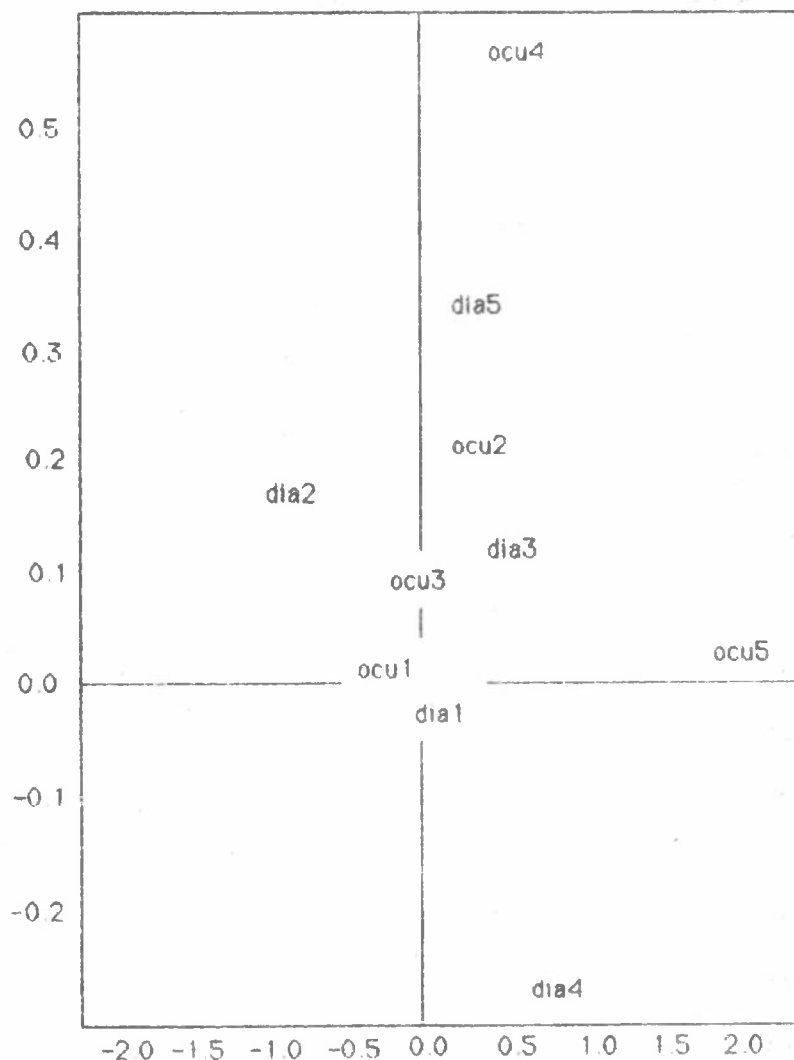
ANALISIS DE CORRESPONDENCIAS: COMPONENTES PRINCIPALES 1 Y 2 (PERIODO 1984-1987)

CENTROS DE GRAVEDAD DE CLASES DE DIAGNOSTICO Y OCUPACION, MUJERES

Eje 1: Componente principal 1

Eje 2: Componente principal 2

EJE 2



EJE 1

CUADRO NO. 5

3 PRIMEROS VALORES PROPIOS E INERCIA EXPLICADA
DIAGNOSTICOS Y OCUPACIONES
MUJERES, PERIODO 1984-1987

	DIAGNOSTICOS			OCUPACIONES		
	VALOR PROPIO	%INERCIA	%ACUMULADO	VALOR PROPIO	%INERCIA	%ACUMULADO
1	0.21	37.1	37.1	0.29	65.06	65.06
2	0.06	11.08	48.18	0.03	5.58	70.64
3	0.04	6.98	55.16	0.02	4.56	75.2

CUADRO NO. 6

100*COS
MUJERES, PERIODO 1984-1987

	COS 1	COS 2	COS 3	QL15
DIA 1	3.52	28.19	12.89	97.33
DIA 2	44.26	6.24	14.60	97.01
DIA 3	16.56	3.37	15.81	98.30
DIA 4	25.11	3.72	0.29	77.07
DIA 5	3.00	15.81	53.92	99.33
OCU 1	99.47	0.03	0.24	99.96
OCU 2	1.35	2.46	2.61	98.44
OCU 3	0.03	0.82	72.30	98.70
OCU 4	0.01	5.78	7.62	73.83
OCU 5	99.91	0.00	0.05	99.98

la V de diagnósticos. Algunas explicaciones sobre la explicación de los ejes podrían derivarse de estas asociaciones, pero su importancia es discutible dada la débil representación de los diagnósticos.

