

ISSN 0378 052 X

Eduardo Piza

Editor científico

Métodos Matemáticos



aplicados a las Ciencias

V y VI Simposio

Indice

Presentación	9
A la memoria de Guy Der Mégréditchian	11
1 El Cielo y la Tierra: un enfoque constructivista del fenómeno de la opinión en Francia (Jean Pierre Pagès)	13
1 Las ideas y los hechos	14
2 Un enfoque constructista y sistémico de la opinión	18
3 El análisis de la diferenciación: recurso al modelo factorial y al análisis en componentes principales	28
4 La encuesta de referencia	33
5 Una interpretación a varios niveles de las primeras dimensiones factoriales	37
2 Comunicación entre encuestas mediante el método del injerto (Carlos Arce)	47
1 Introducción	47
2 Términos básicos	47
3 Criterios para la selección del puente	49
4 Selección optimal de las variables del puente	51
5 Representación de individuos de <i>B</i> en los planos de <i>A</i>	52
6 Ejecución del algoritmo pasarela	53
7 Evaluación y perspectivas	55
3 Ajuste estructural, salud y opinión pública (Nora Garita, Marta Eugenia López Subirós)	57
1 Introducción	57
2 Ajuste estructural y salud	57
3 Estado benefactor y salud	58
4 Problemática actual	60
5 Observaciones finales	63

4	La opinión pública regional y la política económica en Costa Rica, 1988 (<i>Marta Eugenia López Subirós</i>)	65
1	Aspectos teóricos y metodológicos	65
2	La crisis y la política económica, 1979-1986	66
3	La opinión pública y la política económica en 1988	68
4	Análisis en componentes principales (ACP)	71
5	Conclusiones	72
5	Clasificación automática: grupos de opinión pública en Costa Rica (<i>William Castillo</i>)	77
1	Introducción	77
2	Particionamiento de la muestra	78
3	Sobre las variables utilizadas	79
4	Examen interpretativo de las clases de la partición encontrada	81
5	Conclusiones	87
6	Anexos	89
6	Métodos y resultados en Clasificación Automática (<i>William Castillo</i>)	105
1	Notaciones y definiciones	105
2	Descomposición de la inercia total	107
3	Agregación de Ward	108
4	Método de las nubes dinámicas	110
7	Análisis multivariado de conceptos en la educación costarricense (<i>Javier Trejos</i>)	117
1	El cuestionario	117
2	Tratamiento previo de los datos	119
3	Análisis de las relaciones entre las variables de opinión	119
4	Confirmación de resultados comparando con otros métodos	123
5	Conclusiones	126
8	Clasificación por particiones (<i>Jose Luis Espinoza, Javier Trejos</i>)	129
1	Introducción	129
2	Métodos de Nubes Dinámicas	130
3	Algoritmo de transferencias	144
4	Estudio de una multipartición	146
5	Clasificación optimal	151
9	Un enfoque unificado de los métodos de la estadística multivariada (<i>Asdrúbal Duarte, Jorge González</i>)	155
1	Introducción	155
2	Definición del coeficiente RV	157
3	Análisis en Componentes Principales (ACP) de X	159

4	Componentes Principales de Y con respecto a X	161
5	Regresión multivariada	163
10 La métrica inducida por un análisis en componentes principales		
	(Eduardo Piza)	167
1	Introducción	167
2	Los datos disponibles	169
3	Algunos resultados teóricos en ACP	169
4	Estudio de la métrica inducida por un ACP	171
5	Consideraciones finales	178

Desde el año 1975 y dentro del marco de la colaboración bilateral entre los Gobiernos de Costa Rica y Francia, la Escuela de Matemáticas de la Universidad de Costa Rica suscribió un convenio de cooperación científica técnica con algunas instituciones académicas y de desarrollo científico francesas.

Desde entonces se han obtenido a partir de este convenio de cooperación. Así por ejemplo, en el campo de la docencia, varios estudiantes y profesores han cursado estudios superiores en Francia. En el plano de la investigación científica, se han establecido relaciones y desarrollado algunas investigaciones de interés nacional con esta asistencia y asesoría de colegas franceses.

Como parte de la difusión de las actividades del convenio de cooperación, se han realizado en nuestro país siete Simposios de Métodos Matemáticos Aplicados a las Ciencias (SMMAC), y algunas otras actividades análogas en el campo de la Estadística de las Matemáticas. Estos Simposios se han realizado de manera con importantes participaciones de investigadores franceses de primer orden, tales como Yves Schekman, Jacques Hadri, Guy Des Méditerranée, Jean Pierre Pages, André Aron, Yves Lechevalier y Edwige Dany.

Este libro contiene los trabajos y artículos más importantes presentados en los pasados V y VI Simposios de Métodos Matemáticos Aplicados a las Ciencias, actividades que se realizaron en San José, en los meses de diciembre de 1988 y noviembre de 1989 respectivamente.

La temática principal desarrollada en los V y VI SMMAC, la impuso el reconocido investigador francés Jean Pierre Pages, del Centro de Estudios Económicos de París. Pages vino a Costa Rica con sus interesantes trabajos en el campo de las estructuras de la opinión pública en Francia, logrando cautivar con sus métodos y sus disertaciones a un grupo importante de sociólogos, quienes junto con el grupo de Análisis de Datos de la Escuela de Matemáticas se lanzaron de inmediato a la investigación de las correspondientes estructuras de la opinión pública en Costa Rica. El entusiasmo que logró infundir el profesor Pages a este grupo permitió que el grupo de investigadores locales ha rendido sus frutos: se han realizado a la fecha varias encuestas de referencia en opinión pública, se han publicado varios artículos y un libro al respecto.

Presentación

Desde el año 1978 y dentro del marco de la colaboración bilateral entre los Gobiernos de Costa Rica y Francia, la Escuela de Matemáticas de la Universidad de Costa Rica suscribió un convenio de cooperación científica-técnica con algunas instituciones académicas y de desarrollo científico francesas.

Grandes beneficios se han obtenido a partir de este convenio de cooperación. Así por ejemplo, en el plano de la docencia, varios estudiantes costarricenses han cursado estudios superiores en Francia. En el plano de la investigación científica, nos hemos fortalecido al plantear y desarrollar algunas investigaciones de interés nacional con colaboración y asesoría de colegas franceses.

Como parte de la difusión de las actividades del convenio de cooperación, se han realizado en nuestro país siete Simposios de Métodos Matemáticos Aplicados a las Ciencias (SMMAC), y algunas otras actividades análogas en el campo de la Enseñanza de las Matemáticas. Estos Simposios se han realizado en su mayoría con importantes participaciones de investigadores franceses de primera línea, tales como Yves Schektman, Jacques Badia, Guy Der Mégréditchian, Jean Pierre Pagès, André Antibi, Yves Lechevallier y Edwin Diday.

Este libro contiene los trabajos y artículos más importantes presentados en los pasados V y VI Simposios de Métodos Matemáticos Aplicados a las Ciencias, actividades que se realizaron en San José, en los meses de diciembre de 1988 y noviembre de 1989 respectivamente.

La temática principal desarrollada en los V y VI SMMAC la impuso el reconocido investigador francés Jean Pierre Pagès, del Centro de Estudios Atómicos de París. Pagès vino a Costa Rica con sus interesantes trabajos en el campo de las estructuras de la opinión pública en Francia, logrando cautivar con sus modelos y sus disertaciones a un grupo importante de sociólogos, quienes junto con el grupo de Análisis de Datos de la Escuela de Matemáticas se lanzaron de inmediato a la investigación de las correspondientes estructuras de la opinión pública en Costa Rica. El entusiasmo que logró infundir el profesor Pagès a este grupo multidisciplinario de investigadores locales ha rendido sus frutos: se han realizado a la fecha varias encuestas de referencia en opinión pública, se han publicado varios artículos y un libro al respecto.

Los primeros cinco artículos del libro tratan sobre el tema de la opinión pública. Para una mejor comprensión de algunos de estos artículos se recomienda leer el libro *Estructuras de la Opinión Pública en Costa Rica*, de J. Poltronieri y E. Piza (Editorial de la Universidad de Costa Rica, 1989), pues algunos análisis se refieren específicamente a resultados allí expuestos. Los siguientes cinco artículos tratan sobre otras temáticas dentro del campo del Análisis de Datos, entre ellas el campo de la Clasificación Automática.

Deseo manifestar mi especial agradecimiento a Jean Pierre Pages y Yves Schektman, quienes me brindaron gran apoyo en la elaboración preliminar de este libro en Francia, así como a Nora Garita, por su muy valiosa colaboración.

Eduardo Piza
San José, julio de 1991.

A la memoria de Guy Der Mégréditchian¹

Guy Der Mégréditchian murió el 13 de febrero de 1990, luego de una larga y penosa enfermedad. Experto en regresión y en análisis discriminante, fue internacionalmente conocido por sus trabajos realizados para el Meteorológico Nacional de Francia, donde dirigía el Laboratorio de Estadística Matemática desde noviembre de 1971. Profesor de la Escuela de Meteorología, enseñó en el Instituto de Estadística de las Universidades de Paris y la Universidad Paris-Dauphine. Participó en la Escuela Internacional de Verano de Meteorología, en Toulouse. Fue electo miembro del Instituto Internacional de Estadística en 1982.

Todos los que admiraron a Guy Der Mégréditchian conocieron las dificultades que este ser excepcional tuvo que afrontar. Su madre, Blanche Gras, oriunda de Puy-en-Velay, murió en 1936, tres años después de su nacimiento. Su padre, oriundo de Van en Turquía, que había huído del genocidio armenio de 1906 y participado en la guerra del 14 como médico en el ejército francés, se reunió con toda su familia en 1947 en Erevan, Armenia Soviética, movido por las exhortaciones de los responsables comunistas que creyeron poder recuperar la parte turca de la Armenia. Fue así como Guy, que en ese momento no dominaba ni el ruso ni el armenio, se encontró "prisionero" en Rusia a la edad de 15 años. Junto a su padre, que trabajó en el hospital de Erevan, descubrió el hambre, la miseria... y la burocracia. Después de obtener el equivalente al bachillerato en 1951, entró a la Universidad de Erevan, donde obtuvo el título de Físico-Matemático con mención de honor en 1956. Este reconocimiento le permitió seguir sus estudios hacia el doctorado en la Universidad de Moscú y en el Instituto Steklov, donde tuvo como profesores a Gnedenko, Kolmogorov, Prokhorov, Smirnov, y Yaglam. Después de la muerte de Stalin en 1953, hizo los trámites para volver a Francia; tuvo que esperar 17 años para alcanzar su

¹Guy Der Mégréditchian fue el principal expositor en el IV Simposio de Métodos Matemáticos Aplicados a las Ciencias (San José, diciembre de 1985). La presente biografía fue elaborada por Jean Pierre Pagès.

A la memoria de Guy Der Mégréditchian¹

Guy Der Mégréditchian murió el 13 de febrero de 1990, luego de una larga y penosa enfermedad. Experto en regresión y en análisis discriminante, fue internacionalmente conocido por sus trabajos realizados para el Meteorológico Nacional de Francia, donde dirigía el Laboratorio de Estadística Matemática desde noviembre de 1971. Profesor de la Escuela de Meteorología, enseñó en el Instituto de Estadística de las Universidades de París y la Universidad Paris-Dauphine. Participó en la Escuela Internacional de Verano de Meteorología, en Toulouse. Fue electo miembro del Instituto Internacional de Estadística en 1982.

Todos los que admiraron a Guy Der Mégréditchian conocieron las dificultades que este ser excepcional tuvo que afrontar. Su madre, Blanche Gras, oriunda de Puy-en-Velay, murió en 1936, tres años después de su nacimiento. Su padre, oriundo de Van en Turquía, que había huído del genocidio armenio de 1906 y participado en la guerra del 14 como médico en el ejército francés, se reunió con toda su familia en 1947 en Erevan, Armenia Soviética, movido por las exhortaciones de los responsables comunistas que creyeron poder recuperar la parte turca de la Armenia. Fue así como Guy, que en ese momento no dominaba ni el ruso ni el armenio, se encontró "prisionero" en Rusia a la edad de 15 años. Junto a su padre, que trabajó en el hospital de Erevan, descubrió el hambre, la miseria... y la burocracia. Después de obtener el equivalente al bachillerato en 1951, entró a la Universidad de Erevan, donde obtuvo el título de Físico-Matemático con mención de honor en 1956. Este reconocimiento le permitió seguir sus estudios hacia el doctorado en la Universidad de Moscú y en el Instituto Steklov, donde tuvo como profesores a Gnedenko, Kolmogorov, Prokhorov, Smirnov, y Yaglam. Después de la muerte de Stalin en 1953, hizo los trámites para volver a Francia; tuvo que esperar 17 años para alcanzar su

¹Guy Der Mégréditchian fue el principal expositor en el IV Simposio de Métodos Matemáticos Aplicados a las Ciencias (San José, diciembre de 1985). La presente biografía fue elaborada por Jean Pierre Pagès.

propósito. Mientras tanto se convirtió en maestro de investigación dentro del Laboratorio de Estadística Matemática de Bagrov en la Meteorología Nacional de Moscú. Tradujo al francés más de 10 libros de matemática y de física de las Ediciones Mir (su primera traducción fue el curso de matemática en 2 tomos de Piskounov). Luego presentó su tesis de doctorado en 1969. Se casó con Lydia Youdokova (1957) y tuvieron dos hijos, Jacques y Philippe, nacidos en Moscú, y fue hasta 1971 cuando finalmente pudo volver, con escasos 3600 francos en el bolsillo, a su patria, Francia.

El Cielo y la Tierra: un enfoque constructivista del fenómeno de la opinión en Francia

Jean Pierre Pagès*

Introducción

Los gráficos que brindan los análisis factoriales, realizados a partir de datos de encuestas de opinión, se parecen unos con otros. Se tiene la impresión de encontrar siempre, más o menos, el mismo resultado. No hay nada importante, podría decirse... siempre las viejas disputas entre antiguos y modernos, dominantes y dominados, izquierda y derecha. Pero las estructuras subyacentes (dimensiones factoriales) parecen tan fuertes que emergen casi de manera idéntica, aún cuando no se hayan tomado precauciones a nivel de la recolección de los datos; se les encuentra en todos los grupos sociales... y, cuando se les ve de cerca, se percibe que se parecen a ciertas dimensiones de la personalidad aisladas por los psicólogos.

Estos hechos sugieren la idea de **encuesta de referencia**, es decir de una encuesta madre que se podría "injertar" dentro del estudio de un problema específico, para recuperar a un costo menor esas dimensiones que parecen ser constantes de la opinión. Uno se imagina el cambio que podría operarse al nivel de las prácticas en el campo de las encuestas, si tal instrumento y una metodología de utilización con cierta credibilidad pudiese efectivamente ponerse a funcionar. Para tener posibilidad de éxito, no es posible situarse contra corriente y hay que operar con el menor costo...

Para llegar a un sistema de encuestas de opinión de referencia que pueda servir de marco general para el análisis de un problema particular, hay que partir de una teoría de la opinión. Para construirla, la reflexión va a desarrollarse más allá de las tentativas existentes, a partir de dos cuerpos que describen ciertas **ideas** esenciales sobre la opinión que se trata de ordenar, y ciertos **hechos** característicos que se deben tener en cuenta.

*Investigador del Commissariat a l'Energie Atomique (CEA), Centre d'Etudes Nucléaires de Fontenay-aux-Roses, Paris, Francia.

Elaborado por correcciones sucesivas, el **esquema conceptual** constituirá la primera etapa de ordenamiento; se vuelve entonces fácil, habiendo precisado los objetivos, proponer un **modelo**. Es el procedimiento estadístico que se va a tomar prestado para la **búsqueda de los parámetros** del modelo y que va a dar las directrices para construir el **protocolo de la encuesta**. El análisis de los datos recogidos por medio del instrumento (encuesta de opinión) producirá las dimensiones buscadas que se tratarán de **interpretar** tomando en cuenta, no solamente los hechos evidenciados y las ideas iniciales, sino también el procedimiento que permitió obtenerlos.

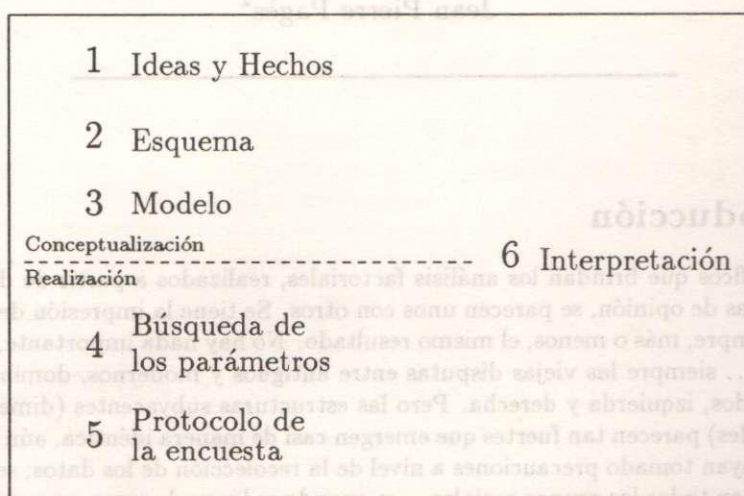


Figura 1: Esquema de las etapas de un sistema de referencia de encuestas de opinión.

1 Las ideas y los hechos

Nada mejor que una máxima para expresar en pocas palabras una idea; nada mejor que una serie de máximas para describir en algunas frases las ideas en voga sobre la opinión. Son estas las ideas que se tratan de poner en claro, teniendo en cuenta los hechos de la opinión. Se buscarán esos hechos en el campo nuclear. La riqueza de informaciones de las que se dispone en ese campo permiten obtener ciertas relaciones o circunvoluciones difíciles de explicar, si se restringen a los enfoques tradicionales.

1.1 Algunas máximas

Estas máximas se inspiran particularmente en Bourdieu, Cathelat, Eysenck, Ferguson, Gaxi, Greimas, Levi-Strauss, Mead, Moscovici, Parlov, Piaget...

- Si la opinión es la respuesta de un individuo ante un estímulo, es también, cuando se pasa de la información bruta al porcentaje (opinión pública), una fuerza que confiere legitimidad al actor.
- Los temas de actualidad, tanto como sus actores, están en competencia frente al público.
- Las representaciones, que aseguran la mediación entre la percepción con predominio sensorial y el concepto de orden puramente intelectual, establecen un puente entre la economía interna, el ser y lo social. Es por su mediación que se ajustan fuerzas psíquicas y fuerzas sociales.
- Se dice que las representaciones son sociales en la medida en que hacen referencia a símbolos cuyo significado es compartido por la colectividad.
- Es explorando las tomas de posición, y no los discursos de argumento, que se puede hacer aflorar lo que difiere en el nivel de las representaciones.
- Quien dice toma de posición dice conflicto; los conflictos de los que se discute en todas las capas sociales se encuentran en los grandes medios de comunicación.
- Los conflictos aparecen como relatos espectaculares. Verdaderas epopeyas o mitos que se desarrollan y se construyen ante nuestros ojos, participan de manera decisiva a la formación y a la transformación de las representaciones; ellas son, para el individuo, espacios privilegiados de la proyección, la identificación, y la diferenciación.
- Salidas de las representaciones, las tomas de posición en los conflictos, retornan y las modifican.
- Actores y público, a través de los conflictos, evolucionan ajustándose. Los medios de comunicación y los sondeos facilitan este ajuste, que todo hecho espectacular puede cuestionar.
- Entre el Cielo (los mitos, los actores) y la Tierra (el público) se efectúa un va y viene incesante; correa de transmisión entre el presente y el futuro, la opinión es como una nave espacial.

1.2 Algunos temas de opinión sobre lo nuclear

Los datos de la encuesta de la que se dispone permiten en general examinar la opinión en dos perspectivas:

- Aquella de la **sincronía**: ¿cuáles relaciones existen entre las tomas de posición sobre diversos temas? ¿En qué esas posiciones son globalmente coherentes? ¿Qué pasa con las diferencias entre grupos sociales?
- Aquella de la **diacronía**: ¿cómo evoluciona la opinión pública? ¿Cómo se transforman las relaciones entre los temas de discusión y las diferencias entre los grupos sociales?

Estas dos maneras de examinar la opinión, no como "en valor absoluto" sino "por comparación" permiten escapar en parte a los problemas difíciles que plantean la escogencia de las preguntas y el muestreo. Se hace a un lado la legitimidad que otorga la opinión a través de un porcentaje, para interesarse solo en las estructuras y en la dinámica.

Juego de correlaciones

En 1989, las reacciones en cadena provocadas en *la psiquis* por los átomos civil y militar, crearon siempre simpatías estadísticas que pueden parecer sorprendentes.

He aquí algunas correlaciones significativas producidas por la opinión entre estos dos átomos y ciertos objetos sociales (conflictos) de los que se discute en el país. Estas correlaciones fueron fuertes frente a las formulaciones de las preguntas planteadas. Nos limitamos aquí a las formulaciones reducidas para designar los conflictos.

"Se debe continuar construyendo centrales nucleares"		"La fuerza de disuasión es indispensable"	
Sostener los ecologistas	-0.234	Reducir los gastos militares	-0.334
Por la fuerza de disuasión	0.226	Suprimir el servicio militar	-0.300
Industria química contaminante	-0.212	Por la patria	0.245
Reducir los gastos militares	-0.207	Respetar las conveniencias	0.231
Confianza en la Justicia	0.166	Construir centrales nucleares	0.226
Garantizar salario mínimo	-0.165	Muchos trabajadores inmigrantes	0.222
Jefes inútiles	-0.157	Controles identidad indispensables	0.210
Cortar a los malos pagadores	0.155	Sostener el desarrollo	0.209
Libertad de precios	0.152	Por la familia	0.207
Por la Patria	0.151	Mariguana en venta libre	-0.192

Figura 2: Principales correlaciones entre los átomos civil y militar.

Las centrales nucleares no se llevan con los ecologistas... pero, ¿qué decir de las relaciones entre el átomo civil y la Justicia, el ingreso mínimo, los jefes, la libertad de precios y la Patria? Lo mismo la fuerza nuclear de disuasión no desentona con los gastos militares, el servicio militar y la patria... pero, ¿qué decir de las afinidades entre el átomo militar y: los trabajadores inmigrados, los controles de identidad, la familia... y la mariguana?

Factores socio-demográficos y socio-culturales

En 1989, la opinión sobre las centrales nucleares ya construídas estaba ligada a los factores:

- **sexo:** las mujeres han sido siempre más reservadas que los hombres;
- **edad:** los más viejos ha sido siempre los más favorables;
- **preferencia partidaria:** extrema izquierda, ecologistas y simpatizantes del partido comunista son los más opuestos;

y en menor grado:

- **religión:** los ateos se oponen a los católicos que son practicantes regulares, quienes son los más favorables.

Todas estas diferencias no podrían explicarse a partir de cualidades objetivas del objeto nuclear (costo del kilowatio-hora, el riesgo...), o del sujeto (ansiedad, saber...). Se sale aquí de los dominios donde opera el técnico, el hombre del mercadeo... o el psicólogo.

Evoluciones y cambios de posición

Las informaciones de las que se dispone desde 1977 permiten distinguir tres períodos en la evolución de las opiniones sobre lo nuclear.

- La primera se extiende hasta las elecciones presidenciales de 1981. De 1977 a 1981, la imagen de lo nuclear se mejora excepto en los jóvenes y los intelectuales. La crisis de la energía y en menor grado la manifestación violenta de Creys-Malville en 1977, ¿serían suficientes para explicar esta evolución?
- La segunda va de mayo 1981 a abril de 1986. La relativa estabilidad de los porcentajes globales esconde una reorganización en profundidad de las opiniones. La imagen de las centrales mejora claramente en los profesores y los estudiantes, y entre los que poseen un diploma de estudios superiores; ésta desciende entre los pensionados y los que se sitúan a la derecha del tinglado político. El acto simbólico que constituye el abandono de Plogoff (cambio de significación del átomo), por un lado, y el anuncio de una sobreproducción de kilowatios-hora en el horizonte de 1991, ¿habrán facilitado este cruce entre grupos sociales de sensibilidades políticas opuestas?

• El tercero empieza en mayo de 1986, luego del accidente de Tchernobyl. El deterioro de la imagen de lo nuclear desde junio de 1986 (descenso medio de más de un 10%) ha afectado únicamente el programa nuclear civil; las imágenes de los expertos y de las instituciones no sufrieron realmente por el accidente. El repliegue casi general ha permitido acentuar la reorganización de las opiniones. Los intelectuales (cuadros superiores, docentes, estudiantes), menos impresionados por el accidente, son actualmente los que estarían menos desfavorables a la construcción de nuevas centrales; por el contrario, los pensionados y los agricultores, que estaban entre los más favorables, se encuentran ahora entre los más críticos (baja de más de un 25%). En 1989, tres años después de Tchernobyl, no se ha constatado ningún cambio en la opinión; se podría incluso considerar que la ligera modificación de la imagen en 1987 ha sido bastante borrada, desde finales de 1988, por una nueva baja en la adhesión al programa nuclear. Este contragolpe de Tchernobyl, ¿no podría explicarse por la renovación de la ecología? Desde fines de 1988 se observa un claro progreso del interés del público por los temas ecológicos. Los ecologistas encuentran ante el público la popularidad que tenían en los años 75.

Estas peripecias de la opinión sobre lo nuclear aparecen claramente en los gráficos de evolución donde se pueden borrar algunos saltos debidos a la imprecisión estadística. Se encuentra, en los gráficos dados, la historia de los efectos "edad", "profesión" y "tendencia política" a través de los desacuerdos expresados en la proposición "Se debe continuar construyendo las centrales nucleares". Estos gráficos muestran que los grupos sociales no evolucionan de manera paralela, ya sea después de la victoria de la izquierda en mayo de 1981 (la imagen de lo nuclear, que progresa entre los docentes y los estudiantes y se degrada entre los pensionados), o después del accidente de Tchernobyl de abril de 1986 (cambio espectacular de actitud entre los agricultores y los más viejos).

Aquí también aparecen hechos que no se pueden explicar sin un punto de vista nuevo sobre la opinión.

2 Un enfoque constructista y sistémico de la opinión

Para describir los mecanismos de la opinión, se busca organizar en un todo coherente las ideas reunidas en el primer cuerpo (las máximas). El esquema simple y admitido del que se partirá será luego confrontado con los hechos, constituyendo el segundo cuerpo (las opiniones sobre lo nuclear desde 1977): ¿qué se ha olvidado? Por añadidos sucesivos, esta prueba conducirá a un nuevo esquema que se refiere a una realidad más compleja. Después de simplificarlo, el recurso de la analogía permitirá entonces encontrarse en

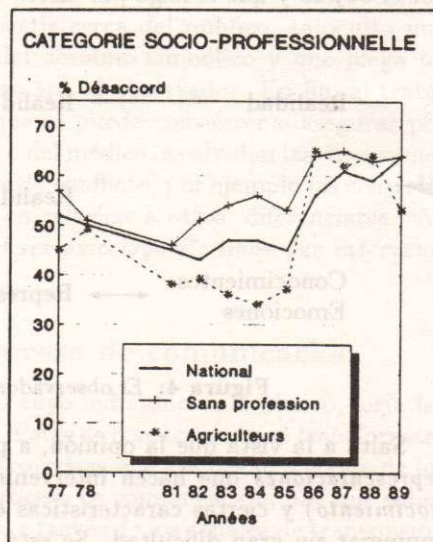
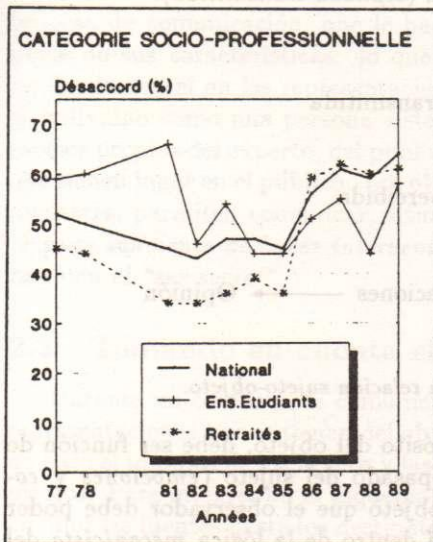
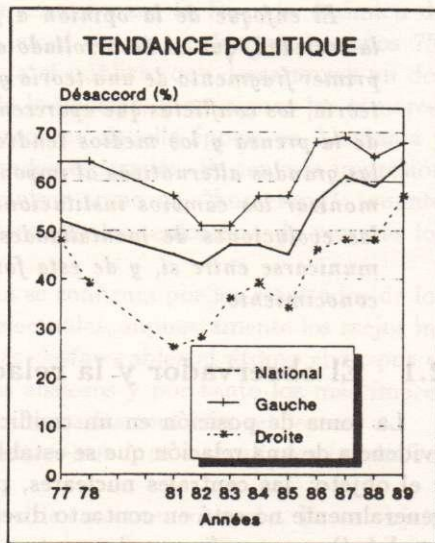
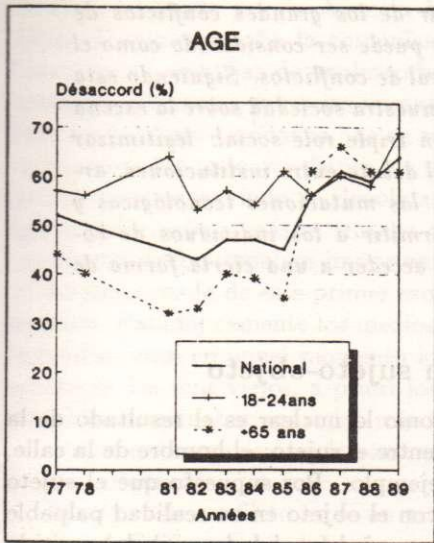


Figura 3: Evolución (1977-1989) del porcentaje de desacuerdos de ciertos grupos sociales con la proposición "Se debe continuar construyendo las centrales nucleares".

un paisaje conocido por los economistas.

El enfoque de la opinión a partir de los grandes conflictos de la sociedad, que es desarrollado aquí, puede ser considerado como el primer fragmento de una teoría general de conflictos. Siguiendo esta teoría, los conflictos que aparecen en nuestra sociedad sobre la escena de la prensa y los medios tendrán un triple rol social: legitimizar las grandes alternativas al imponer el debate entre instituciones; armonizar los cambios institucionales, las mutaciones tecnológicas y las evoluciones de mentalidades; permitir a los individuos de comunicarse entre sí, y de esta forma acceder a una cierta forma de conocimiento.

2.1 El observador y la relación sujeto-objeto

La toma de posición en un conflicto como lo nuclear es el resultado de la evidencia de una relación que se establece entre el sujeto: el hombre de la calle, y el objeto: las centrales nucleares, por ejemplo. Por supuesto que el sujeto generalmente no está en contacto directo con el objeto en su realidad palpable (*realidad*); no es más que el receptor de una señal (*realidad percibida*) emitida por el objeto y que le llega por cierto canal (*realidad transmitida*).

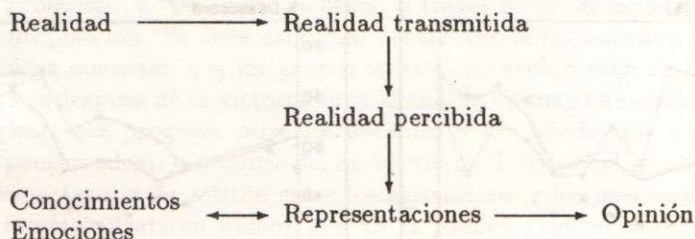


Figura 4: El observador y la relación sujeto-objeto.

Salta a la vista que la opinión, a propósito del objeto, debe ser función de *representaciones* que hacen intervenir el pasado del sujeto (*emociones* y *conocimiento*) y ciertas características del objeto que el observador debe poder enumerar sin gran dificultad. Se está aquí dentro de la lógica *mecanicista* del viejo modelo *SR* (estímulo \Rightarrow respuesta), o *S(P)R* (estímulo \Rightarrow personalidad \Rightarrow respuesta), tan querido por los conductistas.

En el caso de las centrales nucleares, utilizar este esquema viene a establecer la hipótesis de que el individuo realiza un balance entre el pro y el contra en el que ciertos argumentos favorables o desfavorables tienen más peso que otros. Es a través de estas ponderaciones que intervendrá la experiencia, lo vivido u otros factores ligados a la actividad síquica presente del sujeto.

Es este esquema tan simple, del cual se obtuvo el modelo de mercadeo de Fishbein utilizado por la Agencia Internacional para la Energía Atómica de Viena, el que inspiraba la comunicación en el campo nuclear en los años 75. Conducía a campañas de explicación hacia el público, que descansaron en dos consignas: informar sobre las ventajas de lo nuclear, y apaciguar los temores. Pero las costosas campañas que se realizaron en aquella época, en Alemania y en Suecia sobre todo, no dieron los resultados esperados. Se tenía la impresión que el mensaje enviado se devolvía al emisor como un "bumerang": cuanto más se informaba y más se trataba de disminuir los temores, cada vez más los cuestionamientos tomaban importancia.

Lo inadecuado de este primer esquema se confirma por los resultados de los sondeos. Paradójicamente los medios intelectuales, supuestamente los mejor informados, eran en aquel momento los más desfavorables al átomo civil; por el contrario, los más viejos, a priori los más ansiosos y por tanto los más impresionables, se situaban entre los más favorables a las centrales nucleares.

Partir entonces de la relación sujeto-objeto conduce a resultados insatisfactorios. ¿Porqué?

El objeto de conflicto no es un producto de consumo corriente. Es un objeto "mediático". Es aprehendido ante todo a través de los medios de comunicación colectiva o en las discusiones que estos provocan. Al no hacer intervenir el proceso de comunicación, que le hace existir cerca del público, se oculta una parte de sus características: lo que es del dominio simbólico y que juega un rol fundamental en las representaciones escapa al observador. En fin, al tratar al individuo como una persona aislada que se puede convencer o asegurar por medios propios del experto, del publicista o del médico, se olvidan las discusiones que tienen lugar en el público. Los objetos de conflicto, por ejemplo las centrales nucleares, permiten comunicar, situarse en relación a otros, diferenciarse. *No es pues suficiente de hacer intervenir "el ser psicológico"; tiene que intervenir también el "ser social".*

2.2 Tomando en cuenta el proceso de comunicación

Durante un conflicto, la comunicación cuyo individuo es el blanco, forja las representaciones que se tienen del objeto. Estas se construyen y se transforman, no bajo el impulso de los hechos y las controversias que nos llegan a través de los medios. Para integrar en el esquema el proceso de comunicación, hay que saber circular las fuentes hasta los destinatarios, a través de los canales de transmisión de mensajes.

De los actores al público: Las fuentes son los actores portavoces de las instituciones concernidas por lo que está en juego, alrededor de lo que se desarrolla o se desarrolló el conflicto. Respecto a lo nuclear, entre estos actores se encuentran mezclados: el Estado, el Parlamento, el Gobierno, los ministerios, la EDF, el CEA, las empresas como COGEMA, FRAMATOME o NOVATOME, los grupos de expertos, los partidos políticos, los elegidos nacionales o regionales,

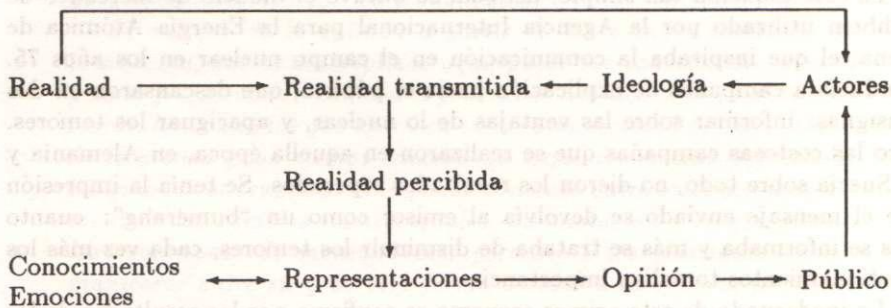


Figura 5: Los procesos de comunicación son tomados en cuenta en este esquema modificado.

los comités, las asociaciones o sociedades que apoyan o se oponen al programa nuclear, la Iglesia, los sindicatos... y los medios, que preocupados por su audiencia, no podrían ser considerados como simples transmisores de la información. Cuando se hace referencia a lo que está en juego, no se habla solo de lo tecnológico o lo económico; son también juegos de poder, ya que cada organización trabaja para defender sus intereses y aumentar su poder.

Los discursos de los actores son realmente *preparados para pensar* (enlatados). Combinando tomas de posición y argumentos que permiten defenderlos o hacerlos parecer razonables, traducen los roles que tienen; y el sociólogo no tiene dificultades para explicar lo que se dice cuando hizo aparecer a la luz del día lo que está en juego en el conflicto. Pero el debate que se inicia no es sigiloso y rasurador, como el de un consejo de administración: *se está en el escenario de un teatro*. En esta escena, que no es otra que la de la prensa y de los otros medios, el actor no es dueño del protocolo; él percibe que la situación es confusa y resiente la angustia de la acusación. Todo contribuye a darle miedo: el lenguaje técnico al que está mal habituado, está mal adaptado a la comunicación del público; no se le pide hablar en nombre de la empresa, sino de argumentar sobre el interés de sus posiciones a nivel del país... y todo esto en directo, frente al público. ¿Sabrá seducir para reforzar o restablecer su posición? ¿Encontrará ahí la legitimidad que creía perdida o que le falta?

El destinatario de los mensajes —el individuo del público— no es, salvo excepción, directamente involucrado en los asuntos en juego; no participa ni en la concepción de los micro-ordinadores o en la definición de la política agrícola, ni en el financiamiento de la seguridad social o en la recolección de impuestos..., ni en la construcción de las centrales nucleares. Él se encuentra también demasiado lejos del lugar donde se toman las decisiones para que pueda tener una influencia determinante sobre el curso de las cosas. Y la mayoría de las veces, estas decisiones no cuestionan más que de manera abstracta, ya que se trata de un

futuro hipotético y controversial lo que posee, su confort, su vida. Le resulta difícil comparar, sopesar... en fin, dar prueba de discernimiento pues, más allá del conocimiento de los protagonistas no puede retirar de la justa al que le asiste más que un saber tan incierto como contradictorio. Este saber no resulta de una experiencia concreta: *es en gran parte la grabación de un espectáculo*. ¿Cómo sacar una opinión mesurada de un conjunto de afirmaciones opuestas, a menudo enfáticas y perentorias? En fin, no se trata de escoger entre lo verdadero y lo falso, sino de aceptar o rechazar alguna cosa que dé cuenta de lo verdadero o de lo falso, incluso del bien o del mal. Si todo está hecho para exitar el interés del hombre de la calle, nada está hecho para incitarlo a investigar, evadiéndose de sus sentimientos o de sus "a priori", un orden desencarnado y universal. Sin embargo, ¿quién no ha dado su opinión e incluso argumentado en favor o en contra de las nacionalizaciones, la igualdad de los ingresos, el desarrollo de las bio-tecnologías, la pena de muerte o el programa nuclear? Obviamente los conflictos interesan en la opinión; campos de enfrentamiento entre instituciones se tornan la ocasión para comunicarse con los próximos. *Espacio de inversión psíquica, lugar mayor de proyección y de implicación, el conflicto, que provoca el debate entre actores, es ante todo para el individuo espectador un tema de expresión.*

Pero entonces, *¿qué hay de común entre el espectáculo sobre la escena, es decir, el debate entre actores sociales, y esas conversaciones animadas que son la prolongación en el público? ¿Cómo los actores se ajustan al público y recíprocamente?*

Del público a los actores: El público es un juego mayor cuando hay conflicto. A nivel de los actores, ¿qué importa que el hombre de la calle esté o no impregnado por el problema que se debate! Lo esencial es que se opere un arbitraje entre las fuerzas, ya sea que les cueste imponerse o que duden, o que no encuentren más, con los portavoces institucionalizados, intermediarios para negociar en un cuadro más restringido... y ese es el rol de la "vox populi" a través de la prensa y los sondeos. Todo ocurre como si hubiese un tribunal público en el que la argumentación producida sería analizada en su justo valor; no hay jueces en el público, pero esta ilusión se le ofrece a toda la sociedad. El combate producido por el problema y esta ilusión van a permitir a los actores aclarar las premisas, afinar la doctrina, ajustar el razonamiento. La prueba es temible y escasos son los responsables que se prestan tranquilos a un ejercicio tan poco confortable; los inquisidores imponen la transparencia, rastrean la contradicción... Así la disimulación se torna difícil y la denuncia de falta de cohesión en un estado mayor puede crear problemas al poder. Los conflictos aportan entonces más que indicaciones sobre las resquebrajaduras que afectan al edificio institucional; relegados al segundo plano en una elección aquel que opone permanentemente las fuerzas políticas, *los conflictos son una manera de administrar las relaciones de fuerza que permiten respetar con suavidad los grandes principios que han conducido al sufragio universal.*

El actor está pues obligado a tomar en cuenta al público; modula su posición

y sus argumentos en función de la información sobre la opinión que le dan los medios y los sondeos. La comunicación establece así un círculo donde fuentes y destinatarios cambian constantemente los roles. Por un aprendizaje continuo y recíproco, los actores y el público se ajustan para tender a un equilibrio, es del compromiso, que podría cuestionarse de un día para otro. Bastaría un acontecimiento importante, un accidente en el caso de lo nuclear, por ejemplo, para que la armonía deje lugar a una cacofonía, productora con el tiempo de un nuevo orden.

El esquema al que se llega es tan mecánico como el precedente; simplemente se ha pasado, al introducir los actores, desde un punto de vista psicológico a un punto de vista psico-social (conductismo social). Al pasar, por la integración de opiniones individuales, de la opinión de un individuo a la opinión pública, se pasa del nivel "micro" al nivel "macro". El actor en el aflujo de porcentajes, el sociólogo y el politólogo en busca de juegos y legitimidades, y el etnólogo limitándose a consideraciones muy globales sobre el funcionamiento de las sociedades, pueden satisfacerse de un macro-análisis. El trabajo del psicólogo o del psicoanalista, por el contrario, tiene que ver con el micro-análisis. En cuanto al psico-sociólogo que busca decir algo más sobre ese ajuste que ocurre entre el mundo interior del sujeto y su medio social, debe continuamente circular entre ambos niveles; ¡y es ahí donde reside toda la dificultad!. Quedan aún por obtenerse las consecuencias de la potencialidad de implicarse, que ofrece el conflicto al individuo.

2.3 La apropiación del objeto: del esquema mecanicista al esquema constructivista

La búsqueda del reconocimiento por el otro, la afirmación de su estatus, bastan para explicar esas animadas conversaciones que tienen lugar en el público. Pero, con esos resortes de orden psicológico se opera una verdadera mutación, ya que el individuo de receptor pasa a ser emisor.

Para alcanzar el conocimiento, dicen algunos con razón, es necesario tener fe. Es el carácter espectacular del conflicto que hace que el individuo se involucre. Es la fe la fuerza que le hace cortar, dentro de la red social que lo solicita, en favor de unos (¿el padre? ¿el modelo?) y en detrimento de otros; esta se expresa en la toma de posición, que puede considerarse como una primera fase afectiva de la respuesta a la situación (el conflicto) que lo interpela. Es la información que graba durante el espectáculo la que torna posible este primer ajuste; sin puntos de referencia, incluso confuso, el conflicto no puede operar como objeto de diferenciación: es entonces la fuga, la indiferencia o la angustia. Aquí no puede recurrirse, como con Dios, al argumento definitivo de la gracia... Es la conversación, y cuanto más animada mejor, la que va a permitir fundar la fe. ¿No es ésta el espacio propicio donde, bajo la presión de los otros, se encuentra la facultad de ordenar sus ideas? Así lo oscuro puede acceder a cierta claridad: los argumentos producidos por los actores son de cierta manera dirigidos para

llegar a un hilo conductor que da a menudo a nuestro individuo social más la apariencia de la razón. El argumento a posteriori es el desenlace de esta segunda fase, más evaluativa, de la respuesta a la situación que impone el conflicto; salida de un examen crítico que no esté separado, ella indica que se ha interiorizado una lógica y por tanto se ha accedido a cierto conocimiento. *En esa fase de apropiación del objeto, el individuo pone a posteriori sus representaciones en armonía con su toma de posición, considerada con un mensaje que él da sobre sí mismo.*

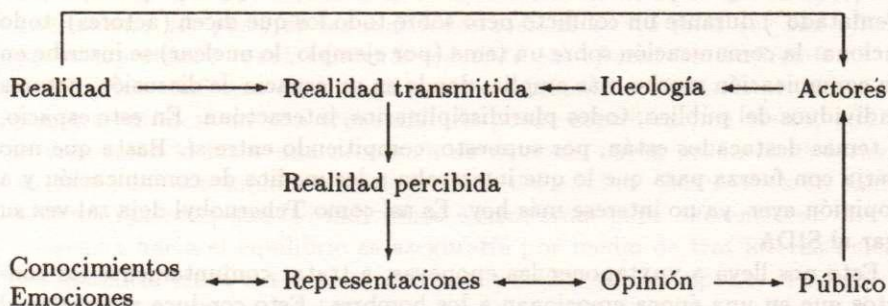


Figura 6: El esquema constructivista.

He aquí pues cómo el individuo devendría en "social". Al asimilarlo pues a un proceso de ajuste entre fuerzas psíquicas y fuerzas sociales (este proceso se llama aquí "representaciones"), se deja el pensamiento racional tal como es visto por el actor para un pensamiento donde sí caben los símbolos y los valores. Pero aún queda por explicar esos lazos estadísticos que se han constatado entre opiniones expresas acerca de temas a priori bastante diferentes.

2.4 El examen de los mensajes

Esos "listos para pensar" o "enlatados" a los que se han asimilado los escritos y las declaraciones de los actores no son transmitidos "tal cual" al individuo. ¿Son puestos en escena por los medios de comunicación que toman en cuenta tanto "lo que se ha dicho" como "quién lo dijo"? La fuente del "listo para pensar" forma parte integrante de los mensajes; desde que es percibida de manera negativa, lo que se ha dicho es sentido como sospechoso. Se observa entonces que los mensajes que asaltan al individuo del público no constituyen simplemente un juego de argumentos. Tomando en cuenta los discursos y los actores a la vez, por su carácter espectacular, *forman un relato épico, una epopeya con múltiples vueltas cuyo desarrollo se efectúa ante nuestros ojos.*

Pero la toma de posición sobre las centrales nucleares ya se ha visto, está correlacionada con las posiciones sobre Dios, las privatizaciones, las huelgas,

la independencia de Nueva Caledonia... De una manera más general, las posiciones sobre los grandes conflictos que mueven la sociedad francesa forman un todo bien estructurado. ¿Por qué?

Entre los actores, muchos son los que se encuentran implicados en más de un conflicto. El Estado, el Gobierno, los ministerios, las grandes empresas, la Iglesia, las sociedades de sabios, los partidos políticos, los elegidos, los sindicatos, las asociaciones en favor de defender la flora y la fauna... participan, de una manera muy general, a este cuestionamiento permanente que propulsa, como máquinas de vapor, nuestras democracias hacia las sorpresas del futuro. Entonces, por medio de mensajes, que hacen intervenir lo que está dicho ("listo para pensar" o "enlatado") durante un conflicto pero sobre todo los que dicen (actores), todo funciona: la comunicación sobre un tema (por ejemplo, lo nuclear) se inscribe en una comunicación mucho más amplia, donde en un espacio de discusión, actores e individuos del público, todos pluridisciplinarios, interactúan. En este espacio, los temas destacados están, por supuesto, compitiendo entre sí. Basta que uno emerja con fuerza para que lo que interesaba a los medios de comunicación y a la opinión ayer, ya no interese más hoy. Es así como Tchernobyl deja tal vez su lugar al SIDA.

Esto nos lleva a yuxtaponer las epopeyas, a tratar conjuntamente los conflictos que en una época emocionan a los hombres. Esto conduce a arreglar el esquema constructivista introducido. *Se pasa de lo singular a lo global tomando en cuenta a todos los actores que intervienen en la actualidad y ampliando el espacio de los temas de discusión para integrar todos los conflictos que cuentan.*

2.5 De la canción de gesta al gran trueque: el esquema de la economía de roles

Cuanto más espectacular es el conflicto, más lo real y la razón, tal como aparecen en las "listas para pensar", ven disminuir su importancia en favor del actor. En el límite el actor solo interviene; es entonces un poco como el héroe o el ídolo. Su presencia magnificada cambia el significado de la confrontación; ésta se transforma en el símbolo del cuestionamiento de algunos de los poderes que participan. Para los fieles, el sello tiene más importancia que la lógica que sustenta el discurso; el debate permite movilizar los convertidos y de revelarse una verdad con todas las apariencias de la razón, para ampliar el círculo de los proselitistas. Es así como se opera un cambio a nivel de las significaciones; se pasa de la denotación a la connotación, de lo real y de los hechos a los símbolos.

Las opiniones y los argumentos producidos frente al público por los actores no son más que migajas de una negociación que se desarrolla permanentemente entre organizaciones antagonistas; a través de sus portavoces, cada una de ellas busca mejorar o defender su posición. Los individuos proceden de manera equivalente: pisarle los talones a su actor, retomar sus argumentos, es decir su rol es también hacer una escogencia estratégica en la relación que se desarrolla con otro. Se trata de confortar su posición; pero el beneficio obtenido no es

evidente: cuesta a estas ruedas interiores que hacen del individuo la búsqueda permanente de una identidad. Es así como los debates entre actores se prolongan en el público, donde se constata la misma diversidad en las tomas de posición y los argumentos; pero la nueva pieza que se actúa solo es idéntica en apariencia a la que se desarrolla en el escenario de los medios.

Resultado de los ajustes que se operan durante la comunicación entre individuos sociales, los roles no están distribuidos al azar; son escogidos en función del lugar que se ocupa o que se desea ocupar en el universo social. Es apropiándose de un rol que se dice al otro dónde uno se sitúa, cómo se diferencia. *Y es así como se efectuaría el cambio de opiniones: coincidiría con la apropiación permanente por parte de los individuos sociales de los roles que son ofrecidos por el espectáculo constantemente renovado que ofrecen los conflictos.*

Con los conflictos considerados, si se admiten las hipótesis precedentes, se penetra entonces en una economía. Se puede considerar, en efecto, que todo ocurre como si fuese una economía de mercado cuyos productores serían los actores y los consumidores los individuos del público. Los actores en conflicto producirían roles que los individuos consumirían para diferenciarse. La convergencia hacia el equilibrio se aseguraría por medio de tres interacciones: la competencia entre actores en el debate público que los opone; el juego de la oferta y la demanda, donde intervienen los medios y los sondeos, que permiten armonizar el rol jugado por el actor y la espera del público; la diferenciación, donde para valorarse en relación a otro, el individuo retoma uno de los roles jugados en el mercado al encanto que constituyen los grandes medios.

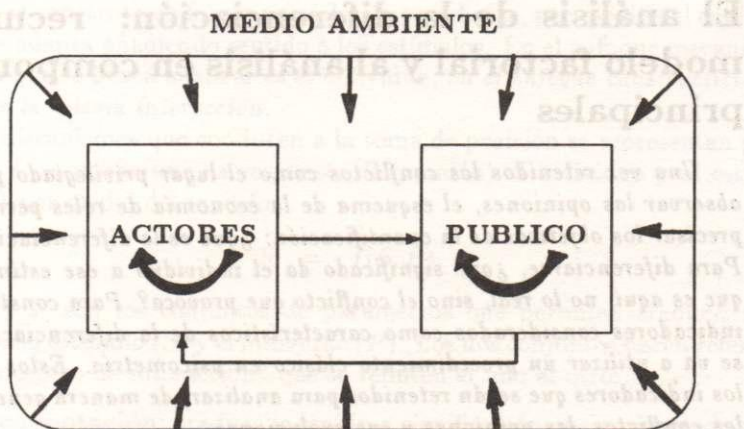


Figura 7: El esquema de la economía de roles.

Este esquema de la economía de roles es pues una respuesta global a la pre-

gunta: "¿Cómo se efectúa la comunicación en países como el nuestro, cuando hay conflicto?" Producido por analogía, parecería simple. *Se ha reducido el fenómeno de la opinión a un sistema, compuesto de dos subsistemas: "los actores" y "el público", abiertos sobre un medio ambiente que teje restricciones y produce los hechos.* De hecho, y los economistas han tomado conciencia de ello hace tiempo, el sistema describe una realidad bastante difícil de dominar. Se autorregula a dos niveles: a un nivel global, por ajustes entre el medio ambiente y los subsistemas de actores y público; a un nivel elemental, por ajustes que se operan entre actores por un lado y por individuos por el otro. Se dispone aquí de un ejemplo claro de embrión de teoría en la que el procedimiento, tan holístico como sea, concuerda con una visión reduccionista de lo real: el elemento (el individuo, el actor) y el todo (la sociedad en su conjunto) entran en resonancia a través de una serie de choques de naturaleza diferentes (conflictos entre elementos, conflictos entre subsistemas) que se propagan, de abajo hacia arriba y de arriba hacia abajo, atravesando las fronteras que corresponden a los diferentes niveles de organización. Si se habla como un economista: ¿adónde están las zonas de equilibrio? Si se habla como un mecánico: ¿adónde están los atractores exógenos?

Hasta aquí conduce el esquema inicial cuando se busca explotar las máximas del primer cuerpo, luego eliminar, por añadidos sucesivos, las paradojas que hace aparecer el segundo cuerpo. *El esquema simplificado al cual se llega reduce la opinión a un proceso dinámico de aprendizaje: es el recurso al lenguaje, fuente inagotable de distinción, que permite a todas las fuerzas manifestarse y así, autorizando los ajustes, asegurar la convergencia hacia un equilibrio.*

3 El análisis de la diferenciación: recurso al modelo factorial y al análisis en componentes principales

Una vez retenidos los conflictos como el lugar privilegiado para observar las opiniones, el esquema de la economía de roles permite precisar los objetivos de la cuantificación: ¿qué es la diferenciación? Para diferenciarse, ¿qué significado da el individuo a ese estímulo que es aquí, no lo real, sino el conflicto que provoca? Para construir indicadores considerados como característicos de la diferenciación, se va a utilizar un procedimiento clásico en psicometría. Estos son los indicadores que serán retenidos para analizar, de manera general, los conflictos, las opiniones y sus evoluciones.

3.1 Un nuevo objeto: la interacción

En el enfoque mecanicista del modelo $S(P)R$ (estímulo-personalidad-respuesta), se considera que la respuesta (opinión) se produce a través del individuo, por una causa (el conflicto-estímulo) que preexiste y es susceptible de ser descrita por un cierto número de características (aquellas del observador). A este acercamiento se le puede oponer un *enfoque constructivista* donde el individuo y el estímulo son tratados de manera simétrica.

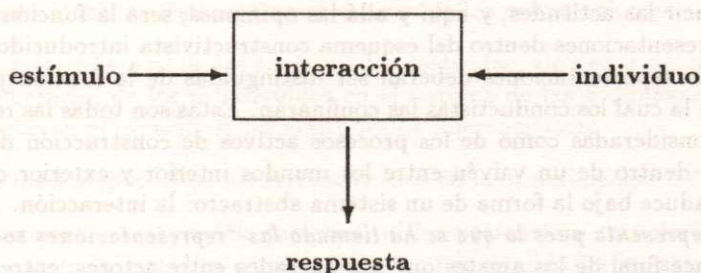


Figura 8: *Enfoque constructivista.*

La respuesta es considerada como resultante, no de características del individuo y del estímulo tomados en forma aislada, sino de ese nuevo objeto que se forma en sus interacciones. Se respeta la hipótesis que la significación para el individuo se construye en forma paralela al aprendizaje, por medio del cual el individuo se adapta añadiendo sentido a los estímulos. En el enfoque mecanicista, el sistema que se busca analizar es el individuo; *en el enfoque constructivista, el sistema es la misma interacción.*

Si los mecanismos que conducen a la toma de posición se representan por el funcional f , se puede unir la respuesta R (opinión) al individuo y al estímulo (conflicto) por medio del modelo:

$$R = f(S, P), \quad (1)$$

donde S y P son los conjuntos de parámetros que permiten traducir en su interacción el estímulo (S) y el individuo (P). Los dos conjuntos hacen referencia a dos "conceptos de interacción" que se remiten el uno al otro:

- **La connotación:** traduce la legítima significación que es aportada por el individuo social al conflicto;
- **La actitud de diferenciación:** es la predisposición a actuar que provoca el conflicto dentro del individuo. Esta predisposición consiste en utilizar la ocasión para expresar su diferencia.

Si el conflicto engendra las actitudes individuales, éstas, recíprocamente brindan su significación al conflicto. Para prever las opiniones individuales, se debe conocer las connotaciones: ¿cuál es la significación social del conflicto? Pero para encontrar la significación de un conflicto, se hace necesario partir de las actitudes de diferenciación: ¿qué indica el hombre de la calle cuando utiliza el conflicto para comunicarse? Es para romper con este círculo vicioso que se toma en cuenta también la interacción: *se va a tratar a la connotación y a la actitud de diferenciación en forma simultánea, en búsqueda de los hechos que matemáticamente se correspondan por dualidad.*

Producir las actitudes, y aquí y allá las opiniones, será la función atribuida a las representaciones dentro del esquema constructivista introducido anteriormente; las representaciones deberán ser distinguidas de la función perceptiva dentro de la cual los conductistas las confinarán. Estas son todas las representaciones, consideradas como de los procesos activos de construcción del sentido funcional dentro de un vaivén entre los mundos interior y exterior del sujeto, que se traduce bajo la forma de un sistema abstracto: la interacción. *Esta interacción representa pues lo que se ha llamado las "representaciones sociales"; es el desenlace final de los ajustes que son operados entre actores, entre actores y público, y entre individuos del público. Dentro de un estado particular, representa las "representaciones" de un individuo particular; es entonces el desenlace del ajuste que se opera entre este individuo y su ambiente.*

Se ve que se establece aquí una conjunción entre la noción de representación tal como es vista en Francia por ejemplo por S. Moscovici, y el modelo del interaccionismo simbólico del hombre, tal como es visto en Estados Unidos por los discípulos de G.M. Mead. Para describir las representaciones sociales a través de la pareja de conjuntos de parámetros (P, S) , se recurre a una modelización de la interacción entre conflictos e individuos sociales, cuyos significados simbólicos P dados a los conflictos son el reflejo de actitudes de diferenciación S ; y recíprocamente (dualidad).

3.2 En búsqueda de los parámetros del modelo

Considerar la interacción como sistema vuelve a asimilar los parámetros que describen el estímulo (S) y el individuo (P) a parámetros de estado. No se conocen ni el funcional f , que permite pasar de la interacción a las opiniones, ni los conjuntos de parámetros S y P . *Una aproximación va a permitir acercarse a un modelo lineal del que falta encontrar los parámetros.*

Perturbemos el sistema cercano de un estado de referencia; esto viene a definir una zona de estados alrededor de un individuo de referencia y del conflicto que le corresponde por dualidad. Si en esta zona, que corresponde a ciertos franceses y a ciertos conflictos, se decide considerar la aproximación del primer orden como suficiente, y la derivada de f como constante, se llega a partir de (1) a un modelo lineal que pone en relación el espacio de los estados y el espacio de las respuestas, es decir, representaciones y opiniones. Este *modelo factorial*

se escribe de manera matricial:

$$X = UC + E, \quad (2)$$

donde son designados por:

- X , la matriz ($p \times n$) de las respuestas x_i^j , centradas alrededor del individuo de referencia, de n individuos (i) en p estímulos (j);
- U , la matriz ($p \times k$) de las coordenadas de p estímulos en el sistema de las k dimensiones que estructuran el espacio de las connotaciones;
- C , la matriz ($k \times n$) de las coordenadas centradas, alrededor del individuo de referencia, de los n individuos en el sistema de las k dimensiones que estructuran el espacio de las actitudes de diferenciación;
- E , la matriz ($p \times n$) de los errores hechos en la aproximación UC de X .

No se conocen, en el modelo factorial, ni los parámetros característicos de las connotaciones, ni aquellos correspondientes por dualidad a las actitudes de diferenciación. Para descubrirlos, se llega a un problema de optimización: *hacer que la matriz de respuestas (opiniones) centradas X sea lo mejor ajustada*, y esto para respetar la simetría entre conflictos e individuos, ya sea que se razone en el espacio \mathbf{R}^n de n dimensiones sobre sus filas (conflictos) o que se razone en el espacio \mathbf{R}^p en p dimensiones sobre sus columnas (individuos).

Tanto Karl Pearson en 1901 como Charles Spearman en 1904 recurrieron a la escritura de las ecuaciones (2) del modelo factorial. Pero para juzgar la adecuación del modelo, Spearman, como la mayoría de los estadísticos, ha exigido independencia entre los residuos: lo que se deja de lado (E) debe parecerse al ruido. Haciendo poco caso de toda consideración estadística (normalidad de las distribuciones), el geómetra traduce esta exigencia al imponer en las ecuaciones (2) la ortogonalidad entre los k componentes factoriales c^l (filas de la matriz C) y las variables residuos e^j (filas de la matriz E) por un lado, y entre residuos e^j por el otro. Este modelo de Spearman o "*modelo en factores comunes y específicos*", al que se llega completando las ecuaciones (2) por las condiciones de ortogonalidad precedentes, sugiere por ejemplo el problema de optimización siguiente: "hacer máximo el determinante de la matriz de correlaciones entre residuos (criterio bien conocido del *determinante máximo*)". Esta manera de minimizar las correlaciones entre residuos conduce a un procedimiento iterativo de extracción de las componentes factoriales c^l , muy usado aún hoy por la escuela psicométrica americana.

Aquí, para ajustar la matriz X , respetando la simetría de los roles entre individuos y variables, se abandonará a Spearman para recurrir, como Pearson, a los mínimos cuadrados. Para esto hay que proporcionar dos distancias euclidianas:

- La primera, en el espacio \mathbf{R}^n , no es más que lo que los probabilistas llaman "distancia en media cuadrática", que corresponde a la dada por una matriz diagonal D_p ($n \times n$), cuyos elementos diagonales son los pesos p_i ($p_i > 0$; $\sum_{i=1}^n p_i = 1$) que se atribuye a los n individuos.
- La segunda, en el espacio \mathbf{R}^p , corresponde a la dada por una matriz diagonal M ($p \times p$) cuyos elementos diagonales son positivos; estos son los pesos que se atribuyen a los conflictos.

Se va a operar por proyección en los dos espacios \mathbf{R}^n y \mathbf{R}^p simultáneamente. En efecto, si se desea respetar la simetría entre conflictos e individuos ajustando UC a X , es necesario entonces:

- si se razona sobre las filas, las coordenadas U de los conflictos-estímulos se obtienen en \mathbf{R}^n por D_p -proyección de las filas de X sobre el sub-espacio de dimensión k engendrado por las filas de C (problema clásico de regresión lineal);
- si se razona sobre las columnas, las coordenadas C de los individuos se obtienen en \mathbf{R}^p por M -proyección de las columnas de X sobre el sub-espacio de dimensión k engendrado por las columnas de U (problema clásico de calibración).

Los expertos en Análisis de Datos conocen la solución de este doble problema; está dado por este método de análisis factorial introducido por Pearson, generalizado a distancias euclidianas arbitrarias: *el análisis en componentes principales del triplete* (X, M, D_p).

El análisis en componentes principales permite la obtención de los parámetros U y C que se corresponden matemáticamente por dualidad. Se puede, en efecto, organizar los objetos matemáticos que participan en la solución según el esquema de dualidad de la Figura 9.

En el esquema de dualidad tenemos que:

- E^* y F^* son los espacios duales de $E = \mathbf{R}^p$ y $F = \mathbf{R}^n$ respectivamente;
- M y D_p son distancias euclidianas definidas sobre E y F respectivamente, asimiladas aquí a isomorfismos entre los espacios y sus duales;
- X es la matriz de datos y X' su traspuesta; los dos se asimilan a aplicaciones que permitan circular respectivamente de F^* a E y de E^* a F ;
- $V = XD_pX'$ y $W = X'MX$.

Las soluciones U y C corresponden entonces a los k primeros vectores propios respectivamente (columnas de U y filas de C) de VM y de WD_p . Si C' designa la traspuesta de C , U y C están ligados entre ellos por las igualdades:

$$C' = X'MU \quad \text{y} \quad U = XD_pC'D_{1/\lambda},$$

donde $D_{1/\lambda}$ es la matriz diagonal cuyos k elementos diagonales son los inversos de los k más grandes valores propios de VM y de WD_p .

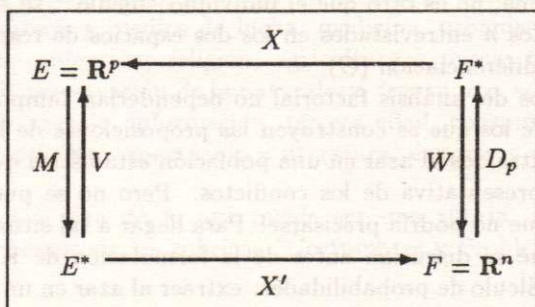


Figura 9: El esquema de dualidad.

4 La encuesta de referencia

Hasta el momento, el procedimiento ha sido completamente determinista: se ha perturbado un sistema: "la interacción individuo-conflicto" en una cierta zona de estados. Queda por indicar cómo se obtiene la matriz X de las respuestas del sistema perturbado, es decir, cómo se realiza la encuesta que permite recoger las tomas de posición de los n individuos en los p conflictos.

Las k dimensiones extraídas por el análisis en componentes principales, que estructuran los dos espacios duales de las connotaciones y de las actitudes de diferenciación, deben ser representativas del sistema que forma la interacción individuo-conflicto (representaciones sociales). A través de los individuos y de los conflictos seleccionados (perturbación) estas no deben depender más que de los dos universos a los que se hace referencia: *el universo de los individuos*, que no es más que el de la población de franceses mayores que viven en alguna metrópoli; *el universo de los conflictos*, que comprende todo lo que es noticia importante en los grandes medios de comunicación. *Esta doble condición de universalidad conduce a tomar dos métodos de selección que podrían considerarse como sorteo al azar.*

Para seleccionar los 1.000 a 2.000 entrevistados (n), se trabajará como en casi todos los sondeos que se realizan en Francia, con un "método estratificado y por cuotas". Los entrevistados podrían entonces considerarse como extraídos al azar de la población de franceses de referencia, y los resultados del análisis factorial sobre las respuestas del cuestionario dependerán, a través de la muestra interrogada, de esta población. Observemos que el procedimiento de selección escogido respeta el principio de igualdad de todos los franceses ante el voto como base del sufragio universal. Se encuentra este principio con la asignación de parte del observador de ponderaciones (distancia D_p) iguales $p_i = 1/n$ a cada uno de los entrevistados. El individuo de referencia, alrededor del cual se

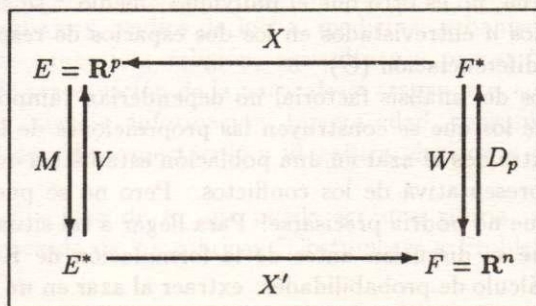


Figura 9: El esquema de dualidad.

4 La encuesta de referencia

Hasta el momento, el procedimiento ha sido completamente determinista: se ha perturbado un sistema: "la interacción individuo-conflicto" en una cierta zona de estados. Queda por indicar cómo se obtiene la matriz X de las respuestas del sistema perturbado, es decir, cómo se realiza la encuesta que permite recoger las tomas de posición de los n individuos en los p conflictos.

Las k dimensiones extraídas por el análisis en componentes principales, que estructuran los dos espacios duales de las connotaciones y de las actitudes de diferenciación, deben ser representativas del sistema que forma la interacción individuo-conflicto (representaciones sociales). A través de los individuos y de los conflictos seleccionados (perturbación) estas no deben depender más que de los dos universos a los que se hace referencia: *el universo de los individuos*, que no es más que el de la población de franceses mayores que viven en alguna metrópoli; *el universo de los conflictos*, que comprende todo lo que es noticia importante en los grandes medios de comunicación. *Esta doble condición de universalidad conduce a tomar dos métodos de selección que podrían considerarse como sorteo al azar.*

Para seleccionar los 1.000 a 2.000 entrevistados (n), se trabajará como en casi todos los sondeos que se realizan en Francia, con un "método estratificado y por cuotas". Los entrevistados podrían entonces considerarse como extraídos al azar de la población de franceses de referencia, y los resultados del análisis factorial sobre las respuestas del cuestionario dependerán, a través de la muestra interrogada, de esta población. Observemos que el procedimiento de selección escogido respeta el principio de igualdad de todos los franceses ante el voto como base del sufragio universal. Se encuentra este principio con la asignación de parte del observador de ponderaciones (distancia D_p) iguales $p_i = 1/n$ a cada uno de los entrevistados. El individuo de referencia, alrededor del cual se

perturba el sistema, no es otro que el individuo "medio": se sitúa en el centro de gravedad de los n entrevistados en los dos espacios de respuestas (X) y de las actitudes de diferenciación (C).

Los resultados del análisis factorial no dependerían tampoco de los p conflictos a partir de los que se construyen las proposiciones de los cuestionarios, si estos fuesen extraídos al azar en una población estadística que se podría considerar como representativa de los conflictos. Pero no se puede tirar al azar en un universo que no podría precisarse! Para llegar a tal situación recordemos las paradojas que se discutían antes de la formulación de Kolmogorov de la axiomática del cálculo de probabilidades: extraer al azar en un universo de conflictos provisto de una medida de probabilidad equivale a dar un procedimiento fijo de producción de conflictos, que siendo errático, respete ciertas reglas. La construcción del cuestionario de nuestra encuesta de referencia va a reposar sobre tal procedimiento. Al hacer trabajar a los expertos se va a operar de la misma manera que Buffon, quien para encontrar el número π soltaba siempre su aguja con la misma torpeza.

El procedimiento retenido para construir el cuestionario, y que puede compararse a una extracción al azar, opera en tres etapas:

- Selección de un gran conjunto de conflictos con ayuda de un método fijo igual para siempre: este conjunto se considera el universo de los conflictos;
- Extracción por un grupo de expertos con intereses diversificados de una muestra representativa de este universo;
- Construcción, alrededor de los conflictos repertoriados, de proposiciones que no tengan ninguna ambigüedad.

4.1 Construcción de un universo de conflictos

Es el análisis sistemático de los grandes títulos de la prensa nacional y regional, realizado un mes antes del desarrollo de la encuesta, el que permite obtener lo que será considerado como el universo de los conflictos. Esta es la lista de los diarios y revistas consultados en marzo de 1981:

- *Los diarios*: Figaro, France-Soir, l'Humanité, Le Matin, Le Monde, Le Parisien Libéré, Le Dauphiné, Les Dernières Nouvelles d'Alsace, Ouest-France, Sud-Ouest, La Voix du Nord;
- *Las revistas*: Bonne Soirée, Elle, L'Express, Minute, Le Novel Observateur, Paris-Match, Le Point, Que Choisir, Télé 7 Jours, La Vie.

Los quinientos o seiscientos conflictos repertoriados se clasifican de acuerdo a una nomenclatura determinada para siempre. He aquí las rúbricas de esta nomenclatura: política nacional, política internacional, asuntos franceses raros,

catástrofes y problemas nacionales e internacionales, economía, seguridad, justicia, familia, trabajo y medios de lucha, medicina, urbanismo y alojamiento, energía, ejército, educación, religión, agricultura y pesca, medio ambiente y lucha en favor de preservación de la naturaleza, transporte, costumbres, problemas de sociedad, ciencia, información, tercera edad, consumo y alimentación, calidad y modo de vida, espectáculo y literatura, deporte y diversión, moda y belleza.

Para tener una idea de lo que puede ser una rúbrica, se presenta en la Figura 10 el contenido de las rúbricas "Costumbres y Problemas de Sociedad" en marzo de 1981.

Costumbres	Problemas de Sociedad
La prostitución masculina y femenina	El vandalismo
La homosexualidad	La criminalidad
La liberación sexual	La inmigración
La transexualidad	El racismo
La violación	Los niños mártires
La unión libre	La droga
Tener un amante	La violencia
Los matrimonios interraciales	La delincuencia juvenil
El matrimonio es algo bueno	Los suicidas
La liberación de la mujer	El robo
El nudismo	La tristeza de algunos jóvenes
Broncearse con los senos desnudos	
El exhibicionismo	
La masturbación	
Los gigolos	

Figura 10: Contenido de las rúbricas "Costumbres y Problemas de Sociedad" en marzo de 1981.

Lo que se llama conflicto sobrepasa obviamente lo que se puede recensar a través de la prensa: la hipótesis que se hace es que operando así se obtiene un conjunto de conflictos explícitos tan diversos como el universo al que se hace implícitamente referencia.

4.2 Muestreo

Un grupo de unos diez expertos se ponen de acuerdo para retener, luego de discusiones a veces acaloradas, más de sesenta conflictos en el conjunto precedente considerado como universo. He aquí algunos de los conflictos que han sido seleccionados en todas las encuestas desde 1977: *el aborto, la fuerza de*

disuasión, la inmigración, la televisión, el crecimiento, la contaminación, la inseguridad.

No puede aquí hablarse explícitamente de sorteo al azar: hay sencillamente producción de cierto número de conflictos, de acuerdo a un procedimiento que permita a cada uno de los expertos expresar sus convicciones. La discusión entre expertos introduce implícitamente sobre el universo de conflictos las ponderaciones que reflejan los intereses medios suscitados (probabilidades subjetivas). Para respetar la decisión de los expertos, el observador va a hacer de tal manera que los conflictos tengan todos la misma importancia en el procedimiento de cálculo. Esto viene a dar a cada conflicto un peso $1/s_j^2$ (distancia M), donde s_j^2 es la variancia de las tomas de posición de los n individuos dentro del conflicto j -ésimo.

4.3 Construcción de las proposiciones "ad referendum"

Las proposiciones sobre las que los entrevistados se pronuncian por medio de una escala de cinco estadios se construyen alrededor de los conflictos precedentes. Por ejemplo, si se retoman los conflictos citados anteriormente, las formulaciones de las proposiciones correspondientes son:

- La liberalización del aborto es algo bueno.
- La fuerza de disuasión es indispensable.
- Hay demasiados trabajadores inmigrados.
- Nos creen tontos en la televisión.
- Se debe hacer todo por el crecimiento económico.
- La contaminación es terriblemente preocupante.
- Ya no nos sentimos seguros.

Se observará que en general la frase construida no comprende más que un solo término fuerte que se refiere al conflicto: para que no sea ambigua es importante retener un solo significado. La proposición debe además hacerse lo más cerca posible de las fórmulas usadas para debatir en el público.

4.4 Balance

Se ve entonces cómo se opera en la búsqueda de las estructuras de la opinión. *El pasaje a la estadística es una manera de preservarse de los sesgos que puede introducir la perturbación de la interacción considerada como sistema.* Las estructuras obtenidas (las primeras dimensiones principales), si pueden considerarse como características del fenómeno de la opinión, son también dependientes, a través del respeto de los diferentes principios (igualdad entre individuos, competencia del experto) de ciertas decisiones del observador.

5 Una interpretación a varios niveles de las primeras dimensiones factoriales

Si el cuestionario ha sido construido respetando el procedimiento establecido de extracción¹, las dimensiones factoriales obtenidas en una época dada de la tabla de respuestas a las proposiciones de referencia describen por construcción, de acuerdo con la teoría desarrollada y el modelo empleado, lo que son las representaciones sociales.

Al explotar las fotografías que ellas brindan de los conflictos (círculos de correlaciones) y de las opiniones (planos principales) intentaremos interpretar, es decir, de asociarles significados que van a permitir abordar una nueva etapa en el análisis de las opiniones. Para llegar a obtener el instrumento, el acercamiento ha sido hasta ahora global; no hemos establecido distinciones entre los conflictos ni entre los individuos sociales: lo esencial era que fueran representativos de sus universos respectivos.

Para interpretar las dimensiones producidas debemos observar más con detalle esos conflictos y los grupos sociales que los toman en cuenta. Producidas a partir de un esquema holístico, las estructuras extraídas permiten ir más lejos; medidores del universo hasta ahora, nos hemos vuelto entomologistas. Si la explicación desde el interior que buscamos ahora nos obliga a tomar un procedimiento que podría ser calificado de reduccionista, no olvidamos sin embargo los grandes espacios inter-sistémicos de donde viene el instrumento.

5.1 Inestabilidad de los indicadores factoriales

La estabilidad es la primera dificultad sobre la que se tropieza cuando uno se interroga sobre el significado de las dimensiones factoriales. Las dimensiones extraídas por el análisis en componentes principales, ¿coinciden de una época a otra? Si éste no es el caso: ¿es posible retener las dimensiones producidas en una época, como dimensiones de referencia?

Cuando se comparan los resultados obtenidos por el análisis en componentes principales desde 1977, se constatan los hechos siguientes:

- Los grupos sociales ocupan posiciones que pueden considerarse, con algunas excepciones, como fijas en los planos principales;
- La disposición de la mayoría de los conflictos sobre los círculos de correlación, estable de 1977 hasta 1981, ha evolucionado a partir de 1982. Pareciera que se ha retornado a una posición estable después de 1988.

¹El procedimiento no puede aplicarse con todo el rigor deseado, pues ciertas preguntas deben conservarse de una encuesta a otra.

Estos resultados no tienen nada de extraño. Las posiciones relativas ocupadas por los individuos y los grupos sociales, los estatus tales como son percibidos, no pueden evolucionar de manera significativa en un período breve. Para que haya evolución, es necesario a la vez que los roles ocupados por los individuos y las clases en las que se insertan hayan sido redistribuidos, y que esta redistribución haya afectado la percepción que se tiene del otro. Pero si el individuo se sitúa frente a los otros más o menos siempre de la misma manera, expresa su diferencia por medio de conflictos cuya significación puede cambiar rápidamente. Esta emerge de un debate entre actores regulado por una lógica de interés; si el rol de algunos de esos actores se modifica, los intereses en juego no son los mismos, el debate se transforma, las significaciones evolucionan. Una modificación así interviene obviamente cuando hay un cambio de mayoría presidencial: no se dice hoy lo mismo que se decía ayer cuando se critica al gobierno, cuando se cuestionan las instituciones que pertenecen al Estado o cuando se expresan dudas sobre lo fundamentado de un proyecto que no podría realizarse sin su voluntad.

La constancia sobre la estabilidad de las posiciones sociales sobre los planos principales no es entonces un descubrimiento; este conforma sencillamente el procedimiento adoptado: pese a cierta falta de vigor en el nivel de la construcción del cuestionario, se controla que el instrumento ha funcionado. Dando eso por un hecho, la comprensión de los mecanismos de la opinión pasa entonces por el análisis de los discursos y de su evolución: lo que se quiere decir a través de los conflictos, siendo más o menos constante, ¿cómo se dice y por qué se lo dice así en cada época?

Para interpretar las dimensiones factoriales se debe reconstituir la interacción, es decir, tomar en cuenta simultáneamente la posición de los conflictos sobre los círculos de correlación y la de los individuos sobre los planos principales; se busca explotar así la dualidad entre conflictos e individuos sociales. No conociendo a los entrevistados que fueron extraídos al azar, no se les hace aparecer sobre los planos principales, bajo la forma de centros de gravedad, sino a los grupos sociales a los que pertenecen. Incluso si las posiciones de esos grupos aparecen como fijas sobre los planos principales, si la mayoría de los conflictos cambian en el círculo de correlación, la significación dada a las dimensiones factoriales, siendo deducidas de las dos series de gráficos, debe cambiar. También hay que considerar que las dimensiones extraídas por el análisis en componentes principales no tienen la misma significación desde 1982, y su significado ha evolucionado hasta en 1988.

Si se vuelve al sistema actores-público, al esquema de la economía de roles, se ve que las referencias factoriales construidas en diferentes épocas no tienen la misma importancia. *De 1977 a 1981, luego de más de veinte años de poder gaullista (De Gaulle) o neo-gaullista, el sistema podría considerarse cercano a un equilibrio.* Los sobresaltos de mayo 1968, el ascenso de las mujeres, de la ecología, la liberalización de las costumbres, la crisis económica y la desocupación, las nuevas tecnologías, ... al provocar tantas discusiones han hecho

evolucionar las mentalidades ... pero el edificio institucional no se encuentra afectado en nada. Los nuevos actores que han aparecido con el desarrollo del movimiento asociativo no han cambiado mucho las cosas: el debate público ha estado siempre regulado por las mismas fuerzas y las mismas relaciones de fuerza. También puede decirse que los actores, al encargarse de los conflictos situados casi en el mismo lugar, el debate que los opone no ha cambiado fundamentalmente incluso si se ha dado en torno a temas en perpetua renovación. Es esto lo que indica la constancia de las estructuras factoriales entre marzo de 1977 y marzo o junio de 1981. Durante este período se hace diferencia siguiendo las mismas leyes, es decir, siguiendo la misma jerarquía de dimensiones, siendo el significado dado a los conflictos el reflejo de un juego estable de actores. El punto de referencia factorial de marzo 1981 es también la referencia de un *sistema en equilibrio estacionario*. En el sistema que evoluciona hay armonía entre actores y público: a la demanda casi constante del consumidor de roles corresponde, si se va más allá de lo que está dicho, una oferta que tampoco evoluciona.

La redistribución de poderes que sigue a la victoria de la izquierda en mayo de 1981 va a romper el equilibrio: los nuevos sitios ocupados por los actores de la arena política van a transformar el debate. Poco cambio aparente al inicio en el discurso de los actores: sin modificar sus proyectos, se calma el juego a la izquierda; se vuelve más incisivo a la derecha. Unificador por necesidad, el herético de ayer aprendía el ritual. En el público, de manera simultánea, se asiste a un reflujo de las ideas del programa común; la menor adhesión de la gente de la izquierda se combina en una cierta rigidez de poca amplitud a la derecha. En este terreno se observa un acercamiento de los puntos de vista entre las dos tendencias. El cambio brutal de roles objetivamente cumplidos provoca al inicio movimientos análogos en el debate entre actores y en las opiniones del público que no indican para nada una evolución profunda de las actitudes. El hecho de que se pueda cambiar de opinión con tal rapidez, de un lado como del otro, confiere al sistema un inercia que le permite evitar virajes destructores. Pero desde que los actos se precisan, a medida que las reformas son propuestas o realizadas, los nuevos argumentos que son producidos permiten al fin el vals de las etiquetas.

La justicia y la televisión de hoy no son ya lo que fueron ayer; sus imágenes evolucionan. Los del átomo militar y el átomo civil van a esbozar un movimiento que se explica de igual manera. Entramos en un período de perturbaciones internas, de desequilibrio, que terminará cuando las contradicciones aparecidas entre los roles atribuidos por un lado y por otro los discursos, las proposiciones e incluso los actos se hayan borrado, los hechos económicos y su impacto en la opinión estarán ahí para recordar al actor que su margen de autonomía es estrecho. Son las *capacidades de auto organización del sistema* las que resultan de las facultades de adaptación de los actores y de los individuos sociales, las que hacen posible la convergencia hacia un nuevo estado de equilibrio. Los puntos de referencia de 1982 a 1987 son entonces las referencias de un sistema desequi-

librado en plena reorganización: traducen la manera en que coyunturalmente funciona la diferenciación, en la espera de que por el juego en cascada de los aprendizajes recíprocos reine una mayor armonía entre actores, entre actores y público, y entre individuos sociales.

Por el momento no nos interesan los mecanismos del aprendizaje que permiten caminar del desorden hacia el orden, sino las representaciones: ¿cómo se opera con los conflictos para indicar al otro adónde se sitúa en relación a él? ¿Cuáles son las significaciones que se dan a tal o cual conflicto cuando se toma posición?

Las significaciones no son independientes del debate entre actores; hemos visto que éste las produce y es a la vez producto de ellas. Cuando hay un equilibrio estacionario, las significaciones de los conflictos de fondo (los que crean el dispositivo institucional y los grandes principios que le corresponden) evolucionan poco; no hay diferencia, es decir causa de desequilibrio, entre los roles jugados por los actores, su discurso y las significaciones acordadas a los conflictos por los individuos sociales. También puede analizarse la diferenciación en el público a la luz del debate que se desarrolla en el mismo momento entre los actores: la manera de diferenciarse refleja el debate y recíprocamente. Cuando ese equilibrio se rompe, para poder explicar lo que se dice en el público hay que tomar en cuenta lo pasado y lo presente; sentado entre dos sillas, uno se encuentra incapaz de explotar la dualidad que hace se correspondan actores y público para interpretar las estructuras puestas en evidencia; estas por otra parte no son más que un sucedáneo, esperando que la fase de transición entre dos estados de equilibrio se termine, lo que pareciera haberse producido en 1988.

Es por eso que *la encuesta de marzo de 1981 es considerada como una encuesta de referencia*; sus estructuras factoriales caracterizan un equilibrio, que permite plantearse interrogantes sobre su interpretación.

5.2 Etiquetaje de las dos primeras dimensiones factoriales

Las dos primeras dimensiones factoriales de marzo de 1981, tales como se las reconstituye en la encuesta de mayo de 1989², dan simultáneamente un mapa del cielo (Figura 11) cuyos conflictos son las estrellas, y un mapa de la tierra (Figura 12) donde se encuentran posicionados los grupos sociales. Los ejes de las connotaciones permiten ubicarse en el cielo (componentes principales) y los ejes de diferenciación que señalan la tierra (ejes principales) están en correspondencia por dualidad; esto hace indispensable contemplar simultáneamente los dos mapas para interpretar las dimensiones extraídas: los cuatro cuadrantes, numerados 1, 2, 3 y 4 en los dos gráficos se corresponden; en las asociaciones: 1

²Cada año se hace necesario recalcular las dimensiones factoriales extraídas de las dimensiones obtenidas los años precedentes (método del injerto). Eso permite borrar los sesgos estadísticos y evitar así su focalización sobre las interpretaciones que no son ciertamente diferentes de aquellas ya obtenidas.

con 1, 2 con 2, ... se unen los grupos sociales a las opiniones sobre los conflictos que los caracterizan.

Primeras impresiones

En el primer cuadrante se encuentran los más viejos y los que tienen menos diplomas; ahí se está como en una fortaleza: se defienden los valores, las reglas, las normas o los estereotipos de los que se ha dotado a la sociedad para resolver sus problemas; al mismo tiempo, a veces de manera paradójica, se denuncian ahí las disfunciones sociales. Este primer cuadrante del *repliegue* se opone al tercero, que es el de las grandes ofensivas y de la *aventura*. Aquí los más jóvenes, los más llenos de diplomas, los que se conocen como intelectuales de izquierda, desarrollan sus ataques contra las prohibiciones; es ahí que en 1981 se fabrican realmente las nuevas ideas. La polémica que opone los dos otros cuadrantes no se centra ni en las reglas ni en los valores. Las proposiciones hechas por los que ocupan el segundo cuadrante, y que en su mayoría se reconocen en 1981 en la oposición de izquierda, constituyen un verdadero *desafío*; más allá de las contradicciones, se pide más en el dominio de la libertad, de la igualdad y de la solidaridad a la vez. El desafío viene del hecho de que la reivindicación parece legítima: se vuelve a las primicias, a la declaración de los derechos del hombre que aparecen como el fundamento del Estado Republicano. Pero, ¿cierta distancia no separa siempre la teoría de la práctica, el ideal de la realidad, la intención de la acción? Las sociedades resultan de un aprendizaje; compromiso dado por la historia, las sociedades se someten permanentemente a la prueba de los hechos. También, siguiendo a Milan Kundera, ¿no sería más bien la idea de "la Gran Marcha"³ a la cual se adheriría apoyando este vasto programa que podría renovarse hasta el infinito? Lo que es cierto, es que aquellos del cuarto cuadrante, que se sitúan en mayoría al centro o a la derecha en la encuesta de marzo de 1981, aparecen insensibles a esta idea fabulosa! Frente al ideal desestabilizador, se oponen el pragmatismo y el realismo: proponiendo las medidas que no podrían comprometer los equilibrios, adoptando una actitud de la más grande *sumisión* a la realidad.

Estas cuatro categorías de actividades que corresponden a los cuatro cuadrantes aparecen como combinaciones de actitudes elementales que parecieran caracterizar las dos dimensiones factoriales: la primera dimensión "de la diferenciación" será aquella que opone a la *estabilidad* del *movimiento*; la segunda opondrá a la *dramatización* del *compromiso*.

Estas primeras impresiones dan un juego de ocho etiquetas que permiten etiquetar el primer plano factorial (ver Figura 13).

³ "La idea de la Gran Marcha, por la que se deja embriagar Franz, es el *kitsch* político que une a las personas de izquierda de todas las épocas y corrientes. La Gran Marcha es ese hermoso camino hacia adelante, el camino hacia la fraternidad, la igualdad, la justicia, la felicidad y aún más allá, a través de todos los obstáculos, porque ha de haber obstáculos si la marcha debe ser una Gran Marcha." (Milan Kundera, *La insoportable levedad del ser*, Tusquets, 1984).