6.3 Comparación de las tendencias según el sexo

Tratemos de resumir los hechos más relevantes de los apuntados en 5.1 y 5.2, estableciendo semejanzas y divergencias entre los sexos. Estas comparaciones deben tomarse con precaución por la forma diferente en que se conformaron las clases para cada sexo y dada a diferente calidad de representación en los planos.

No se puede dejar de mencionar que la tendencia común más clara es la de la importancia creciente de ciertas clases de ocupaciones para la explicación de los ejes. En los hombres hemos mencionado la clase II y en las mujeres las clases V y VI. La proyección de la clase de desocupados mostró que esta clase es igualmente importante en el caso de los hombres en el segundo período. En ambos sexos se trata en realidad de los grupos más marginales de las fuerzas productivas, los que explican progresivamente mejor las diferencias globales del cuadro de morbilidad de la población en estudio. En el caso de las mujeres hay que agregar un grupo relativamente pequeño (religiosas), cuyo modo de vida podría explicar su singularidad de perfil de morbilidad.

La estructura de los diagnósticos parece relativamente estable en los varones, no siendo así en las mujeres. En el caso de las mujeres este fenómeno podría ser vinculado con la débil explicación de los diagnósticos. El resultado de los varones demostraría la buena explicación de los diagnósticos en el plano de correspondencias.

Las clases promedio e mantienen en las mujeres mientras que en los hombres tienden a aparecer otras nuevas.

La significancia de estos resultados de tendencias debe ser evaluados según la calidad de las clases obtenidas para cada sexo. La caracterización relativamente satisfactoria de las clases en los varones contrasta con la presencia de clases masivas en las mujeres, casi residuales. De ahí la limitación de los alcances de una interpretación para el sexo femenino, que hace solo resaltar los cambios para grupos pequeños, excepción hecha de las desocupadas.



VII CONCLUSIONES

El estudio multivariado sobre los datos brutos hospitalarios ha revelado ciertos resultados sobre fenómenos no advertidos en los estudios preliminares realizados sobre el mismo tema, más concretamente, los estudios exploratorios comentados en la sección II. Es necesario entonces aclarar en qué sentido los resultados son novedosos.

La principal ventaja de los métodos multivariados es la de poder establecer en un cúmulo de datos empíricos los aspectos más relevantes de dichos datos. Es lo que originalmente, y en términos muy generales, llamamos estructura de los datos y cuyo significado fue aclarado en el transcurso del análisis. Gracias a la formalización matemática del concepto perfil, tantas veces empleado en la literatura pero en forma ambigua, se ha logrado traducir en términos operacionables, las variables más importantes del fenómeno en estudio.

Un logro importante es el de haber reducido al mínimo las hipótesis a priori. Esto se reflejó, primeramente, en la clasificación jerárquica de las clases de ocupaciones y diagnósticos, en la que no se excluyó a priori ninguna de las variables, aun las de muy débil efectivo. Es solo después de los primeros análisis que se efectúan algunas exclusiones. Los métodos multivariados fueron además valiosos en la búsqueda de una interpretación de las clases. Gracias a la inspección de los planos de correspondencias se llegó a concluir, que sin lugar a dudas, la clase que reúne a las enfermedades mentales de origen orgánico por un lado y la que reúne a las ocupaciones más definidas por otro son las que mejor explican las dispersiones de los perfiles, en ambos sexos. Se determinó además, que estas clases estaban muy asociadas, lo que permite caracterizar una por medio de la otra. Una diferencia esencial entre los sexos es el hecho de que la interpretación de las clases de los varones sea más nítida y satisfactoria que en el de las mujeres en las que es inevitable la presencia de clases masivas y por otro lado de clases de efectivo pequeño que resultan ser las más explicativas. Esto parece confirmar la tesis del diferente grado de inserción en el campo laboral de los sexos, tesis que se había planteado en los estudios exploratorios.

El alcance de los estudios exploratorios son mucho más limitados, pues desde el momento en que el análisis de los datos se efectúa sobre tablas de efectivos totales es inevitable la exclusión de muchas ocupaciones y diagnósticos que, aunque de efectivo débil, resultaron ser finalmente muy explicativas. Por otro lado, la inspección somera de estas tablas no permite establecer clases a falta de un cruce entre las variables. Así por ejemplo en el caso de los egresos femeninos era de esperar que las causas de tipo ginecológico tuvieran un peso muy importante en cuanto a efectivo total, pero no muestra en absoluto cuáles son las causas de egresos que mejor explican diferencias a la media global.

La omisión es todavía más clara en el caso de las ocupaciones puesto que las estadísticas tienden a agregarlas por ramas de actividad, es decir asumiendo una clasificación a priori de las ocupaciones que no es necesariamente la más conveniente para los objetivos del estudio. El estudio multivariado confirmó la hipótesis sobre la importancia de la variable ocupación como criterio diferenciador entre los sexos, que se destaca de la variable actividad económica, empleada en los estudios exploratorios.

El estudio multivariado comprendió, además, un análisis de tendencias, que se concretó en la comparación de los planos de correspondencias principales entre periodos. Aunque los alcances aquí son más limitados a falta de una metodología más desarrollada, debe mencionarse el fenómeno del progresivo poder explicativo de ciertas clases en la dispersión de los perfiles, que puede interpretarse como un fenómeno de polarización de las clases.

El trabajo que hemos presentado puede ser enriquecido con la incorporación de otros datos y técnicas analíticas. Se podría por ejemplo pensar en: la inclusión de los datos de egresos de épocas más antiguas a fin de obtener un panorama temporal más amplio del fenómeno salud-fuerza laboral; estudiar la interrelación entre otras variables (residencia, nivel de escolaridad, condición de asegurado); considerar otras fuentes de datos (encuestas de morbilidad de la CCSS); aplicación de otros métodos del análisis multivariado como los recientes estudios sobre comparación de tablas, etc. Una posibilidad interesante sería en la actualización de dicho estudio con los datos que periódicamente ofrece la CCSS sobre egresos hospitalarios, permitiendo una revisión continua y actualizada de los problemas que nos hemos planteado.

El interés de este tipo de estudio puede valorarse desde diferentes ángulos. Por un lado se podrá valorar mejor los aportes y alcances de los métodos multivariados en el estudio de datos complejos como los de las ciencias sociales. El beneficio obtenido con la aplicación de estos métodos es la de un conocimiento más exacto y profundo del fenómeno en estudio, que como en nuestro caso, no se hubiera logrado con la simple confección de tablas estadísticas. Finalmente dicho conocimiento podrá ser aprovechado de alguna manera por los responsables de las políticas de salud de nuestro país ayudándoles a replantear y modificar algunos programas actuales.

APENDICE

Ciertas nociones del análisis multivariado

Explicamos aquí muy someramente algunas nociones del análisis multivariado que empleamos en el estudio. Para exposiciones más claras y extensas el lector puede referirse a la vasta literatura que existe sobre el análisis multivariado. Recordemos [3] y [4] por su valor pedagógico, y a [5] y [6] por contener una exposición completa del tema.