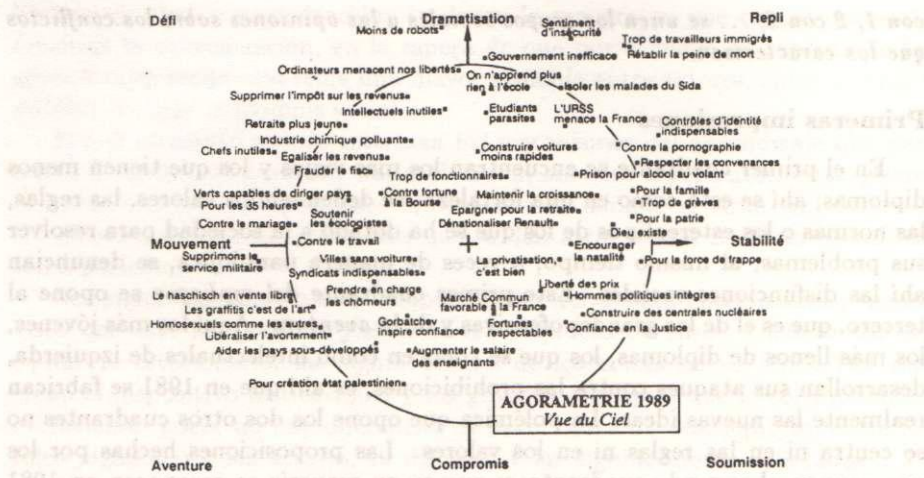


Figura 11: Vista del cielo

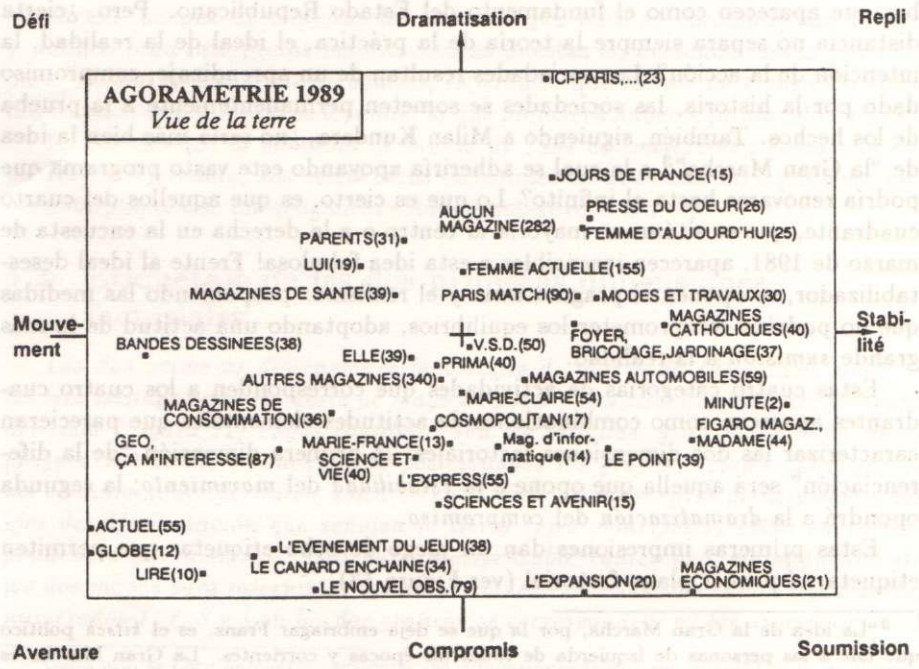


Figura 12: Vista de la tierra



Figura 13: Las regiones de la tierra y los ejes de la diferenciación: primer plano principal del ACP.

Estas etiquetas de las que una verdadera álgebra asegura la coherencia, no tienen otro interés que en la medida en que inspiran al usuario. Si este último quisiera seguir a René Girard, preferirá etiquetar de otra manera (Figura 14).

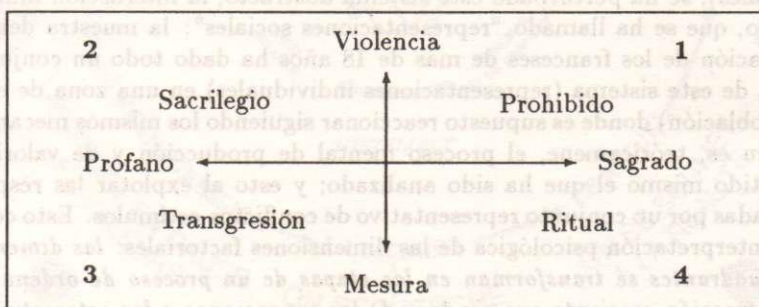


Figura 14: Otro etiquetaje de las regiones de la tierra, a la René Girard.

Aquí la *violencia* y lo *sagrado* son las dimensiones principales que determinan el conjunto de las actitudes: lo *prohibido* y el *ritual* son respectivamente la forma violenta y la forma medida de lo sagrado; estas formas se oponen al *sacrilegio* y a la *transgresión* que son a su vez, la expresión violenta y la expresión medida de lo profano. Se observa el interés —y el peligro!— de estas nuevas etiquetas que sugieren profundizar en René Girard para explicar lo que ocurre en el nivel de las opiniones.

Se puede así asociar juegos de etiquetas, que no se contradicen entre sí, a las primeras dimensiones factoriales; para lo sistémico, todos estos juegos son equivalentes. Se tiende a fijar el sistema en el primer cuadrante, a desestabi-

lizarlo en el segundo, a hacerlo transitar hacia un nuevo estado de equilibrio en el tercero; a preservar el equilibrio en el que se encuentra en el cuarto cuadrante.

Hacia una interpretación estructuralista a varios niveles

Hemos visto que imponiendo ciertas reglas se llegaría a escoger un juego coherente de etiquetas para designar las dos primeras dimensiones factoriales y los cuatro cuadrantes que delimitan. ¿Es posible orientar esta selección teniendo en cuenta el procedimiento que fue usado para construir el barómetro?

Al escoger los conflictos para observar la evolución de las opiniones nos hemos inspirado en Claude Lévi-Strauss; el mito perenne siendo reemplazado por el actor efímero, se ha retomado la idea que es una discusión entre el cielo (los mitos, los Dioses) y la tierra (los hombres) que son el desenlace de costumbres y opiniones. Lo que es estadístico en una tribu cortada desde hace tiempo del mundo exterior, está en equilibrio dinámico en una sociedad abierta donde las influencias se ejercen sin tapujos. Sin embargo, no es esta escogencia lo que hace que el procedimiento seguido pueda ser calificado de estructuralista; es ante todo porque las estructuras han sido obtenidas al explotar no las opiniones sino las diferencias entre las opiniones.

Para llegar a un problema de ajuste estadístico (el análisis en componentes principales), se ha perturbado este sistema abstracto, la interacción individuo-conflicto, que se ha llamado "representaciones sociales": la muestra dentro de la población de los franceses de más de 18 años ha dado todo un conjunto de estados de este sistema (representaciones individuales) en una zona de estados (esta población) donde es supuesto reaccionar siguiendo los mismos mecanismos. También es, teóricamente, el proceso mental de producción y de valorización del sentido mismo el que ha sido analizado; y esto al explotar las respuestas provocadas por un conjunto representativo de conflictos-estímulos. Esto conduce a una interpretación psicológica de las dimensiones factoriales: *las dimensiones y los cuadrantes se transforman en las etapas de un proceso de ordenamiento por abstracción creciente que conduce de las sensaciones a los estereotipos.*

Se ha hecho un esfuerzo en el gráfico de la Figura 15 de hacer coherente con las primeras impresiones esta traducción estructuralista de las dimensiones factoriales y de los cuatro cuadrantes. La interpretación de muchos niveles que se propone aparece a través de series de palabras clave dispuestas siguiendo círculos concéntricos: se pasa de los mecanismos mentales a los componentes al alejarse del centro. Esta interpretación en la que se establece una correspondencia entre instancias psíquicas y comportamientos sociales, es eminentemente especulativa; permite la comunicación entre los que discuten sobre los resultados de la encuesta, pues ella puede asimilarse a un sistema lingüístico coherente con el procedimiento seguido.

Se ve que las representaciones sociales han sido analizadas como un proceso dinámico que regula el comportamiento del individuo social.

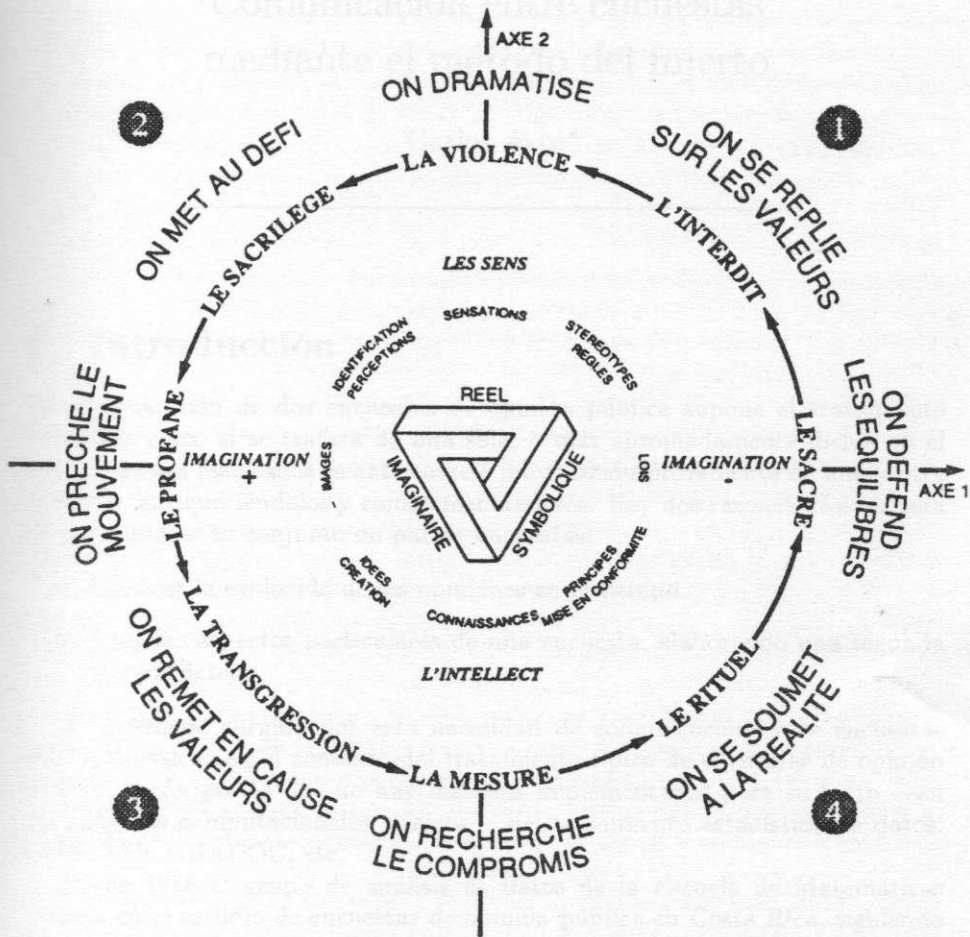


Figura 15: Interpretación estructuralista.

Comunicación entre encuestas mediante el método del injerto

Carlos Arce*

1 Introducción

La comunicación de dos encuestas de opinión pública supone el tratamiento de ambas como si se tratara de una sola, o más apropiadamente dicho, en el análisis de sus resultados se entremezcla información proveniente de una y otra encuesta, enriqueciéndolos y complementándolos. Hay dos razones básicas para querer estudiar en conjunto un par de encuestas:

- a. Conocer la evolución de las opiniones en el tiempo.
- b. Ampliar aspectos particulares de una encuesta, elaborando una segunda que los detalle.

El problema surgido por esta necesidad de comunicación entre encuestas resulta novedoso en el contexto del tratamiento típico de encuestas de opinión pública, razón por la que no hay métodos implementados para su logro —en los paquetes computacionales clásicos— del tratamiento estadístico de datos: SPSS, SAS, CHADOC, etc.

Desde 1988 el grupo de análisis de datos de la Escuela de Matemáticas trabaja en el estudio de encuestas de opinión pública en Costa Rica, siguiendo la metodología de *Agorametría*. Y para enfrentar las tareas de comunicación entre encuestas se ha implementado computacionalmente la técnica del injerto. En adelante se hace una exposición de ese trabajo.

2 Términos básicos

Encuesta de referencia: Es aquella diseñada para determinar la estructura de la opinión pública: verdaderas radiografías de las dimensiones principales de los mecanismos que originan la opinión en una sociedad.

*Profesor de la Escuela de Matemáticas de la Universidad de Costa Rica.

Encuesta particular: La que trata temas de interés más específico, por ejemplo: la problemática del SIDA, el perfil comercial de un detergente, o la evaluación de la función pública de alguna entidad.

VARIABLES DE SEÑALIZACIÓN: Son las variables que permiten particionar la población en grupos sociales, las cuales pueden ser muy diversas. Por ejemplo: nivel de estudio, sexo, edad, ingresos, etc. Se distinguen de las variables de conflicto, que recogen opiniones sobre temas de interés público que polarizan la opinión.

Individuos: Se refiere a las personas entrevistadas, pero también, a grupos sociales determinados por las variables de señalización.

Componentes principales o factoriales: Son los vectores C^1, C^2, \dots, C^t que resultan de un análisis en componentes principales ACP(X, M, D_p), con X la matriz de temas de conflictos y M y D_p ciertas métricas, los cuales determinan:

- 1) Los *planos principales*: regiones donde tiene cabida la representación de individuos o grupos sociales y a los que se identifica como planos de la *tierra*.
- 2) Los *círculos de correlación*: lugares de las representaciones de temas de conflicto, prejuicios e ideales, mitos e ideologías, donde los "dioses" gobiernan la opinión pública. A estos mundos se les reconoce como *cielo* y permiten calificar diferentes regiones de la *tierra* y los individuos representados en ella.

Estructuras de la opinión pública: Se refiere a los planos de la *tierra* y *cielo* donde se proyectan los distintos grupos sociales, según las opiniones de estos sobre los temas de conflicto. Los gráficos resultantes son como fotos de los grupos sociales, reflejados por sus opiniones, donde es posible observar sus cercanías y diferencias, o reconocer las características principales.

La comunicación de encuestas pretende tender los puentes necesarios para observar una encuesta particular, como información complementaria de una encuesta de referencia. De manera que los datos de interés específico sean más bien el resultado de "afinar la lente" para obtener detalles, dentro del marco más amplio y rico de la encuesta de referencia. También se propone estudiar encuestas de referencia a una misma población en tiempos distintos y observar la evolución de la opinión de los diferentes núcleos sociales.

Planteamiento del problema

El aprovechar las estructuras de la opinión generadas en la encuesta de referencia A , para el análisis de los datos de una encuesta particular B , lleva a la

problemática de representar los individuos de B en los planos de la *tierra* de A , como se ilustra en el gráfico de la Figura 1.

Las etiquetas que califican las regiones o ejes del plano (**Reto, Seguridad, Inconformidad, Intranquilidad, Mesura, Violencia, Disenso, Consenso**) son resultados de la encuesta de referencia A , al igual que algunos de los grupos representados. Pero, por ejemplo, otra encuesta B puede aportar la información que atañe a los grupos religiosos. Se obtiene así un gráfico que enriquece el análisis del comportamiento de los diferentes sectores del credo.

Para lograr el objetivo expuesto debe ser posible, a partir de los datos de la encuesta particular B , reconstruir las componentes factoriales de la encuesta de referencia A . Y para ello se requiere obtener cierta información básica de los individuos de B , acerca de las variables $\{x^1, x^2, \dots, x^p\}$, de las que nacieron las estructuras de la opinión en A .

Sea A una encuesta de referencia, con t variables de señalización y p variables de conflicto $\{s^1, s^2, \dots, s^t, x^1, x^2, \dots, x^p\}$ y B una encuesta particular, con variables de señalización según se requieran. Se propone elegir un subconjunto $\{z^1, z^2, \dots, z^q\}$ de variables entre x^1, x^2, \dots, x^p , las más representativas en cierto sentido que se precisa adelante y que deberán incluirse en la encuesta particular B . Estas variables constituirán un *punte* o pasarela que posibilitará la comunicación entre A y B .

3 Criterios para la selección del puente

Las siguientes bases definen criterios de calidad para una elección óptima de la pasarela:

1. El *punte* $\{z^1, z^2, \dots, z^q\}$ debe ser un conjunto de variables poco numeroso, en relación al número de variables iniciales $\{x^1, x^2, \dots, x^p\}$.
2. Las variables deben permitir la reconstrucción de la información aportada por la encuesta de referencia A que sea de interés, de la mejor manera posible.
3. Las variables deben tener la misma significación en las dos encuestas consideradas.
4. Las variables deben ser poco redundantes o lo suficientemente diferentes las unas de las otras.

A partir de 2 y dada la dependencia de *cielo* y *tierra* de las componentes factoriales, se establece que las variables seleccionadas deben reconstruir lo mejor posible las componentes principales significativas. Por tanto, hay que elegir variables muy correlacionadas con estos vectores y además, según 4, deben ser lo menos redundantes posible.

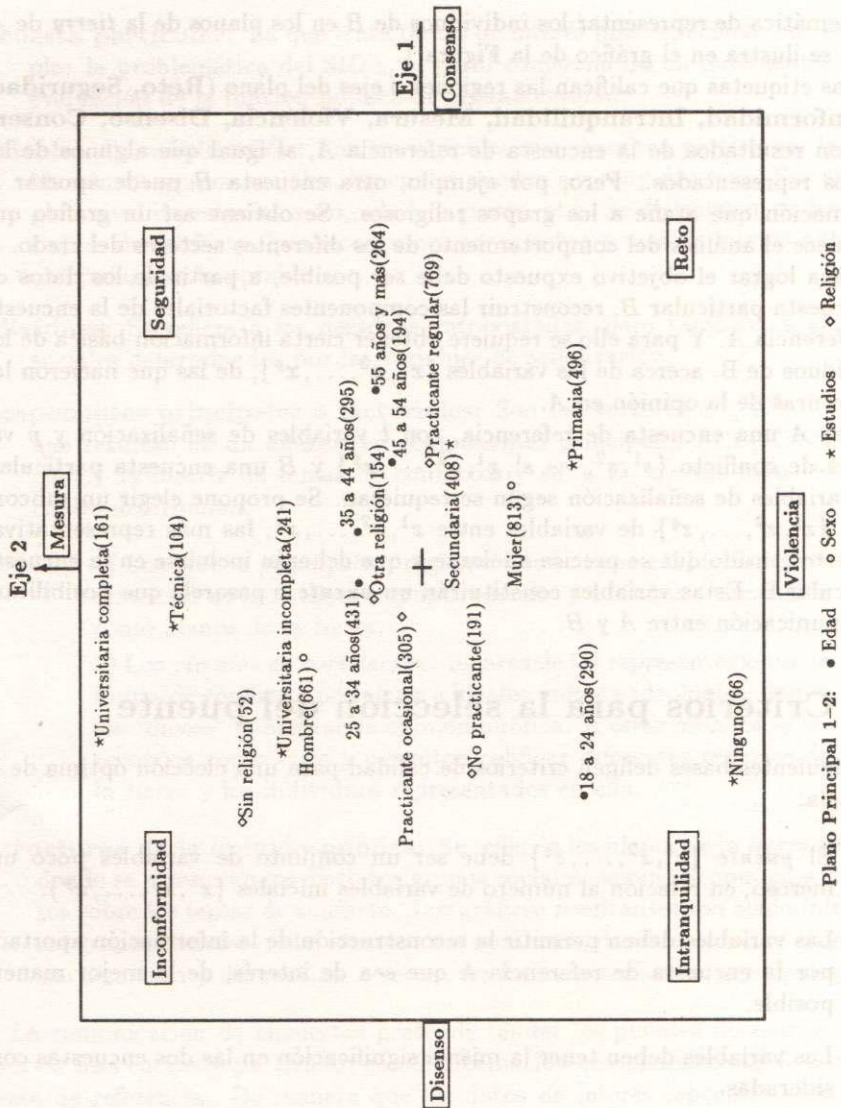


Figura 1: Plano de la tierra.

4 Selección optimal de las variables del puente

Sean $C = (C^1 C^2 \dots C^k)$ la matriz $n \times k$ cuyas columnas son las k componentes principales más significativas y $X = (x^1 x^2 \dots x^p)$ la matriz $n \times p$ con columnas iguales a las variables x^1, x^2, \dots, x^p de la encuesta de referencia, centradas y reducidas.

La idea anterior sugiere seleccionar el *puente* $\{z^1, z^2, \dots, z^q\}$ en la siguiente forma:

1. Sean $C_0^\ell = C^\ell$ para $\ell = 1, 2, \dots, k$.

2. Para $m = 1, 2, \dots, q$:

A. Buscar la variable j^* tal que:

$$R^2(x^{j^*}; C_{m-1}^1, \dots, C_{m-1}^k) \geq R^2(x^j; C_{m-1}^1, \dots, C_{m-1}^k) \quad \forall j$$

B. Elegir $z^m = x^{j^*}$.

C. Definir nuevos C_m^ℓ , eliminado de cada C_{m-1}^ℓ la parte de la correlación con z^m , previsible por regresión lineal,

donde:

$$R^2(x^j; C_{m-1}^1, \dots, C_{m-1}^k) = (x^j)^t [C_{m-1} (C_{m-1}^t C_{m-1})^{-1} C_{m-1}^t] x^j.$$

La parte A garantiza variables muy correlacionadas con las componentes importantes, rescatando aquellos temas conflictivos más relevantes en la definición del *cielo y tierra*. En C se procura eliminar la información redundante en las variables elegidas.

Para la implantación del proceso de selección del *puente*, se elaboró un programa en lenguaje C, implementado para ordenadores con procesadores 8088 y 80286. Los detalles del procedimiento se omiten, pero en resumen obedecen al siguiente planteamiento:

Algoritmo pasarela

Como antes $C = (C^1 C^2 \dots C^k)$ es la matriz $n \times k$ con columnas iguales a las k componentes principales más significativas y $X = (x^1 x^2 \dots x^p)$ la matriz $n \times p$ cuyas columnas son las variables x^1, x^2, \dots, x^p de la encuesta de referencia, centradas y no reducidas. Inicialmente $C_0 = C$, $X_0 = X$ y $S_0 = \{1, 2, \dots, p\}$.

Para $n = 0, 1, \dots, q$, se repite recursivamente:

1. Calcular $W_n = (C_n^t C_n)^{-1}$ mediante la descomposición en valores singulares de $C_n^t C_n$.

2. Calcular $Y = (y^1 y^2 \dots y^{p-n}) = C_n^t X_n$.

3. $\forall j \in S_n$ hacer $a^j = W_n y^j$ y calcular $R_j^2 = \frac{a^j \cdot y^j}{x^j \cdot x^j}$.

4. Determinar j^* tal que $R_{j^*}^2$ sea máxima.

5. Eliminar la influencia de la variable seleccionada j^* en cada componente principal. Para $j = 1, 2, \dots, k$

$$C_{n+1}^j = C_n^j - \frac{C_n^j \cdot X^{j^*}}{X^{j^*} \cdot X^{j^*}} X^{j^*}.$$

6. Definir $S_{n+1} = S_n - \{j^*\}$ y X_{n+1} como la matriz X_n eliminado de ella la columna j^* .

7. Si $(n \bmod 5) = 0$ determinar en cuanto se han explicado las componentes principales:

A. Definir $\mathcal{X} = (x^{i1} x^{i2} \dots x^{in})$ donde $x^{i1}, x^{i2}, \dots, x^{in}$ son las variables seleccionadas hasta el momento.

B. Calcular $(\mathcal{X}^t \mathcal{X})^{-1}$ y $\mathcal{X}^t C$.

C. Calcular $R_j^2(C^j; x^{i1}, x^{i2}, \dots, x^{in})$ porcentaje de varianza de la componente principal j explicada por las variables seleccionadas.

5 Representación de individuos de B en los planos de A

La manera de analizar los individuos de B , sobre la base de la representación factorial de la encuesta de referencia A , consiste en proyectarlos como puntos suplementarios sobre los ejes factoriales señalados. Para ello se requiere, a partir de las variables del *punteo* en B , reconstruir por regresión las componentes factoriales de A . Así, se podrá "injetar" los individuos de B , en los gráficos de la encuesta A .

La coordenada de un individuo i de B , sobre el indicador factorial C^ℓ , es aproximada por

$$\hat{C}_i^\ell = \sum_{j=1}^q b_j^\ell \frac{z_i^j - m^j}{s^j},$$

donde b_j^ℓ es el coeficiente de regresión asociado a la j -ésima variable del *punteo*, en la reconstrucción de la componente $C^\ell = b_1^\ell z^1 + b_2^\ell z^2 + \dots + b_q^\ell z^q + \epsilon^\ell$. Y m^j, s^j son respectivamente la media y desviación estandar de la j -ésima variable. $Z_i^t = (z_i^1, z_i^2, \dots, z_i^q)$ es el vector de respuestas del individuo i , no centradas.

**VARIABLES SELECCIONADAS PARA EL PUENTE.
ENCUESTA COSTA RICA, 1988**

Conflicto seleccionado	Correlación con				
	C ¹	C ²	C ³	C ⁴	C ⁵
14 Financiar campañas de vacunación					
06 Adoptar la semana de 35 horas					
34 Los refugiados deben salir					
51 Gorbatchov inspira confianza					
05 Se debe dar capacitación al desempleado	.462	.452	.165	.328	.162
13 El seguro social debe privatizarse					
20 Hay que fortalecer la banca privada					
52 Está bien que los hombres se maquillen					
36 La policía no cumple su misión					
23 Hay que defender la banca nacionalizada	.572	.584	.542	.527	.512
54 Las amas de casa deben recibir salario					
37 Se debe acortar la campaña electoral					
40 Homosex. deben integrarse a la sociedad					
16 Introd. sector privado en los hospitales					
33 Se puede confiar en la justicia	.656	.657	.616	.632	.609
35 Luchar enérg. contra la pornografía					
01 Hay que fomentar la creación de empresas					
19 El FMI ayuda a resolver la crisis					
32 Dios existe					
11 Extranjeros no deben beneficiarse del S.S.	.740	.721	.708	.721	.764
31 Se debe extraditar a los extranjeros					
46 El precarismo es un mal social					
03 Hay que crear empleo en el sector público					
43 Hay demasiadas huelgas					
55 Igualdad de h. y a. en puestos elegibles	.819	.781	.811	.803	.823

Tabla 1: Selección de las variables del puente.

6 Ejecución del algoritmo pasarela

El procedimiento de selección de variables del *puente* fue aplicado a los datos de la encuesta de referencia de 1988, con el propósito de elegir veinticinco variables que explicaran de la mejor manera posible las primeras cinco componentes principales. Los resultados obtenidos se muestran en la Tabla 1.

En la Figura 2 se observa la distribución de estas variables en el círculo de correlaciones de los ejes 1 y 2, primer plano del *cielo*.

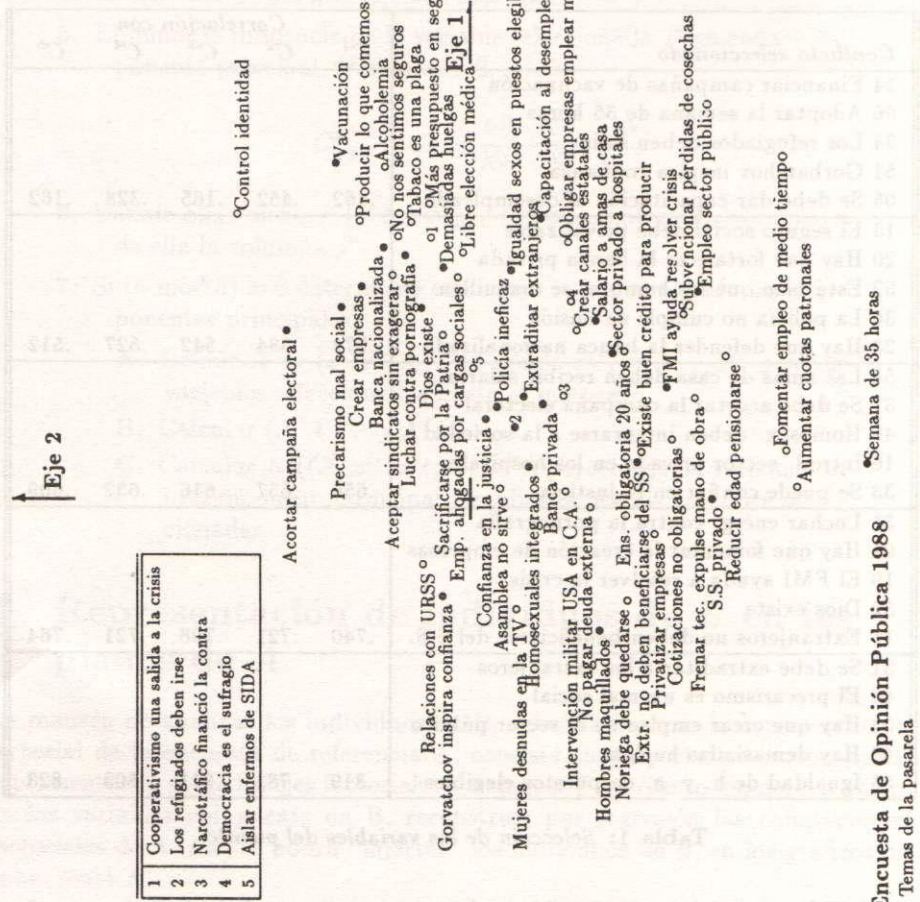


Figura 2: Círculo de correlaciones de los ejes 1 y 2.

7 Evaluación y perspectivas

Con la implementación del algoritmo pasarela se dispone de un primer recurso para obtener mayor provecho de las encuestas de referencia realizadas. Los resultados de estos sondeos pueden analizarse en mayor detalle hacia diversas temáticas de interés específico mediante las encuestas particulares.

La elección de un *punte* termina luego de haber efectuado la encuesta particular *B*, y con los nuevos datos someter las variables seleccionadas a procesos de calibración, con el propósito de verificar que la significación de estas variables no se ha modificado efectivamente. Eventualmente en esa etapa, puede reducirse el número de variables del *punte*, para ajustarlo a los criterios de calidad definidos. En esta dirección deberá continuar el trabajo.

Paralelamente, la solución del entronque de encuestas necesitó resolver el cálculo de matrices inversas de una manera eficiente. Algoritmo logrado mediante la descomposición en valores singulares de la matriz, el cual es un proceso estable aun con matrices de tamaños 100×100 o 150×150 . Esto a la vez posibilita la implementación de otros algoritmos importantes en el análisis de datos, que involucran matrices de tales tamaños, como los mismos cómputos del ACP. Se vislumbra así la posibilidad real de salirse del marco que definen los paquetes estadísticos clásicos, para los propósitos de investigar variantes en las metodologías o para responder de mejor manera a necesidades propias y más específicas.

Referencias

- [1] Cailliez, F. et Pagés, J.P.: *Introduction à l'Analyse des Données*. Editorial SMASH, Paris, 1975.
- [2] Greenere, Michael J.: *Applications of Correspondence Analysis*. Academic Press, Orlando Fda, 1984.
- [3] Iliakopoulos, Anastassios: *Etude de l'Opinion et enquêtes de référence*. Thesis Doctorat, Université Paris- Dauphine, 1988.
- [4] Poltronieri, Jorge y Piza, Eduardo: *Estructuras de la Opinión Pública en Costa Rica*. Editorial Universidad de Costa Rica, San José, 1989.
- [5] Press, William H. y Otros: *Numerical Recipes, the Art of Scientific Computing*. Cambridge University Press, New York, 1986.

Ajuste estructural, salud y opinión pública

Nora Garita Bonilla
Marta Eugenia López Subirós*

1 Introducción

En un momento histórico en el que se redefine el Estado costarricense, creemos importante conocer lo que piensa la opinión pública al respecto. Dentro del juego democrático los gobernantes deben escucharla y respetarla, como un elemento considerable en la toma de decisiones políticas.

Es nuestra intención realizar estudios de opinión pública en la óptica del conocimiento de lo que ocurre en *el ágora*, en la plaza pública, en el diálogo democrático de *la polis*. Medir la opinión, para escucharla y respetarla.

Un aspecto fundamental en el proceso de re-estructuración del Estado se refiere al campo de la salud, por el rol protagónico en la salud pública que ha asumido el Estado en Costa Rica.

2 Ajuste estructural y salud

El *ajuste estructural*, política económica vigente, pretende eliminar las distorsiones económicas que obstaculizan mejoras en la estructura productiva. Entre sus principales medidas (descritas en la carta de intenciones del Banco Mundial) están [1]:

- Reducción arancelaria a un máximo de 40%.
- Reducción de los depósitos previos a la importación.
- Nuevos incentivos a las exportaciones.
- Disminución de los subsidios del Consejo Nacional de Producción (CNP), al maíz, arroz y frijoles.

*Profesoras de la Escuela de Sociología y Antropología de la Universidad de Costa Rica.

- Reducción del diferencial de precios nacionales de granos básicos con las cotizaciones internacionales.
- Libre importación de granos básicos, ante escasez nacional.
- Mejoras en la administración financiera de los bancos estatales.
- Límites cuantitativos al crédito subsidiado.
- Aumentos en el ahorro y la inversión pública.

Es decir, ajuste en los sectores productivos, del Estado y el sector externo. Para los neoliberales, la racionalización del gasto y el tamaño del Estado implica entonces una redefinición de su papel y el abandono del *paternalismo del Estado* [1].

La pregunta que nos planteamos es: —La seguridad social, tal como se fue perfilando en Costa Rica desde los años 40 y cobró mayor cobertura y servicios en los años 50 y 60, ¿es considerada como paternalista por parte de los economistas neoliberales? ¿Será parte de los “privilegios” que hay que eliminar? [2].

3 Estado benefactor y salud

El modelo anterior de desarrollo se implementa en Costa Rica a partir de 1948 y llega al límite con la crisis de 1980 ([4], capítulo 1).

El desarrollo costarricense de estas tres décadas es considerado por Rovira como la resultante histórica de la contraposición y el conflicto entre dos principales proyectos políticos, alrededor de los cuales han quedado articuladas diferentes coaliciones de intereses burgueses que han aspirado a dirigir el desarrollo del país ([4], página 17).

El primero es el proyecto del partido Liberación Nacional y el otro, impulsado por diversos partidos que fundamentalmente “ha procurado resistir las orientaciones emanadas del primero” ([4], capítulo 1).

Los rasgos del proyecto liberacionista podrían resumirse así:

- Diversificación de la estructura productiva.
- Medidas de distribución y redistribución del ingreso.
- Profundización de dos valores fundamentales para el costarricense: seguridad social y educación.
- Constitución de un Estado intervencionista (por ejemplo, nacionalización bancaria, del Instituto Nacional de Electricidad (ICE), y de Acueductos y Alcantarillados (A y A).

Es pues, un período en el que el Estado asume en Costa Rica muchas funciones de bienestar público y funciones de desarrollo.

Todo esto repercutió, entre otras cosas, en los componentes demográficos. Tomándolos como indicadores de bienestar en la salud y bienestar económico en general, los componentes demográficos evolucionaron de manera positiva en el período 1950-1980.

Veamos dos componentes demográficos importantes: la esperanza de vida y la tasa de mortalidad infantil.

La esperanza de vida al nacer tiene mucho que ver con las posibilidades de desarrollo económico de un país, al estar ligado a las condiciones que van a permitir a un individuo vivir teóricamente cierto número de años.

En 1950 en Costa Rica la esperanza de vida cuando alguien nacía era de 57.26 años. En los hombres era algo más baja que las mujeres (56 años). Esto fue evolucionando; se llega al quinquenio 1975-1980, con un aumento de 13 años para ambos sexos, porque en 1980 la esperanza de vida ya es de 70.80 años. En los hombres era de 68.63 años, en las mujeres de 73.08 años.

Otra de las tasas que se utiliza es la mortalidad infantil. En 1950 era de 97.4 muertes por 1000 nacimientos. Al llegar a 1980 se ha reducido a una tasa de 35.6 por 1000. Quiere decir entonces que ese año primero de la vida (determinante como indicador de las condiciones que favorecen que los niños vivan) ha tenido una notable mejoría en estos 30 años de desarrollo costarricense.

Respecto a las instituciones que tienen que ver con la salud fundamentalmente preventiva o asistencial, el papel central lo juega la Caja Costarricense del Seguro Social (CCSS). En 1950 se crea la Dirección General de Asistencia Médico Social, como responsable, por parte del Estado, de la coordinación y administración de los fondos destinados a los centros hospitalarios. Aquí es donde surgen las Juntas de Protección que anteriormente eran conocidas como Juntas de Caridad. En 1956 se da un cambio importante, al modificarse el "Reglamento de Enfermedad y Maternidad" para introducir la cobertura familiar obligatoria a la esposa o compañera y a los hijos menores de 12 años, así como a los padres en caso de que fueran dependientes del asegurado directo. Es decir, empieza a beneficiarse la familia costarricense con los servicios del Seguro Social. A pesar de que en 1956 se incluía solo a los hijos menores de 12 años, eso se va a ir ampliando en el futuro.

En el año 1960 se amplía también la cobertura del "Seguro de Invalidez, Vejez y Muerte". En 1966 se extiende el seguro de "Enfermedad y Maternidad" a las zonas rurales que se habían visto aisladas hasta ese momento de este tipo de beneficios.

Al finalizar la década de los 60 ya la CCSS tenía una cobertura del 45% de la población total del país. En 1971 se da una modificación en la Asamblea Legislativa, y se vuelve obligatoria la extensión a toda la población del país el "Régimen de Enfermedad y Maternidad", y se asigna un plazo de 10 años para que la CCSS haga una realidad la universalización. Se empieza también el traspaso de hospitales a la CCSS, hospitales que antes estaban en manos de la

Junta de Protección Social o del Ministerio de Salud.

Cuando el Seguro Social se inició en Costa Rica no cubría a los trabajadores agrícolas y tampoco a los no asalariados. El Hospital San Juan de Dios funcionaba como un lugar donde se daban servicios de salud a los no asalariados, y los servicios de la CCSS que fueron creciendo por muchas poblaciones, que fueron teniendo sus clínicas, y que tenían ya sus instalaciones médicas adecuadas, cubrían a los asegurados. Este es pues un esfuerzo más que se hizo en esos años, ampliando la cobertura a la población costarricense que tenía derecho a los servicios de salud.

Esto es importante entenderlo en el modelo de desarrollo que venía impulsándose en esta época. Se quería desarrollar la industria. Hay que tomar en cuenta que Costa Rica carecía de una fuerza de trabajo abundante (a diferencia de otros países centroamericanos). Era de suma importancia, para el modelo de desarrollo, levantar los niveles de vida y rebajar los índices de mortalidad, para poder tener un contingente de fuerza de trabajo que implementara el proyecto de desarrollo que se impulsaba en esa época.

4 Problemática actual

La crisis de 1980 nos lleva en la presente década a la modificación de las políticas económicas, en una conjugación de razones internas del agotamiento del modelo anterior (evidenciado por la crisis) y por propuestas de organismos internacionales.

El programa de ajuste estructural tiene como elemento central la búsqueda de la eficiencia. Esto a nivel del Estado significa redefinir sus funciones e Instituciones, con tendencia creciente a la privatización en los servicios que antes estaban en sus manos.

Entre los sectores que serán incorporados al ajuste está el sector salud. Se ha querido privatizar ya el servicio de farmacia y ciertos servicios hospitalarios. Paralelo a esto, las medidas de liberalización de la economía encarecen las medicinas y los equipos.

En este contexto toma relevancia la opinión pública. A partir de la encuesta nacional, realizada en 1988 por el grupo de investigadores del proyecto Estructuras de la Opinión Pública en Costa Rica [3], se puede analizar dicha opinión referente a los temas de salud, los cuales se resumen en la Tabla 1.

Ante la pregunta *¿Se debe privatizar el Seguro Social?*, un alto porcentaje de la población se muestra en desacuerdo (69.9%). Es decir, pese a las quejas orquestadas por los grandes medios de comunicación respecto a la consulta externa, a la lentitud de la farmacia hospitalaria, etc., el 69.9% de los costarricenses mantienen como un valor el *"Seguro Social en manos del Estado"*.

En esta pregunta, solo un 14.3% está de acuerdo en privatizar el Seguro Social. Las personas que se dijeron no tener religión, o los que eran no practicantes o de otra religión diferente a la católica, también se muestran radicalmente en

Opinión pública nacional y salud por provincia, 1988.

Temas de conflicto	Promedio	SJ	A	C	H	G	P	L
Reducir edad pensionarse (acuerdo)	33.6	30.2	38.0	36.1	37.1	32.6	40.8	32.1
Aumento cuotas patron. (desacuerdo)	55.9	56.0	52.0	61.3	54.1	51.1	51.2	68.6
Extran. no deben benef. (desacuerdo)	52.4	56.5	46.9	54.2	51.3	59.1	39.8	62.7
Cotiz. no deben ser oblig. (desacuerdo)	44.4	44.6	37.4	45.8	51.7	47.9	43.3	48.5
Finan. campañas vacun. (acuerdo)	88.6	87.0	93.0	90.4	79.7	94.7	87.2	91.3
Introd. hospitales sector privado (acuerdo)	39.2	41.6	46.5	43.5	35.6	36.8	32.3	20.8
SS debe privatizarse (desacuerdo)	69.9	68.4	66.1	66.9	70.1	87.4	64.1	86.4
Libre elección médica (acuerdo)	68.8	70.1	71.7	69.4	62.7	65.9	67.7	65.7
Enf. SIDA deben aislarse (acuerdo)	61.1	58.7	56.7	66.7	55.5	59.7	74.9	64.7
Privatizar medicinas SS (desacuerdo)	74.1	73.1	67.2	79.5	81.7	81.7	71.3	69.3

Tabla 1: Porcentajes de respuesta a los temas propuestos. Fuente: Encuesta opinión pública 1988 [3].

Significado de las etiquetas

Reducir edad pensionarse	Hay que reducir la edad de pensionarse.
Aumento cuotas patron.	Se debe aumentar las cuotas patronales.
Extran. no deben benef.	Los extranjeros no deben beneficiarse del Seguro Social.
Cotiz. no deben ser oblig.	Las cotizaciones del Seguro Social no deben ser obligatorias.
Finan. campañas vacun.	Hay que favorecer el financiamiento de las campañas de vacunación.
Introd. hospitales sector privado	Hay que introducir al sector privado en los hospitales.
SS debe privatizarse	El Seguro Social debe privatizarse.
Libre elección médica	Hay que propiciar la libre elección médica.
Enf. SIDA deben aislarse	Los enfermos del SIDA deben aislarse.
Privatizar medicinas SS	Para salir de la crisis estaría Ud. dispuesto a aceptar, al menos por un tiempo, la privatización de las medicinas del Seguro Social.

contra. En cuanto a educación se refiere, las personas con ninguna educación, o solamente con educación primaria, se muestran más favorables que los otros niveles. En este caso la educación es significativa. Podría pensarse que cuanto más alto sea el nivel educativo, la capacidad de valoración positiva de la seguridad social pública es también mayor.

A continuación se comentan los resultados obtenidos en otros temas de conflicto que se incluyeron en la encuesta, relacionados con el tema de la seguridad social y la salud pública [3].

- “*Hay que reducir la edad de pensionarse*”: En este tema la provincia no es un factor determinante para tomar posición. Puntarenas es la provincia que más a favor se manifiesta, con un 40.8%.
- “*Se deben aumentar las cuotas patronales del Seguro Social*”: En este tema la región tampoco incide fuertemente. Sin embargo encontramos el mayor desacuerdo en Limón con un 68.6% y el menor con un 51.5%.
- “*Los extranjeros no deben beneficiarse del Seguro Social*”: Al respecto, la opinión se manifiesta muy determinada por la provincia. Siendo Limón junto con Guanacaste, los más abiertos a que los extranjeros hagan uso del Seguro Social. En el otro extremo, se ubica la provincia de Puntarenas, con un 39.8% de la opinión en contra de la afirmación.
- “*Las cotizaciones del Seguro Social no deben ser obligatorias*”: El comportamiento por provincia es muy homogéneo, sobresaliendo levemente en el desacuerdo.
- “*Hay que favorecer el financiamiento para las campañas de vacunación*”: El porcentaje nacional de acuerdo es muy alto (88.6%). Se observan leves variaciones en las provincias siendo Guanacaste y Alajuela las que más lo favorece.
- “*El Seguro Social debe privatizarse*”: En este tema podríamos afirmar que la tendencia es el desacuerdo, ya que un 69.9% de la población se opone a la privatización del Seguro. Las provincias de Guanacaste y Limón, con grandes problemas socio-económicos y por ende de salud, son las que más oposición presentan (87.4% y 86.4% respectivamente).
- “*Hay que introducir al sector privado en los hospitales*”: Este tema está muy determinado por las provincias, pues a nivel nacional se da una repartición bastante homogénea en la toma de posición. Así por ejemplo podríamos ubicar a Heredia, Guanacaste y Puntarenas en una posición intermedia. Siendo más favorables en orden de importancia Alajuela (46.5%), Cartago (43.5%), y San José (41.6%). Situándose en el otro extremo se ubica Limón, donde solo un 20.8% aprueba esta medida.

- “Hay que propiciar la libre elección médica”: Este tema no está determinado por ninguna de las variables sociodemográficas, por lo que podemos afirmar que, a pesar de ser un conflicto, no constituye una de las mayores preocupaciones del costarricense.
- “Los enfermos de SIDA deben aislarse”: La opinión ante este tema refleja un desconocimiento de la enfermedad en mención. Pareciera que el temor al contagio es alto (61.1% de acuerdo con aislarlos), siendo Puntarenas la provincia que más fuertemente apoya esta medida (74.9%).
- “Para salir de la crisis estaría Ud. dispuesto a aceptar, al menos por un tiempo, la privatización de medicinas del Seguro Social”: Este tema no constituye un tema de conflicto, sino un tema de sacrificio temporal ante la crisis. Pese a lo transitorio supuesto del enunciado, la población opina no estar dispuesta a la privatización (74.1% en contra). Existen variaciones por provincia (ver Tabla 1).

Es importante destacar que los temas anteriores la opinión no varía de acuerdo al sexo.

Pese a la constante queja en este sentido aparecida en los medios de comunicación, el 74% de la población costarricense está en contra de privatizar las medicinas del Seguro Social y sólo un 11.1% está de acuerdo en privatizarlas.