Tablas de contingencia, perfiles y análisis de correspondencias

Sobre una población de n individuos se dispone del censo de dos caracteres cualitativos X , Y (por ejemplo los caracteres de la ocupación y del diagnóstico en nuestro estudio). Dada una codificación de los caracteres X , Y (q códigos para X y p para Y), el censo sobre los dos caracteres se representa bajo la forma de tabla de frecuencia: la casilla (i,j) de esta tabla está dada por el cociente $f(i,j) = (\text{número de individuos pertenecientes a la categoría } i \text{ de } X \text{ y la } j \text{ de } Y) / n$ o sea la frecuencia relativa de la casilla (i,j) de la matriz original., La frecuencia marginal del carácter X está dada por los números $f(i) = f(i,1) + \dots + f(i,p)$, $i=1, \dots, q$. Las frecuencias relativas del carácter Y entre los individuos de la clase i de X es dada por los cocientes: $f(i,j)/f(i)$, $j=1, \dots, p$. De esta manera se obtienen q perfiles relativos del carácter Y y p del carácter X . En el caso concreto de nuestro estudio el concepto de perfil permite traducir el concepto de similitud de ocupaciones según su perfil de morbilidad y de la misma manera para los diagnósticos. En el caso de matrices de dimensión 10×100 como las de nuestro estudio hallamos 100 perfiles de diagnósticos (uno por ocupación) y 100 de ocupaciones (uno por diagnóstico). Puesto que un perfil es en realidad una frecuencia relativa, la similitud entre dos tipos de diagnóstico se traduce en la similitud de sus perfiles ocupacionales respectivos. Es por esto que se representa por un vector en un espacio euclídeo de cierta dimensión y la disimilitud de perfiles se expresa aquí en términos de distancia entre dos vectores. Como es ya usual para el caso de tablas de contingencia, tomamos como definición de distancia la del chi-cuadrado (o índice del chi-cuadrado). La representación de las ocupaciones así como la de los diagnósticos se hace por medio de una nube de puntos en un espacio de cierta dimensión.

En el contexto de este trabajo se explotan dos propiedades del análisis factorial de correspondencias. La primera se refiere a la posibilidad de representar en un plano, vectores (diagnósticos y ocupaciones) que se encuentran en espacios de alta dimensión donde el ojo humano no puede "ver". La segunda se refiere al hecho de que un análisis factorial de correspondencias puede ser usado como etapa previa a una clasificación de Ward. Ambos objetivos han de ser obtenidos con la optimización de un criterio de calidad, que en nuestro vaso mide la dispersión total (o inercia total) de la nube

de perfiles (fila o columna). Para medir las proximidades entre perfiles se utiliza la bien conocida distancia euclídea del chi-cuadrado, de modo tal que podemos establecer las siguientes definiciones y propiedades:

1. El perfil promedio: denotado por $x(i)$ el perfil correspondiente a la clase i de x , se define el perfil medio de los perfiles $x(i)$, $i=1, \dots, q$ como el baricentro de estos perfiles afectados de los pesos $f(1), f(2), \dots, f(q)$ respectivamente.
2. Inercia total (o dispersión total): para cada perfil $x(i)$ se define la distancia del chi-cuadrado entre este y el perfil medio. La dispersión o inercia total de los perfiles $x(i)$ sería entonces un cierto promedio ponderado de estas distancias.
3. El criterio de calidad: dicho intuitivamente y evitando la formulación matemática, los vectores originales, es decir los perfiles relativos a diagnósticos y ocupaciones, son proyectados perpendicularmente (en el sentido de la distancia chi-cuadrado), sobre rectas por el origen, planos por el origen, o sobre subespacios de más alta dimensión, de modo tal que la dispersión de los perfiles así proyectados, sea máxima.
4. El primer eje principal (o factor): La representación más simple que podemos hacer, de un conjunto de puntos, como los perfiles de diagnóstico o de ocupaciones, es sobre una recta. Se justifica desde un punto de vista matemático que existe una recta única (o una dirección única, salvo por la orientación), de modo que los puntos que se obtienen al proyectar perpendicularmente los perfiles sobre dicha recta, tienen dispersión máxima entre todas las rectas que pasen sobre el origen. En este sentido decimos que es la dirección de máxima dispersión de la nube original de puntos.
5. Segundo eje principal (o segundo factor) y primer plano principal (o primer plano factorial): la práctica habitual es representar los perfiles en un plano (donde el ojo humano puede ver), para estudiar las proximidades entre ellos. Este primer plano principal es también óptimo en el mismo sentido que el primer factor principal. Es decir, entre todos los planos que pasan por el origen, el plano principal es aquel sobre el cual los perfiles se proyectan (perpendicularmente) con máxima dispersión. Se puede probar matemáticamente que para determinar este plano es suficiente determinar la recta de máxima dispersión, perpendicular al primer eje principal, que pasa por el origen. Este es el segundo eje principal y el plano determinado por el primer y segundo ejes principales, es el primer plano principal.