En relación con la posición política, la izquierda (cuadrante de la *inconformidad*) se declara en contra. La derecha y la extrema derecha se muestran más favorables. Esto pareciera tener que ver con el apoyo dado por la derecha y la extrema derecha al proyecto de ajuste estructural (PAE).

Otro tema de sacrificio es —“¿estaría usted dispuesto a dedicar tiempo y dinero para investigar en salud?”. En este, un 59.7% de la población está de acuerdo.

Al preguntársele si estaría dispuesto, aunque fuese en forma temporal para salir de la crisis, a perder los derechos que ya tiene en cuanto a protección social, el 89.9% se manifestó en contra. Solo un 5% de la población estuvo de acuerdo.

Pareciera haber conciencia en la mayoría de los costarricenses de la importancia de las campañas de salud preventiva, si se interpreta de esta manera la respuestas a la pregunta sobre el financiamiento de las campañas de vacunas. El 88.6% está de acuerdo en financiar compañías de vacunas, solo un 2.7% en desacuerdo.

5 Observaciones finales

Recapitulando respecto a los datos a nivel nacional, el ajuste estructural es un proceso que se está llevando a cabo desde la administración del presidente Luis Alberto Monge, un proceso que se está consolidando con el transcurso de los años. El costarricense pareciera entonces no querer sacrificar los logros en salud.

La opinión pública es diferencial con respecto al tema de salud según la provincia. Esto podría deberse al hecho de que la realidad socio-económica también es diferencial por provincia, esto es, a menos desarrollo más apego a la salud como un servicio público y más apego al tipo de Estado Benefactor.

El análisis de la opinión pública nacional de otros temas relacionados con el PAE, que no son tratados en este trabajo, arroja resultados similares a los del tema salud. Esto podría reflejar el bajo nivel organizativo de los sectores sociales afectados por la política económica que no han logrado proponer un proyecto alternativo a sus intereses.

Referencias

- [1] Guardia Quirós, Jorge: *Apoyo al ajuste estructural*. Periódico La Nación, 12 de mayo de 1989, pág. 16-A.
- [2] Lizano, Eduardo: *No es fácil eliminar privilegios*. Periódico La Nación, Foro dominical, 12 de marzo de 1989, pág. 15-A.
- [3] Poltronieri, Jorge; Piza, Eduardo; et. al.: *Estructuras de la Opinión Pública en Costa Rica*. Editorial de la Universidad de Costa Rica, San José, 1989.
- [4] Rovira, Jorge: *Costa Rica en los años 80*. Editorial Porvenir, Tercera Edición, San José, 1989.

La opinión pública regional y la política económica en Costa Rica, 1988.

Marta Eugenia López Subirós*

Resumen

El presente trabajo pretende incorporar la opinión pública como elemento a tomar en cuenta antes de la formulación y ejecución de la política económica.

Se realiza un estudio de la opinión pública regional, pues se considera que la política económica incide diferencialmente en las regiones.

La perspectiva regional incorpora la dimensión territorial, tan ajena a los estudios tradicionales que se han hecho de la crisis y de la política económica.

Para cumplir con este cometido se utiliza una metodología novedosa en el país: el *análisis en componentes principales* (ACP).

1 Aspectos teóricos y metodológicos

Se entenderá la crisis como la interrelación entre acumulación y conflicto capital-trabajo, es decir, la crisis como "cambios estructurales" refleja el conflicto social.

La crisis se da cuando hay un cambio en las relaciones de fuerza entre capital y trabajo.

El conflicto se institucionaliza. Este proceso es también una relación dialéctica entre conflicto y pacto (procesos de control sobre las relaciones de trabajo). Los conflictos son un modo de administración de la correlación de fuerzas, generan debates que son retomados por los medios de comunicación colectiva. Estos debates los realizan un conjunto de instituciones (estado, empresas, sindicatos, asociaciones...), que se llamarán *actores sociales*. El público juega un papel de árbitro omnipresente.

La opinión pública se genera como resultado de la interiorización de una lógica y del acceso a un cierto conocimiento. Las opiniones pueden considerarse

*Profesora de la Escuela de Sociología y Antropología de la Universidad de Costa Rica.

como reflejo de un equilibrio análogo al existente en el mercado, donde también hay dos clases de protagonistas: los actores sociales y el público. La competencia entre actores se evidencia en el debate que se da, en parte, en los medios de comunicación colectiva. El debate se convierte en un espectáculo para el público y se da el juego de *la oferta y la demanda*: interviene la diferenciación.

Lo anterior permite introducir el concepto de economía de roles, que será el triple juego de las siguientes interacciones:

- entre los actores sociales: *el debate*.
- entre actores y el público: *oferta y demanda de roles*.
- entre individuos sociales: *la diferenciación*.

La diferenciación lleva a nuevos equilibrios bajo el impulso de los conflictos, donde el debate entre instituciones y las opiniones quedan en correspondencia con sus diversidades.

Se tratará la crisis como el escenario donde se dan las diversas políticas económicas para resolverla. Estas políticas a su vez generan conflictos que son tratados por los medios de comunicación colectiva, de donde se extraerán para conocer la opinión pública regional.

El camino metodológico es la obtención de indicadores que se puedan considerar características de la diferenciación, provistas por el modelo factorial y el análisis en componentes principales.

Tomando en cuenta la interacción, se puede hacer corresponder por dualidad conflictos e individuos sociales. Si el conflicto engendra actitudes individuales, éstas recíprocamente dan significado al conflicto.

Del mismo modo, el concepto de las desigualdades regionales, a nivel metodológico, posibilita un modo de aproximación de los fenómenos sociales consecuente con una noción más totalizadora.

El instrumento fue una encuesta a nivel nacional-regional a una población de 1.500 personas costarricenses y mayores de 18 años. Para la elaboración de la misma y la extracción de los conflictos, se analizaron todos los medios de comunicación colectiva que difunden noticias (radio, prensa escrita y televisión) durante un mes, previo al trabajo de campo.

2 La crisis y la política económica, 1979-1986

La crisis para el período 1979-1986 ha sido caracterizada, por diferentes autores, fundamentalmente a partir de indicadores económicos: caída del ritmo de crecimiento del producto interno bruto (PIB), caída del ingreso por habitante en términos reales, aumento de la inflación, y aumento del desempleo abierto entre otros.

Las causas se han encontrado en la apertura de la economía y en la diversificación limitada de las exportaciones, así como al proteccionismo y al rápido y

no planificado crecimiento del sector público. Otra causa citada es la falta de integración vertical del sistema productivo y su excesiva dependencia económica, tecnológica y de insumos.

La crisis se profundiza por el alza en los hidrocarburos, la recesión mundial, la situación política centroamericana, el servicio de la deuda pública externa y el incremento del déficit del sector público. Esta situación se logra controlar en el período 1983-1986 y hacer un poco más manejable la crisis.

Partiendo del concepto de crisis como un cambio en la correlación de fuerzas entre capital y trabajo, se encuentra una doble contradicción: el gasto público crece más que el PIB y la política de expansión de la demanda tiende a ser inflacionista, ya que el mecanismo clásico del mercado deja de funcionar. La crisis de acumulación tiene efectos en las relaciones de producción afectando al capital y al trabajo. Por ello la crisis económica como emergencia de nuevas estructuras productivas y desestructuración de las antiguas, comporta efectos de descomposición sobre la fuerza de trabajo presente en los sectores no renovados y unos efectos de recomposición de la fuerza de trabajo necesaria para los sectores en alza.

Es con este escenario que se implementan una serie de medidas de política económica para contrarrestar la crisis. Se parte del problema de balanza de pagos y del déficit fiscal.

A inicios del período en estudio, el desmantelamiento de las actividades que requieren recursos externos y no dan beneficios suficientes al capital, se convierte en objetivo de la política económica de estabilización.

Se nota la intencionalidad por impulsar el desarrollo agrario en nuevas áreas de explotación que se vinculen con los mercados de exportación. Se opta por devaluar la moneda y por una política de contracción del gasto público. En materia de política salarial, las medidas adoptadas fueron insuficientes para recuperar la pérdida del poder adquisitivo.

Se empieza a expresar la necesidad de privatizar más la economía y de liberalizar las relaciones de intercambio. Se dan aumentos de los precios, de las tarifas del sector público y de las tasas de interés reales, así como un reajuste de precios a los artículos de consumo popular.

El comportamiento contractivo del sistema productivo nacional se reflejó en el campo del empleo. El desempleo se afectó tanto por el incremento en la oferta como por la contracción en la demanda de mano de obra.

El eje de la política económica se da en el aumento de las exportaciones y en la restricción de la demanda interna. Se insiste en reducir y cambiar el tamaño y el papel del Estado, priorizando así los criterios de rentabilidad privada.

Para 1983 los objetivos de política económica eran lograr la estabilidad en precios, en el tipo de cambio y en la posición externa. Se aplicó una política tendiente a recuperar el deterioro salarial (30% en términos reales). Se sigue con la medida de ajustes paulatinos de la paridad oficial del colón respecto del dólar, y se da un repunte en cuanto al empleo.

3 La opinión pública y la política económica en 1988

Estos grandes lineamientos de política económica arriba expuestos, generan debate, conflictos, los que son retomados por los medios de comunicación colectiva. A partir de éstos se seleccionaron 34 temas de conflicto para trabajar y aplicar el análisis en componentes principales.

3.1 La opinión pública nacional

Para un mejor análisis se harán bloques temáticos de los conflictos asociados a la política económica.

En lo referente a ...

empleo, el costarricense está a favor de la creación y estabilidad del mismo. La actitud ante el uso de tecnología que expulse mano de obra está repartida.

ajuste estructural, la población mantiene una posición de defensa del tipo de Estado que se ha conocido en los últimos años; sin embargo, las opiniones se reparten más en lo relacionado con la banca privada. Se cree en la autosuficiencia alimentaria pero se considera que no existen buenas condiciones de crédito para producir.

En otros temas, son pocos los costarricenses que ven en el Fondo Monetario Internacional (FMI) una ayuda para resolver la crisis. La población tiene una posición de avanzada al sustentar que el ama de casa debe percibir salario, de hecho se reconoce al oficio doméstico como un trabajo.

3.2 La opinión pública regional

Región Central.

Presenta una serie de particularidades que la hacen ocupar un lugar privilegiado con respecto a las otras regiones por razones históricas.

Es la única región con crecimiento económico autosostenido. Ocupa el 15% de la superficie total del país y contaba en 1973 con el 63% de la población total y en 1984 el 64% lo que denota un crecimiento intercensal y una tendencia a la concentración poblacional. Contaba en 1984 con el 66.99% de la población económicamente activa (PEA) nacional.

Hacia ella se dirigen y se localizan los flujos migratorios más importantes de capitales y hombres, lo que implica que esta región absorbe los excedentes económicos de las otras regiones del país y es punto de atracción estacional de la fuerza de trabajo.

A su interior se diferencia la aglomeración metropolitana de San José, que focaliza el proceso de concentración económica y de decisión política. Es una

subregión que genera preocupación en cuanto a los cambios del uso del suelo, a la forma de apropiación de la renta urbana, a la generación de la inversión privada, a la ampliación de su superpoblación relativa, y sus presiones por empleo y vivienda. Se da un debilitamiento de los gobiernos locales como consecuencia del proceso de metropolización.

Respecto del empleo, la población de dicha región está a favor de la creación de empleo en el sector público, de la capacitación al desempleado por parte del gobierno, y de obligar a las empresas a emplear minusválidos.

Se da una fuerte oposición a la privatización de empresas estatales así como a la liberalización de los alquileres. Hay una percepción de que la coexistencia de la banca nacionalizada y la privada debe darse.

Es la región que tiene menor frecuencia a cuanto a la creación de empleo en el sector público. En los otros temas, la posición de la población está muy cercana a la media nacional.

Pacífico Sur.

Es la región que más cambios ha sufrido después del último censo, debido al abandono de la bananera, lo que incidió en el empleo modificando las categorías ocupacionales y pudiendo afectar los saldos migratorios.

En 1984 contaba con el 9.3% de la población nacional, con una tasa de crecimiento de 2.7% y con un saldo migratorio de -5.2%. El 46.2% de su población es activa y se concentra en la agricultura. La industria absorbe poca fuerza de trabajo (3.4%). El desempleo abierto alcanzó un 7.2%.

Se da una sobreutilización de las tierras para la ganadería (136%) y una subutilización de las tierras para la agricultura (solamente 36%).

Esta región es la que más a favor está en frenar la tecnología que expulse mano de obra. Creen que el país debe producir lo que come y que el Estado debe subvencionar las pérdidas de las cosechas.

Pacífico Central.

Cuenta con el 6.19% de la población nacional y es la segunda región en importancia como expulsadora de población a nivel nacional.

La PEA regional es de 44.8%. El índice de desocupación es de 9.7% y la actividad económica se concentra en el Gran Puntarenas y Esparza, donde la ganadería aporta el 38% del empleo generado por las actividades agropecuarias.

Supera por amplio margen el promedio nacional en los temas referentes al empleo, sin embargo, se encuentra entre las poblaciones con posición más favorable a la privatización de las instituciones estatales y la que más se opone a la disminución de las libertades sindicales.

Pacífico Norte.

Se encuentra en una categoría intermedia en lo que concierne a contexto urbano. La tendencia a la expulsión de población se ha acentuado (tasa migratoria de -9.59 para el período 1968-1973, y de -13.52 para el de 1973-1984).

La PEA alcanza el 44.10% de su población, localizada en las ramas de la agricultura, silvicultura y caza (53.05%). La manufactura apenas ocupa al 4.92%, mostrando uno de los índices más bajos de desarrollo industrial. La desocupación es del orden del 10.1% y el analfabetismo es del 10.9%.

El desarrollo de la ganadería ha provocado un proceso de reconcentración de la pequeña y mediana propiedad. El área potencial para la reactividad agrícola alcanza el 11.5% del total, los granos básicos ocupan el 80% de esa superficie.

Su población se opone a la privatización de las empresas estatales y está a favor de la creación de empleo en el sector público. Es la que más cree que el cooperativismo es una salida a la crisis y que se debe defender a la banca nacionalizada.

Norte.

Es la región con más baja densidad demográfica del país (14.74 habitantes por km²) con el 5.74% de la población total. En 1984 la PEA era el 49% de la población regional mayor de 12 años.

Al igual que otras regiones las actividades del sector primario absorben la mayoría de la PEA (67.7%). La actividad industrial está ligada a los productos agropecuarios y forestales concentrados en Ciudad Quesada.

Esta región se diferencia en mucho de las restantes: a pesar de tener solo un 21.4% a favor de la privatización del Seguro Social, es la que más a favor está. Del mismo modo para todo lo que a privatizaciones se refiere. Paralelamente, es la región que más a favor está del no pago de la deuda externa.

En lo referente al empleo, también es la que más a favor está por la creación de empleo en el sector público y la que menos está a favor del control de la tecnología que expulsa mano de obra.

Es interesante observar que en cuanto a la intervención militar de los EEUU en Centroamérica, el promedio nacional es de 16.1% a favor y las provincias no son determinantes, pero en el análisis regional, la región Norte sobrepasa casi en el doble la media nacional (31.7% a favor).

En términos generales, se observa que la región Norte es la que más se aparta del resto de las regiones, asumiendo una actitud más cercana a las políticas neoliberales.

Atlántica.

En 1984 representaba al 6.95% de la población nacional con un 69.78% de población rural y una PEA del orden del 47.3% de su fuerza de trabajo. La

agricultura y los servicios son las dos ramas que más empleo generaron. En este mismo año alcanzó un índice de desocupación de 8.83%

La región Atlántica se caracteriza por un alto grado de concentración de la tierra como del número de grupos campesinos identificados por el Instituto de Desarrollo Agrario, IDA, (37.8% del total, el más alto del país). El 8.8% del área nacional dedicada a los granos básicos se encuentra en dicha región y tiene el 40% de los productores de plátano.

Es su población la que más fuertemente opina a favor de la capacitación al desempleado por parte del gobierno.

Se opone a las medidas de privatización de las instituciones estatales en forma general, siendo la más radical de todas las regiones, así como en lo que a fomentar la creación de empresas.

4 Análisis en componentes principales (ACP)

Se parte del esquema de la Figura 1 para construir un conjunto de indicadores universales característicos de la opinión particular.

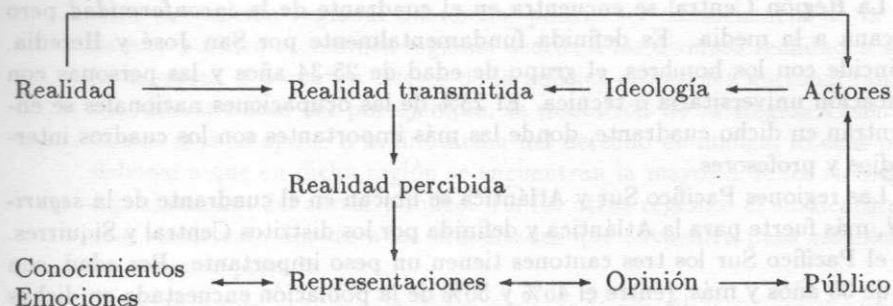


Figura 1: Esquema constructivista del fenómeno de la opinión.

El ACP permite:

- obtener dimensiones significativas que determinen la estructura de la opinión pública,
- analizar la parte común a todos los conflictos que es la parte simbólica, el *cielo* o diccionario de conflictos que se debe analizar junto con la *tierra*, conjunto de variables de señalización,

lo que posibilita construir las estructuras.

Los ejes principales son, por definición y por construcción, los ejes de la diferenciación máxima. Casi la totalidad de las variables tienen correlación po-

sitiva con el primer factor (*consenso-disenso*) y el segundo factor (*movimiento-estabilidad*). Las regiones de la tierra del primer plano principal del ACP se muestran en la Figura 2.

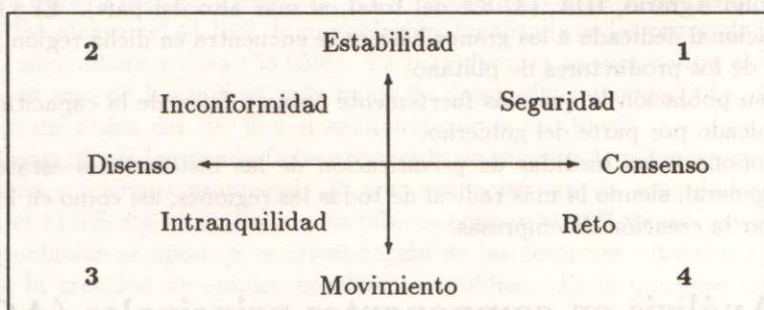


Figura 2: Las regiones de la tierra y los ejes de la diferenciación: primer plano principal del ACP.

La Región Central se encuentra en el cuadrante de la *inconformidad* pero cercana a la media. Es definida fundamentalmente por San José y Heredia. Coincide con los hombres, el grupo de edad de 25-34 años y las personas con educación universitaria o técnica. El 23% de las ocupaciones nacionales se encuentran en dicho cuadrante, donde las más importantes son los cuadros intermedios y profesores.

Las regiones Pacífico Sur y Atlántica se ubican en el cuadrante de la *seguridad*, más fuerte para la Atlántica y definida por los distritos Central y Siquirres. En el Pacífico Sur los tres cantones tienen un peso importante. Por edad, son los de 35 años y más, (entre el 45% y 50% de la población encuestada en dichas regiones). En este cuadrante se ubica la derecha.

El Pacífico Central se ubica en el tercer cuadrante: la *intranquilidad*, determinada por Puntarenas. Estas posiciones las representan las personas con ningún nivel educativo y de edades entre 18 y 24 años. No hay preferencia por algún partido ni candidato político.

El Pacífico Norte y la región Norte se localizan en el cuarto cuadrante: *reto* representa a las mujeres, a las personas con nivel educativo de primaria y secundaria así como a las profesiones intermedias, amas de casa y peones agrícolas. Son los separados, divorciados y en unión libre.

5 Conclusiones

A partir del presente trabajo se pueden extraer las siguientes conclusiones:

1. La política económica impulsada en los últimos años encuentra resistencia —aunque ésta no sea organizada— en las diferentes regiones. La concepción interventora del Estado sigue estando arraigada en las poblaciones regionales. Solo en el tema relacionado con la banca privada se nota una tendencia a su aceptación, pero en coexistencia con la banca nacionalizada.
2. La política económica es percibida de manera diferencial por la población de las regiones. Dicho de otro modo, a pesar de que la muestra se basa en los centros de concentración poblacional, las diferencias regionales se conservan, en lo que a opinión pública se refiere. Esta diferenciación se manifiesta especialmente en los temas relacionados con las privatizaciones, (*CCSS, empresas estatales, medicinas del Seguro Social*), con el empleo y salarios, (*capacitación al desocupado, emplear a minusválidos, amas de casa deben recibir salarios, disposición a baja temporal de los salarios*), medidas para salir de la crisis, (*cooperativismo, limitar derecho de huelga, sacrificarse por la Patria, disminución del poder adquisitivo*), así como en los temas referentes a la producción ligados a la concepción proteccionista del Estado (*producir lo que el país se come, subvencionar pérdidas de las cosechas, el precarismo es un mal social*).
3. Esta percepción regional diferencial podría ser consecuencia de la falta de una política económica regional acorde a las diversas realidades socio-económicas, y a que los efectos de la política económica nacional también son diferenciales; así por ejemplo, la población de la Región Central es la que más se opone a la limitación del derecho de huelga, lo cual puede deberse a que en dicha región se encuentran la mayoría de los sindicatos, especialmente del sector público. En las otras regiones el sindicalismo es casi inexistente debido a las dificultades que encuentra para establecerse en el sector privado.

El Pacífico Sur, al ser una de las regiones que sufre fuertemente el problema del desempleo, se opone a la introducción de tecnología que afecte el nivel del empleo. Al igual que otras regiones periféricas, la falta de servicios y las bajas condiciones de vida, hacen que su población se oponga a la privatización de servicios, tales como el de salud, lo que podría significarles el verse excluidos de los mismos, por las dificultades para obtener mayores ingresos y así pagar dichos servicios. Podría considerarse que su posición favorable a la seguridad alimentaria se relaciona con la composición de su PEA, la que se concentra en la agricultura (64,3%).

La Región Atlántica sobrepasa la tasa de desempleo nacional por lo que su posición ante la capacitación para el desempleo responde a la problemática regional, así como la necesidad de fomentar la creación de empresas de manera desconcentrada, contrariamente a lo actuado hasta el momento y casi ausente en la política económica teórica. Su posición para salvaguardar el modelo vigente del sistema de salud es la más radical del país,

y su explicación posible sería similar a la antes citada para la región del Pacífico Sur. Se aprecia que ambas regiones (Pacífico Sur y Atlántica) se acercan entre sí. Los problemas que viven, y la falta de una política económica específica para enfrentarlos, hacen de ellas regiones olvidadas por los diferentes gobiernos en esta materia, por lo que tienden a adoptar posiciones que les garanticen un mínimo de seguridad con una actitud de sumisión.

4. La estructura de la opinión pública regional, es decir, aquello que hace a la población pronunciarse de cierta forma ante los conflictos (las variables de señalización), es igualmente diferencial, por lo que la política económica debería tomar en cuenta a qué sectores va dirigida. Dicha estructura regional es comparable con la estructura nacional.

5. A pesar de que la teoría dice que la estructura es independiente de las personas y de las preguntas, para refinar aún más la estructura de opinión pública regional, sería necesario realizar encuestas específicas en cada región, aún si la rotación de los ejes va a ser mínima.

6. El método matemático-estadístico brinda, en un reducido volumen, un máximo de informaciones estadísticas fáciles de aprehender con un vistazo, brindando informaciones de base que permiten juzgar las posiciones de los habitantes de las regiones de Costa Rica, en junio de 1988, con respecto a los grandes temas de la política económica.

7. Los documentos estadísticos, como un todo, permiten una interpretación que se puede realizar desde una doble perspectiva:

- Una perspectiva microscópica: el énfasis puesto en los mecanismos que permiten el ajuste entre el individuo y su medio.
- Una perspectiva macroscópica: la atención centrada sobretudo en el actor y las relaciones que él sostiene con las opiniones, lo cual puede jugar un papel esencial en las relaciones de fuerza.

Ambas perspectivas muy importantes tanto para la formulación de la política económica por los actores políticos, como para aquellos actores sociales que quieran contrarrestarla o apoyarla.

Referencias

- [1] Aglietta, M.: *Regulación y Crisis del Capitalismo*. Siglo XXI, Madrid, 1976.
- [2] Arrighi, Giovanni: *La crisis en relación con la fuerza estructural de la clase obrera*. Ed. Pablo Iglesias, Madrid, 1980.

- [3] Artiles, Antonio M. y Jodar, M. Pere,: *Breve ensayo sobre Crisis Económica y Relaciones Industriales*. Ed. Grupo Cultural Zero, Madrid, 1984.
- [4] Brenot, Jean y Pagés, Jean-Pierre: *Les Structures de l'Opinion Publique*. Centre d'Etudes Nucléaires de Fontanay-aux-Roses, France, 1986.
- [5] CEPAL: *Centroamérica: Crisis y Políticas de Ajuste, 1979-1986*. CEPAL-ONU, México, 1988.
- [6] Céspedes, Víctor et al.: *Costa Rica: Una economía en Crisis*. Editorial Studium, San José, 1983.
- [7] Chouchan, Dominique: *L'Opinion Publique Radiographiée*. En *Sciences et Techniques* No. 8, octubre 1984, París, 1984.
- [8] Del Cid, José Rafael: *Honduras: Industrialización, empleo y explotación de la fuerza de trabajo*. En *El fracaso social de la integración centroamericana: capital, tecnología, empleo*, Colección Seis, EDUCA, San José, 1979.
- [9] Fonseca P., Edmur: *El concepto de espacio y las regiones: algunas consideraciones de orden metodológico*. MOPT e Instituto Geográfico Nacional, San José, 1975.
- [10] Gallardo, María E.: *La Iniciativa para la Cuenca del Caribe: su impacto en Centroamérica*. FLACSO, San José, 1988.
- [11] Garnier, Leonardo: *Crisis, desarrollo y democracia en Costa Rica: Crisis y desafíos*. DEI, San José, 1987.
- [12] Humbert, Marc: *L'Industrialisation sous contraintes*. En *Révue Tiers Monde*, Vol. XXII, No. 87, juillet-septembre 1981, París, 1981.
- [13] Lanzarotti, Mario: *Quelle industrialisation?*. En *Révue Tiers Monde*, Vol. XXI, No. 81, París, 1980.
- [14] Jodar, Pere y Artiles, Antonio M.: *Crisis, Clases y Estado: Elementos de debate sindical*. En *Crisis Económica y relaciones industriales*, Ed. Grupo Cultural ZERO, Madrid, 1984.
- [15] Lizano, Eduardo: *Política económica: Reflexiones sobre el caso reciente de Costa Rica*. En *Antología de Maestría en Política Económica*, agosto de 1989, San José, 1989.
- [16] López Subirós, Marta: *Le Planning Familial a Costa Rica: Ideologie et Réalité*. Tesis de Maestría en Sociología, Université Toulouse le Mirail, Francia, 1978.

- [17] MIDEPLAN: *Informes anuales de coyuntura de los años de 1980 a 1986*. MIDEPLAN, San José, 1980-1986.
- [18] Morales Alvarez, Miguel: *Redistribución espacial de Población: Objetivos, Limitaciones e Impacto de la Acción Estatal en Centroamérica: el caso de Costa Rica*. CSUCA-UNA-PISPAL-IDRC, San José, 1983.
- [19] Nuhn, Helmut: *Regionalización para la Planificación*. En Seminario "Planificación para el Desarrollo Regional", CEDAL, San José, 1983.
- [20] Ominami, Carlos: *L'ajustement contre l'industrie: Etude des tendances récentes de l'industrie latinoamericaine*. En *Révue Tiers Monde*, Vol. XXVII, No. 107, juillet-septembre 1986, París, 1986.
- [21] Palloix, Christian: *Proceso de producción y crisis del Capitalismo*. H. Blume Ediciones, Madrid, 1980.
- [22] Poltronieri, J., Piza, E., et al.: *Estructuras de la Opinión Pública en Costa Rica*. Editorial Universidad de Costa Rica, San José, 1989.
- [23] Rivera, Eugenio: *Centroamérica: política económica y crisis*. DEI-ICADIS-UNA, San José, 1986.
- [24] Rojas Bolaños, Manuel: *Costa Rica: el final de una era*. En *La Crisis Centroamericana*, EDUCA, San José, 1984.
- [25] Rosanvallon, Pierre: *Crise du Capitalisme et stratégies Syndicales en Europe*. CFDT Aujourd'hui, No. 27, París, 1977.
- [26] Rovira Mas, Jorge: *Costa Rica en los años ochenta*. Editorial Porvenir, San José, 1987.
- [27] Rovira Mas, Jorge y Trejos S., Luis Diego: *El curso de la crisis de Costa Rica y las opciones de política económica en el segundo lustro de los años ochenta*. Avances de Investigación No. 53, IIS-UCR, San José, 1985.
- [28] Solís, Manuel y Esquivel, Francisco: *Las Perspectivas del Reformismo en Costa Rica*. DEI, San José, 1979.
- [29] Solís, Otón: *El ajuste estructural*. Mesa redonda de la Cátedra Introducción a la Sociología, UCR, San José, 1989.
- [30] Zúñiga, Guillermo: *Estabilización y Ajuste Estructural en Costa Rica durante los períodos de Carazo y Monge (1978-1986)*. En *El debate sobre la Reforma Económica*, CIERA, Managua, 1988.

Clasificación Automática: Grupos de Opinión Pública en Costa Rica

William Castillo*

1 Introducción

En el presente trabajo se propone hacer un análisis de clasificación automática de la encuesta de opinión pública de 1988, realizada por el *Grupo de Análisis de Datos* de la Escuela de Matemáticas de la Universidad de Costa Rica, con la colaboración de un equipo de investigadores de la Escuela de Sociología. Los resultados de dicha encuesta se incluyen en [2].

Específicamente el estudio pretende determinar grupos homogéneos de entrevistados que se forman obedeciendo a ciertas tendencias de opinión, presumiblemente existentes en el seno de nuestra sociedad. Dichas tendencias se configuran por las respuestas dadas por los entrevistados sobre 56 "*temas de conflicto*" sometidos a su consideración. Así, los grupos deberán ser lo más homogéneos posible, de acuerdo con sus respuestas. En la sección 2 nos referiremos explícitamente a este punto.

El método de análisis de encuestas de opinión por medio de la clasificación automática, que aquí hemos empleado, ha sido utilizado con éxito por la AGORAMETRIE en sus encuestas [1]. Básicamente se desarrolla en dos fases. La primera consiste en usar el *método de los centros móviles* y el *método de Ward* a fin de determinar una partición óptima de la muestra. La segunda se refiere al proceso de caracterización de los grupos de individuos en que ha sido separada la muestra, todo ello previa elaboración de una serie de documentos sobre los que se basa dicha caracterización.

El estudio se organiza por secciones. La sección 2 contiene una descripción del procedimiento seguido para determinar una partición de la muestra. La sección 3 contiene un breve comentario sobre las variables que intervienen con mayor importancia en el estudio de clasificación, mientras que en la sección 4

*Profesor de la Escuela de Matemáticas de la Universidad de Costa Rica.

se incluye un análisis interpretativo de la partición encontrada. Asimismo la sección 4 contiene algunas conclusiones y comentarios finales.

2 Particionamiento de la muestra

Nuestro estudio se fundamenta, originalmente, en las respuestas dadas por los entrevistados sobre 56 *temas de conflicto* que se les proponen para responder en una escala de 1 a 5, del modo siguiente:

1. En total desacuerdo
2. No tan de acuerdo
3. Podría estar de acuerdo
4. Bastante de acuerdo
5. Totalmente de acuerdo

Estas respuestas forman una gran tabla de datos rectangular de tamaño 1477×56 que llamaremos *tabla de respuestas*.

La otra información proporcionada por la encuesta —y que es relevante para efectos del estudio clasificatorio— son las llamadas *variables de señalización*, a las cuales nos referiremos en la sección 3.

El proceso de búsqueda de una partición de la muestra compuesta de 1477 entrevistados, es como sigue:

- (a) A fin de trabajar con una tabla no tan grande como la tabla de respuestas, se recurre al Análisis en Componentes Principales de la tabla de respuestas. De ese modo se construyen un conjunto de 7 nuevas “variables”, llamadas *componentes principales*, que sustituyen las 56 variables de conflicto, de suerte que se conserve el máximo de la “información”. Es decir, las tendencias de opinión más fuertes se mantienen en las nuevas variables. De este modo obtenemos una tabla más pequeña, de tamaño 1477×7 , que llamaremos *tabla reducida*.
- (b) Sobre la tabla reducida se aplica el método de los centros móviles, inicializado con una partición de la muestra en 9 clases, escogida al azar. Este método calcula los centros de gravedad de la partición inicial; luego asigna cada entrevistado al centro de gravedad más próximo, creando de esta manera una nueva partición. Los centros de gravedad son recalculados, respecto de la nueva partición, y se reasignan los entrevistados según su cercanía a estos nuevos centros, creando una nueva partición. El proceso sigue hasta que la inercia interclases no cambie entre dos particiones sucesivas. Ello significa que con la inicialización escogida no es posible encontrar otra partición con mayor inercia interclases. Es decir,

el método encuentra un máximo relativo del criterio de calidad (la inercia interclases).

Aplicando 2 veces este método se obtienen dos particiones P_1 y P_2 en forma de variables cualitativas, con 9 modalidades. Para ubicar los grupos de individuos que fueron clasificados juntos respecto de P_1 y P_2 (son los grupos más estables), se intersecan ambas particiones, dando origen a una tercera partición P_3 con 55 clases.

- (c) Usando la tabla reducida se calculan los centros de gravedad de las clases de P_3 , a partir de los cuales se establece una jerarquía de particiones por el método de Ward y el algoritmo de los vecinos recíprocos. Por este método se construye una secuencia de particiones empezando con P_3 , de modo que la primera partición se obtiene de P_3 por fusión de las dos clases de ésta, que produzcan el máximo incremento de la inercia interclases en la nueva partición. De igual manera se define la segunda partición respecto de la primera, y así sucesivamente hasta tener la partición más "gruesa" cuya única clase es el conjunto de los 1477 entrevistados.

La representación arborescente parcial de dicha jerarquía se presenta en la Figura 1. Cortando verticalmente este árbol, como se indica, se obtiene una partición final de la muestra en 10 clases.

En lo que sigue nos abocamos al problema de interpretación de esta partición, teniendo en cuenta que la presente es solo una entre muchas posibilidades de particionar la muestra.

3 Sobre las variables utilizadas

La posibilidad de establecer una caracterización de las opiniones —cuya primera expresión es la partición encontrada— se fundamenta en la idea de "tipificar" las clases de acuerdo con la opinión de sus miembros frente a los conflictos de la sociedad, y de acuerdo con sus rasgos socio-culturales dominantes. Las variables que sirven a este objetivo son respectivamente, los 56 temas de conflicto y las 8 variables socioculturales (también llamadas variables de señalización).

Las 8 variables socioculturales son las siguientes:

- **Región** (por provincia)
- **Edad** (cinco grupos de edad)
- **Sexo**
- **Condición Socio-profesional** (10 categorías)
- **Escolaridad** (seis niveles)

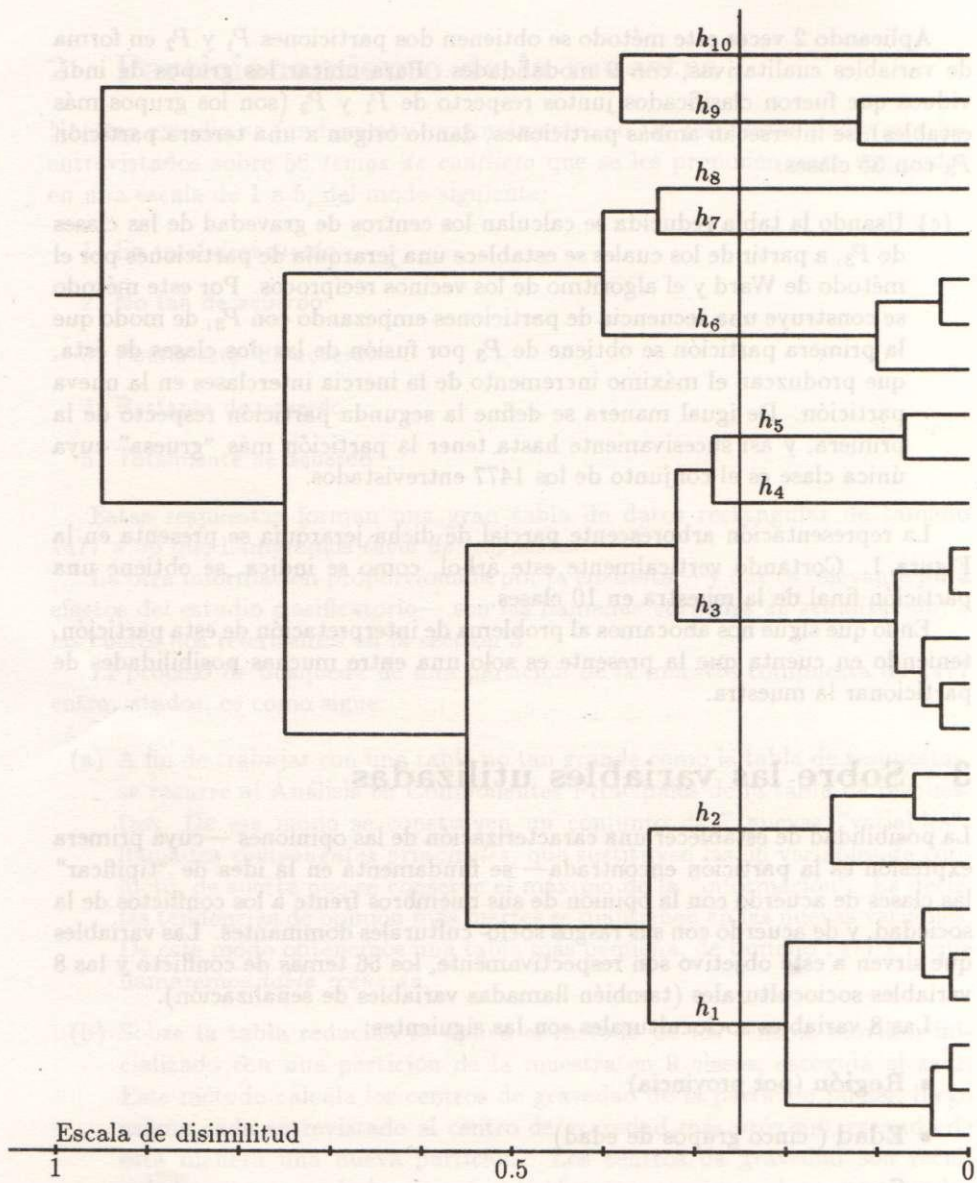


Figura 1: Representación parcial de la jerarquía obtenida. Se obtienen los 10 grupos h_1, h_2, \dots, h_{10} al cortar el árbol verticalmente, como se ilustra en la figura.

- Religión (seis modalidades)
- Preferencia Sindical (12 modalidades)
- Tendencia Política (5 modalidades)

A partir de los *perfiles columna* (ver anexo), se nota que la constitución de casi todas las clases está determinada, para cada variable, por una o dos modalidades. En ese sentido, observamos que el Sindicato mayoritariamente preferido es APSE, excepto en la clase 6. Las amas de casa y los empleados son mayoritarios en todas las clases, salvo la 5 y la 10. El nivel de escolaridad que domina es educación media y primaria, con algunos matices entre clases. La fe católica es una constante, con leves diferencias en orden al grado con que se practica, y la posición política se reparte de modo casi similar entre las clases, entre centro, derecha, e indiferentes.

En tales circunstancias, las variables de señalización constituyen un elemento de diferenciación de los diversos grupos, severamente limitado. Es decir, el componente socio-cultural es muy similar en casi todas las clases en que fue particionada la muestra.

Por otra parte, para localizar las opiniones que han determinado la formación de los grupos de individuos, se recurre al estudio comparativo de las medias de los temas de conflicto a nivel de la muestra y de las clases. Para ello se ordenan los temas de conflicto en cada clase, de acuerdo con el valor decreciente del estadístico T (sin tomar en cuenta su signo). Se nota que las diferencias entre las medias no son por lo general tan marcadas. Por otro lado, la diferenciación de una clase se ve limitada por el hecho de que frecuentemente sus miembros tienen opiniones singulares solamente respecto de unos pocos temas, los cuales algunas veces se repiten de clase en clase. Se dispone así de escasos elementos para la interpretación (ver anexo y sección 4).

Esta situación, si bien no impide, al menos dificulta el objetivo de dar algún sentido a las agrupaciones de individuos encontradas. Pese a lo anterior es posible identificar —aunque no en todos los casos—, ciertos rasgos dominantes relativos a la opinión y a aspectos socio-culturales de las clases, que nos permiten discernir sobre algunas características de las clases de la partición encontrada, y con ello realizar un primer acercamiento al espectro de las opiniones en el seno de la sociedad costarricense.

4 Examen interpretativo de las clases de la partición encontrada

Los comentarios que siguen se fundamentan en la información cuantitativa que se presenta en el anexo. Se tiene en cuenta además que entre más grande sea la media de una variable en la clase, más factible es que ésta tenga una opinión favorable a la proposición planteada. El estadístico T es una apreciación de

la desviación relativa de la media de la clase respecto de la media muestral. Entre más grande (sin importar el signo) sea el valor de T , mejor caracteriza la variable a la clase.

Considerando que es materia más compleja formular conclusiones cualitativas acerca de las diferentes clases, hemos omitido toda referencia en esa dirección. Sin embargo ulteriores análisis de esta u otra encuesta podrían conducir a conclusiones de ese tipo.

Clase 1 (efectivo: 275)

El 53.8% de sus miembros tienen edades entre 25 y 44 años, mientras el 17.8% son personas con más de 55 años. De manera que los miembros de esta clase —como los de la siguiente—, son de edades intermedias: ni tan viejos ni tan jóvenes.

El nivel de escolaridad es relativamente bajo. El 38.2% tienen estudios primarios y el 28.4% estudios secundarios (hay una situación similar en la clase 9). En cuanto al sexo, son mayoritarias las mujeres (63%) y se presume que serán especialmente amas de casa en razón de que el 23% del total pertenecen a esta clase, conformando el 40.7% de sus miembros.

En cuanto a sus opiniones, se nota un sentimiento de inseguridad más fuerte que en las otras clases. Hay una leve tendencia a creer que la policía no cumple bien su misión. Se inclinan por aceptar a los sindicatos y son contrarios a las privatizaciones. Opinan a favor del fortalecimiento del sector estatal (crear canales de televisión y defender la banca nacionalizada).

En torno a ciertas cuestiones sobre las cuales se produce globalmente un gran consenso —en el sentido del acuerdo—, esta clase opina de una manera más definida y en el mismo sentido. Tal es el caso de temas como: la necesidad de los controles de identidad, el país debe producir los alimentos básicos de consumo interno y el gobierno debe capacitar al desempleado.

Se aprecia una actitud que podría evolucionar hacia el rechazo de algunos sectores de extranjeros. Ello se nota en relación con la extradición y con la idea de que los refugiados deben irse.

Clase 2 (efectivo: 193)

Esta clase tiene escolaridad ligeramente superior a la anterior en el sentido de que sus miembros se distribuyen por igual entre los que tienen educación primaria y educación secundaria (una situación similar prevalece en las clases 3 y 8). La edad es similar a la clase 1 con un poco más de acumulación de gente en los intervalos señalados (56.5% entre 25 y 44, y 19.2% con más de 55 años). Respecto a otros aspectos socio-culturales no se nota nada que diferencie esta clase.

Lo más sobresaliente de sus opiniones se refleja en la actitud crítica en relación con el hecho de identificar la democracia con el sufragio. Asimismo

existe cierta tendencia a descalificar (o desconfiar) de ciertas instituciones como la Asamblea Legislativa, la Policía y la Justicia. No son tan definidos en favor de la banca nacionalizada y creen que no hay buenas condiciones de crédito para quien trabaja. El sentimiento de inseguridad es más débil que en la clase 1.

Aún cuando es un hecho que globalmente hay consenso contra la pornografía, contra el tabaco y por acortar la campaña electoral, en esta clase estos rasgos se expresan con más fuerza en razón de los altos promedios alcanzados y porque las desviaciones son pequeñas.

Clase 3 (efectivo: 154)

El 59% de sus miembros tienen edades que varían de los 18 a los 34 años, lo cual evidencia una acumulación de individuos de menor edad que en las clases 1 y 2. Una situación similar ocurre en las clases 5, 8, 9 y 10. Son las clases más "jóvenes".

El nivel de educación es como en la clase anterior. Por otra parte, los "empleados" son el 39% de los miembros de esta clase (algo similar ocurre en las clases 7 y 8).

Por otra parte, es una clase con fuerte predominio del sexo masculino: el 61.7% de sus miembros son hombres.

Es, desde el punto de vista de la religión, la clase más heterogénea. El 15.6% de sus miembros son de "otra religión" (el mayor porcentaje por clase). Al mismo tiempo se registra el menor porcentaje de católicos practicantes regulares (37.7%), excepto por la clase 10.

En cuanto a sus opiniones, notamos que esta clase manifiesta una actitud un tanto contraria a la posibilidad de que las amas de casa reciban un salario. En cuanto a la posibilidad de que las mujeres accedan a puestos elegibles en igualdad de condiciones con los hombres, no tienen una opinión favorable, como sí la tienen en promedio el total de los entrevistados.

Respecto a las siguientes proposiciones con las cuales los entrevistados se manifestaron en promedio de acuerdo, la opinión de esta clase es menos favorable: "el tabaco es una plaga", "hay que propiciar la libre elección médica" y, "hay que acortar la campaña electoral".

Por otra parte tienen una opinión contraria a la privatización en los hospitales.

Clase 4 (efectivo: 124)

El 71% de sus miembros se distribuyen casi uniformemente en los intervalos de 25 a 34 años, de 35 a 44 años y de 55 años y más.

En cuanto al nivel de estudios, el 87.9% de sus miembros se reparten casi por igual entre los que tienen educación primaria, secundaria y universitaria.

Es por tanto una clase muy heterogénea respecto a la edad y al nivel de estudios.

Es la clase con mayor porcentaje de pensionados (14.5%), sin llegar a quebrar la tendencia de predominio de las amas de casa y los empleados.

Veamos lo que ocurre con respecto a las opiniones. Esta clase se opone a la clase 2 en el sentido de que valora positivamente la función de la Asamblea Legislativa y en menor grado, de la Policía y la Justicia.

Esta clase se diferencia de la clase 7 en varios aspectos. Se pronuncia en desacuerdo con las privatizaciones del Seguro Social, del ICE y en menor grado con las privatizaciones en los hospitales. Por otra parte son contrarios al establecimiento de ciertos beneficios para los asalariados, como la semana laboral de 35 horas y el empleo de medio tiempo.

Estiman que no se debe aumentar las cuotas patronales y que la cotización del seguro social debe ser obligatoria. De modo menos decidido son favorables a la incorporación de tecnología que pueda desplazar mano de obra.

Clase 5 (efectivo: 99)

El 55.5% de sus miembros tienen entre 18 y 34 años de edad y políticamente el 49.5% declararon ser de centro.

En cuanto al nivel de estudios, el 74.5% de sus miembros han cursado por lo menos los estudios secundarios, de ellos el 52.5% tienen estudios universitarios. De modo que se trata de una clase "joven", políticamente de centro y con alta escolaridad.

La categoría socio-profesional "patrón de empresa o comercio" representa el 20.2% de los miembros de la clase.

Son contrarios a medidas proteccionistas como la subvención por pérdidas de cosechas y que el gobierno capacite al desempleado. Un poco más débil es su oposición a que se obligue a las empresas a dar trabajo a los minusválidos.

Esta clase tiene una opinión más claramente definida que la anterior, en contra de más oportunidades de empleo y de ciertas ventajas para los asalariados. En ese sentido hay una actitud opuesta a la creación de empleo estatal y de empleo de medio tiempo, a la reducción de la edad de pensionarse y a la semana laboral de 35 horas.

Son más favorables a la incorporación de tecnología que la clase anterior y contrarios a aumentar las cuotas patronales.

Como la clase 3, son contrarios a que las amas de casa reciban salario y no son definidos a favor de la participación de las mujeres en puestos elegibles, en igualdad de condiciones que los hombres.

Clase 6 (efectivo: 116)

Respecto a la edad y al nivel de educación, es —como la clase 4— muy heterogénea. No hay otras características socio-culturales dignas de mención.

Cabe la hipótesis de que debido a esta heterogeneidad, surge una opinión en alguna medida controversial, y en apariencia desarmónica, respecto de ciertos

temas. Veamos:

Son más abiertos a confiar en las reformas en la URSS y a aceptar las relaciones comerciales con este país. Son susceptibles a aceptar la privatización de las instituciones estatales de salud, el ICE, etc. Hay una tendencia a creer que el FMI contribuye a resolver la crisis y a favor de fortalecer la banca privada. Tienen alguna confianza en la justicia.

Aunque tenuemente, surge en esta clase una actitud menos prejuiciada que la totalidad de los entrevistados, en relación con temas como la integración social de los homosexuales, los desnudos en la televisión y que los hombres se maquillen. Asimismo, no se definen abiertamente contrarios a una intervención militar de Estados Unidos en Centro América, como sí ocurre a nivel de la muestra.

Clase 7 (efectivo: 135)

La residencia de sus miembros se distribuye uniformemente entre San José, Alajuela y Puntarenas.

La edad en esta clase es muy heterogénea pero es un poco más joven que las clases 4 y 6. En efecto, el 26% de sus miembros tienen edades entre 18 y 24 años, el 20.7% tienen edades entre 25 y 34 años. Mientras tanto el 24.4% son mayores de 55 años.

El 40.7% de los miembros de esta clase se ubican en la categoría socio-profesional "empleados", seguido de cerca por las amas de casa con el 37%. Además domina el sexo femenino, el cual representa el 66% del total de la clase. Sus miembros poseen baja escolaridad: el 85.2% tienen hasta estudios secundarios, de los cuales el 56.3% tienen a lo sumo enseñanza primaria.

El 40% de los miembros de esta clase se declaran políticamente como indiferentes, el resto se distribuyen por partes iguales entre la derecha y el centro.

Esta clase tiene una opinión favorable a medidas que consideran beneficiosas para los asalariados. Así, ellos están por la semana laboral de 35 horas, por reducir la edad de pensionarse, por la no aceptación de la tecnología que expulse mano de obra.

Están a favor de más oportunidades de empleo en el sector público y por el empleo de medio tiempo. Son propensos a aceptar un incremento en las cuotas patronales.

También tienen una opinión favorable en relación con otras medidas ventajosas para ciertos sectores sociales, tales como que las cotizaciones del Seguro Social no deben ser obligatorias y que se debe subvencionar las pérdidas por cosechas.

Valoran positivamente la educación primaria al estimar conveniente que todo ciudadano participe de ella obligatoriamente, hasta los 20 años de edad.

Por otro lado, mientras globalmente hay una opinión contraria a la privatización de instituciones como el ICE o el Seguro Social, en esta clase la

respuesta promedio indica que se podría eventualmente generar una actitud favorable a dichas privatizaciones. Su opinión es un poco más definida —pero no claramente— a favor de fortalecer la banca privada.

Clase 8 (efectivo: 156)

Es una clase “joven” y de escolaridad “media”. De sus miembros, el 59.6% tienen entre 18 y 34 años de edad, mientras que el 72.5% se distribuyen por igual entre los que tienen estudios secundarios y primarios. Los “empleados” habitan en esta clase en mayor porcentaje (38.5%) seguidos por las amas de casa (33.3%). Políticamente se declaran de derecha (34.6%) e indiferentes (36.5%).

La opinión en esta clase sobre varios temas, es como en la clase 7, pero surge más débilmente que en ésta. Es el caso de las oportunidades de empleo, de las medidas a favor de los asalariados y de otros sectores sociales.

Tienen una actitud más comprensiva hacia los refugiados y hacia los extranjeros “extraditables”. Asimismo tienen un concepto positivo de la Asamblea Legislativa.

Clase 9 (efectivo: 163)

El 40.5% de sus miembros tienen solamente estudios primarios, y el 25.2% tienen estudios secundarios. Es por tanto una clase de baja escolaridad. Sus miembros son también mayoritariamente jóvenes.

El 38% de sus miembros se declaran políticamente indiferentes y el resto se distribuyen casi por igual entre el centro y la derecha.

Aparentemente, esta es la clase cuya opinión es más indefinida, en razón de que no participa del gran consenso que se produce en torno a ciertos temas, sin que ello configure una clara disensión. Es el caso de: la necesidad de los controles de identidad, la prueba de alcoholemia, el financiamiento de campañas de vacunación, fomentar la creación de empresas, producir los alimentos básicos de consumo interno, capacitación gubernamental a los desempleados, la libre elección médica, lucha contra la pornografía, disminuir el tiempo de campaña electoral y que el tabaco es dañino. En menor grado hay una situación similar respecto de la defensa de la banca estatal, el sentimiento de inseguridad, aumentar el presupuesto en seguridad, la aceptación de los sindicatos, la importancia de las cooperativas como salida a la crisis y la obligatoriedad de las empresas de dar empleo a los minusválidos.

Clase 10 (efectivo: 62)

Es una clase “joven” y con predominio del sexo masculino. De sus miembros, el 29% son empleados y el 21% son estudiantes.

El 56.5% de sus miembros tienen estudios universitarios, mientras que sólo el 27.5% tienen a lo más estudios primarios.

El 53.2% de sus miembros católicos practican la religión ocasionalmente o no la practican, mientras que sólo el 27.4% son practicantes regulares. Además el 8% son no religiosos.

La opinión en esta clase es favorable al establecimiento de vínculos comerciales con la URSS y son susceptibles de confiar en las reformas de Gorbachov. Son contrarios a la banca privada y estiman que el FMI no ayuda a resolver la crisis. Creen que no hay demasiadas huelgas y que Estados Unidos no debe intervenir militarmente en Centro América.

Se nota una tendencia favorable a la integración social de los homosexuales y por el no aislamiento de los enfermos de SIDA. De modo similar se puede decir que hay una tendencia a creer que las empresas no tienen grandes problemas por culpa de las cargas sociales, que no se debería pagar la deuda externa ni aumentar el presupuesto en seguridad. Se muestran críticos sobre la identificación de la democracia con el sufragio.

5 Conclusiones

Pese a las limitaciones apuntadas en la sección 3, con el presente estudio se muestra la importancia de la clasificación automática en el campo del análisis de encuestas de opinión. Nos ha permitido separar la muestra en 10 clases para estudiar sus características en los términos de la sección 4.

Notamos que las clases 1, 5, 7 y 10 poseen características mejor definidas, mientras que en los otros casos, los rasgos surgen con menos fuerza. Así por ejemplo la clase 8 tiene una opinión análoga a la 7, excepto porque en ésta las características se marcan más claramente. ¿Porqué entonces no fundir algunas clases para llegar a una partición con menos clases? Desde luego esa posibilidad es válida, pero al hacerlo formamos clases más grandes y posiblemente más heterogéneas, perdiendo un poco del potencial interpretativo.

De modo que trabajar con 10 clases no significa que existan 10 opiniones diferenciadas. Solamente fue una opción experimental que nos permitió exhibir ciertos grupos de entrevistados con opiniones bien marcadas respecto de algunos temas.

Por otra parte conviene hacer notar que una partición distinta se pudo obtener variando un poco el procedimiento. En efecto, por ejemplo inicializando el método de los centros móviles con particiones en menos clases. También la jerarquía pudo ser distinta si hubiésemos fusionado en una sola las clases de P_3 con muy poco efectivo (1, 2 y 3 entrevistados por clase).

En todo caso creemos que los rasgos fundamentales de la opinión salieron a flote, con lo cual fue posible presentar un primer estudio en el campo del análisis de encuestas de opinión, sirviéndonos de la clasificación automática.

Referencias

- [1] Drouet, Marion: *Les francais face aux conflits de societe: vers une typologie operationnelle des opinions* (rapport de stage), 1989.
- [2] Poltronieri, Jorge y Piza, Eduardo: *Estructuras de la Opinión Pública en Costa Rica*. Editorial de la Universidad de Costa Rica, San José, 1989.

Conclusiones

Para las conclusiones apuntadas en la sección 3, con el presente estudio se muestra la importancia de la clasificación automática en el campo del análisis de actitudes de opinión. Nos ha permitido separar la muestra en 10 clases para estudiar sus características en los términos de la sección 4.

Notamos que las clases 1, 2, 7 y 10 poseen características muy definidas, mientras que en los otros casos, las características son vagas. Así por ejemplo la clase 8 tiene una opinión análoga a la 7, excepto porque en esta clase características se marcan más claramente. Por los mismos motivos, las clases para llegar a una partición con menos clases. Desde luego es necesario validar, pero el hecho de haber formado clases más grandes y poderosas más heterogéneas, perdiendo un poco del potencial interpretativo.

De modo que trabajar con 10 clases no significa que existan 10 opiniones diferenciadas. Simplemente fue una opción experimental que nos permitió estudiar ciertos grupos de entrevistados con opiniones bien marcadas respecto de algunos temas.

Por otra parte conviene hacer notar que una partición distinta se pudo obtener variando un poco el procedimiento. En efecto, por ejemplo iniciando el estudio de los temas móviles con particiones en menos clases. También se podría haber intentado el procedimiento realizado en una sola clase de 10 con muy poco éxito (1, 2 y 3 entrevistados por clase).

En todo caso creemos que los rasgos fundamentales de la opinión se refieren a lo cual fue posible presentar un primer esbozo en el campo del análisis de actitudes de opinión, sirviéndonos de la clasificación automática.

Anexos

Se incluyen en el orden respectivo los siguientes documentos:

1. Los temas de conflicto ordenados de acuerdo con la magnitud decreciente —sin importar el signo— del estadístico T . Para calcularlo se usó la fórmula:

$$T = \frac{M_c - M}{\{(n_c - 1)S_c^2 + (n - 1)S^2\} \cdot \{(n_c + n)/n_c n(n_c + n - 2)\}}$$

donde:

- M y M_c son respectivamente la media muestral y la media en la clase (primera y tercera columna respectivamente).
 - S y S_c son respectivamente la desviación estándar en la muestra y en la clase (segunda y cuarta columna respectivamente).
 - n y n_c son respectivamente el tamaño de la muestra y el efectivo de la clase.
2. Los perfiles fila y columna de las 8 variables de señalización, respecto a la partición.

Clase 1 (efectivo: 275)

Tema	\bar{x}	σ	$\bar{x} c_1$	σc_1	T
Aceptar sindicatos, sin exagerar	3.61	1.06	4.09	0.80	7.14
Ya no nos sentimos seguros	3.84	1.12	4.34	0.83	7.05
Los refugiados deben irse	3.06	1.18	3.60	1.20	6.94
Defender a la banca nacionalizada	3.75	0.97	4.16	0.76	6.64
Necesario crear canales TV estatales	3.15	1.25	3.68	1.17	6.51
Se debe extraditar a los extranj.	2.89	1.16	3.38	1.15	6.44
El SS debe privatizarse	2.02	1.19	1.56	0.93	-6.07
Controles de identidad indispensables	4.13	0.91	4.47	0.68	5.89
El país debe producir lo que se come	4.43	0.86	4.74	0.50	5.79
Privatizar emp. estatales como ICE	2.06	1.17	1.64	1.00	-5.58
El precarismo es un mal social	3.82	1.11	4.20	0.87	5.37
La policía no cumple su misión	3.02	1.11	3.40	1.06	5.24
Emp. ahogadas por cargas sociales	3.34	1.15	3.73	1.08	5.21
Gob. debe capacitar al desemp.	4.02	0.95	4.33	0.70	5.15
Cooperat. es una salida a la crisis	3.69	1.03	4.01	0.92	4.80
Narcotráfico financia a la contra	2.92	1.20	3.30	1.24	4.79
Favorecer la prueba de alcoholemia	4.16	0.91	4.43	0.74	4.64
Obligar emp. a emplear minusválidos	3.77	1.13	4.10	0.97	4.54
Aumentar presup. en seguridad	3.73	1.05	4.03	0.87	4.46
Favorecer financ. de camp. de vacunac.	4.37	0.79	4.59	0.56	4.41
Luchar contra la pornografía	4.20	1.04	4.49	0.88	4.34
Hay demasiadas huelgas	3.53	1.17	3.86	1.10	4.33
Acortar campaña electoral	4.00	1.17	4.32	1.07	4.21
El tabaco es una plaga	3.97	1.18	4.29	1.04	4.20
Subvencionar pérdidas por cosechas	3.32	1.24	3.66	1.27	4.15
Igualdad de sexos en puestos elegibles	3.79	1.18	4.10	1.12	4.03
USA debe intervenir militarmente en C.A.	2.15	1.27	1.82	1.19	-3.99
La democracia es el sufragio	3.44	1.35	3.77	1.33	3.73
Fortalecer la banca privada	3.07	1.22	2.77	1.34	-3.68
Esta bien que los hombres se maquillen	1.34	0.79	1.16	0.56	-3.61
Esta bien ver mujeres desnudas en TV	1.86	1.17	1.59	1.02	-3.58
Gorbachov inspira confianza	2.19	1.12	1.93	1.06	-3.56
Enfermos de SIDA deben aislarse	3.69	1.32	3.99	1.29	3.47
Intr. sector priv. en los hosp.	3.05	1.28	2.77	1.37	-3.29
Asamb. Legislativa no sirve para nada	2.82	1.24	3.08	1.32	3.16
Hay buenas cond. de crédito para prod.	2.53	1.27	2.27	1.30	-3.10
Fomentar creación de empresas	4.01	0.97	4.20	0.88	3.02
Aumentar las cuotas patronales	2.43	1.23	2.19	1.19	-2.98
Amas de casa deben recibir salario	3.46	1.34	3.72	1.31	2.96
Ens. obligatoria hasta los 20 años	3.21	1.37	3.47	1.36	2.89
Establecer relac. com. con la URSS	2.59	1.30	2.36	1.38	-2.66
Se puede confiar en la justicia	3.21	1.12	3.02	1.20	-2.55
Dios existe	4.79	0.65	4.89	0.47	2.43
Crear empleo sector público	3.41	1.33	3.62	1.33	2.40
Propiciar libre elección médica	3.91	1.03	4.07	1.01	2.37
Reducir edad de pensionarse	2.80	1.37	3.01	1.44	2.31
El FMI ayuda a resolver la crisis	2.72	1.16	2.55	1.16	-2.23
No hay que pagar la deuda externa	2.52	1.37	2.71	1.50	2.08
Uno debe sacrificarse por la patria	4.08	1.06	4.21	1.03	1.87
Cotizac. del SS no deben ser oblig.	2.67	1.25	2.82	1.39	1.79

Clase 2 (efectivo: 193)

Tema	\bar{x}	σ	\bar{x}_{c_2}	σ_{c_2}	T
La democracia es el sufragio	3.44	1.35	2.82	1.33	-6.01
Gorbachov inspira confianza	2.19	1.12	1.76	0.89	-5.12
Ya no nos sentimos seguros	3.84	1.12	4.27	0.89	5.12
Defender a la banca nacionalizada	3.75	0.97	3.37	1.05	-5.06
Luchar contra la pornografía	4.20	1.04	4.59	0.71	5.05
Hay buenas cond. de crédito para prod.	2.53	1.27	2.07	1.05	-4.82
Necesario crear canales TV estatales	3.15	1.25	2.69	1.25	-4.80
El tabaco es una plaga	3.97	1.18	4.39	0.83	4.79
Aumentar las cuotas patronales	2.43	1.23	1.99	0.99	-4.77
Asamb. Legislativa no sirve para nada	2.82	1.24	3.27	1.23	4.74
Acortar campaña electoral	4.00	1.17	4.40	0.83	4.60
La policía no cumple su misión	3.02	1.11	3.39	1.07	4.37
Se puede confiar en la justicia	3.21	1.12	2.84	1.03	-4.35
Propiciar libre elección médica	3.91	1.03	4.23	0.81	4.15
Noriega debe quedarse	1.67	0.98	1.38	0.65	-3.99
Hay demasiadas huelgas	3.53	1.17	3.88	1.02	3.96
Intr. sector priv. en los hosp.	3.05	1.28	3.43	1.05	3.95
Cooperat. es una salida a la crisis	3.69	1.03	3.39	1.04	-3.80
Gob. debe capacitar al desemp.	4.02	0.95	3.74	1.09	-3.78
Subvencionar pérdidas por cosechas	3.32	1.24	2.97	1.17	-3.71
Amas de casa deben recibir salario	3.46	1.34	3.83	1.23	3.64
Establecer relac. com. con la URSS	2.59	1.30	2.27	1.17	-3.25
Aceptar sindicatos, sin exagerar	3.61	1.06	3.36	1.09	-3.07
El FMI ayuda a resolver la crisis	2.72	1.16	2.45	1.07	-3.06
Por la semana laboral de 35 hrs.	2.22	1.30	1.92	1.10	-3.06
Esta bien ver mujeres desnudas en TV	1.86	1.17	1.59	1.02	-3.05
Narcotráfico financia a la contra	2.92	1.20	2.65	1.14	-2.95
Uno debe sacrificarse por la patria	4.08	1.06	3.85	1.19	-2.79
Fomentar empleo de medio tiempo	2.61	1.32	2.33	1.26	-2.78
Esta bien que los hombres se maquillen	1.34	0.79	1.18	0.52	-2.73
Controles de identidad indispensables	4.13	0.91	4.28	0.78	2.18
Igualdad de sexos en puestos elegibles	3.79	1.18	3.98	1.08	2.12
Dios existe	4.79	0.65	4.89	0.47	2.06
El SS debe privatizarse	2.02	1.19	2.20	1.21	1.97
Fomentar creación de empresas	4.01	0.97	3.87	0.94	-1.89
Extr. no deben beneficiarse del SS	2.57	1.25	2.39	1.20	-1.89
Crear empleo sector público	3.41	1.33	3.22	1.31	-1.87
Ens. obligatoria hasta los 20 años	3.21	1.37	3.02	1.41	-1.80
Frenar tec. que expulse mano de obra	2.94	1.33	2.76	1.26	-1.77
Se debe extraditar a los extranj.	2.89	1.16	3.03	1.10	1.58
Cotizac. del SS no deben ser oblig.	2.67	1.25	2.53	1.18	-1.47
Favorecer la prueba de alcoholemia	4.16	0.91	4.26	0.76	1.46
El precarismo es un mal social	3.82	1.11	3.94	1.11	1.41
Obligar emp. a emplear minusválidos	3.77	1.13	3.89	1.02	1.40
Fortalecer la banca privada	3.07	1.22	2.95	1.19	-1.28
Reducir edad de pensionarse	2.80	1.37	2.69	1.36	-1.05
Favorecer financ. de camp. de vacunac.	4.37	0.79	4.43	0.68	1.00
No hay que pagar la deuda externa	2.52	1.37	2.42	1.33	-0.95
Enfermos de SIDA deben aislarse	3.69	1.32	3.78	1.35	0.88
Homosex. deben integrarse socialmente	2.34	1.30	2.26	1.30	-0.80

Clase 3 (efectivo: 154)

Tema	\bar{x}	σ	$\bar{x} c_3$	σc_3	T
Amas de casa deben recibir salario	3.46	1.34	2.35	1.22	-9.86
El tabaco es una plaga	3.97	1.18	3.05	1.33	-9.09
Igualdad de sexos en puestos elegibles	3.79	1.18	3.04	1.26	-7.45
Intr. sector priv. en los hosp.	3.05	1.28	2.41	1.15	-5.95
Obligar emp. a emplear minusválidos	3.77	1.13	3.21	1.28	-5.77
Propiciar libre elección médica	3.91	1.03	3.41	1.15	-5.66
Acortar campaña electoral	4.00	1.17	3.46	1.39	-5.34
Aumentar presup. en seguridad	3.73	1.05	3.29	1.09	-4.93
Homosex. deben integrarse socialmente	2.34	1.30	1.87	0.99	-4.35
Ens. obligatoria hasta los 20 años	3.21	1.37	2.76	1.41	-3.86
Luchar contra la pornografía	4.20	1.04	3.86	1.11	-3.83
Hay buenas cond. de crédito para prod.	2.53	1.27	2.13	1.08	-3.76
Privatizae emp. estatales como ICE	2.06	1.17	1.70	0.91	-3.70
El FMI ayuda a resolver la crisis	2.72	1.16	2.39	0.95	-3.41
Uno debe sacrificarse por la patria	4.08	1.06	3.77	1.21	-3.40
Se debe extraditar a los extranj.	2.89	1.16	2.56	1.07	-3.38
Frenar tec.que expulse mano de obra	2.94	1.33	2.58	1.35	-3.19
La democracia es el sufragio	3.44	1.35	3.10	1.28	-2.98
Esta bien que los hombres se maquillen	1.34	0.79	1.15	0.49	-2.92
Necesario crear canales TV estatales	3.15	1.25	2.85	1.13	-2.85
Se puede confiar en la justicia	3.21	1.12	2.94	1.08	-2.85
Esta bien ver mujeres desnudas en TV	1.86	1.17	2.14	1.22	2.81
Crear empleo sector público	3.41	1.33	3.71	1.15	2.69
Por la semana laboral de 35 hrs.	2.22	1.30	1.93	1.03	-2.68
El SS debe privatizarse	2.02	1.19	1.76	0.92	-2.63
Hay demasiadas huelgas	3.53	1.17	3.27	1.24	-2.60
Asamb. Legislativa no sirve para nada	2.82	1.24	3.09	1.19	2.58
Gov. debe capacitar al desemp.	4.02	0.95	4.19	0.85	2.13
Aumentar las cuotas patronales	2.43	1.23	2.21	1.14	-2.12
Controles de identidad indispensables	4.13	0.91	3.97	1.02	-2.05
Aceptar sindicatos, sin exagerar	3.61	1.06	3.43	0.93	-2.02
Establecer relac. com. con la URSS	2.59	1.30	2.37	1.18	-2.01
No hay que pagar la deuda externa	2.52	1.37	2.29	1.23	-2.00
Favorecer la prueba de alcoholemia	4.16	0.91	4.02	0.86	-1.82
Subvencionar pérdidas por cosechas	3.32	1.24	3.13	1.23	-1.81
Narcotráfico financia a la contra	2.92	1.20	2.74	1.19	-1.77
Fortalecer la banca privada	3.07	1.22	2.90	1.16	-1.65
El precarismo es un mal social	3.82	1.11	3.67	1.13	-1.59
Cotizac. del SS no deben ser oblig.	2.67	1.25	2.82	1.22	1.42
Ya no nos sentimos seguros	3.84	1.12	3.71	1.12	-1.37
Gorbachov inspira confianza	2.19	1.12	2.08	0.99	-1.17
Cooperat. es una salida a la crisis	3.69	1.03	3.59	1.09	-1.14
Fomentar empleo de medio tiempo	2.61	1.32	2.49	1.25	-1.07
Noriega debe quedarse	1.67	0.98	1.59	0.89	-0.97
Favorecer financ. de camp. de vacunac.	4.37	0.79	4.43	0.64	0.91
Los refugiados deben irse	3.06	1.18	2.97	1.14	-0.90
Reducir edad de pensionarse	2.80	1.37	2.72	1.25	-0.69
El país debe producir lo que se come	4.43	0.86	4.47	0.80	0.55
La policía no cumple su misión	3.02	1.11	3.07	1.11	0.53
USA debe itervenir militarmente en C.A.	2.15	1.27	2.18	1.22	0.28

Clase 4 (efectivo: 124)

Tema	\bar{x}	σ	\bar{x}_{c_4}	σ_{c_4}	T
Asamb. Legislativa no sirve para nada	2.82	1.24	2.09	0.82	-6.43
Por la semana laboral de 35 hrs.	2.22	1.30	1.50	0.77	-6.07
El SS debe privatizarse	2.02	1.19	1.37	0.67	-6.00
La policía no cumple su misión	3.02	1.11	2.45	0.92	-5.56
Fomentar empleo de medio tiempo	2.61	1.32	1.94	1.11	-5.49
Controles de identidad indispensables	4.13	0.91	4.56	0.54	5.18
Privatizar emp. estatales como ICE	2.06	1.17	1.53	0.76	-4.95
Cotizac. del SS no deben ser oblig.	2.67	1.25	2.12	1.10	-4.74
Uno debe sacrificarse por la patria	4.08	1.06	4.51	0.72	4.43
Aumentar las cuotas patronales	2.43	1.23	1.94	0.92	-4.33
Favorecer financ. de camp. de vacunac.	4.37	0.79	4.68	0.50	4.29
Defender a la banca nacionalizada	3.75	0.97	4.13	0.70	4.26
Frenar tec. que expulse mano de obra	2.94	1.33	2.42	1.23	-4.20
Intr. sector priv. en los hosp.	3.05	1.28	2.58	1.31	-3.92
Extr. no deben beneficiarse del SS	2.57	1.25	2.12	1.11	-3.88
Luchar contra la pornografía	4.20	1.04	4.56	0.61	3.79
Acortar campaña electoral	4.00	1.17	4.40	0.81	3.73
Se puede confiar en la justicia	3.21	1.12	3.59	0.99	3.66
Reducir edad de pensionarse	2.80	1.37	2.35	1.30	-3.52
Ens. obligatoria hasta los 20 años	3.21	1.37	2.77	1.26	-3.45
Aumentar presup. en seguridad	3.73	1.05	4.06	0.85	3.40
Los refugiados deben irse	3.06	1.18	2.69	1.00	-3.39
Narcotráfico financia a la contra	2.92	1.20	2.55	1.06	-3.32
Cooperat. es una salida a la crisis	3.69	1.03	3.99	0.87	3.15
No hay que pagar la deuda externa	2.52	1.37	2.14	1.26	-2.98
USA debe itervenir militarmente en C.A.	2.15	1.27	1.80	1.06	-2.98
Noriega debe quedarse	1.67	0.98	1.41	0.63	-2.90
Aceptar sindicatos, sin exagerar	3.61	1.06	3.89	0.95	2.84
Se debe extraditar a los extranjer.	2.89	1.16	2.60	1.05	-2.69
El tabaco es una plaga	3.97	1.18	4.24	1.07	2.46
Esta bien ver mujeres desnudas en TV	1.86	1.17	1.60	1.03	-2.39
Dios existe	4.79	0.65	4.93	0.26	2.38
Subvencionar pérdidas por cosechas	3.32	1.24	3.06	1.24	-2.24
Crear empleo sector público	3.41	1.33	3.14	1.38	-2.16
Esta bien que los hombres se maquillen	1.34	0.79	1.19	0.43	-2.08
Favorecer la prueba de alcoholemia	4.16	0.91	4.33	0.73	2.02
El país debe producir lo que se come	4.43	0.86	4.59	0.68	2.01
Fomentar creación de empresas	4.01	0.97	4.18	0.77	1.90
Emp. ahogadas por cargas sociales	3.34	1.15	3.15	1.18	-1.76
Fortalecer la banca privada	3.07	1.22	2.87	1.20	-1.75
Homosex. deben integrarse socialmente	2.34	1.30	2.13	1.14	-1.74
Gob. debe capacitar al desemp.	4.02	0.95	3.89	0.97	-1.46
El FMI ayuda a resolver la crisis	2.72	1.16	2.57	1.10	-1.38
Propiciar libre elección médica	3.91	1.03	3.78	1.17	-1.33
Enfermos de SIDA deben aislarse	3.69	1.32	3.53	1.38	-1.29
Establecer relac. com. con la URSS	2.59	1.30	2.74	1.23	1.23
Ya no nos sentimos seguros	3.84	1.12	3.72	1.04	-1.15
Obligar emp. a emplear minusválidos	3.77	1.13	3.65	1.11	-1.13
Amas de casa deben recibir salario	3.46	1.34	3.33	1.34	-1.03
Igualdad de sexos en puestos elegibles	3.79	1.18	3.90	1.14	1.00

Clase 5 (efectivo: 99)

Tema	\bar{X}	σ	\bar{X}_{c5}	σ_{c5}	T
Crear empleo sector público	3.41	1.33	1.90	1.15	-11.02
Subvencionar pérdidas por cosechas	3.32	1.24	2.29	0.94	-8.10
Reducir edad de pensionarse	2.80	1.37	1.67	0.96	-8.07
Amas de casa deben recibir salario	3.46	1.34	2.53	1.27	-6.70
Por la semana laboral de 35 hrs.	2.22	1.30	1.34	0.73	-6.66
Necesario crear canales TV estatales	3.15	1.25	2.36	1.11	-6.12
Gob. debe capacitar al desemp.	4.02	0.95	3.43	1.11	-5.91
Obligar emp. a emplear minusválidos	3.77	1.13	3.11	1.24	-5.59
Frenar tec. que expulse mano de obra	2.94	1.33	2.19	1.09	-5.48
Igualdad de sexos en puestos elegibles	3.79	1.18	3.12	1.29	-5.43
Fomentar empleo de medio tiempo	2.61	1.32	1.89	1.19	-5.28
Aumentar las cuotas patronales	2.43	1.23	1.80	0.79	-5.02
Cooperat. es una salida a la crisis	3.69	1.03	3.22	1.01	-4.40
Gorbachov inspira confianza	2.19	1.12	2.70	1.06	4.40
Fortalecer la banca privada	3.07	1.22	3.61	0.98	4.31
Esta bien ver mujeres desnudas en TV	1.86	1.17	2.36	1.28	4.09
Luchar contra la pornografía	4.20	1.04	3.81	1.07	-3.60
Defender a la banca nacionalizada	3.75	0.97	3.39	1.07	-3.55
Establecer relac. com. con la URSS	2.59	1.30	3.05	1.25	3.41
Privatizar emp. estatales como ICE	2.06	1.17	2.46	1.24	3.28
Cotizac. del SS no deben ser oblig.	2.67	1.25	2.25	1.15	-3.25
Acortar campaña electoral	4.00	1.17	4.37	0.82	3.09
El país debe producir lo que se come	4.43	0.86	4.15	1.14	-3.06
Fomentar creación de empresas	4.01	0.97	4.31	0.82	3.00
La policía no cumple su misión	3.02	1.11	2.69	1.03	-2.87
La democracia es el sufragio	3.44	1.35	3.06	1.45	-2.69
No hay que pagar la deuda externa	2.52	1.37	2.14	1.24	-2.68
Ens. obligatoria hasta los 20 años	3.21	1.37	2.84	1.32	-2.60
Noriega debe quedarse	1.67	0.98	1.42	0.65	-2.50
El precarismo es un mal social	3.82	1.11	4.10	0.90	2.45
Asamb. Legislativa no sirve para nada	2.82	1.24	2.51	1.12	-2.42
El FMI ayuda a resolver la crisis	2.72	1.16	2.47	1.12	-2.08
USA debe intervenir militarmente en C.A.	2.15	1.27	2.41	1.38	1.96
Extr. no deben beneficiarse del SS	2.57	1.25	2.35	1.14	-1.70
Enfermos de SIDA deben aislarse	3.69	1.32	3.48	1.29	-1.53
Esta bien que los hombres se maquillen	1.34	0.79	1.46	0.88	1.45
Ya no nos sentimos seguros	3.84	1.12	3.68	1.17	-1.37
Se debe extraditar a los extranj.	2.89	1.16	2.73	1.15	-1.32
Hay demasiadas huelgas	3.53	1.17	3.37	1.23	-1.31
Narcotráfico financia a la contra	2.92	1.20	2.76	1.30	-1.27
Controles de identidad indispensables	4.13	0.91	4.25	0.83	1.27
Propiciar libre elección médica	3.91	1.03	3.78	1.13	-1.20
Hay buenas cond. de crédito para prod.	2.53	1.27	2.39	1.16	-1.06
Homosex. deben integrarse socialmente	2.34	1.30	2.20	1.20	-1.04
Aceptar sindicatos, sin exagerar	3.61	1.06	3.50	1.22	-0.99
Dios existe	4.79	0.65	4.85	0.56	0.89
Uno debe sacrificarse por la patria	4.08	1.06	3.99	1.10	-0.81
Favorecer la prueba de alcoholemia	4.16	0.91	4.09	0.85	-0.74
Los refugiados deben irse	3.06	1.18	2.98	1.17	-0.65
Favorecer financ. de camp. de vacunac.	4.37	0.79	4.32	0.81	-0.60

Clase 6 (efectivo: 116)

Tema	\bar{x}	σ	\bar{x}_{c_6}	σ_{c_6}	T
Gorbachov inspira confianza	2.19	1.12	3.23	1.24	9.55
Privatizar emp. estatales como ICE	2.06	1.17	2.91	1.29	7.47
Fortalecer la banca privada	3.07	1.22	3.92	1.06	7.29
El FMI ayuda a resolver la crisis	2.72	1.16	3.50	1.14	6.98
El SS debe privatizarse	2.02	1.19	2.72	1.42	6.00
Establecer relac. com. con la URSS	2.59	1.30	3.30	1.29	5.66
Esta bien que los hombres se maquillen	1.34	0.79	1.78	1.26	5.47
USA debe intervenir militarmente en C.A.	2.15	1.27	2.81	1.40	5.34
Intr. sector priv. en los hosp.	3.05	1.28	3.67	1.20	5.04
Se puede confiar en la justicia	3.21	1.12	3.71	1.05	4.65
Esta bien ver mujeres desnudas en TV	1.86	1.17	2.35	1.43	4.26
Hay buenas cond. de crédito para prod.	2.53	1.27	3.05	1.32	4.23
Homosex. deben integrarse socialmente	2.34	1.30	2.85	1.39	4.04
Fomentar creación de empresas	4.01	0.97	4.38	0.85	3.98
La democracia es el sufragio	3.44	1.35	3.93	1.14	3.80
El precarismo es un mal social	3.82	1.11	4.22	0.88	3.78
Aceptar sindicatos, sin exagerar	3.61	1.06	3.98	0.83	3.67
Cooperat. es una salida a la crisis	3.69	1.03	4.04	0.91	3.55
Aumentar presup. en seguridad	3.73	1.05	4.06	0.87	3.29
Acortar campaña electoral	4.00	1.17	4.33	0.90	2.96
Narcotráfico financia a la contra	2.92	1.20	3.26	1.30	2.92
Controles de identidad indispensables	4.13	0.91	4.38	0.72	2.88
Hay demasiadas huelgas	3.53	1.17	3.83	1.14	2.66
Propiciar libre elección médica	3.91	1.03	4.17	0.85	2.64
Subvencionar pérdidas por cosechas	3.32	1.24	3.01	1.28	-2.58
Reducir edad de pensionarse	2.80	1.37	2.46	1.37	-2.57
Fomentar empleo de medio tiempo	2.61	1.32	2.29	1.31	-2.51
El tabaco es una plaga	3.97	1.18	4.25	0.99	2.48
Defender a la banca nacionalizada	3.75	0.97	3.98	0.97	2.45
Ens. obligatoria hasta los 20 años	3.21	1.37	3.53	1.32	2.42
Necesario crear canales TV estatales	3.15	1.25	3.43	1.31	2.31
Uno debe sacrificarse por la patria	4.08	1.06	4.31	0.91	2.27
Por la semana laboral de 35 hrs.	2.22	1.30	1.95	1.22	-2.16
Se debe extraditar a los extranj.	2.89	1.16	3.13	1.17	2.14
Emp. ahogadas por cargas sociales	3.34	1.15	3.55	1.14	1.89
Enfermos de SIDA deben aislarse	3.69	1.32	3.90	1.20	1.66
La policía no cumple su misión	3.02	1.11	2.87	1.26	-1.38
El país debe producir lo que se come	4.43	0.86	4.54	0.84	1.32
Crear empleo sector público	3.41	1.33	3.25	1.45	-1.23
Luchar contra la pornografía	4.20	1.04	4.08	1.09	-1.19
Obligar emp. a emplear minusválidos	3.77	1.13	3.89	1.13	1.10
Frenar tec. que expulse mano de obra	2.94	1.33	3.07	1.36	1.01
Favorecer la prueba de alcoholemia	4.16	0.91	4.22	0.95	0.68
No hay que pagar la deuda externa	2.52	1.37	2.43	1.37	-0.68
Cotizac. del SS no deben ser oblig.	2.67	1.25	2.59	1.25	-0.66
Extr. no deben beneficiarse del SS	2.57	1.25	2.49	1.25	-0.66
Ya no nos sentimos seguros	3.84	1.12	3.91	1.08	0.65
Amas de casa deben recibir salario	3.46	1.34	3.39	1.37	-0.54
Los refugiados deben irse	3.06	1.18	3.12	1.12	0.52
Favorecer financ. de camp. de vacunac.	4.37	0.79	4.41	0.81	0.52

Clase 7 (efectivo: 135)

Tema	\bar{F}	σ	\bar{F}_{c7}	σ_{c7}	T
Por la semana laboral de 35 hrs.	2.22	1.30	3.76	1.13	13.31
El SS debe privatizarse	2.02	1.19	3.12	1.32	10.18
Aumentar las cuotas patronales	2.43	1.23	3.51	1.19	9.79
Fomentar empleo de medio tiempo	2.61	1.32	3.70	1.03	9.33
Reducir edad de pensionarse	2.80	1.37	3.81	1.11	8.31
Privatizar emp. estatales como ICE	2.06	1.17	2.88	1.25	7.74
Frenar tec. que expulse mano de obra	2.94	1.33	3.80	1.02	7.31
Ens. obligatoria hasta los 20 años	3.21	1.37	4.05	1.06	6.93
Cotizac. del SS no deben ser oblig.	2.67	1.25	3.42	1.08	6.74
Crear empleo sector público	3.41	1.33	4.15	0.89	6.33
Intr. sector priv. en los hosp.	3.05	1.28	3.75	1.07	6.16
Subvencionar pérdidas por cosechas	3.32	1.24	3.99	0.85	6.14
El FMI ayuda a resolver la crisis	2.72	1.16	3.35	1.10	6.06
Hay buenas cond. de crédito para prod.	2.53	1.27	3.20	1.24	5.87
Extr. no deben beneficiarse del SS	2.57	1.25	3.23	1.28	5.86
Fortalecer la banca privada	3.07	1.22	3.64	0.99	5.27
La democracia es el sufragio	3.44	1.35	4.00	0.98	4.70
Amas de casa deben recibir salario	3.46	1.34	4.01	1.16	4.61
Narcotráfico financia a la contra	2.92	1.20	3.39	1.14	4.37
USA debe intervenir militarmente en C.A.	2.15	1.27	2.65	1.32	4.36
Obligar emp. a emplear minusválidos	3.77	1.13	4.19	0.91	4.19
Gob. debe capacitar al desemp.	4.02	0.95	4.36	0.61	4.08
Necesario crear canales TV estatales	3.15	1.25	3.60	1.10	4.04
Noriega debe quedarse	1.67	0.98	2.02	1.20	3.89
No hay que pagar la deuda externa	2.52	1.37	2.97	1.36	3.65
Enfermos de SIDA deben aislarse	3.69	1.32	4.09	1.04	3.42
La policía no cumple su misión	3.02	1.11	3.35	1.18	3.28
Emp. ahogadas por cargas sociales	3.34	1.15	3.67	0.98	3.22
Igualdad de sexos en puestos elegibles	3.79	1.18	4.12	0.91	3.16
Controles de identidad indispensables	4.13	0.91	3.88	0.87	-3.06
Los refugiados deben irse	3.06	1.18	3.38	1.10	3.03
Se debe extraditar a los extranj.	2.89	1.16	3.19	1.12	2.88
Fomentar creación de empresas	4.01	0.97	4.24	0.83	2.66
Aumentar presup. en seguridad	3.73	1.05	3.97	0.97	2.55
Hay demasiadas huelgas	3.53	1.17	3.77	1.01	2.30
El tabaco es una plaga	3.97	1.18	4.20	0.96	2.19
Favorecer la prueba de alcoholemia	4.16	0.91	4.33	0.71	2.11
Acortar campaña electoral	4.00	1.17	3.83	1.17	-1.61
Esta bien que los hombres se maquillen	1.34	0.79	1.45	0.94	1.52
Ya no nos sentimos seguros	3.84	1.12	3.99	1.04	1.49
Favorecer financ. de camp. de vacunac.	4.37	0.79	4.47	0.61	1.43
Dios existe	4.79	0.65	4.87	0.40	1.40
Asamb. Legislativa no sirve para nada	2.82	1.24	2.96	1.27	1.25
Cooperat. es una salida a la crisis	3.69	1.03	3.79	0.99	1.08
Esta bien ver mujeres desnudas en TV	1.86	1.17	1.97	1.33	1.03
Establecer relac. com. con la URSS	2.59	1.30	2.48	1.26	-0.94
El precarismo es un mal social	3.82	1.11	3.73	1.06	-0.90
Homosex. deben integrarse socialmente	2.34	1.30	2.43	1.50	0.76
Se puede confiar en la justicia	3.21	1.12	3.27	1.15	0.59
Propiciar libre elección médica	3.91	1.03	3.96	0.97	0.54

Clase 8 (efectivo: 156)