6. Los siguientes ejes principales (tercero, cuarto, etc.), se determinan de modo análogo. Es decir el tercer eje principal es la recta que maximiza la dispersión de los perfiles perpendicularmente proyectados sobre ella, entre todas las rectas perpendiculares al primer plano principal, que pasan por el origen. Y así sucesivamente.

Siendo así, podemos asegurar con una representación plana (plano de correspondencias), en un par de ejes principales una imagen muy fiel de la nube original. La fidelidad de la representación puede ser medida por el porcentaje de inercia inicial. Los planos de correspondencias ofrecen además la posibilidad de representar simultáneamente ambas variables, diagnóstico y ocupación, permitiendo en algunos casos establecer relaciones entre estas dos variables. Se obtiene además el siguiente resultado de dualidad: las inercias de los caracteres X, Y con respecto a sus r primeros ejes principales, (r no siendo mayor a ninguna de las dimensiones de los espacios de perfiles de X y Y), son iguales. Gracias a esto en los planos de correspondencias la representación simultánea de ambos perfiles se da con la misma fidelidad. Se pueden establecer en estos planos las semejanzas entre los perfiles de X y de Y siguiendo reglas muy estrictas. Así, las proximidades entre puntos son tanto más significativas cuanto más alejados estén del origen. En la periferia del gráfico podemos tener una idea del perfil de Y examinando la posición relativa de este último con respecto a los perfiles de X y viceversa.

Clasificaciones jerárquicas aglomerativas

El análisis de correspondencias lo hemos usado como un paso previo a la realización de una clasificación jerárquica de las variables. Primeramente se determina un cierto número de ejes principales y en estos la nube original se representa con pérdida mínima de "información", donde es posible aplicar un algoritmo que realice en ambas una clasificación en clases bien distinguidas y separadas. En el presente caso hemos optado por un método llamado jerárquico aglomerativo en el que los grupos o clases se aglomeran siguiendo la escogencia de un índice de aglomeración y los grupos se jerarquizan en forma de árbol binario. Un índice idóneo para este tipo de estudios es el de Ward, el cual adoptamos. La jerarquía producida estará además afectada de una indexación que mide algo así como la "importancia" de cada clase en el árbol. En esta etapa de la clasificación es posible mejorar la clasificación escogida después de la aplicación del algoritmo de nubes dinámicas. La estabilidad de la clasificación inicial puede medirse por la similitud encontrada con las nubes dinámicas. Damos a continuación una formulación más precisa de estos conceptos.

Representaciones arbóreas y jerarquías: En las clasificaciones jerárquicas las representaciones se presentan en la forma de árboles. Una jerarquía sobre un conjunto de objetos $\Omega = \{1, 2, \dots, n\}$ es un conjunto de subconjuntos T de Ω satisfaciendo las condiciones: $\Omega \in T$, $\emptyset \in T$, $\{i\} \in T$ para todos los $i \in \Omega$, y $A \cap B \in \{\emptyset, A, B\}$ para todos $A, B \in T$. Un ejemplo de árbol de una jerarquía es el del gráfico 1, donde cada uno de los nodos internos especifica un subconjunto no siguiente de T comprendiendo todos los objetos que se hallan bajo este nodo en el árbol. Un dendograma, o jerarquía indexada, es una jerarquía T provista además de una función h de altura sobre los nodos internos del árbol, donde si $A \cap B \neq \emptyset$, se tiene $h(A) \leq h(B)$ si y solo si A está contenido en B . Para cada par de objetos (i, j) , $h(i, j)$ es definido como la altura del nodo interno especificando la clase más pequeña en la cual los dos objetos pertenecen; cuanto más pequeño es el valor de $h(i, j)$ mayor será la similitud entre los objetos i, j . El árbol del gráfico 1 es además binario, en el sentido de que cada nodo interno posee exactamente dos ramas saliendo de él.