Tema	\bar{x}	σ	\bar{x}_{c8}	σ_{c8}	T
Crear empleo sector público	3.41	1.33	4.12	0.83	6.53
Fomentar empleo de medio tiempo	2.61	1.32	3.27	1.26	5.96
Los refugiados deben irse	3.06	1.18	2.48	1.08	-5.88
Subvencionar pérdidas por cosechas	3.32	1.24	3.92	1.12	5.79
Se debe extraditar a los extranj.	2.89	1.16	2.33	1.04	-5.78
Asamb. Legislativa no sirve para nada	2.82	1.24	2.25	1.03	-5.54
Aumentar las cuotas patronales	2.43	1.23	2.99	1.27	5.39
El FMI ayuda a resolver la crisis	2.72	1.16	3.22	1.11	5.14
Reducir edad de pensionarse	2.80	1.37	3.36	1.23	4.90
Por la semana laboral de 35 hrs.	2.22	1.30	2.75	1.37	4.81
Acortar campaña electoral	4.00	1.17	3.54	1.31	-4.61
Obligar emp. a emplear minusválidos	3.77	1.13	4.20	0.87	4.61
Gob. debe capacitar al desemp.	4.02	0.95	4.38	0.70	4.60
La policía no cumple su misión	3.02	1.11	2.61	1.02	-4.42
Amas de casa deben recibir salario	3.46	1.34	3.94	1.14	4.31
Igualdad de sexos en puestos elegibles	3.79	1.18	4.20	0.94	4.20
Se puede confiar en la justicia	3.21	1.12	3.59	1.08	4.04
Hay buenas cond. de crédito para prod.	2.53	1.27	2.96	1.32	4.00
Esta bien ver mujeres desnudas en TV	1.86	1.17	1.48	0.75	-3.97
El precarismo es un mal social	3.82	1.11	3.45	1.28	-3.89
Gorbachov inspira confianza	2.19	1.12	1.83	0.92	-3.87
La democracia es el sufragio	3.44	1.35	3.87	1.24	3.81
Intr. sector priv. en los hosp.	3.05	1.28	3.44	1.17	3.64
Favorecer la prueba de alcoholemia	4.16	0.91	4.43	0.65	3.60
Fortalecer la banca privada	3.07	1.22	3.41	1.08	3.34
Frenar tec. que expulse mano de obra	2.94	1.33	3.31	1.37	3.29
Cooperat. es una salida a la crisis	3.69	1.03	3.97	0.94	3.25
Luchar contra la pornografía	4.20	1.04	4.47	0.83	3.13
Defender a la banca nacionalizada	3.75	0.97	3.99	0.78	2.99
Esta bien que los hombres se maquillen	1.34	0.79	1.15	0.54	-2.93
Uno debe sacrificarse por la patria	4.08	1.06	4.33	0.83	2.85
Narcotráfico financia a la contra	2.92	1.20	2.64	1.13	-2.78
Propiciar libre elección médica	3.91	1.03	4.14	0.91	2.68
No hay que pagar la deuda externa	2.52	1.37	2.25	1.26	-2.35
Establecer relac. com. con la URSS	2.59	1.30	2.34	1.24	-2.29
Dios existe	4.79	0.65	4.91	0.43	2.25
Favorecer financ. de camp. de vacunac.	4.37	0.79	4.51	0.59	2.15
Ya no nos sentimos seguros	3.84	1.12	3.64	1.27	-2.09
Privatizar emp. estatales como ICE	2.06	1.17	1.87	1.10	-1.94
Aumentar presup. en seguridad	3.73	1.05	3.90	1.05	1.92
Necesario crear canales TV estatales	3.15	1.25	3.35	1.17	1.91
Enfermos de SIDA deben aislarse	3.69	1.32	3.48	1.40	-1.87
El país debe producir lo que se come	4.43	0.86	4.56	0.58	1.84
Ens. obligatoria hasta los 20 años	3.21	1.37	3.40	1.35	1.65
Hay demasiadas huelgas	3.53	1.17	3.41	1.16	-1.21
Cotizac. del SS no deben ser oblig.	2.67	1.25	2.78	1.22	1.04
Emp. ahogadas por cargas sociales	3.34	1.15	3.24	1.04	-1.04
USA debe intervenir militarmente en C.A.	2.15	1.27	2.04	1.28	-1.02
Extr. no deben beneficiarse del SS	2.57	1.25	2.47	1.21	-0.95
Fomentar creación de empresas	4.01	0.97	3.94	1.00	-0.85

Clase 9 (efectivo: 163)

Tema	\bar{X}	σ	\bar{X}_{C9}	σ_{C9}	T
Favorecer financ. de camp. de vacunac.	4.37	0.79	3.41	1.07	-14.15
Favorecer la prueba de alcoholemia	4.16	0.91	3.24	1.14	-11.91
Controles de identidad indispensables	4.13	0.91	3.26	0.99	-11.48
Defender a la banca nacionalizada	3.75	0.97	2.88	0.85	-10.99
Fomentar creación de empresas	4.01	0.97	3.21	1.02	-9.94
El país debe producir lo que se come	4.43	0.86	3.71	1.18	-9.72
Ya no nos sentimos seguros	3.84	1.12	3.04	1.03	-8.72
Aumentar presup. en seguridad	3.73	1.05	2.99	0.98	-8.59
Gob. debe capacitar al desemp.	4.02	0.95	3.37	1.04	-8.21
Aceptar sindicatos, sin exagerar	3.61	1.06	2.93	1.02	-7.80
Cooperat. es una salida a la crisis	3.69	1.03	3.04	0.96	-7.69
Acortar campaña electoral	4.00	1.17	3.28	1.12	-7.48
Luchar contra la pornografía	4.20	1.04	3.55	1.22	-7.43
El tabaco es una plaga	3.97	1.18	3.36	1.21	-6.24
Propiciar libre elección médica	3.91	1.03	3.42	0.95	-5.80
Obligar emp. a emplear minusválidos	3.77	1.13	3.24	1.12	-5.68
Dios existe	4.79	0.65	4.47	1.00	-5.59
Emp. ahogadas por cargas sociales	3.34	1.15	2.82	0.90	-5.58
Igualdad de sexos en puestos elegibles	3.79	1.18	3.27	1.06	-5.39
Hay demasiadas huelgas	3.53	1.17	3.03	0.98	-5.25
El precarismo es un mal social	3.82	1.11	3.37	0.99	-4.96
Esta bien que los hombres se maquillen	1.34	0.79	1.64	0.96	4.49
Uno debe sacrificarse por la patria	4.08	1.06	3.71	1.09	-4.21
Por la semana laboral de 35 hrs.	2.22	1.30	2.65	1.18	4.04
Necesario crear canales TV estatales	3.15	1.25	2.76	1.06	-3.83
Noriega debe quedarse	1.67	0.98	1.97	0.99	3.70
Fortalecer la banca privada	3.07	1.22	2.74	0.94	-3.34
Privatizar emp. estatales como ICE	2.06	1.17	2.33	0.98	2.83
Intr. sector priv. en los hosp.	3.05	1.28	2.79	1.01	-2.50
Crear empleo sector público	3.41	1.33	3.14	1.15	-2.49
Ens. obligatoria hasta los 20 años	3.21	1.37	2.94	1.23	-2.41
Enfermos de SIDA deben aislarse	3.69	1.32	3.43	1.24	-2.40
Aumentar las cuotas patronales	2.43	1.23	2.66	1.09	2.29
La democracia es el sufragio	3.44	1.35	3.20	1.21	-2.17
La policía no cumple su misión	3.02	1.11	2.83	0.90	-2.11
Hay buenas cond. de crédito para prod.	2.53	1.27	2.73	1.04	1.94
El SS debe privatizarse	2.02	1.19	2.20	1.12	1.84
Amas de casa deben recibir salario	3.46	1.34	3.27	1.12	-1.74
Se debe extraditar a los extranj.	2.89	1.16	2.75	1.11	-1.46
Los refugiados deben irse	3.06	1.18	2.92	1.01	-1.45
No hay que pagar la deuda externa	2.52	1.37	2.68	1.20	1.43
Establecer relac. com. con la URSS	2.59	1.30	2.44	1.02	-1.42
USA debe intervenir militarmente en C.A.	2.15	1.27	2.29	1.04	1.35
Subvencionar pérdidas por cosechas	3.32	1.24	3.19	1.09	-1.28
Fomentar empleo de medio tiempo	2.61	1.32	2.73	1.02	1.12
Frenar tec. que expulse mano de obra	2.94	1.33	2.82	1.06	-1.11
Extr. no deben beneficiarse del SS	2.57	1.25	2.68	1.14	1.07
Homosex. deben integrarse socialmente	2.34	1.30	2.23	1.16	-1.03
Narcotráfico financia a la contra	2.92	1.20	2.82	1.04	-1.02
Esta bien ver mujeres desnudas en TV	1.86	1.17	1.95	1.13	0.93

Clase 10 (efectivo: 62)

Tema	\bar{x}	σ	\bar{x}_{c10}	σ_{c10}	T
Dios existe	4.79	0.65	3.98	1.31	-9.07
Establecer relac. com. con la URSS	2.59	1.30	3.96	1.12	8.17
Hay demasiadas huelgas	3.53	1.17	2.40	1.14	-7.45
Noriega debe quedarse	1.67	0.98	2.61	1.33	7.27
Fortalecer la banca privada	3.07	1.22	1.98	0.96	-6.94
Gorbachov inspira confianza	2.19	1.12	3.20	1.35	6.89
Homosex. deben integrarse socialmente	2.34	1.30	3.42	1.25	6.41
Emp. ahogadas por cargas sociales	3.34	1.15	2.45	1.10	-5.98
Aumentar presup. en seguridad	3.73	1.05	2.92	1.14	-5.93
El FMI ayuda a resolver la crisis	2.72	1.16	1.90	1.00	-5.48
No hay que pagar la deuda externa	2.52	1.37	3.48	1.40	5.40
Enfermos de SIDA deben aislarse	3.69	1.32	2.79	1.32	-5.26
USA debe intervenir militarmente en C.A.	2.15	1.27	1.36	0.83	-4.85
La democracia es el sufragio	3.44	1.35	2.59	1.41	-4.84
Esta bien que los hombres se maquillen	1.34	0.79	1.79	1.11	4.31
El SS debe privatizarse	2.02	1.19	1.37	0.79	-4.26
El precarismo es un mal social	3.82	1.11	3.24	1.44	-3.97
Defender a la banca nacionalizada	3.75	0.97	4.24	0.79	3.92
Ya no nos sentimos seguros	3.84	1.12	3.27	1.32	-3.89
Esta bien ver mujeres desnudas en TV	1.86	1.17	2.45	1.28	3.87
Luchar contra la pornografía	4.20	1.04	3.68	1.28	-3.81
Aumentar las cuotas patronales	2.43	1.23	3.01	1.25	3.63
Reducir edad de pensionarse	2.80	1.37	2.18	1.10	-3.51
El tabaco es una plaga	3.97	1.18	3.47	1.30	-3.25
Intr. sector priv. en los hosp.	3.05	1.28	2.52	1.35	-3.18
Privatizar emp. estatales como ICE	2.06	1.17	1.58	1.06	-3.17
Cotizac. del SS no deben ser oblig.	2.67	1.25	2.17	1.14	-3.09
Amas de casa deben recibir salario	3.46	1.34	3.98	1.10	3.01
Crear empleo sector público	3.41	1.33	2.92	1.36	-2.83
Los refugiados deben irse	3.06	1.18	2.71	1.26	-2.28
Controles de identidad indispensables	4.13	0.91	3.91	0.95	-1.86
Por la semana laboral de 35 hrs.	2.22	1.30	1.91	1.01	-1.85
Gob. debe capacitar al desemp.	4.02	0.95	4.23	0.73	1.71
Hay buenas cond. de crédito para prod.	2.53	1.27	2.27	1.35	-1.57
Frenar tec. que expulse mano de obra	2.94	1.33	3.21	1.31	1.56
Favorecer financ. de camp. de vacunac.	4.37	0.79	4.52	0.73	1.46
Se debe extraditar a los extranj.	2.89	1.16	2.68	1.07	-1.40
Igualdad de sexos en puestos elegibles	3.79	1.18	4.00	1.08	1.37
Aceptar sindicatos, sin exagerar	3.61	1.06	3.44	1.12	-1.23
Fomentar empleo de medio tiempo	2.61	1.32	2.82	1.24	1.23
La policía no cumple su misión	3.02	1.11	2.86	1.03	-1.11
Fomentar creación de empresas	4.01	0.97	3.87	1.05	-1.11
Se puede confiar en la justicia	3.21	1.12	3.05	1.13	-1.10
Necesario crear canales TV estatales	3.15	1.25	3.32	1.29	1.04
Subvencionar pérdidas por cosechas	3.32	1.24	3.48	1.05	1.00
Propiciar libre elección médica	3.91	1.03	3.97	0.90	0.45
Cooperat. es una salida a la crisis	3.69	1.03	3.75	0.90	0.45
Asamb. Legislativa no sirve para nada	2.82	1.24	2.89	1.32	0.43
Ens. obligatoria hasta los 20 años	3.21	1.37	3.15	1.33	-0.33
Extr. no deben beneficiarse del SS	2.57	1.25	2.62	1.15	0.31

Perfiles columna

Provincia	1	2	3	4	5	6	7	8	9	10
San José	34.9	45.1	38.3	36.3	44.4	29.3	24.4	34.6	40.5	32.3
Alajuela	16.0	15.5	16.2	16.9	17.2	23.3	25.9	18.6	14.1	11.3
Cartago	13.1	10.4	10.4	8.1	7.1	13.8	10.4	14.7	11.7	12.9
Heredia	6.2	7.8	8.4	9.7	10.1	10.3	5.2	3.8	11.7	11.3
Guanacaste	7.6	4.7	3.9	11.3	3.0	2.6	6.7	9.0	6.1	9.7
Puntarenas	11.6	15.5	13.6	8.9	9.1	13.8	23.7	11.5	10.4	16.1
Limón	10.5	1.0	9.1	8.9	9.1	6.9	3.7	7.7	5.5	6.5

Edad	1	2	3	4	5	6	7	8	9	10
de 18 a 24	14.9	11.9	20.8	12.1	23.2	8.6	25.9	28.2	31.3	25.8
de 25 a 34	30.5	28.0	38.3	24.2	32.3	27.6	20.7	31.4	26.4	32.3
de 35 a 44	23.3	28.5	18.2	21.8	18.2	20.7	15.6	15.4	16.6	11.3
de 45 a 54	12.7	12.4	11.0	14.5	10.1	19.8	13.3	11.5	13.5	14.5
de 55 y más	17.8	19.2	11.7	26.6	16.2	23.3	24.4	13.5	12.3	16.1

Sexo	1	2	3	4	5	6	7	8	9	10
masculino	37.1	37.8	61.7	47.6	59.6	58.6	34.1	38.5	40.5	58.1
femenino	62.9	62.2	38.3	52.4	40.4	41.4	65.9	61.5	59.5	41.9

Profesión	1	2	3	4	5	6	7	8	9	10
Agricultor	0.7	3.1	3.9	3.2	2.0	0.9	0.7	0.0	1.8	1.6
Pat. Emp.Com.	9.5	9.8	11.7	9.7	20.2	16.4	5.2	6.4	5.5	9.7
Prof. Lib.	1.5	4.1	3.9	0.8	7.1	3.4	0.7	1.3	1.8	3.2
Empleados	25.1	26.9	39.0	20.2	17.2	27.6	40.7	38.5	31.3	29.0
Prof. Int.	6.2	4.1	2.6	9.7	6.1	1.7	1.5	3.2	3.7	12.9
Maest.y Prof.	6.5	5.7	5.2	7.3	6.1	3.4	2.2	3.2	4.3	4.8
Estudiantes	4.4	2.6	7.8	5.6	13.1	2.6	5.9	9.6	8.0	21.0
Pensionado	3.3	4.1	1.3	14.5	7.1	6.0	4.4	3.2	4.3	1.6
Amas de casa	40.7	37.3	21.4	28.2	20.2	33.6	37.0	33.3	37.4	16.1
Desempleados	1.8	1.6	3.2	0.0	0.0	4.3	1.5	1.3	1.8	0.0

Educación	1	2	3	4	5	6	7	8	9	10
Ninguna	1.8	3.6	7.1	0.8	2.0	5.2	9.6	4.5	6.1	6.5
Primaria	38.2	28.5	29.2	27.4	13.1	37.9	46.7	37.2	40.5	21.0
Secundaria	28.4	31.1	29.9	28.2	22.2	22.4	28.9	35.3	25.2	9.7
Técnica	7.3	7.8	7.8	11.3	10.1	5.2	3.7	5.1	6.1	6.5
Univ. Incomp.	13.1	16.1	15.6	12.1	34.3	18.1	8.9	14.7	13.5	37.1
Univ. Comp.	11.3	13.0	10.4	20.2	18.2	10.3	2.2	3.2	8.6	19.4

Perfiles columna (continuación)

Religión	1	2	3	4	5	6	7	8	9	10
Sin Religión	2.9	3.1	5.8	3.2	5.1	1.7	3.7	1.3	3.7	8.1
Cat.no Pract.	10.9	13.0	13.6	8.1	14.1	15.5	14.1	12.2	11.7	25.8
Cat. no reg.	20.4	20.7	27.3	19.4	14.1	13.8	20.0	20.5	22.7	27.4
Cat.prac.reg.	56.7	50.3	37.7	62.1	52.5	62.1	56.3	53.8	49.1	27.4
Otra. relig.	8.7	11.9	15.6	7.3	14.1	6.0	5.9	11.5	12.9	9.7
Prof. sindic.	1	2	3	4	5	6	7	8	9	10
Anep	0.4	1.0	0.0	0.0	0.0	0.9	0.0	0.6	0.0	1.6
Cut	10.2	10.4	8.4	6.5	2.0	8.6	8.1	9.0	4.9	4.8
Catd	2.9	3.1	3.2	4.0	3.0	2.6	3.7	0.6	2.5	12.9
Cctd	0.4	0.0	0.0	0.8	1.0	0.9	0.0	0.0	0.6	1.6
Ande	1.8	0.5	1.9	2.4	0.0	30.2	4.4	2.6	2.5	3.2
Apse	28.7	23.8	20.8	29.8	28.3	6.0	30.4	29.5	21.5	16.1
Upanacinal	6.5	4.1	7.1	8.9	7.1	3.4	6.7	3.8	7.4	3.2
Fedeagro	5.8	5.7	3.9	5.6	6.1	0.9	1.5	7.1	3.7	4.8
Upagra	1.8	0.0	0.6	0.8	0.0	0.0	2.2	2.6	0.6	8.1
Solidarismo	4.0	2.6	4.5	3.2	0.0	15.5	8.1	14.7	1.8	16.1
Ninguno	11.6	16.6	16.9	16.9	25.3	13.8	14.8	10.9	11.0	11.3
N.R.	14.5	16.6	16.2	12.1	16.2	18.1	20.0	19.2	18.4	17.7
Pos. polít.	1	2	3	4	5	6	7	8	9	10
Izquierda	1.8	0.5	1.3	3.2	5.1	2.6	3.0	0.6	0.6	19.4
Centro	34.5	40.4	38.3	45.2	49.5	40.5	25.9	21.2	28.8	50.0
Derecha	26.5	26.4	23.4	22.6	24.2	24.1	25.9	34.6	27.6	12.9
Ext. Derecha	4.0	2.6	5.2	29.0	3.0	6.9	4.4	5.8	4.3	14.5
Indiferentes	32.0	29.5	31.2	0.0	16.2	24.1	40.0	36.5	38.0	3.2
N.R.	1.1	0.5	0.6	0.0	2.0	1.7	0.7	1.3	0.6	0.0

Perfiles fila

Provincia	1	2	3	4	5	6	7	8	9	10
San José	17.8	16.2	11.0	8.4	8.2	6.3	6.1	10.0	12.3	3.7
Alajuela	17.1	11.6	9.7	8.1	6.6	10.5	13.6	11.2	8.9	2.7
Cartago	21.3	11.8	9.5	5.9	4.1	9.5	8.3	13.6	11.2	4.7
Heredia	14.4	12.7	11.0	10.2	8.5	10.2	5.9	5.1	16.1	5.9
Guanacaste	22.1	9.5	6.3	14.7	3.2	3.2	9.5	14.7	10.5	6.3
Puntarenas	16.3	15.3	10.7	5.6	4.6	8.2	16.3	9.2	8.7	5.1
Limón	28.2	1.9	13.6	10.7	8.7	7.8	4.9	11.7	8.7	3.9

Edad	1	2	3	4	5	6	7	8	9	10
de 18 a 24	14.1	7.9	11.0	5.2	7.9	3.4	12.1	15.2	17.6	5.5
de 25 a 34	19.5	12.5	13.7	7.0	7.4	7.4	6.5	11.4	10.0	4.6
de 35 a 44	21.7	18.6	9.5	9.2	6.1	8.1	7.1	8.1	9.2	2.4
de 45 a 54	18.0	12.4	8.8	9.3	5.2	11.9	9.3	9.3	11.3	4.6
de 55 y más	18.6	14.0	6.8	12.5	6.1	10.2	12.5	8.0	7.6	3.8

Sexo	1	2	3	4	5	6	7	8	9	10
masculino	15.4	11.0	14.3	8.9	8.9	10.2	6.9	9.0	9.9	5.4
femenino	21.3	14.8	7.3	8.0	4.9	5.9	10.9	11.8	11.9	3.2

Profesión	1	2	3	4	5	6	7	8	9	10
Agricultor	7.7	23.1	23.1	15.4	7.7	3.8	3.8	0.0	11.5	3.8
Pat.Emp.Com.	17.8	13.0	12.3	8.2	13.7	13.0	4.8	6.8	6.2	4.1
Prof.Liberal	10.5	21.1	15.8	2.6	18.4	10.5	2.6	5.3	7.9	5.3
Empleados	15.7	11.8	13.7	5.7	3.9	7.3	12.5	13.7	11.6	4.1
Prof.Intrmed.	24.3	11.4	5.7	17.1	8.6	2.9	2.9	7.1	8.6	11.4
Maest.y Prof.	24.3	14.9	10.8	12.2	8.1	5.4	4.1	6.8	9.5	4.1
Estudiantes	11.9	5.0	11.9	6.9	12.9	3.0	7.9	14.9	12.9	12.9
Pensionados	12.9	11.4	2.9	25.7	10.0	10.0	8.6	7.1	10.0	1.4
Amas de casa	23.1	14.9	6.8	7.2	4.1	8.1	10.3	10.7	12.6	2.1
Desempleados	20.0	12.0	20.0	0.0	0.0	20.0	8.0	8.0	12.0	0.0

Educación	1	2	3	4	5	6	7	8	9	10
Ninguna	7.6	10.6	16.7	1.5	3.0	9.1	19.7	10.6	15.2	6.1
Primaria	21.2	11.1	9.1	6.9	2.6	8.9	12.7	11.7	13.3	2.6
Secundaria	19.1	14.7	11.3	8.6	5.4	6.4	9.6	13.5	10.0	1.5
Técnica	19.2	14.4	11.5	13.5	9.6	5.8	4.8	7.7	9.6	3.8
Univ.Incomp.	14.9	12.9	10.0	6.2	14.1	8.7	5.0	9.5	9.1	9.5
Univ.Completa	19.3	15.5	9.9	15.5	11.2	7.5	1.9	3.1	8.7	7.5

Perfiles fila (continuación)

Religión	1	2	3	4	5	6	7	8	9	10
Sin religión	15.4	11.5	17.3	7.7	9.6	3.8	9.6	3.8	11.5	9.6
Cat.No pract.	15.7	13.1	11.0	5.2	7.3	9.4	9.9	9.9	9.9	8.4
Cat. no reg.	18.4	13.1	13.8	7.9	4.6	5.2	8.9	10.5	12.1	5.6
Cat.prac.Reg.	20.3	12.6	7.5	10.0	6.8	9.4	9.9	10.9	10.4	2.2
Otra Religión	15.6	14.9	15.6	5.8	9.1	4.5	5.2	11.7	13.6	3.9
N.R.	16.7	33.3	0.0	0.0	0.0	16.7	0.0	16.7	0.0	16.7

Prof. sindic.	1	2	3	4	5	6	7	8	9	10
Anep	23.9	17.1	11.1	6.8	1.7	8.5	9.4	12.0	6.8	2.6
Cut	16.7	12.5	10.4	10.4	6.3	6.3	10.4	2.1	8.3	16.7
Catd	16.7	0.0	0.0	16.7	16.7	16.7	0.0	0.0	16.7	16.7
Cctd	7.9	1.6	4.8	4.8	0.0	55.6	9.5	6.3	6.3	3.2
Ande	21.9	12.7	8.9	10.2	7.8	1.9	11.4	12.7	9.7	2.8
Apse	20.5	9.1	12.5	12.5	8.0	4.5	10.2	6.8	13.6	2.3
Upanacinal	23.2	15.9	8.7	10.1	8.7	1.4	2.9	15.9	8.7	4.3
Fedeagro	25.0	0.0	5.0	5.0	0.0	0.0	15.0	20.0	5.0	25.0
Upagra	12.0	5.4	7.6	4.3	0.0	19.6	12.0	25.0	3.3	10.9
Solidarismo	15.0	15.0	12.1	9.8	11.7	7.5	9.3	7.9	8.4	3.3
Ninguno	16.2	13.0	10.1	6.1	6.5	8.5	10.9	12.1	12.1	4.5
N.R.	21.1	21.1	16.4	7.2	7.2	0.0	0.0	0.0	27.0	0.0

Pos. polít.	1	2	3	4	5	6	7	8	9	10
Izquierda	13.2	2.6	5.3	10.5	13.2	7.9	10.5	2.6	2.6	31.6
Centro	17.9	14.7	11.1	10.6	9.2	8.9	6.6	6.2	8.9	5.8
Derecha	19.1	13.4	9.4	7.3	6.3	7.3	9.2	14.1	11.8	2.1
Ext.Derecha	10.8	4.9	7.8	35.3	2.9	7.8	5.9	8.8	6.9	8.8
Indiferentes	21.4	13.8	11.7	0.0	3.9	6.8	13.1	13.8	15.0	0.5
N.R.	23.1	7.7	7.7	0.0	15.4	15.4	7.7	15.4	7.7	0.0

1. Notaciones y definiciones

Este estudio se realizó en el marco de un convenio suscrito entre el Instituto de Estadística y Censos del Perú y el Banco Mundial. El estudio se realizó en el marco de un convenio suscrito entre el Instituto de Estadística y Censos del Perú y el Banco Mundial. El estudio se realizó en el marco de un convenio suscrito entre el Instituto de Estadística y Censos del Perú y el Banco Mundial.

Los datos fueron recolectados a través de encuestas realizadas en el territorio nacional. El estudio se realizó en el marco de un convenio suscrito entre el Instituto de Estadística y Censos del Perú y el Banco Mundial. El estudio se realizó en el marco de un convenio suscrito entre el Instituto de Estadística y Censos del Perú y el Banco Mundial.

Los resultados de este estudio se presentan en el presente informe. El estudio se realizó en el marco de un convenio suscrito entre el Instituto de Estadística y Censos del Perú y el Banco Mundial. El estudio se realizó en el marco de un convenio suscrito entre el Instituto de Estadística y Censos del Perú y el Banco Mundial.

Métodos y resultados en Clasificación Automática

William Castillo*

Resumen

El objetivo de este trabajo es exponer los fundamentos teóricos de la Clasificación Automática, específicamente lo que concierne a la *agregación de Ward* y los métodos de clasificación del tipo *nubes dinámicas*, preocupándonos especialmente por los desarrollos matemáticos que sustentan dicha teoría.

Nos interesa esta materia justamente por las amplias posibilidades de aplicación conjunta de los métodos de Ward (Clasificación Aglomerativa) y de los *centros móviles*, para determinar una buena partición de un conjunto de individuos.

Nos ocuparemos de los resultados que conducen de manera natural a la definición de la agregación de Ward usada en Clasificación Jerárquica Ascendente (sección 3). Luego presentamos, con bastante generalidad, el método iterativo de las nubes dinámicas y discutimos su convergencia.

1 Notaciones y definiciones

Sea Ω un conjunto de n *unidades estadísticas* (llamadas también *objetos* o *individuos*), caracterizadas por p *variables cuantitativas*. Para cada $i \in \{1, \dots, n\}$, el i -ésimo objeto ω_i se identifica con el vector de las p observaciones realizadas sobre él: $\omega_i = (\omega_i^1, \dots, \omega_i^p)$.

Análogamente, cada variable se identifica con el vector de las n observaciones realizadas sobre los n objetos. De este modo los datos de base se presentan en la forma de una matriz de n objetos (filas de la matriz) por p variables (columnas de la matriz). Lo anterior permite trabajar en el espacio euclídeo p -dimensional \mathbb{R}^p , en el cual los objetos ω_i son vectores en dicho espacio.

A cada objeto $\omega \in \Omega$ se le asigna un *peso* $p_\omega > 0$, el cual es un indicador de la importancia relativa del mismo dentro de la población Ω en estudio. Se tiene que $\sum_{\omega \in \Omega} p_\omega = 1$. Por otra parte, para cada subconjunto A de Ω , se define el

*Profesor de la Escuela de Matemáticas de la Universidad de Costa Rica.

peso de A , $p(A)$, mediante $p(A) = \sum_{\omega \in A} p_{\omega}$. Los centros de gravedad de A y de Ω se denotan respectivamente por g_A y g , y se definen mediante

$$g_A = \sum_{\omega \in A} \frac{p_{\omega}}{p(A)} \cdot \omega \quad ; \quad g = \sum_{\omega \in \Omega} p_{\omega} \cdot \omega.$$

El objetivo central de la Clasificación Automática es determinar una *partición* de Ω , de modo tal que sus clases sean tan diferenciadas y homogéneas como sea posible. Dichas cualidades serán definidas en términos de índices que expresen el grado de proximidad entre objetos y entre clases. Para este efecto supondremos definido un producto interno $\langle x, y \rangle$ en \mathbf{R}^p , y para medir la proximidad entre objetos emplearemos la norma euclídea correspondiente: $\|x\|$. La proximidad entre clases se puede definir de diversas formas. Una de ellas es la agregación de Ward, de la cual nos vamos a ocupar en la sección 3.

Las siguientes definiciones resultan esenciales en el desarrollo posterior.

1.1 Inercia

Sea A un subconjunto no vacío de Ω . La *inercia* de A con respecto a un punto v del espacio euclídeo \mathbf{R}^p se denota por $I_v(A)$, y se define mediante

$$I_v(A) = \sum_{\omega \in A} p_{\omega} \cdot \|\omega - v\|^2.$$

$I_v(A)$ es un índice que representa una medida de la dispersión de los objetos de A alrededor del "objeto hipotético" v . Es conocido que la mínima dispersión se alcanza cuando v es el centro de gravedad g_A de A (Propiedad 5). En tal caso ($v = g_A$), escribimos $I(A)$ en vez de $I_{g_A}(A)$. En particular, la *inercia total* de Ω es

$$I = I(\Omega) = \sum_{\omega \in \Omega} p_{\omega} \cdot \|\omega - g\|^2.$$

1.2 Inercia según una partición

Sea $P = (\Omega_1, \dots, \Omega_r)$ una partición de Ω en r clases disjuntas. Se definen la *inercia interclases* y la *inercia intraclases*, de la siguiente forma :

- **Inercia interclases:** la inercia interclases de Ω respecto de la partición P , denotada por $\text{Inter}(P)$, se define por la fórmula

$$\text{Inter}(P) = \sum_{i=1}^r p(\Omega_i) \|g - g_i\|^2,$$

donde g_i es el centro de gravedad de Ω_i . Se observa que $\text{Inter}(P)$ representa una medida de la dispersión de los centros de Ω_i respecto del centro de Ω . Entre más grande sea $\text{Inter}(P)$, más diferenciadas son las clases de la partición.

- **Inercia intraclases:** la inercia intraclases de Ω respecto de la partición P , denotada por $\text{Intra}(P)$, se define mediante la fórmula

$$\text{Intra}(P) = \sum_{i=1}^r I(\Omega_i).$$

La inercia intraclases es una medida global de la dispersión interna de las clases. Entre más pequeña sea $\text{Intra}(P)$, más homogéneas serán las clases de la partición.

2 Descomposición de la inercia total

Existe una relación entre la inercia total y las inercias interclases e intraclases, análoga a la relación existente en análisis de la varianza. En efecto, se demuestra que la inercia total es la suma de las inercias interclases e intraclases (parte (a) de la Propiedad 1).

De modo que $\text{Intra}(P)$ es mínimo si y sólo si $\text{Inter}(P)$ es máximo. Así, cuando las condiciones de homogeneidad y diferenciación de las clases se asocian a las magnitudes $\text{Intra}(P)$ e $\text{Inter}(P)$ respectivamente, el problema consistente en encontrar una partición P de Ω , cuyas clases sean simultáneamente bien homogéneas y diferenciadas, equivale a minimizar $\text{Intra}(P)$ o maximizar $\text{Inter}(P)$.

Propiedad 1 Sean $\Omega_1, \dots, \Omega_r$ un conjunto de partes de Ω , no vacías, y disjuntas dos a dos.

- La inercia de $A = \bigcup_{i=1}^r \Omega_i$ con respecto a $v \in \mathbb{R}^p$, se expresa como la suma de dos componentes: $I_v(A) = \sum_{i=1}^r I(\Omega_i) + \sum_{i=1}^r p(\Omega_i) \|v - g_i\|^2$, donde g_i es el centro de la clase Ω_i .
- La descomposición de la inercia total de Ω , según una partición P , está dada por la fórmula: $I = \text{Intra}(P) + \text{Inter}(P)$.

Demostración:

- Por definición tenemos:

$$I_v(A) = \sum_{\omega \in A} p_\omega \|\omega - v\|^2 = \sum_{i=1}^r \sum_{\omega \in \Omega_i} p_\omega \|\omega - v\|^2.$$

Además,

$$p_\omega \|\omega - v\|^2 = p_\omega \|\omega - g_i\|^2 + p_\omega \|g_i - v\|^2 + 2p_\omega \langle \omega - g_i, g_i - v \rangle.$$

Por otra parte,

$$\begin{aligned} \sum_{i=1}^r \sum_{\omega \in \Omega_i} p_{\omega} \langle \omega - g_i, g_i - v \rangle &= \sum_{i=1}^r \langle \sum_{\omega \in \Omega_i} p_{\omega} (\omega - g_i), g_i - v \rangle \\ &= \sum_{i=1}^r \langle \sum_{\omega \in \Omega_i} p_{\omega} \omega - p(\Omega_i) g_i, g_i - v \rangle \\ &= \sum_{i=1}^r \langle 0, g_i - v \rangle = 0. \end{aligned}$$

Se deduce entonces que

$$I_v(A) = \sum_{i=1}^r \sum_{\omega \in \Omega_i} p_{\omega} \|\omega - g_i\|^2 + \sum_{i=1}^r \sum_{\omega \in \Omega_i} p_{\omega} \|g_i - v\|^2,$$

es decir,

$$\begin{aligned} I_v(A) &= \sum_{i=1}^r I(\Omega_i) + \sum_{i=1}^r \|g_i - v\|^2 \sum_{\omega \in \Omega_i} p_{\omega} \\ &= \sum_{i=1}^r I(\Omega_i) + \sum_{i=1}^r p(\Omega_i) \|g_i - v\|^2. \end{aligned}$$

- (b) La fórmula se obtiene poniendo $v = g$ en la parte (a) de esta proposición, donde se supone que $P = (\Omega_1, \dots, \Omega_r)$. \square

3 Agregación de Ward

Veamos en esta sección los resultados que nos permiten deducir el significado de la agregación de Ward.

Propiedad 2 Sean A y B dos partes disjuntas y no vacías de Ω . Entonces:

$$(a) g_{A \cup B} = \frac{p(A) g_A}{p(A) + p(B)} + \frac{p(B) g_B}{p(A) + p(B)}.$$

$$(b) I(A \cup B) = I(A) + I(B) + \frac{p(A) p(B) \|g_A - g_B\|^2}{p(A) + p(B)}.$$

Demostración:

(a) Por definición se cumple que

$$\begin{aligned} g_{A \cup B} &= \sum_{\omega \in A \cup B} \frac{p_{\omega} \cdot \omega}{p(A \cup B)} = \sum_{\omega \in A} \frac{p_{\omega} \cdot \omega}{p(A) + p(B)} + \sum_{\omega \in B} \frac{p_{\omega} \cdot \omega}{p(A) + p(B)} \\ &= \frac{p(A)}{p(A) + p(B)} \sum_{\omega \in A} \frac{p_{\omega}}{p(A)} \omega + \frac{p(B)}{p(A) + p(B)} \sum_{\omega \in B} \frac{p_{\omega}}{p(B)} \omega \\ &= \frac{p(A)}{p(A) + p(B)} g_A + \frac{p(B)}{p(A) + p(B)} g_B. \end{aligned}$$

(b) Usando la Propiedad 1, se puede escribir:

$$I(A \cup B) = I(A) + I(B) + p(A) \|g_{A \cup B} - g_A\|^2 + p(B) \|g_{A \cup B} - g_B\|^2.$$

Sustituyendo $g_{A \cup B}$ en esta identidad por su equivalente deducido en la parte (a) que recién probamos, y haciendo un poco de cálculo, se obtiene el resultado requerido. \square

Propiedad 3 Sea $P_1 = (\Omega_1, \dots, \Omega_k)$ una partición de Ω . Se define una nueva partición P_2 de Ω , a partir de P_1 , uniendo dos clases, por ejemplo Ω_t y Ω_q . La partición P_2 es $P_2 = \{\Omega_t \cup \Omega_q\} \cup \{\Omega_i \mid i \neq t, q\}$. Entonces se deducen las siguientes identidades:

$$(a) \text{ Intra}(P_2) = \text{Intra}(P_1) + \frac{p(\Omega_t)p(\Omega_q)}{p(\Omega_t) + p(\Omega_q)} \|g_t - g_q\|^2.$$

$$(b) \text{ Inter}(P_2) = \text{Inter}(P_1) - \frac{p(\Omega_t)p(\Omega_q)}{p(\Omega_t) + p(\Omega_q)} \|g_t - g_q\|^2.$$

donde g_t y g_q son los centros de Ω_t y Ω_q respectivamente.

Demostración: Usando la Propiedad 1, parte (b), las inercias totales, según P_1 y P_2 , se escriben respectivamente como:

$$P_1 : \quad I = \text{Intra}(P_1) + \text{Inter}(P_1)$$

$$P_2 : \quad I = \sum_{i \neq q, t} I(\Omega_i) + I(\Omega_q \cup \Omega_t) + \text{Inter}(P_2)$$

Sustituyendo $I(\Omega_q \cup \Omega_t)$ por su equivalente de acuerdo con la Propiedad 2, se obtienen las fórmulas deseadas. \square

3.1 Agregación de Ward

La agregación δ de Ward se define por la fórmula

$$\delta(A, B) = \frac{p(A)p(B)}{p(A) + p(B)} \|g_A - g_B\|^2,$$

donde A y B son partes de Ω no vacías y disjuntas. La Propiedad 3 establece que el valor de la agregación de Ward es el incremento (decremento) de la inercia intraclases (interclases), cuando se pasa de P_1 a P_2 . Así, es lógico que esta agregación se use en Clasificación Aglomerativa Ascendente para construir una secuencia de particiones, uniendo en cada paso las dos clases más próximas, en el sentido de la agregación δ .

4 Método de las nubes dinámicas (MND)

El método de las nubes dinámicas (MND) es un proceso de búsqueda de una clasificación de Ω , que se desarrolla por iteraciones, construyendo cada vez una nueva partición de mejor calidad. Para formular este proceso vamos a definir previamente la manera como se crean las nuevas particiones y como se representan; asimismo debemos precisar un criterio de calidad para medir la adecuación entre una partición y su correspondiente representación. Luego discutiremos la convergencia del proceso iterativo.

En la presente sección no es necesario asumir todo el contexto descrito en la sección 1. Nos limitaremos únicamente a admitir la existencia de...

- (a) Un conjunto \mathcal{N} , llamado conjunto de *núcleos admisibles*.
- (b) Una medida de la disimilitud entre elementos de Ω y núcleos admisibles. Esto es: una función d , de valores no negativos, tal que $d: \Omega \times \mathcal{N} \rightarrow \mathbf{R}$, y de modo que, identificando los conjuntos $\Omega \times \mathcal{N}$ y $\mathcal{N} \times \Omega$, se cumpla

$$d(\omega, \eta) = d(\eta, \omega) \quad ; \quad \forall (\omega, \eta) \in \Omega \times \mathcal{N}.$$

Bajo las hipótesis anteriores es posible definir la disimilitud D , entre partes no vacías de Ω y núcleos admisibles:

$$D(A, \eta) = \sum_{\omega \in A} d(\omega, \eta).$$

4.1 Fases de representación y asignación

Cada iteración en el MND consta de dos fases que vamos a describir por separado a continuación.

Fase de representación

El objetivo de esta fase es definir la forma como se va a representar una partición. Primeramente debemos formular la siguiente hipótesis:

Hipótesis (unicidad de la representación): Para cada parte no vacía A de Ω existe exactamente un núcleo $\eta_A \in \mathcal{N}$ (representación de A) tal que

$$D(A, \eta_A) = \min\{D(A, \eta) \mid \eta \in \mathcal{N}\}.$$

La adecuación entre la partición $P = (\Omega_1, \dots, \Omega_r)$ y un sistema de núcleos admisibles $\Xi = (\eta_1, \dots, \eta_r) \in \mathcal{N}^r$ —también llamada *criterio de calidad*—, se define mediante

$$C(P, \Xi) = \sum_{i=1}^r D(\Omega_i, \eta_i) = \sum_{i=1}^r \sum_{\omega \in \Omega_i} d(\omega, \eta_i).$$

En virtud de la existencia y unicidad de la representación de una parte no vacía A de Ω , tenemos que para cualquier partición $P = (\Omega_1, \dots, \Omega_r)$ de Ω , existe una *representación única* $\Xi = (\eta_1, \dots, \eta_r) \in \mathcal{N}^r$, de modo tal que

$$C(P, \Xi) = \min\{C(P, S) : S \in \mathcal{N}^r\},$$

donde η_i es la representación de Ω_i , $i = 1, \dots, r$.

Fase de asignación

En esta fase, como su nombre lo indica, se construye una partición de Ω , asignando cada objeto al núcleo más próximo. Para cada sistema de r núcleos $L = (\eta_1, \dots, \eta_r) \in \mathcal{N}^r$, se definen las r clases correspondientes $\Omega_1, \dots, \Omega_r$ de la siguiente manera:

$$\Omega_i = \{\omega \in \Omega \mid d(\omega, \eta_i) \leq d(\omega, \eta_j)\} \quad ; \quad j = 1, \dots, r.$$

Si para $i < s$ ocurre que

$$d(\omega, \eta_i) = d(\omega, \eta_s) \leq d(\omega, \eta_j) \quad ; \quad \forall j$$

entonces el individuo ω se asigna a Ω_i . Es claro que $P = (\Omega_1, \dots, \Omega_r)$ es una partición de Ω . En la formación de las nuevas clases, puede ocurrir que algunas de éstas sean vacías, en cuyo caso el número de clases disminuye.

4.2 Descripción del método

El proceso algorítmico que vamos a describir enseguida se puede representar por medio del esquema siguiente:

$$\begin{array}{ccccccc} \Xi_0 & \longrightarrow & \underbrace{P_0 \rightarrow \Xi_1}_{\text{iteración 1}} & \longrightarrow & \underbrace{P_1 \rightarrow \Xi_2}_{\text{iteración 2}} & \longrightarrow \dots \longrightarrow & \underbrace{P_{n-1} \rightarrow \Xi_n}_{\text{iteración } n} \longrightarrow \dots \\ \text{inicio} & & & & & & \end{array}$$

Por iteraciones el proceso se describe como sigue: sea

$$\Xi_0 = (\eta_0(1), \dots, \eta_0(k)) \in \mathcal{N}^k$$

un sistema de k núcleos escogido al azar o por algún otro método.

Iteración 1:

A partir de Ξ_0 y por medio de la fase de asignación, se define una partición

$$P_0 = (\Omega_0(1), \dots, \Omega_0(k_1)),$$

donde $k_1 \leq k$. Luego, de la fase de representación resulta que

$$\Xi_1 = (\eta_1(1), \dots, \eta_1(k_1)) \in \mathcal{N}^{k_1}$$

es la representación de P_0 . Esta iteración da como resultado la creación del par (P_0, Ξ_1) . El valor del criterio de calidad es

$$C(P_0, \Xi_1) = \sum_{i=1}^{k_1} D(\Omega_0(i), \eta_1(i)) = \sum_{i=1}^{k_1} \sum_{\omega \in \Omega_0(i)} d(\omega, \eta_1(i)).$$

Si la condición inicial es una partición (en vez de un conjunto de núcleos), entonces se define Ξ_0 por medio de la fase de representación, luego el algoritmo procede como si Ξ_0 fuese la condición inicial.

Iteración n :

Sea (P_{n-2}, Ξ_{n-1}) el par construido por el algoritmo en la iteración $n - 1$. Partiendo de Ξ_{n-1} se define P_{n-1} por medio de la fase de asignación:

$$P_{n-1} = (\Omega_{n-1}(1), \dots, \Omega_{n-1}(k_n)),$$

donde $k_n \leq \dots \leq k_1 \leq k$. Luego, por medio de la fase de representación se determina Ξ_n :

$$\Xi_n = (\eta_n(1), \dots, \eta_n(k_n)) \in \mathcal{N}^{k_n}.$$

El valor del criterio de calidad en el par (P_{n-1}, Ξ_n) es

$$C(P_{n-1}, \Xi_n) = \sum_{i=1}^{k_n} D(\Omega_{n-1}(i), \eta_n(i)) = \sum_{i=1}^{k_n} \sum_{\omega \in \Omega_{n-1}(i)} d(\omega, \eta_n(i)).$$

4.3 Convergencia del método

La convergencia del MND comprende la estabilización del criterio de calidad $C(P_{n-1}, \Xi_n)$, y de las sucesiones $\{\Xi_n\}_n$ y $\{P_n\}_n$, creadas por el algoritmo.

Propiedad 4 (Convergencia del método)

(a) El criterio de calidad se estabiliza decreciendo: esto significa que existe un entero positivo m , mínimo, tal que:

$$C(P_0, \Xi_1) \geq C(P_1, \Xi_2) \geq \dots \geq C(P_{m-1}, \Xi_m) = C(P_m, \Xi_{m+1}) = \dots$$

(b) Los núcleos y particiones se estabilizan: Esto es, $\forall j \in \{0, 1, \dots\}$, se tiene

$$P_{m+j} = P_{m+j+1} \quad y \quad \Xi_{m+j} = \Xi_{m+j+1}.$$

Demostración:

(a) Para probar que

$$C(P_{n-1}, \Xi_n) \geq C(P_n, \Xi_{n+1}) \quad ; \quad \forall n \in \{1, 2, \dots\}$$

se utiliza la cantidad $C(P_n, \Xi_n) = \sum_{i=1}^{k_{n+1}} D(\Omega_n(i), \eta_n(i))$ como intermediario. Se probará la doble desigualdad

$$C(P_{n-1}, \Xi_n) \geq C(P_n, \Xi_n) \geq C(P_n, \Xi_{n+1}) \quad ; \quad \forall n \in \{1, 2, \dots\}.$$

En efecto, es claro que $\forall i = 1, \dots, k_{n+1}$ se cumple que

$$D(\Omega_n(i), \eta_{n+1}(i)) = \min\{D(\Omega_n(i), \eta) \mid \eta \in \mathcal{N}\} \leq D(\Omega_n(i), \eta_n(i)),$$

de donde $C(P_n, \Xi_{n+1}) \leq C(P_n, \Xi_n)$. Por otra parte, para cada $\omega \in \Omega$, existen enteros h y s tales que $\omega \in \Omega_n(h)$ y $\omega \in \Omega_{n-1}(s)$. Por la definición de $\Omega_n(h)$ se tiene entonces que

$$d(\omega, \eta_n(h)) \leq d(\omega, \eta_n(s)).$$

Es decir, el sumando $d(\omega, \eta_n(h))$ de $C(P_n, \Xi_n)$ no es mayor que el sumando correspondiente $d(\omega, \eta_n(s))$ de $C(P_{n-1}, \Xi_n)$. De aquí se deduce que

$$\begin{aligned} C(P_{n-1}, \Xi_n) &= \sum_{i=1}^{k_n} \sum_{\omega \in \Omega_{n-1}(i)} d(\omega, \eta_n(i)) \\ &\geq \sum_{j=1}^{k_{n+1}} \sum_{\omega \in \Omega_n(j)} d(\omega, \eta_n(j)) = C(P_n, \Xi_n). \end{aligned}$$

Ahora, como Ω es finito, también lo es la sucesión $\{(P_n, \Xi_{n+1})\}_n$. Luego la sucesión $\{C(P_n, \Xi_{n+1})\}_n$ es finita y además es creciente. Por tanto existe un entero positivo m , mínimo, tal que:

$$C(P_{m-1+j}, \Xi_{m+j}) = C(P_{m+j}, \Xi_{m+j+1}) \quad ; \quad \forall j = 0, 1, \dots$$

(b) Utilizando la doble desigualdad arriba demostrada podemos concluir que

$$C(P_{m+j}, \Xi_{m+j}) = C(P_{m+j}, \Xi_{m+j+1}),$$

lo cual, junto con las desigualdades

$$D(\Omega_{m+j}(i), \eta_{m+j+1}(i)) \leq D(\Omega_{m+j}(i), \eta_{m+j}(i)),$$

válidas para $i = 1, \dots, k_{m+j+1}$, justifican la validez de las siguientes igualdades, para $i = 1, \dots, k_{m+j+1}$:

$$D(\Omega_{m+j}(i), \eta_{m+j+1}(i)) = D(\Omega_{m+j}(i), \eta_{m+j}(i)),$$

en donde $\eta_{m+j+1}(i)$ y $\eta_{m+j}(i)$ son representaciones de $\Omega_{m+j}(i)$, $i = 1, \dots, k_{m+j+1}$. En virtud de la unicidad de la representación debe cumplirse que $\eta_{m+j+1}(i) = \eta_{m+j}(i)$, para $i = 1, \dots, k_{m+j+1}$. De lo anterior se sigue, para $i = 1, \dots, k_{m+j+1}$, que $\Omega_{m+j}(i) \subseteq \Omega_{m+j+1}(i)$. Consecuentemente se concluye, puesto que P_{m+1} y P_{m+i+1} son particiones de Ω ,

$$\Omega_{m+j}(i) = \Omega_{m+j+1}(i) \quad ; \quad \forall i = 1, \dots, k_{m+j+1},$$

es decir, $P_{m+j} = P_{m+j+1}$, para todo valor de j . \square

Un caso particular del MND es el *algoritmo de los centros móviles*, el cual corresponde a escoger \mathbf{R}^p como conjunto de núcleos admisibles y la distancia euclídea usual como medida de disimilitud entre elementos de Ω y núcleos admisibles. Así, se llega a que la disimilitud $D(A, \eta)$, entre una parte A de Ω y un núcleo admisible $\eta \in \mathbf{R}^p$, resulta ser la inercia de A con respecto a η . Por lo tanto la representación de la parte A de Ω , es un núcleo η_A tal que

$$D(A, \eta_A) = I_{\eta_A}(A) = \min\{I_{\eta}(A) \mid \eta \in \mathbf{R}^p\}.$$

El siguiente resultado conocido como *teorema de Huyghens* garantiza que la representación de A es justamente el centro de gravedad de A .

Propiedad 5 Sea A una parte no vacía de Ω , y sea $\eta \in \mathbf{R}^p$. Entonces:

$$(a) \quad I_{\eta}(A) = I(A) + p(A) \|\eta - g_A\|^2.$$

$$(b) \quad D(A, g_A) = \min\{I_{\xi}(A) \mid \xi \in \mathbf{R}^p\}.$$

Demostración:

(a) Se trata de un caso particular de la Propiedad 1.

(b) El resultado es una consecuencia de la parte (a). En efecto:

$$I_{\xi}(A) = I(A) + p(A) \|\xi - g_A\|^2 \geq I(A) = D(A, g_A); \quad \forall \xi \in \mathbf{R}^p.$$

Además g_A es la única representación de A , puesto que $I_{\xi}(A) = I(A)$ si y solo si $\xi = g_A$. \square

Se concluye entonces que el criterio de calidad $C(P, \Xi)$ —en el caso del algoritmo de los centros móviles— se reduce a la inercia intraclases de Ξ respecto a la partición P . Así, tanto este algoritmo como el método de Ward optimizan el mismo criterio de calidad.

Es debido a lo anterior que cuando se busca una clasificación de un conjunto de individuos descritos por variables continuas, tiene sentido enlazar el método de Ward (construcción ascendente jerárquica) y el algoritmo de los centros móviles, para encontrar una partición de buena calidad.

Referencias

- [1] Cailliez, F. y Pages, J. P.: *Introduction a l'analyse des données*. SMASH, Paris; 1976.
- [2] Diday, Edwin; Pouget, Jean; Lemaire, Jacques; Testu, Françoise: *Eléments d'analyse de données*. Bordas, Paris; 1982.
- [3] Govaert, Gérard y Lechevallier, Yves.: *Classification automatique non hiérarchique*. Conferenciers, Ecole d'été; 1987.

Análisis multivariado de conceptos sociológicos en la educación costarricense

Javier Trejos*

Resumen

Se estudian las actitudes de los estudiantes costarricenses de secundaria respecto a ciertos conceptos sociológicos escogidos, tales como *empresa privada-libertad*, y *democracia-ejército*.

Se analizan las relaciones entre estas variables. La metodología empleada es el análisis multivariado de datos, especialmente el *Análisis de Correspondencias Múltiples* (ACM), el *Análisis de Correspondencias* y la *Clasificación Jerárquica*.

1 El cuestionario

Se redactó un cuestionario con dos tipos de preguntas. Por un lado, lo que llamaremos variables socioeconómicas, y por otro, la opinión de los estudiantes acerca de una serie de aspectos relacionados con la libertad, la democracia, la paz, etc.

Pese a que el presente estudio tiene por objetivo el análisis de las relaciones entre algunas variables de opinión, como se señala más abajo, para efectos del tratamiento previo al análisis multivariado (sección 2) se tomaron en cuenta también algunas variables socioeconómicas. Ellas son:

- *sexo*: sexo del estudiante.
- *cole*: tipo de colegio.
- *ingr*: ingreso familiar.
- *reli*: religión del estudiante.
- *poli*: partido político por el que simpatiza.
- *edpa*: grado de educación del padre.

*Profesor de la Escuela de Matemáticas de la Universidad de Costa Rica.

- *edma*: grado de educación de la madre.

En cuanto al fenómeno de la opinión, cada tema tiene asociadas cinco preguntas: una principal, dos características y dos de control. De los doce temas del cuestionario escogimos dos: *libertad* y *democracia*. En primer lugar se estudian las relaciones entre *libertad* y *empresa privada*. Para ello se analizaron las siguientes preguntas:

- *var1*: la empresa privada es necesaria para que exista libertad.
- *var2*: se puede entrar en un sindicato sin ser perseguido o despedido.
- *var3*: si se trabaja en la empresa privada no hay tiempo para ocuparse de sus propias cosas.
- *var4*: los trabajadores deben respetar las órdenes de su patrono sin criticarlas.
- *var5*: mi padre debe dedicarle más tiempo a la empresa o institución donde trabaja.

Y para analizar las relaciones entre *democracia* y *ejército*:

- *var6*: un país democrático no necesita ejército.
- *var7*: un país con ejército es un país totalitario.
- *var8*: es necesario que Costa Rica tenga ejército para defenderse de amenazas externas.
- *var9*: la existencia de un ejército en Costa Rica podría llevar a una dictadura militar en corto plazo.
- *var10*: Estados Unidos no es un país democrático porque tiene ejército.

Cada una de estas preguntas tiene cinco modalidades de respuesta:

- *fde*: fuertemente en desacuerdo.
- *des*: en desacuerdo.
- *ind*: indeciso.
- *dac*: de acuerdo.
- *fac*: fuertemente de acuerdo.

El cuestionario fue pasado en dos colegios del área metropolitana, uno público y otro privado. Se entrevistaron un total de 232 estudiantes de cuarto y quinto año, durante el primer semestre de 1987.

2 Tratamiento previo de los datos

Entre las variables socioeconómicas, las preguntas relativas al ingreso familiar (*ingr*) y al partido político de sus simpatías (*poli*) fueron las que presentaron el mayor número de respuestas faltantes.

Partiendo de la hipótesis que la no respuesta refleja principalmente la actitud estudiantil, más o menos consciente, de no dar la información, hemos conservado la no respuesta como una modalidad de las variables *ingr* y *poli*. Sin embargo cabe observar que en el caso del ingreso familiar (*ingr*), la no respuesta podría obedecer además a que los estudiantes estiman que no tienen una información precisa acerca del ingreso familiar.

Para los otros casos se recurre a una asignación al azar de las no respuestas, entre el resto de las modalidades (Morineau, [2]). Las variables involucradas y las asignaciones realizadas son: *edpa*, 12 asignaciones; *edma*, 3 asignaciones; *var3*, *var5* y *var7*, una asignación cada una.

Finalmente las modalidades con muy poco efectivo fueron reunidas con otras. En todo caso se fusionaron modalidades similares o cercanas, obteniéndose los resultados siguientes:

- *edpa*: ninguna educación y educación primaria (1+33).
- *poli*: partido Vanguardia Popular, partido del Pueblo Costarricense y otros (3+1+5).
- *var6*: *des6* y *fde6* (31+9).
- *var2*: *dac2* y *fac2* (19+5).
- *var3*: *fde3* y *des3* (1+9).
- *var4*: *fde4* y *des4* (7+51).
- *var5*: *fde5* y *des5* (12+29).
- *var9*: *des9* y *fde9* (9+25).

De esta manera, las tablas de frecuencias para las variables socioeconómicas y las variables de opinión se presentan en las Tablas 1 y 2 respectivamente.

3 Análisis de las relaciones entre las variables de opinión

Primero se hizo un análisis global sobre las diez variables de opinión que abarcan los temas *empresa privada-libertad* y *democracia-ejército*. La metodología utilizada es el *Análisis de Correspondencias Múltiples* (ACM), que nos permite obtener representaciones planas de las distintas modalidades de las variables (ver

Variable	Modalidades	código	frec.	%
Tipo de colegio	público	publ	124	53
	privado	priv	108	47
Sexo	masculino	masc	101	44
	femenino	feme	131	56
Nivel de ingreso de la familia (en miles de colones)	$I < 7$	I1	24	10
	$7 \leq I < 14$	I2	44	19
	$14 \leq I < 21$	I3	35	15
	$21 \leq I < 28$	I4	30	13
	$28 \leq I < 35$	I5	34	15
	$35 \leq I$	I6	47	20
	no responde	I7	18	8
Grado de educación del padre	primaria incompleta	P1	35	15
	primaria completa	P2	30	13
	secundaria incompleta	P3	34	15
	secundaria completa	P4	35	16
	universit. incompleta	P5	29	12
	universit. completa	P6	68	29
Grado de educación de la madre	primaria incompleta	M1	37	16
	primaria completa	M2	41	18
	secundaria incompleta	M3	43	18
	secundaria completa	M4	41	17
	universit. incompleta	M5	18	8
	universit. completa	M6	52	23
Religión	católica	cato	127	55
	evangélica	evan	75	32
	no creyente	nocr	8	3
	otra	otra	22	10
Partido político	PUSC	pusc	82	35
	PLN	pln	129	56
	otro	otro	8	4
	no responde	nore	12	5

Tabla 1: Frecuencias para las variables socioeconómicas.

Cód.	Variable	Mod.	frec.	%
var1	La empresa privada es necesaria para que exista libertad	fde1	60	26
		des1	74	32
		ind1	39	17
		dac1	32	14
		fac1	27	11
var2	Se puede entrar en un sindicato sin ser perseguido o despedido	fde2	60	26
		des2	119	51
		ind2	29	13
		dac2	24	10
var3	Si se trabaja en la empresa privada no hay tiempo para ocuparse de sus propias cosas	des3	10	4
		ind3	36	16
		dac3	145	62
		fac3	41	18
var4	Los trabajadores deben respetar las órdenes de su patrono sin criticarlas	des4	58	25
		ind4	22	9
		dac4	123	53
		fac4	29	13
var5	Mi padre debe dedicarle más tiempo a la empresa o institución donde trabaja	des5	32	18
		ind5	93	40
		dac5	60	26
		fac5	37	16
var6	Un país democrático no necesita ejército	des6	40	17
		ind6	24	11
		dac6	56	24
		fac6	112	48
var7	Un país con ejército es un país totalitario	fde7	19	8
		des7	45	19
		ind7	106	46
		dac7	44	19
		fac7	18	8
var8	Es necesario que Costa Rica tenga ejército para defenderse de amenazas externas	fde8	15	6
		des8	23	10
		ind8	36	16
		dac8	41	18
		fac8	117	50
var9	La existencia de un ejército en Costa Rica podría llevar a una dictadura...	des9	97	41
		ind9	67	29
		dac9	34	15
		fac9	34	15
var10	Estados Unidos no es un país democrático porque tiene ejército	fde0	23	10
		des0	41	18
		ind0	118	51
		dac0	50	21

Tabla 2: Frecuencias de las variables de opinión.

Figura 1). Las relaciones encontradas como resultado de este primer análisis son corroboradas mediante el *Análisis de Correspondencias Simples* (sólo se analizan simultáneamente dos variables) y la *Clasificación Ascendente Jerárquica*.

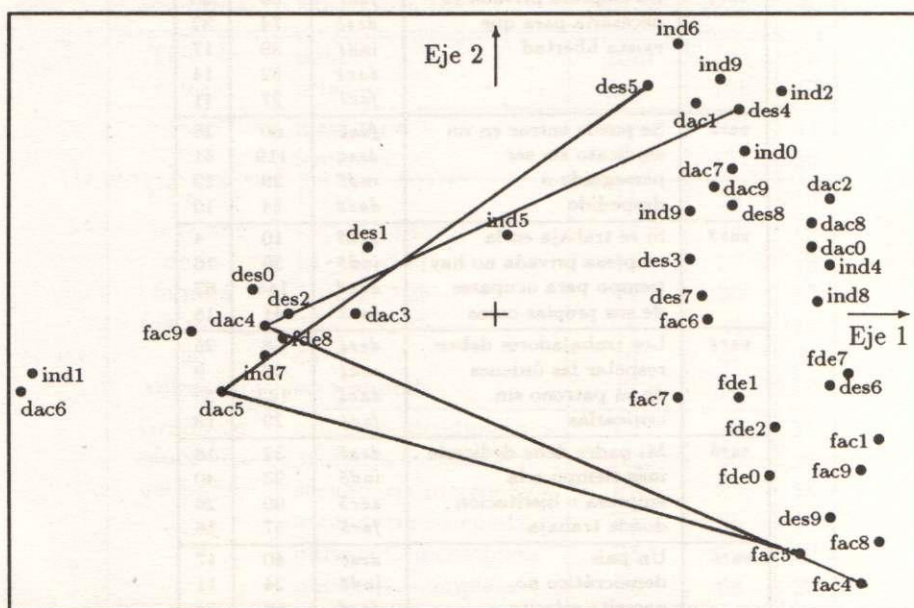


Figura 1: ACM de las diez variables de opinión: plano principal (25% de inercia).

3.1 ACM de las 10 variables de opinión

En la Figura 1 se muestra el plano principal del ACM de las variables *var1* a *var10*, con un 25% de inercia explicada. En él sobresalen los hechos siguientes:

- Se muestra una relación directa entre las variables 4 y 5, excepto por la modalidad "indecisos". Es decir, la opinión de los entrevistados respecto de la proposición "los trabajadores deben respetar las órdenes de los patronos" (*var4*), es básicamente la misma que tienen respecto de "mi padre debe dedicarle más tiempo a la empresa" (*var5*), excepto por los indecisos. Dicha relación se indica en la Figura 1 por medio de trayectorias siguiendo el orden de las modalidades de una misma variable.
- Sobre este plano también se observa que la opinión favorable (*fac1*) respecto de la proposición "la empresa privada es necesaria para que exista libertad", está asociada con una opinión desfavorable (*des6*) respecto de la proposición "un país democrático no necesita ejército".

El plano definido por los ejes 1 y 3 no fue reportado en este trabajo. Sin embargo este plano, con 23% de inercia explicada, confirma la fuerte relación entre *var1* y *var6*, que corresponde a dos tipos de preguntas diferentes. La correspondencia entre modalidades es: *fac1-des6*, *dac1-ind6*, *ind1-dac6*, *fde1-fac6*. Se observa además una trayectoria ordenada de acuerdo a las modalidades. Sin embargo, no se confirma la relación entre las variables 8 y 9. Sólo se observa una cercanía entre *fac9* y *fde8*. La modalidad *dac8* parece estar mal ubicada, tanto aquí como en el plano principal.

3.2 ACM de las preguntas “empresa privada-libertad”

Al realizar el ACM entre las modalidades de las variables 1 a 5 se obtiene como plano principal el mostrado en la Figura 2. En ésta se ve claramente la fuerte relación directa que hay entre *var4* y *var5*, observándose también la trayectoria ordenada de la que antes hablábamos. En el plano generado por el primer y tercer ejes principales se confirma esta relación (sin observar la trayectoria), y como aquí, no se ve ninguna otra relación importante.

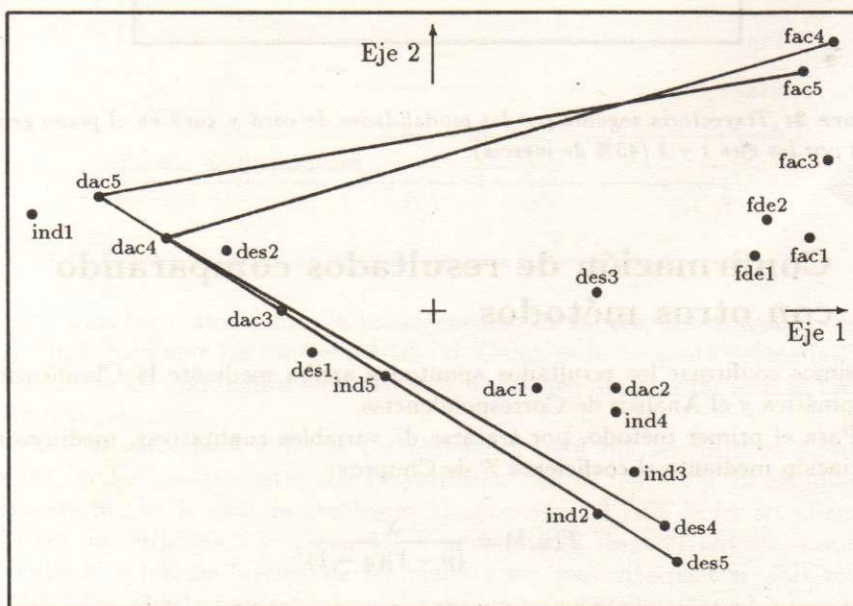


Figura 2: ACM de los temas “empresa privada-libertad”: plano principal (49% de inercia explicada).

3.3 ACM de las preguntas “democracia–ejército”

Al hacer el análisis entre las variables *var6* a *var10*, se confirma el agrupamiento de algunas modalidades de *var8* y *var9*, pero en cierto desorden, y no es sino en el gráfico que cruza los ejes principales 1 y 3 en que se puede ver una cierta trayectoria, con *dac8* mejor ubicado para que ésta tenga un sentido, como se ilustra en la Figura 3.

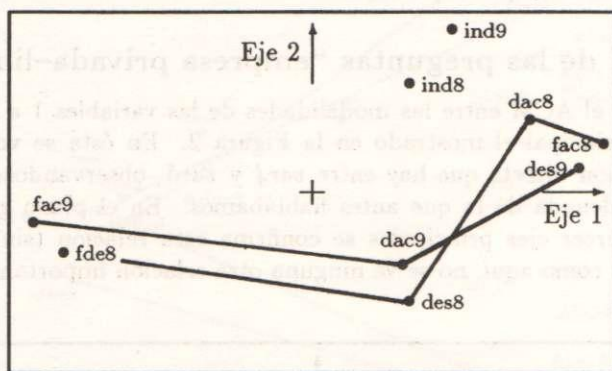


Figura 3: Trayectoria seguida por las modalidades de *var8* y *var9* en el plano generado por los ejes 1 y 3 (43% de inercia).

4 Confirmación de resultados comparando con otros métodos

Quisimos confirmar los resultados apuntados arriba mediante la Clasificación Automática y el Análisis de Correspondencias.

Para el primer método, por tratarse de variables cualitativas, medimos su asociación mediante el coeficiente *T* de Chuprov:

$$T(a, b) = \frac{\chi^2}{(p-1)(q-1)},$$

donde *p* y *q* son la cantidad de modalidades de cada variable, y χ^2 el chi-cuadrado de contingencia. Empleamos la agregación del promedio de las distancias:

$$\delta(A, B) = \sum_{a \in A} \sum_{b \in B} \frac{d(a, b)}{\text{card}(A) \text{card}(B)},$$

donde $d(a, b) = 1 - T(a, b)$.

Aplicando el algoritmo de clasificación jerárquica ascendente [1], se obtiene el árbol de clasificación mostrado en la Figura 4. Véase que se confirma la fuerte relación entre *var1* y *var6* (índice de agregación igual a 0.48) por un lado, y entre *var4* y *var5* (índice de agregación igual a 0.49) por el otro. Sigue la agregación entre *var8* y *var9* con índice de agregación igual a 0.85, y el resto se encuentra por encima del valor 0.92. Con los índices de agregación del ligamen simple y del ligamen máximo se obtuvieron resultados similares.

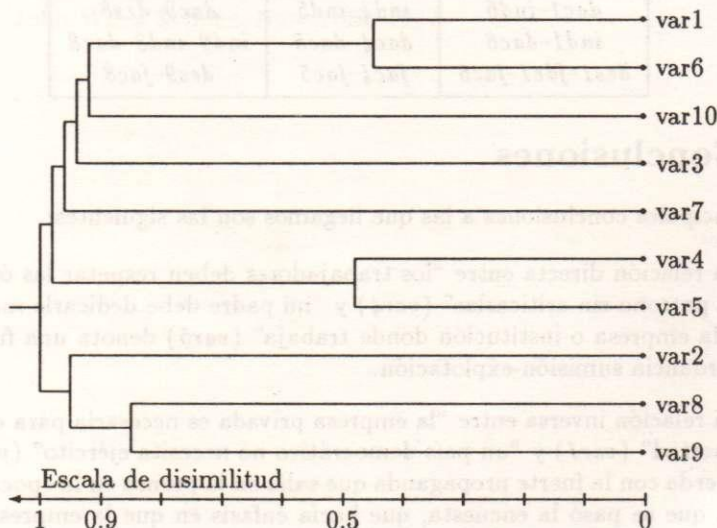


Figura 4: Árbol de clasificación para las variables de opinión.

Por otro lado, analizamos la homogeneidad de las respuestas dadas a estas preguntas, mediante las formas fuertes [1]. Como cada pregunta determina una partición de los estudiantes entrevistados, se pueden analizar las formas fuertes entre dos preguntas.

Para las variables 4 y 5, con 4 modalidades cada una, se obtienen 8 formas fuertes, de las cuales cuatro son mayoritarias con 60, 54, 41 y 29 estudiantes respectivamente, lo cual da una buena clasificación del 79% de los estudiantes.

Para las variables 1 y 6, con 5 y 4 modalidades respectivamente, también se obtienen 8 formas fuertes, de las cuales 4 son mayoritarias con efectivos de 60, 52, 34 y 27 estudiantes, lo cual resulta en una distribución coherente para el 75% de los estudiantes.

Lo anterior permite concluir que hay grupos bastante grandes con tendencias de opinión respecto a estas preguntas.

Por su parte, el estudio de las formas fuertes para las variables 8 y 9 no permite sacar ninguna información, pues hay 20 formas fuertes, de las que una sola, ($fac8 \cap des9$, con 82 elementos: el 35% de la población) es suficientemente

grande.

Al formar la tabla de contingencia entre las modalidades de las variables retenidas, esto es, *var1-var6*, *var4-var5*, *var8-var9*, se puede hacer el análisis de correspondencias sobre cada tabla. Así, pudimos confirmar las fuertes relaciones siguientes sobre el primer plano principal:

<i>var1 y var6</i>	<i>var4 y var5</i>	<i>var8 y var9</i>
<i>fac1-des6</i>	<i>des4-des5</i>	<i>fac9-fde8</i>
<i>dac1-ind6</i>	<i>ind4-ind5</i>	<i>dac9-des8</i>
<i>ind1-dac6</i>	<i>dac4-dac5</i>	<i>ind9-ind8-dac8</i>
<i>des1-fde1-fac6</i>	<i>fac4-fac5</i>	<i>des9-fac8</i>

5 Conclusiones

Las principales conclusiones a las que llegamos son las siguientes:

1. La relación directa entre “los trabajadores deben respetar las órdenes de su patrono sin criticarlas” (*var4*) y “mi padre debe dedicarle más tiempo a la empresa o institución donde trabaja” (*var5*) denota una fuerte concordancia sumisión-explotación.
2. La relación inversa entre “la empresa privada es necesaria para que exista libertad” (*var1*) y “un país democrático no necesita ejército” (*var6*) concuerda con la fuerte propaganda que salía en la prensa en la época anterior en que se pasó la encuesta, que hacía énfasis en que la empresa privada produce libertad, y que insistía a la vez en que Costa Rica debía armarse para preservar su democracia. Aquí se muestra la gran permeabilidad de los estudiantes de secundaria a la propaganda masiva e ideológica.
3. La relación inversa entre “Costa Rica necesita ejército para defenderse de amenazas externas” (*var8*) y “un ejército en Costa Rica podría llevar a una dictadura militar a corto plazo” (*var9*) parece muy lógica: quienes quieren la implantación de un ejército en Costa Rica hacen creer que en nuestro país no se correría el riesgo de una dictadura militar.
4. La ausencia de otras relaciones entre las otras preguntas puede ser debida a la falta de consistencia en la opinión de los estudiantes de educación media.
5. Respecto de la metodología estadística empleada, que dicho sea de paso tiene un carácter descriptivo por excelencia, cabe mencionar la importancia de los gráficos obtenidos a partir del ACM, en la medida que nos permitieron explorar, no sólo las asociaciones entre variables, sino también su carácter (directo o inverso), propiedad que no es posible observar mediante el clásico chi-cuadrado.

Referencias

- [1] Diday, E.; Lemaire, J.; Pouget, J. y Testu, F.: *Eléments d'Analyse de Données*. Editorial Dunod, Paris, 1982.
- [2] Morineau, A.: *L'Analyse des Correspondances*. Ecole d'Eté en Analyse de Données, CIMPA, Niza, 1987.
- [3] Lebart, L.; Morineau, A. y Warwick: *Multivariate Statistical Data Analysis*. John Wiley & Sons, New York, 1987.

Clasificación por Particiones

José Luis Espinoza

Javier Trejos*

Resumen

En el presente trabajo se presentan las principales técnicas para obtener, para un conjunto de individuos, una clasificación en una sola partición a la vez. Los métodos de *nubes dinámicas* se introducen de una manera casi exhaustiva en los casos más conocidos, dándole una presentación unificada. Asimismo se presenta el método de transferencias de Régnier. Luego, se proponen las técnicas actuales para el tratamiento de una multipartición que resulte de la aplicación sucesiva de alguno de los métodos anteriores. Finalmente, se presenta el algoritmo de Fisher, que da una clasificación optimal de un conjunto en el caso más simple.

1 Introducción

Los diferentes métodos de Clasificación Automática tratan de buscar particiones de un conjunto Ω sobre el que se han hecho una serie de observaciones. Los métodos más conocidos son los de *clasificación jerárquica*, que dan una serie de particiones encajadas en forma de árbol. Estos tienen la ventaja de poseer una fácil y rápida interpretación.

En contraste con los métodos jerárquicos, están los de *particionamiento*. En estos más bien se trata de obtener una sola partición de Ω , de tal forma que los conjuntos sean bastante homogéneos internamente, y suficientemente distintos entre sí. Estos algoritmos pueden ser particularmente útiles cuando Ω tiene cardinalidad muy grande y no es posible implementar los algoritmos jerárquicos. Los más conocidos son los métodos del tipo *nubes dinámicas* y los de *transferencias*.

En adelante llamaremos Ω el conjunto de n individuos a particionar, sobre el que se han medido p variables. Además, sobre Ω se define un índice de disimilitud d .

*Profesores de la Escuela de Matemáticas de la Universidad de Costa Rica.

2 Métodos de Nubes Dinámicas

Se quiere obtener una partición de Ω en k clases bien agregadas, bien separadas y de intersección vacía. El número k de clases es dado a priori.

Este método fue introducido por Diday [6], generalizando algunos algoritmos particulares, como el " k -means" de Forgy. Se basa en que cada clase debe tener una representación (llamada núcleo), y luego se hace una búsqueda iterada de núcleos y de particiones, hasta optimizar un cierto criterio.

El algoritmo general de Nubes Dinámicas es el siguiente:

1. Se da una partición inicial de Ω .
2. Se calculan los núcleos, mediante una función g de representación.
3. Se forma otra partición, afectando cada elemento al núcleo más próximo, mediante una función f de afectación.
4. Se recalculan los núcleos de la nueva partición.
5. *Se repiten los pasos 3 y 4 hasta que las clases se estabilicen.*

Como se verá más adelante se puede probar que —en los casos que presentaremos—, dado un cierto criterio W que mida la adecuación de la representación, en cada iteración se mejora W y además se converge a una clase estable.

2.1 La representación

Llamaremos \mathcal{L} al espacio de representación. Un elemento $L_i \in \mathcal{L}$ se llama núcleo admisible. Así \mathcal{L}_k será el conjunto de k -tuples de núcleos:

$$(L_1, \dots, L_k) \in \mathcal{L}_k.$$

Llamaremos \mathcal{P} al conjunto de particiones en k clases de Ω . Así, L_i será el núcleo que representa a la clase P_i .

2.2 Ejemplos de núcleos

- (a) En el caso euclídeo, el núcleo de una clase puede ser su centro de gravedad, como se ilustra en la Figura 1.
- (b) Un núcleo puede ser también un grupo de elementos de la clase, como se ilustra en la Figura 2.
- (c) También podría ser una recta de regresión, como se ilustra en la Figura 3.
- (d) Pueden ser conceptos conjuntistas, como se ilustra en la Figura 4.

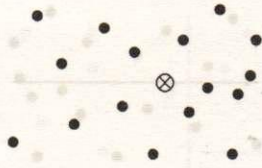


Figura 1: El núcleo de una clase es el centro de gravedad.

- (e) O bien, un plano factorial, un hiperplano discriminante, una función de distribución, etc.

Dependiendo del tipo de representación que se utilice, así será el método de Nubes Dinámicas a tratar.

Esquema General del Método de Nubes Dinámicas

- Ω : conjunto de individuos.
- d : disimilitud sobre Ω .
- k : número de clases a formar.
- \mathcal{P}_k : conjunto de particiones de Ω en k clases.
- \mathcal{L} : espacio de núcleos admisibles.
- $L \in \mathcal{L}$: núcleo admisible.
- \mathcal{L}_k : espacio de representación (observación: no siempre $\mathcal{L}_k = \mathcal{L}^k$).
- Función de afectación: $f: \mathcal{L}_k \mapsto \mathcal{P}_k$.
- Función de representación: $g: \mathcal{P}_k \mapsto \mathcal{L}_k$.
- Criterio a optimizar: $W(L, P)$, con $W: \mathcal{L}_k \times \mathcal{P}_k \mapsto \mathbb{R}^+$.

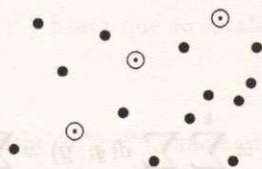


Figura 2: El núcleo de una clase es un grupo de elementos de la clase.

2 Métodos de Nubes Dinámicas

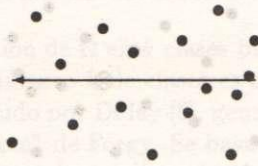


Figura 3: El núcleo de una clase es una recta de regresión.

• Algoritmo: $\mathcal{P}_k \xrightarrow{g} \mathcal{L}_k \xrightarrow{f} \mathcal{P}_k \xrightarrow{g} \mathcal{L}_k \xrightarrow{f} \mathcal{P}_k \dots$

• Solución: L^*, P^* , tal que:

$$W(L^*, P^*) = \min\{W(L, P) : L \in \mathcal{L}_k, P \in \mathcal{P}_k\} \quad (\text{o maximizar})$$

2.3 El núcleo es el centro de gravedad

Esta es la variante más difundida del método de Nubes Dinámicas y la más fácil de implementar. Fue originalmente introducida por Forgy [10] y desarrollada luego por diversos autores. En las referencias en lengua inglesa se le denomina por "k-means". Sólo es aplicable en el caso en que la distancia sea euclídea cuadrática.

Situación:

Los elementos de Ω son elementos de \mathbf{R}^p , entre los que se mide una distancia euclídea cuadrática $d(x, y) = \|x - y\|$.

Núcleos:

El núcleo de la clase P_i es su centro de gravedad (o promedio) $g_i \in \mathbf{R}^p$, donde:

$$g_i = \frac{1}{|P_i|} \sum_{x \in P_i} x.$$

Es decir, $\mathcal{L} = \mathbf{R}^p$, $\mathcal{L}_k = \mathcal{L}^k$.

Criterio:

$$W(L, P) = \sum_{i=1}^k \sum_{y \in P_i} d(g_i, y) = \sum_{i=1}^k I(P_i),$$

donde $L = (L_1, \dots, L_k)$; $P = (P_1, \dots, P_k)$; $I(P_i)$ es la inercia de la clase P_i .

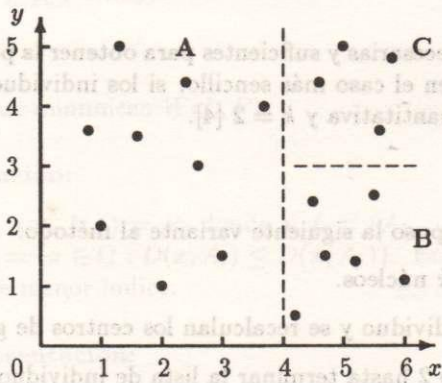


Figura 4: Los núcleos son conceptos conjuntistas: $A = [x < 4]$; $B = [x \geq 4][y < 3]$; $C = [x \geq 4][y \geq 3]$.

Función de Afectación:

Sea $f: \mathcal{L}_k \mapsto \mathcal{P}_k$, con $f(L) = P$, donde si $L = (g_1, \dots, g_k)$ entonces $P = (P_1, \dots, P_k)$, con $P_i = \{x \in \Omega : d(x, g_i) \leq d(x, g_j)\}$. En caso de igualdad, se afecta x a la clase de menor índice.

Función de Representación:

Sea $g: \mathcal{P}_k \mapsto \mathcal{L}_k$, con $g(P) = L$, donde si $P = (P_1, \dots, P_k)$ entonces $L = (g_1, \dots, g_k)$, con g_i el centro de gravedad de P_i .

Algoritmo:

1. Se escoge una partición P^0 al azar.
2. Se calcula, para cada P_i^0 , su centro de gravedad g_i^0 .
3. Se afectan todos los individuos de Ω al centro de gravedad más cercano, formándose la partición P_i^1 .
4. Se reiteran los pasos 2 y 3 hasta que se establezca la partición.

Convergencia:

El algoritmo converge hacia (L^*, P^*) , haciendo decrecer el criterio W en cada iteración [6,8].

Optimalidad:

Las condiciones necesarias y suficientes para obtener la partición óptima sólo han sido estudiadas en el caso más sencillo: si los individuos son descritos por una única variable cuantitativa y $k = 2$ [4].

Variante:

Thorndike [6] propuso la siguiente variante al método:

1. Se escogen los k núcleos.
2. Se afecta un individuo y se recalculan los centros de gravedad.
3. Se sigue el paso 2 hasta terminar la lista de individuos.
4. Se detienen las reiteraciones hasta que haya estabilidad en las clases.

Comentario:

Como se verá más adelante, este método tiende a dar clases de forma esférica con mismo efectivo.

2.4 El núcleo esta formado por individuos

Se trata del primer aporte original de E. Diday. Trata de generalizar el algoritmo anterior a otras situaciones: cuando la disimilitud sobre Ω no es una distancia euclidiana. Se busca entonces una manera de dar una representación adecuada a cada clase.

Situación:

Los elementos de Ω son objetos cualesquiera, entre los que se mide una disimilitud d (no necesariamente euclidiana).

Núcleo:

El núcleo es formado por elementos de Ω . Cada núcleo tendrá cardinalidad q . Lerman [15] recomienda tomar q como la parte entera de $\text{card}(\Omega)$ por un factor de proporción α , entre k : $q = \lfloor \alpha \text{card}(\Omega) / k \rfloor$.

Criterio:

$W(L, P) = \sum_{i=1}^k D(A_i, P_i)$, donde $L = (A_1, \dots, A_k)$, $A_i \in \mathcal{L}$, $|A_i| = q$; $P = (P_1, \dots, P_k)$. Aquí $D(A_i, P_i)$ mide la disimilitud de A_i con la clase P_i , y suponemos que D es aditiva en el siguiente sentido:

$$D(X, Y) = \sum_{x \in X} \sum_{y \in Y} d(x, y) \quad ; \quad \text{si } X, Y \subseteq \Omega.$$

Además, $D(X, y) = \sum_{x \in X} d(x, y)$.

Problema:

Buscar L^* , P^* que minimicen $W(L, P)$.

Función de Afectación:

Sea $f: \mathcal{L}_k \mapsto \mathcal{P}_k$, con $f(L) = P$, donde si $L = (A_1, \dots, A_k)$ entonces $P = (P_1, \dots, P_k)$, con $P_i = \{x \in \Omega : D(x, A_i) \leq D(x, A_j)\}$. En caso de igualdad, se afecta x a la clase de menor índice.

Función de Representación:

Sea $g: \mathcal{P}_k \mapsto \mathcal{L}_k$, con $g(P) = L$, donde si $P = (P_1, \dots, P_k)$ entonces $L = (A_1, \dots, A_k)$, donde

$$D(A_i, P_i) = \min\{D(B, P_i) : B \subseteq \Omega, |B| = q\}$$

Si tal mínimo se alcanza para dos conjuntos, en el desarrollo del algoritmo se toma uno de ellos al azar (por ejemplo, el primero que aparezca). Obviamente, esto afecta al resultado de la partición final.

Algoritmo:

1. Se escoge una partición P^0 al azar.
2. Se calcula, para cada P_i^0 , su núcleo A_i^0 .
3. Se afectan todos los individuos de Ω al núcleo más cercano.
4. Se reiteran los pasos 2 y 3 hasta que se establezca la partición.

Convergencia:

El algoritmo converge hacia (L^*, P^*) , haciendo decrecer el criterio W en cada iteración [6,8].

2.5 Clasificación con distancias adaptativas

El método de Nubes Dinámicas con centros de gravedad tiene la tendencia de formar clases esféricas con mismo efectivo. Por ello, no son útiles cuando se trata de identificar clases que tengan una misma forma de dispersión, quizá no necesariamente esférica, pero con una o varias direcciones de prolongamiento (sobre un eje discriminante, por ejemplo). Por tanto, en este caso se quita la restricción de que la medida de disimilitud (métrica) sea la misma durante todo el algoritmo. Más bien se trata de buscar iterativamente la que mejor se adapte a los datos.

A. Una única distancia adaptativa.

Estamos en el caso en que todas las clases tengan el mismo tipo de dispersión, pero con direcciones de prolongamiento desconocidas a priori. Se tratará entonces de asociar a cada partición una métrica M tal que la suma de las distancias de cada individuo al centro de gravedad sea lo más pequeña posible.

Situación:

$\Omega = \mathbb{R}^p$, se inicia con una métrica cualquiera, por ejemplo, con $M = I$.

Núcleo:

Sea \mathcal{M} el conjunto de métricas $p \times p$. El espacio de representación estará formado por k núcleos y una métrica $M \in \mathcal{M}$: $\mathcal{L}_k = (\mathbb{R}^p) \times \mathcal{M}$.

Criterio:

$$W(L, P) = \sum_{i=1}^k \sum_{y \in P_i} d_M^2(g_i, y),$$

donde $L = (g_1, \dots, g_k, M) \in \mathcal{L}_k$; $P = (P_1, \dots, P_k)$.

Problema:

Buscar L^* , P^* que minimicen $W(L, P)$.

Función de Afectación:

Sea $f: \mathcal{L}_k \mapsto \mathcal{P}_k$, con $f(L) = P$, donde si $L = (g_1, \dots, g_k, M)$ entonces $P = (P_1, \dots, P_k)$, con $P_i = \{x \in \Omega : d_M(x, g_i) \leq d_M(x, g_j)\}$. En caso de igualdad, se afecta x a la clase de menor índice.

Función de Representación:

Sea $g: \mathcal{P}_k \mapsto \mathcal{L}_k$, con $g(P) = L$, donde si $P = (P_1, \dots, P_k)$ entonces $L = (g_1, \dots, g_k, M)$ es tal que se minimiza $W(\cdot, P)$, con la restricción $|M| = 1$. Se obtiene que g_i es el centro de gravedad de P_i , mientras que $M = |D|^{1/p} D^{-1}$, con:

$$D = \sum_{i=1}^k \sum_{y \in P_i} (g_i - y)(g_i - y)^t$$

= Matriz de suma de inercias internas

Obsérvese que se trata de la misma solución que se obtiene en Análisis Discriminante Lineal [11,12].

Algoritmo:

1. Se escoge una partición P^0 al azar.
2. Se calcula, para cada P_i^0 , sus centros de gravedad g_i^0 .
3. Se determina la métrica que mejor se adapta a la partición.
4. Se afectan todos los individuos de Ω al centro de gravedad más cercano, de acuerdo con la métrica calculada en el paso 2.
5. Se reiteran los pasos 2, 3 y 4 hasta que se establezca la partición.

Convergencia:

El algoritmo converge hacia (L^*, P^*) , haciendo decrecer el criterio W en cada iteración [11].

B. Una distancia adaptativa por clase.

El método precedente no permite reconocer clases con direcciones de prolongamiento diferentes y desconocidas. Para lograr eso, se asocia a cada clase de la partición una métrica tal que la suma de las distancias de cada individuo al centro de gravedad de su clase (en el sentido de la métrica de esa clase) sea lo más pequeña posible. Esto es, se buscan las k métricas que mejor se adapten a las k clases.

Situación:

$\Omega = \mathbb{R}^p$. Se inicia con una métrica cualquiera, por ejemplo $M = I$.

Núcleo:

La partición P de Ω será representada por los centros de gravedad g_i y k matrices M_1, \dots, M_k de \mathcal{M} . Luego, $\mathcal{L}_k = (\mathbb{R}^p)^k \times \mathcal{M}^k$.

Criterio:

$$W(L, P) = \sum_{i=1}^k \sum_{y \in P_i} d_{M_i}^2(g_i, y),$$
donde $L = (g_1, \dots, g_k, M_1, \dots, M_k) \in \mathcal{L}_k$; $P = (P_1, \dots, P_k)$. El criterio representa la suma de las inercias de cada clase respecto a su centro de gravedad, en el sentido de la distancia $d_{M_i}^2$ que tiene asociada.

Problema:

Buscar L^* , P^* que minimicen $W(L, P)$.