Algoritmos aglomerativos de clasificación: Para obtener una jerarquía indexada sobre el conjunto de objetos (en nuestro caso el conjunto de ocupaciones o bien el de diagnósticos), en este estudio se utilizó un algoritmo de tipo aglomerativo, el de Ward. En un algoritmo aglomerativo inicialmente hay n clases de un solo objeto cada una. En cada paso del algoritmo los pares de clases más similares son amalgamadas; este par y por lo tanto la jerarquía producida no son necesariamente únicos. El criterio para aglomerar clases es el de calcular en cada paso del algoritmo el índice de aglomeración $\delta(n, n')$ para cada par de clases n, n' , y el de escoger un $n \cup n'$ como nueva clase de la jerarquía para el par que minimiza $\delta(n, n')$ entre todas las clases. En el método de Ward el índice de aglomeración es definido por $\delta(n, n') = I(n \cup n') - I(n) - I(n')$, donde $I(\cdot)$ denota la inercia de un grupo de objetos. Puesto que la jerarquía producida por dicho método es binaria, es posible construir una indexación para la jerarquía T a partir del índice de aglomeración δ mediante el esquema recursivo siguiente: $h(\{i\}) = 0$ para $i \in \Omega$, $h(n) = (n', n'')$ para todo n', n'' tales que $n = n' \cup n''$. Se llama la indexación producida por la agregación δ .

Algoritmo de nubes dinámicas

El algoritmo de nubes dinámicas es un algoritmo de clasificación por partición, y su fin es el de obtener clases de objetos bien agregados, separados y de intersección vacía. necesita primero la definición de un modo de representación simbólico de todo grupo de objetos. Este modo de representación es llamado "núcleo". Cuando se tiene un gran número de puntos, el algoritmo parte de una tirada al azar de k núcleos; cada punto de la población es enseguida asociado al núcleo que le es más cercano. Se obtiene así una partición en K clases cuyos núcleos son ahora

calculados. Se comienza de nuevo el procedimiento con los nuevos núcleos y así sucesivamente. Se demuestra que bajo ciertas condiciones el algoritmo converge a una posición estacionaria mejorando en cada iteración un criterio dado, lo que garantiza que se obtenga una clasificación estable. En nuestro estudio se ha utilizado este algoritmo de dos maneras: en una, a partir de una clasificación previa en k grupos, se calculan los centros de gravedad de cada grupo que constituirán los núcleos; en otra, se parte de una tirada al azar de k núcleos y, utilizando la intersección de varias particiones así obtenidas, cambiando la tirada de partida, se obtiene los "tipos" o "formas fuertes".

BIBLIOGRAFIA

- [1]: Mayra Achío, Jaime Lobo: "Apuntes para un estudio de la morbilidad de la fuerza de trabajo en Costa Rica", en avances de investigación, NO.77, año 1991, publicado por el Instituto de Investigaciones Sociales, Universidad de Costa Rica
- [2]: Mayra Achío: "Fuerza de trabajo y enfermedad en Costa Rica", publicado por la CCSS, mayo 1992
- [3]: Diday E., Lebart L.: "L'analyse des donnés", revista La Recherche, No. 74, enero 1977
- [4]: Cibois Philippe: "L'analyse factorielle", colección Que sais-je, Presses Universitaires de France, No. 2095, 1983
- [5]: Cooley, Lohnes: "Multivariate Data Analysis", Wiley, 1971

Impreso en el Taller del
INSTITUTO DE INVESTIGACIONES SOCIALES

Responsable: Jorge Oconitrillo