Función de Afectación:

Sea $f: \mathcal{L}_k \mapsto \mathcal{P}_k$, con $f(L) = P$, donde si $L = (g_1, \dots, g_k, M_1, \dots, M_k)$ entonces $P = (P_1, \dots, P_k)$, con $P_i = \{x \in \Omega : d_{M_i}(x, g_i) \leq d_{M_j}(x, g_j)\}$. En caso de igualdad, se afecta x a la clase de menor índice.

Función de Representación:

Sea $g: \mathcal{P}_k \mapsto \mathcal{L}_k$, con $g(P) = L$, donde si $P = (P_1, \dots, P_k)$ entonces $L = (g_1, \dots, g_k, M_1, \dots, M_k)$ es tal que se minimiza $W(\cdot, P)$, con la restricción $|M_i| = 1$.

1. Se obtiene que g_i es el centro de gravedad de P_i , mientras que

$$M_i = |D_i|^{1/p} D_i^{-1},$$

donde $D_i = \sum_{y \in P_i} (g_i - y)(g_i - y)^t$ es la matriz de inercias internas de la clase P_i .

Algoritmo:

1. Se escogen los P_i^0 al azar.
2. Se calcula, para cada P_i^0 , su centro de gravedad g_i^0 .
3. Se determina la métrica M_i que mejor se adapta a la clase P_i^0 .
4. Se afectan todos los individuos de Ω al centro de gravedad más cercano, en el sentido de cada métrica, para obtener una nueva partición.
5. Se reiteran los pasos 2, 3 y 4 hasta que se establezca la partición.

Convergencia:

El algoritmo converge hacia (L^*, P^*) , haciendo decrecer el criterio W en cada iteración [11].

Comentario:

Debe tenerse cuidado en la implementación de estos dos métodos, pues en cada iteración se deben invertir matrices, por lo que se recomienda hacer las primeras iteraciones con el algoritmo de los centros de gravedad. Además, debe tenerse la certeza de que un algoritmo más simple no da resultados satisfactorios.

2.6 Mezclas de distribución de probabilidad

Se dispone de una muestra Ω cuyos elementos siguen distintas distribuciones de probabilidad. Se quiere estimar los parámetros de tales distribuciones.

Este es uno de los problemas más viejos de la Estadística Inferencial, que aquí se aborda desde el punto de vista de la Clasificación Automática, en particular con el método de Nubes Dinámicas.

Situación:

Se tiene una muestra $\Omega = \{x_1, \dots, x_n\}$ de una variable aleatoria X en \mathbf{R}^s , cuya ley admite la densidad $f(x) = \sum_{i=1}^k p_i f(x, a_i)$, donde $p_i > 0, \forall i, \sum p_i = 1$.

Se supone que $f(\cdot, a_i)$ es una densidad que depende del parámetro $a_i \in \mathbf{R}^s$ (donde s es el número de componentes del parámetro), y p_i es la probabilidad de que un punto de la muestra siga la ley $f(\cdot, a_i)$. Se quieren estimar las k componentes y los parámetros desconocidos p_i, a_i . Tomamos $p_i = \text{card}(P_i)/n$ como estimador de p_i .

Núcleo:

Como se busca una partición (P_1, \dots, P_k) tal que la clase P_i sea asimilable a la ley $f(\cdot, a_i)$, el núcleo de la clase i -ésima es el parámetro $a_i \in \mathbf{R}^s$. Luego $\mathcal{L}_k = (\mathbf{R}^s)^k$.

Criterio:

$W(L, P) = \sum_{i=1}^k \ln V(P_i, a_i)$, donde:

- $L = (a_1, \dots, a_k) \in \mathcal{L}_k$
- $P = (P_1, \dots, P_k) \in \mathcal{P}_k$
- $V(P_i, a_i) = \prod_{x \in P_i} f(x, a_i)$ es la función de verosimilitud de la submuestra P_i para la ley $f(\cdot, a_i)$.

Función de Afectación:

Sea $f: \mathcal{L}_k \mapsto \mathcal{P}_k$, con $f(L) = P$, donde si $L = (a_1, \dots, a_k)$ entonces $P = (P_1, \dots, P_k)$, con $P_i = \{x \in \Omega : f(x, a_i) \geq f(x, a_j)\}$. En caso de igualdad, se afecta x a la clase de menor índice.

Función de Representación:

Sea $g: \mathcal{P}_k \mapsto \mathcal{L}_k$, con $g(P) = L$, donde si $P = (P_1, \dots, P_k)$ entonces $L = (a_1, \dots, a_k)$ es tal que se maximiza la verosimilitud del parámetro de la densidad

asociada a la muestra P_i . Por ejemplo, en el caso Gaussiano tendremos $a_i = (\mu_i, \Gamma_i)$, donde:

$$\mu_i = \frac{1}{\text{card}(P_i)} \sum_{x \in P_i} x$$

$$\Gamma_i = \frac{1}{\text{card}(P_i)} \sum_{x \in P_i} (x - \mu_i)(x - \mu_i)^t.$$

Problema:

Se busca L^*, P^* tal que $W(L^*, P^*)$ sea máximo.

Algoritmo:

1. Se escoge una partición P^0 al azar.
2. Se estiman los parámetros a_1, \dots, a_k de P^0 .
3. Se afectan los individuos según la función de afectación.
4. Se reiteran los pasos 2 y 3 hasta la estabilización de la partición.

Convergencia:

Si la familia de densidades $f(\cdot, a_i)$ está acotada superiormente $\forall a_i \in \mathcal{L}$, el algoritmo anterior converge hacia una solución estable L^*, P^* , haciendo crecer el criterio en cada iteración [7,12].

Comentario:

Hay que hacer notar que, a pesar de que este método converge rápidamente, la precisión es mejor con los métodos clásicos de estimación de parámetros, con los algoritmos tipo EM y SEM [12].

Caso Gaussiano:

Los parámetros son $a_i = (\mu_i, \Gamma_i)$, con μ_i la esperanza de la i -ésima componente y Γ_i la matriz de varianzas de la i -ésima componente. El criterio será entonces:

$$W(L, P) = \beta - \frac{1}{2} \sum_{i=1}^k \sum_{x \in P_i} \{ (x - \mu_i)^t \Gamma_i^{-1} (x - \mu_i) + \ln |\Gamma_i| \},$$

donde β es una constante. Por tanto, maximizar $W(L, P)$ se reduce a minimizar $C(L, P)$, con:

$$C(L, P) = \sum_{i=1}^k \sum_{x \in P_i} \{ (x - \mu_i)^t \Gamma^{-1} (x - \mu) + \ln |\Gamma_i| \}.$$

Hacemos notar el caso siguiente: si $\Gamma_i = \Gamma, \forall i$, con Γ una matriz conocida, se debe maximizar entonces la cantidad

$$C(L, P) = \sum_{i=1}^k \sum_{x \in P_i} \{ (x - \mu_i)^t \Gamma^{-1} (x - \mu_i) \}.$$

Fijando la partición, los estimadores de máxima verosimilitud de los μ_i son:

$$\mu_i = \frac{1}{\text{card}(P_i)} \sum_{x \in P_i} x.$$

Como Γ es simétrica, existe T tal que $\Gamma = T \cdot T^t$. Haciendo el cambio de variable $y = Tx, \Gamma = I$ obtenemos por lo tanto:

$$C(L, P) = \sum_{i=1}^k \sum_{x \in P_i} \{ (x - \mu_i)^t (x - \mu_i) \}.$$

Obsérvese que se trata del criterio de la versión de nubes dinámicas cuando los núcleos son los centros de gravedad. De aquí se desprende que este método tiende a dar clases esféricas del mismo volumen.

2.7 Regresión tipológica

Se pretende detectar k comportamientos locales de regresión y los hiperplanos asociados, de manera que se minimice un criterio de inadecuación de las muestras con sus representaciones lineales (criterio de mínimos cuadrados) [5].

La idea es dar una partición del espacio Ω de n individuos a los que se han medido m variables explicativas X^j , y una variable a explicar y , así como los hiperplanos de regresión asociados a cada una de las clases de tal partición.

Modelo de Regresión Lineal:

En su forma matricial, puede escribirse como $Y = X \cdot b + e$, donde $Y_{n \times 1}$, $X_{n \times m}$, $b_{m \times 1}$, $e_{n \times 1}$, $E(e) = 0_{n \times 1}$, $V(e) = \sigma^2 \cdot I_n$, siendo Y y X vector y matriz conocidos. Se desea estimar b , un vector a calcular que minimice el criterio

$$\sum_{i=1}^n (y_i - \sum_{j=1}^m x_{ij} b_j)^2,$$

o bien, en forma matricial, hallar el vector b tal que se minimice

$$S(b) = (Y - Xb)(Y - Xb)^t.$$

Para que tal problema tenga solución deben cumplirse las llamadas *ecuaciones normales*:

$$X^t X b = X^t Y.$$

Se presentan dos casos:

- (i) Si $X^t X$ es invertible y bien condicionada, la solución en este caso es:
 $b = (X^t X)^{-1} X^t Y$ (resultado clásico).
- (ii) Si $X^t X$ es no invertible o es mal condicionada, se usa una inversa generalizada de tal matriz $m \times m$. Denotemos $A = X^t X$, $P = X^t Y$. Sea A^- una inversa generalizada de A (cumple que $AA^-A = A$). Para resolver las ecuaciones normales se tiene que $b = A^-P + (I - A^-A)h$, donde h es cualquier vector de \mathbf{R}^m [5].

El algoritmo:

Se basa en la construcción de tres funciones f , g , W , donde:

- f : A partir de una representación, da una partición P de Ω .
- g : Dada una partición P , construye las regresiones lineales de cada clase.
- W : Criterio a minimizar.

Debe iterarse la composición $W \circ g \circ f$ hasta obtener estabilidad.

Notación:

- Los datos se presentan en la forma $(x^1, \dots, x^m; y)$, donde x^1, \dots, x^m son las variables explicativas, mientras que y es la variable a explicar (vectores columna $n \times 1$).

- El espacio de representación es $\mathcal{L} = \{(L_1, \dots, L_k)\}$, donde

$$L_j = (b_j^1, \dots, b_j^m)^t = B_j$$

es el vector de coeficientes de regresión. Es así como $\mathcal{L}_k = (\mathbf{R}^m)^k$.

- $\mathcal{P}_k = \{(P_1, \dots, P_k)\}$ es el espacio de las particiones en k clases de las observaciones de Ω .
- $n_j = \text{card}(P_j)$.
- $Z_i^j = i$ -ésimo elemento de P_j : vector de $m + 1$ componentes.

Además tendremos las siguientes matrices:

$$Z^j = \begin{pmatrix} Z_1^j \\ \vdots \\ Z_{n_j}^j \end{pmatrix} \quad X^j = \begin{pmatrix} x_1^{1,j} & \dots & x_1^{m,j} \\ \vdots & & \vdots \\ x_{n_j}^{1,j} & \dots & x_{n_j}^{m,j} \end{pmatrix} \quad Y^j = \begin{pmatrix} y_1^j \\ \vdots \\ y_{n_j}^j \end{pmatrix}$$

Criterio:

$$W(L, P) = \sum_{j=1}^k \|Y^j - X^j B_j\|_{\mathbf{R}^{n_j}}^2 = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_i^j - x_i^j b_i^j)^2$$

Objetivo:

Hallar un par (P^*, L^*) que minimice W .

Función de Afectación:

Sea $f: \mathcal{L}_k \mapsto \mathcal{P}_k$, con $f(L) = P$, donde si $L = (L_1, \dots, L_k)$ entonces $P = (P_1, \dots, P_k)$, con $P_i = \{z \in \Omega : d(z, B_i) \leq d(z, B_j)\}$, donde

$$z = (x_1^1, \dots, x_1^m; y_1) = (x_i; y_i) \quad ; \quad d(z, B_i) = (y_i - x_i B_i)^2.$$

En caso de igualdad, se afecta z a la clase de menor índice.

Función de Representación:

Sea $g: \mathcal{P}_k \mapsto \mathcal{L}_k$, con $g(P) = L$, donde si $P = (P_1, \dots, P_k)$ entonces $L = (B_1, \dots, B_k)$ es tal que para todo índice j , $S(B_j) = \inf_{\mathbf{R}^m} S(b)$, con $S(b)$ definido como antes para la matriz X^j , y por lo tanto,

$$B_j = B(h) = [(X^j)^t(X^j)]^{-1} (X^j)^t Y^j + (I_m - [(X^j)^t(X^j)]^{-1} (X^j)^t(X^j)) h,$$

con h un vector cualquiera de \mathbf{R}^m .

Algoritmo:

Se definen las sucesiones (u_n) y (v_n) mediante el esquema:

$$\begin{aligned} v_0 &= (P^{(0)}, L^{(0)}) \\ v_{n+1} &= (P^{(n+1)}, L^{(n+1)}) \\ u_n &= W(P^{(n)}, L^{(n)}) = W(v_n), \end{aligned}$$

donde alguno de $P^{(0)}$ o $L^{(0)}$ es escogido al azar o estimado, y además $L^{(n+1)} = g(P^{(n)})$ y $P^{(n+1)} = f(L^{(n)})$. En [5] se prueba que las sucesiones (u_n) y (v_n) convergen.

2.8 Comentarios generales

- El número k de clases debe ser dado a priori, por lo que no se sabe si tal número es "natural", en el sentido de reflejar la cantidad de clases reales de que se forma Ω .
- La solución depende de la partición inicial, dada al azar. Por tanto es recomendable hacer varias corridas del método, para obtener un conjunto de particiones. En la sección 4 detallamos las principales técnicas para el estudio de tal conjunto de particiones.
- Durante el transcurso del algoritmo, una clase se puede vaciar. Es decir, en la etapa de afectación, ningún individuo es afectado a un núcleo. Se puede continuar el algoritmo considerando menos clases.

3 Algoritmo de transferencias

Un segundo tipo de métodos de particionamiento son los algoritmos del tipo de *transferencias*, originalmente propuestos por Régnier. Consisten en hacer la transferencia entre una clase y otra, de un único elemento de Ω a la vez, haciendo mejorar algún criterio en cada iteración.

Digamos que se quiere particionar Ω en k clases. El criterio es una función $W: \mathcal{P}_k \mapsto \mathbf{R}^+$, que mide la adecuación de la partición $P \in \mathcal{P}_k$ de Ω .

Algoritmo:

- Se da una partición inicial $P = (P_1, P_2, \dots, P_k)$ de Ω .
- Se toma un elemento $x \in \Omega$ arbitrario, con $x \in P_i$. Llamamos $P_j^i(x)$ la partición de \mathcal{P}_k consistente en transferir x de P_i hacia P_j en la partición P y dejar las demás clases iguales.
- Sea P^* tal que $W(P^*) = \min\{W(P_j^i(x)) : j = 1, \dots, k\}$. Entonces ponemos $P := P^*$.
- Se repiten los pasos 2 y 3 para todos los elementos $x \in \Omega$.
- Se detiene cuando al aplicar 4 no ocurre ninguna nueva transferencia.

3.1 Caso Euclídeo

Tenemos n individuos descritos por p caracteres cuantitativos y \mathbf{R}^p está provisto de una distancia euclídea.

El criterio a usar será: si $P = (P_1, \dots, P_k)$, entonces $W(P) = \sum_{i=1}^k I(P_i)$, donde $I(A)$ es la inercia de A . Por tanto, al pasar x de P_i a P_j se debe minimizar

$$W(P_j^i(x)) = \sum_{h \notin \{i,j\}} I(P_h) + I(P_i \setminus \{x\}) + I(P_j \cup \{x\}).$$

Sea ω_h el peso de la clase P_h . Sea ω el peso del individuo x , y sea $G(P_h)$ el centro de gravedad de P_h . Luego, para implementar el algoritmo se debe:

- i) Calcular $W(P) = \sum_{h=1}^k I(P_h) = \sum_{h \notin \{i,j\}} I(P_h) + I(P_i) + I(P_j)$.
- ii) Calcular $I(P_i \setminus \{x\})$.
- iii) Calcular $I(P_j \cup \{x\})$.

Entonces, en memoria deben guardarse $I(P_i)$ e $I(P_j)$. Ahora bien,

$$\begin{aligned} I(P_j \cup \{x\}) &= I(P_j) + I(\{x\}) + \frac{\omega_j \omega}{\omega_j + \omega} \|G(P_j) - x\|^2 \\ &= I(P_j) + \frac{\omega_j \omega}{\omega_j + \omega} \|G(P_j) - x\|^2 \end{aligned}$$

Por otra parte,

$$\begin{aligned} I(P_i) &= I(P_i \setminus \{x\} \cup \{x\}) \\ &= I(P_i \setminus \{x\}) + I(\{x\}) + \frac{(\omega_i - \omega)\omega}{\omega_i} \|G(P_i \setminus \{x\}) - x\|^2, \end{aligned}$$

de donde

$$I(P_i \setminus \{x\}) = I(P_i) - \frac{(\omega_i - \omega)\omega}{\omega_i} \|G(P_i \setminus \{x\}) - x\|^2.$$

Se concluye entonces que para implementar el algoritmo en cada paso se debe calcular, para $j = 1, \dots, k$,

$$\|G(P_j) - x\|^2 ; \quad G(P_i \setminus \{x\}) ; \quad \|G(P_i \setminus \{x\}) - x\|^2.$$

3.2 Caso general

En este caso Ω es arbitrario, con d un índice de disimilitud sobre Ω . El criterio W que se define sobre la partición P toma en cuenta la relación de equivalencia \mathcal{R} asociada a P :

$$W(P) = \sup\{d(i, j) : i \mathcal{R} j; i, j \in \Omega\}.$$

Si se tiene $P = (P_1, \dots, P_k)$, con $x \in P_i$, para transferir x a P_j es necesario que:

$$\sup\{d(x, y) : y \in P_j\} < \sup\{d(x, y) : y \in P_i\}.$$

Observaciones:

1. Véase que al igual que en los métodos de Nubes Dinámicas, aquí también la partición final P^* depende de la partición inicial.
2. El número k de clases es dado a priori. Sin embargo las clases también se pueden vaciar en el transcurso del algoritmo. Igualmente, ese número k puede no ser un número "natural" de clases para Ω .
3. Para dar las k clases iniciales en el caso euclídeo, también se puede usar el método de *Polos de Atracción* [15].

4 Estudio de una multipartición

Sea Ω un conjunto de n elementos sobre el que se han medido p variables. Suponga que se han obtenido m particiones P_1, \dots, P_m de Ω , cada una en k clases, mediante algún método de particionamiento (Nubes Dinámicas, Transferencias, Polos de Atracción, Clasificaciones Jerárquicas con distintos métodos, etc.).

El conjunto $\{P_1, \dots, P_m\}$ se llama una *multipartición* de Ω . Queremos saber cuál de las particiones de Ω retener, o bien, determinar cómo se relacionan entre sí, para sacar alguna conclusión. Para esto hay varios puntos de vista:

- a) Retener la partición de mayor inercia inter-clases.
- b) Estudiar las "formas fuertes".
- c) Obtener una "partición central", que sea un consenso entre las distintas particiones.

Veamos con detalle cómo trabajan estos distintos criterios.

4.1 Retención de la partición de mayor inercia

Sea $\{P_1, \dots, P_m\}$ una multipartición de Ω . Nos situamos en el caso euclídeo. Esto es, que $\Omega \in \mathbf{R}^p$ y d es una distancia euclidiana cuadrática. Sea $I(P)$ la inercia inter-clases de la partición $P \in \mathcal{P}_k$:

$$I(P) = \sum_{i=1}^k \sum_{x \in P_i} d^2(x, g_i).$$

Sea P^* definido de tal forma que

$$I(P) = \max\{I(P_i) : i = 1, \dots, m\}.$$

En caso que se dé la igualdad $I(P_i) = I(P_j)$, se escoge $P_0^* = P_i \cap P_j$. Sobre P_0^* se crea una jerarquía —con el índice de agregación de Ward, por ejemplo— y se corta el árbol de jerarquía en el nivel apropiado para obtener k clases. Tal partición será P^* .

Observaciones:

- a) Esta escogencia es sesgada, pues en la mayoría de los métodos conocidos de particionamiento la partición final depende de una escogencia inicial arbitraria.
- b) Podría ser que el número k de clases no sea el más apropiado para Ω . Sin embargo, no hay manera de conocer esto con el método.

4.2 Análisis de las Formas Fuertes

Este criterio se basa en el estudio de aquellas clases que fueron más estables en las distintas particiones.

Sea $\{P_1, \dots, P_m\}$ una multipartición de Ω , con $P_i \in \mathcal{P}_k$. Sea $P_0 = \bigcap_{i=1}^m P_i$. Las clases de P_0 se llaman las *formas fuertes* de la multipartición. En P_0 aparecen los conjuntos formados por los elementos de Ω que en todas las particiones quedaron en la misma clase.

Cabe aquí señalar que —para los métodos de clasificación— la práctica demuestra que habrá varios individuos que siempre queden en la misma clase, por lo que no se debe temer que las clases de P_0 sean sólo los conjuntos unitarios.

El estudio de las formas fuertes trata de mostrar las clases estables mediante una multipartición. Para ello se usa el método de *conexidad descendente*.

Método de Conexidad Descendente

- a) En P_0 se define una matriz $S = (s_{ij})$ tal que s_{ij} cuenta el número de veces que dos formas fuertes i, j han estado en las mismas clases. Esto da una medida de similitud entre las formas fuertes.
- b) Sea $D = (d_{ij})$ una matriz tal que $d_{ij} = \max\{s_{ij}\} - s_{ij}$. Esto da una medida de disimilitud entre las formas fuertes.
- c) Sobre P_0 , con la disimilitud D , se construye una jerarquía.

En el caso euclídeo se podría construir la jerarquía sobre los centros de gravedad de las formas fuertes.

Ejemplo:

Sea $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ y $k = 3$. Considere las cuatro particiones siguientes:

$$P_1 = \{(1, 2, 3), (4), (5, 6, 7, 8, 9)\}$$

$$P_2 = \{(1, 2, 3, 4), (5, 6), (7, 8, 9)\}$$

$$P_3 = \{(1, 2), (3), (4, 5, 6, 7, 8, 9)\}$$

$$P_4 = \{(1, 2, 3), (4, 5, 6), (7, 8, 9)\}$$

Entonces las formas fuertes son:

$$P_0 = \{(1, 2), (3), (4), (5, 6), (7, 8, 9)\}.$$

Las matrices S y D se muestran a continuación:

$$S = \begin{matrix} (1, 2) \\ (3) \\ (4) \\ (5, 6) \\ (7, 8, 9) \end{matrix} \begin{pmatrix} 4 & 3 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 & 0 \\ 1 & 1 & 4 & 2 & 1 \\ 0 & 0 & 2 & 4 & 2 \\ 0 & 0 & 1 & 2 & 4 \end{pmatrix} \quad D = \begin{matrix} (1, 2) \\ (3) \\ (4) \\ (5, 6) \\ (7, 8, 9) \end{matrix} \begin{pmatrix} 0 & 1 & 3 & 4 & 4 \\ 1 & 0 & 3 & 4 & 4 \\ 3 & 3 & 0 & 2 & 3 \\ 4 & 4 & 2 & 0 & 2 \\ 4 & 4 & 3 & 2 & 0 \end{pmatrix}$$

El árbol jerárquico obtenido con la agregación d_3 se muestra en la Figura 5. Del análisis de la jerarquía obtenida, si se quiere retener una partición en \mathcal{P}_3 , ésta podría ser:

$$P^* = \{(1, 2, 3), (4), (5, 6, 7, 8, 9)\}.$$

Debe notarse que al agregar la clase (5,6) con (7,8,9), la distancia entre esas dos clases era 2, igual a la distancia entre las clases (4) y (5,6). Debido a ello se hizo una escogencia arbitraria en la agregación. De haber agregado la clase (4) con la clase (5,6) primero, el árbol hubiera resultado distinto y se hubiera obtenido

$$P^* = \{(1, 2, 3), (4, 5, 6), (7, 8, 9)\}.$$

Con esto se nota que siempre quedan dudas sobre la bondad del método.

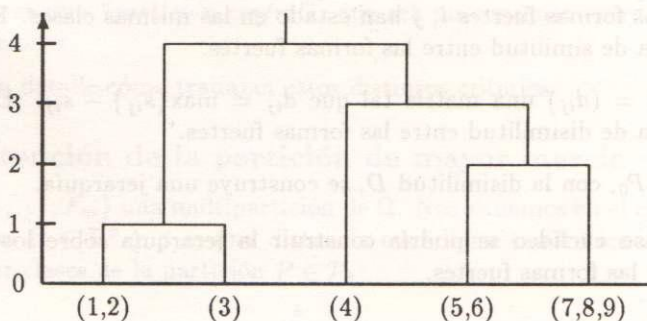


Figura 5: Jerarquía obtenida en el ejemplo.

4.3 Particiones Centrales

Se busca una partición que sea un "consenso" entre las particiones de una multipartición [3,15]. Sea $\{P_1, \dots, P_m\}$ una multipartición de Ω , con $P_i \in \mathcal{P}_k$.

Para cada partición P_l existe una relación de equivalencia v^l tal que

$$v^l(i, i') = \begin{cases} 1 & \text{si } i, i' \in P_l \\ 0 & \text{en otro caso.} \end{cases}$$

Definimos la distancia d de la diferencia simétrica entre las particiones:

$$d(P, Q) = \frac{1}{2} \sum_{i, i' \in \Omega} |u(i, i') - v(i, i')|,$$

donde u y v son las relaciones asociadas con P y Q respectivamente.

Se llama *partición central* de la multipartición $\{P_1, \dots, P_m\}$ a la partición P^* cuya relación de equivalencia u^* minimiza el criterio $C(u^*) = \min C(u)$, donde

$$\begin{aligned} C(u) &= \frac{1}{2} \sum_{l=1}^m \sum_{i, i' \in \Omega} |u(i, i') - v^l(i, i')| \\ &= \sum_{l=1}^m d(P^*, P_l). \end{aligned}$$

La partición central minimiza el número de desacuerdos promedio entre las m particiones P_1, \dots, P_m .

Observación:

Una primera aplicación que hizo Régnier sobre este concepto fue la siguiente: sobre Ω se miden p variables cualitativas, cada variable v_j define una partición P_j según sus modalidades. Entonces se busca la partición central P^* que minimice

$$C(P) = \frac{1}{p} \sum_{j=1}^p d(P, P_j),$$

donde d es la distancia entre las respectivas relaciones de equivalencia.

4.4 Heurística de Celeux

El principal obstáculo que presenta el cálculo de una partición central es que es un problema NP-completo [1] y los algoritmos que se han propuesto son muy costosos. Por ello, Celeux [3] propuso el siguiente algoritmo bastante rápido y que da buenos resultados.

Sea $A = (a_{ii'})$ la matriz $n \times n$, donde su entrada $a_{ii'}$ se define como

$$a_{ii'} = a(i, i') = \sum_{l=1}^m v^l(i, i'),$$

esto es, el número de veces en que i e i' fueron clasificados juntos en la multi-partición.

Supongamos que m es impar: $m = 2p + 1$, con $p \in \mathbb{Z}$. El caso en que m es par es totalmente análogo. Definimos las siguientes relaciones binarias sobre Ω :

$$u^0(i, i') = \begin{cases} 1 & \text{si } a(i, i') \geq p + 1 \\ 0 & \text{si } a(i, i') < p + 1. \end{cases}$$

Para $j = 1, \dots, p$ definimos w^j :

$$w^j(i, i') = \begin{cases} 1 & \text{si } a(i, i') \geq p + 1 + j \\ 0 & \text{si } a(i, i') < p + 1 + j. \end{cases}$$

Para $j = 1, \dots, p + 1$ definimos u^{-j} :

$$u^{-j}(i, i') = \begin{cases} 1 & \text{si } a(i, i') \geq p + 1 - j \\ 0 & \text{si } a(i, i') < p + 1 - j. \end{cases}$$

Obsérvese que en general estas relaciones no son transitivas, salvo u^p que define las formas fuertes y $u^{-(p+1)}$ que es Ω , la partición gruesa. Si fueran transitivas, es de esperarse que la solución óptima fuese u^0 , pero ésta en general no cumple la transitividad. Puede verse [3,8] que

$$\begin{array}{ccccccc} C(u^1) & \leq & C(u^2) & \leq & \dots & \leq & C(u^p) \\ C(u^{-1}) & \leq & C(u^{-2}) & \leq & \dots & \leq & C(u^{-(p+1)}). \end{array}$$

Dado w^j , se define γ^j como su cerradura transitiva¹ y Γ^j su partición asociada.

Algoritmo:

1. Calcular $a(i, i')$.
2. Para $j = -(p + 1), \dots, 0, \dots, p$, determinar w^j .
3. Para $j = -(p + 1), \dots, 0, \dots, p$, determinar γ^j . Si se llega a la partición gruesa Ω , detenerse.
4. Calcular $C(\gamma^j)$.
5. La partición central Γ^* está asociada a la relación de equivalencia γ^* que verifica $C(\gamma^*) = \min\{C(\gamma^j) : j = -(p + 1), \dots, p\}$.

¹La *cerradura transitiva* de una relación reflexiva y simétrica es la intersección de todas las relaciones de equivalencia que la contienen. Es claro que la cerradura transitiva es a su vez una relación de equivalencia.

5 Clasificación optimal

En vista de que el problema de clasificar un grupo de individuos Ω puede verse como el de optimizar un criterio W , real, definido sobre particiones de Ω , poco práctico sería evaluar todas y cada una de las particiones de Ω para hallar una partición óptima P^* en la que se minimice (o maximice, según sea el caso) W , pues ello puede conducir a una cantidad monstruosa de cálculos.

Para tener una mejor idea de esto último, si denotamos por $S(n, k)$ el número de particiones de n objetos en k clases no vacías (números de Stirling de segundo orden), y $B(n)$ el número de particiones no vacías de n objetos (números de Bell), se tienen las conocidas fórmulas

$$S(n, k) = \frac{1}{k!} \sum_{i=0}^k (-1)^{k-i} \frac{k!}{i!(k-i)!} k^i$$

$$B(n) = e^{-1} \sum_{k=1}^{\infty} \frac{k^n}{k!}$$

En la Figura 6 se ilustra el desmesurado crecimiento de los números $B(n)$ [12], mientras que por ejemplo, $S(60, 2) \approx 10^{18}$, y $S(100, 5) \approx 10^{68}$.

n	$B(n)$
4	15
10	115945
15	1382958545
20	$\approx 5.7 \times 10^{13}$
40	$\approx 1.57 \times 10^{35}$
50	$\approx 1.85 \times 10^{47}$

Figura 6: Crecimiento de los números $B(n)$ de particiones de un conjunto Ω de n objetos (números de Bell).

Tal limitación hace necesarios métodos con capacidad de dar clasificaciones óptimas trabajando apenas con un cierto número de particiones posibles de Ω .

La mayor parte de las veces, el criterio a optimizar W depende tan sólo de P , en cuyo caso puede denotarse por $W(P)$. Por ejemplo, el criterio asociado a las nubes dinámicas con centros de gravedad es la inercia intra-clase de la partición. De todas maneras, el problema de optimización no está resuelto más que en algunos casos simples [4].

Observación:

La búsqueda de un árbol de longitud mínima (Ω, A) permite resolver este problema cuando $W(P_i) = \max\{d(x, y) : x, y \in P_i\}$, con

$$W(P) = \sum_{i=1}^k W(P_i).$$

En este caso se resuelve el problema extrayendo de A las $k - 1$ aristas más grandes [12].

Interesan aquí los casos en que $W(P)$ se pueda descomponer como en la fórmula anterior, con $W(P_i)$ que no dependa más que de las distancias $d(x, y)$, con $x, y \in P_i$.

Es bastante útil, por tanto, definir y usar métodos que reduzcan el conjunto de particiones. Uno de estos métodos es el algoritmo de Fisher (1958) [9], que puede ubicarse dentro de la *programación dinámica*.

5.1 Algoritmo de Fisher

Usa el criterio intra-clase y se fundamenta en el orden total asociado a la variable cuantitativa (o cualitativa-ordinal) inducido por ésta sobre el conjunto de individuos. La partición óptima deberá ser compatible con éste orden.

Más específicamente, si $v: \Omega \mapsto \mathbf{R}$ es la variable, sean $v_1 < v_2 < \dots < v_m$ los m valores diferentes que toma v ($m \leq n$). Naturalmente suponemos $k < m$, donde k es el número de clases de la partición óptima buscada.

Si dos individuos $\omega, \omega' \in \Omega$ son tales que $v(\omega) = v(\omega')$, estos deben quedar en una misma clase y , y para efectos de clasificación pueden considerarse como uno solo objeto. De esta manera, el problema se reduce a hallar una partición $P = (P_1, \dots, P_k)$ de $\{\omega_1, \dots, \omega_n\}$ que minimice la inercia intra-clase de P :

$$W(P) = \sum_{i=1}^k \sum_{v_j \in P_i} p_j (v_j - g(P_i))^2 = \sum_{i=1}^k I(P_i),$$

con p_j el número de individuos ω tales que $v(\omega) = v_j$, y $g(P_i)$ el centro de gravedad del conjunto de números v_j . Para simplificar, en lugar de $\{\omega_1, \dots, \omega_n\}$ escribimos $\{1, \dots, m\}$.

De manera recursiva se calcula una sucesión de particiones optimales P_l^i del conjunto inicial $\{1, \dots, m\}$ en l clases, para finalizar con P_k^1 , la partición óptimal de Ω buscada.

Algoritmo:

1. Para $i = 1, \dots, m$, se define $P_1^i = \{i, i + 1, \dots, n\}$.

2. Para $l = 2, \dots, k - 1$, se calcula la partición $P_l^i = (\{i, i + 1, \dots, j\}, P_{l-1}^{j+1})$ de $\{i, \dots, m\}$, donde P_{l-1}^{j+1} ha sido calculada en el paso $j - 1$ éximo, $j \in \{i, \dots, m - l + 1\}$ y que minimice $I(\{1, \dots, j\}) + W(P_{l-1}^{j+1})$.
3. Finalmente debe construirse una partición $P_k^1 = (\{1, \dots, i\}, P_{k-1}^{i+1})$ y que minimice $I(\{1, \dots, j\}) + W(P_{k-1}^{i+1})$.

La demostración de la optimalidad del algoritmo de Fisher descansa en los dos lemas que siguen, que presentamos sin demostración.

Lema 1 Una partición P^* que minimiza W está formada por clases contiguas. Es decir, si dos individuos ω_i, ω_j de una misma clase de P^* toman los valores $v(\omega_i), v(\omega_j)$, entonces no hay individuos externos a esta clase que tomen su valor de v en el intervalo $[v(\omega_i), v(\omega_j)]$.

Lema 2 Si una partición $P^* = (P_1^*, P_2^*, \dots, P_k^*)$ optimiza W en Ω , entonces (P_2^*, \dots, P_k^*) optimiza W en $\Omega \setminus P_1^*$.

La siguiente proposición, basada en los lemas 1 y 2 asegura que tal algoritmo construye una partición optimal de Ω [6,8].

Proposición 1 Para $l = 1, \dots, k$, P_l^1 es una partición de $\{1, \dots, m\}$ en l clases de inercia intra-clase mínima.

Comentarios:

1. No siempre tiene sentido el criterio cuando las contribuciones de cada clase son muy pequeñas, comportamiento que tienden a tener las clases formadas por muy pocos elementos. En este caso podría pedirse un mínimo de elementos por clase [12].
2. Este algoritmo es de complejidad $(k - 1)m^2$, y la solución que da no es necesariamente única.

Referencias

- [1] Benzêcri: *La Taxonomie*. Ed. Dunod, 4a. edición, Paris, 1984.
- [2] Caillez y Pagés: *Introduction à l'Analyse de Données*. SMASH, Paris, 1976.
- [3] Celeux: *Approximation rapide et interprétation d'une partition centrale pour les algorithmes de partitionnement*. Rapport de Recherche 301, INRIA, Le Chesnay, 1984.

- [4] Celeux: *Etude exhaustive de l'algorithme de reallocation-recentrage dans un cas simple*. RAIRO, Recherche Operationnelle, vol. 30, pág. 229-243, 1986.
- [5] Charles: *Régression typologique et reconnaissance des formes*. Thèse de 3ème cycle, Université Paris IX, 1977.
- [6] Diday et al.: *Eléments d'analyse des données*. Ed. Dunod, Paris, 1982.
- [7] Diday y Schroeder: *A new approach in mixed distributions detection*. Rapport de Recherche 52, INRIA, Le Chesnay, 1974.
- [8] Espinoza; Mora; Trejos: *Clasificación automática*. Memorias de Seminario de Graduación, Universidad de Costa Rica, San José, 1987.
- [9] Fisher, W. D.: *On grouping for maximum homogeneity*. Journal American Statistics Association, No. 53, pág. 789-798, 1958.
- [10] Forgy: *Cluster analysis of multivariate data: efficiency versus interpretability of classifications*. Biometrics No. 21, pág. 768-769, 1965.
- [11] Govaert: *Classification automatique et distances adaptatives*. Thèse de 3ème cycle, Université Paris VI, 1975.
- [12] Govaert y Lechevallier: *Classification automatique non hiérarchique*. Ecole d'Eté en Analyse de Données, CIMPA, Nice, 1987.
- [13] Hartigan: *Clustering algorithms*. John Wiley & Sons, New York, 1975.
- [14] Jambu: *Classification automatique pour l'Analyse des Données*. Ed. Dunod, Paris, 1978.
- [15] Lerman: *Classification et analyse ordinale de données*. Ed. Dunod, Paris, 1981.

Un enfoque unificado de los métodos de estadística multivariada

Asdrubal Duarte
Jorge González*

Resumen

En estadística multivariada encontramos una serie de métodos que nos permiten abordar problemas de naturaleza un tanto diversa. Es usual encontrar presentaciones, con detalle, de métodos como: *regresión lineal multivariada*, *análisis en componentes principales*, *análisis canónico*, *análisis discriminante*, etc.; presentaciones que, aparentemente, abordan los diferentes métodos desde perspectivas teóricas un tanto diferenciadas, o que por lo menos no dejan entrever una ligazón estrecha entre esas distintas técnicas.

El objetivo central de este trabajo es hacer una presentación unificada de algunos de estos métodos clásicos de la estadística multivariada. Esta unificación consiste, fundamentalmente, en evidenciar que todos estos métodos mencionados anteriormente pueden interpretarse, desde el punto de vista teórico, como la búsqueda de una transformación lineal sobre la tabla de datos originales que minimice, bajo ciertas restricciones particulares a cada técnica, las *cercanías* entre las configuraciones de puntos correspondientes a la tabla X y la tabla transformada $Y = L'X$.

1 Introducción

Antes de entrar propiamente en lo que es la presentación de los distintos métodos con esta óptica unificadora, trataremos de esclarecer el concepto de *cercanía* entre conformaciones, y muy especialmente nos daremos a la tarea de presentar con todo detalle un coeficiente que permita medir esta *cercanía* o similitud entre configuraciones.

Generalmente cuando se miden p variables cuantitativas sobre una muestra dada de n individuos, los datos se presentan mediante la forma de una tabla o matriz $X = (x_{ij})$ de tamaño $p \times n$, como se muestra en la Figura 1.

*Profesores de la Escuela de Matemáticas de la Universidad de Costa Rica.

	1	...	i	...	n
1	x_1^1	...	x_1^i	...	x_1^n
\vdots					
j	x_j^1	...	x_j^i	...	x_j^n
\vdots					
p	x_p^1	...	x_p^i	...	x_p^n

Figura 1: Tabla de datos.

El vector

$$x^i = \begin{pmatrix} x_1^i \\ x_2^i \\ \vdots \\ x_j^i \\ \vdots \\ x_p^i \end{pmatrix}$$

es la columna i -ésima de X , y representa geoméricamente en \mathbf{R}^p el individuo i -ésimo.

Con esta interpretación el conjunto de individuos queda representado en el espacio \mathbf{R}^p como una nube o conjunto de n puntos, a la cual denominaremos como la configuración asociada a la matriz de datos $X_{p \times n}$ y la denotaremos $C(X)$. En múltiples ocasiones conviene particionar los individuos en clases de elementos próximos, razón por la cual precisar el sentido de la frase "el individuo i está próximo al individuo j " se hace estrictamente necesario. La manera de medir proximidades entre individuos puede variar de manera sustancial de acuerdo al objetivo que se persigue. Es así como, en el caso de la *clasificación automática*, esta proximidad se mide con la ayuda de índices de disimilitud o similitud. Sin embargo para el interés nuestro, es decir, para el análisis lineal, las proximidades se medirán utilizando distancias euclídeas.

La distancia entre el i -ésimo individuo y el k -ésimo se define utilizando una matriz semi-definida positiva Q , de la manera siguiente:

$$d(x^i, x^k) = \sqrt{(x^i - x^k)' Q (x^i - x^k)}.$$

Usualmente Q es una matriz diagonal que tiene como entradas las inversas

de las varianzas:

$$Q = \begin{pmatrix} 1/\sigma_{11} & & \\ & \ddots & \\ & & 1/\sigma_{pp} \end{pmatrix} = D_{1/\sigma},$$

o bien,

$$Q = V^{-1} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix},$$

lo cual hace que la distancia sea independiente de una transformación lineal de los datos.

De acuerdo a un resultado clásico de matrices, y dado que Q es semi-definida positiva, existen $L_{p \times q}$ tal que $Q = LL'$. Realizando el cambio de variable $Y = L'X$, permite usar la distancia euclídea clásica, esto es, $Q = I$. En efecto:

$$\begin{aligned} \mathbb{R}^p &\longrightarrow \mathbb{R}^q \\ x^j &\longmapsto L'x^j = y^j \end{aligned}$$

de donde,

$$\begin{aligned} [d(x^j, x^k)]^2 &= (x^j - x^k)'Q(x^j - x^k) = (x^j - x^k)'LL'(x^j - x^k) \\ &= [L'(x^j - x^k)]' [L'(x^j - x^k)] = (y^j - y^k)'(y^j - y^k). \end{aligned}$$

2 Definición del coeficiente RV

Consideremos una muestra dada de n individuos y dos grupos de p y q variables respectivamente, que pueden ser total o parcialmente distintas, medidas sobre la muestra dada. Sean $X_{p \times n}$ y $Y_{q \times n}$ las matrices de datos correspondientes, y $C(X)$, $C(Y)$ las configuraciones asociadas en \mathbb{R}^p y \mathbb{R}^q respectivamente.

Como vimos anteriormente, para medir proximidades o posiciones relativas entre puntos de una misma configuración, por ejemplo $C(X)$, utilizamos la matriz $D(X) = (d_{jk})$, donde $d_{jk} = \frac{(x^j - x^k)'(x^j - x^k)}{\sqrt{(x^j - x^k)'(x^j - x^k)}}$, la cual resulta invariante bajo traslaciones y rotaciones. Sin embargo, en el caso en que las variables están centradas (media 0) es preferible utilizar la matriz $S(x) / \text{tr } S^2(X)$, donde $S(X) = X'X$, matriz que también resulta ser invariante bajo rotaciones y traslaciones y en la que el término que aparece en el denominador hace que sea independiente de los cambios de escala.

Debido a que $S(X) = X'X = (S_{ik})$, resulta que

$$\text{tr } S^2(X) = \sum_i (S^2(X))_{ii} = \sum_i \sum_k S_{ik} S_{ki} = \sum_{i,k} S_{ik}^2.$$

Usando el producto escalar entre matrices, definido por $\langle A, B \rangle = \text{tr}(A'B)$, se tiene que $\|A\| = \sqrt{\text{tr} A'A}$, con lo cual resulta que:

$$\left\| \frac{S(X)}{\text{tr} S^2(X)} \right\| = \left\| \frac{S(X)}{\text{tr} S'(X)S(X)} \right\| = 1. \quad (1)$$

Por otra parte, a cualquier configuración $C(X)$ se le puede asociar de manera única la matriz de distancia entre individuos $S(X)/\text{tr} S^2(X)$. Así el problema de medir la distancia entre dos configuraciones se transforma en un problema de medir la distancia entre matrices. Con esta identificación en mente, podemos definir:

$$d(C(X), C(Y)) = \left\| \frac{S(X)}{\text{tr} S^2(X)} - \frac{S(Y)}{\text{tr} S^2(Y)} \right\|$$

Haciendo $A = S(X)/\text{tr} S^2(X)$ y $B = S(Y)/\text{tr} S^2(Y)$, se obtiene de la fórmula (1):

$$\begin{aligned} d(C(X), C(Y)) &= \|A - B\| = \sqrt{\text{tr}(A - B)'(A - B)} \\ &= \sqrt{\text{tr} A'A - 2\text{tr} AB + \text{tr} B'B} \\ &= \sqrt{2 - 2\text{tr} AB} = \sqrt{2} \sqrt{1 - \text{tr} AB} \\ &= \sqrt{2} \sqrt{1 - \frac{\text{tr} S(X)S(Y)}{\text{tr} S^2(X) \text{tr} S^2(Y)}} \end{aligned}$$

Definiendo $\text{RV}(X, Y)$ como

$$\text{RV}(X, Y) = \frac{\text{tr} S(X)S(Y)}{\text{tr} S^2(X) \text{tr} S^2(Y)},$$

obtenemos:

$$d(C(X), C(Y)) = \sqrt{2} \sqrt{1 - \text{RV}(X, Y)}. \quad (2)$$

Detallando un poco más $\text{RV}(X, Y)$, se obtiene:

$$\begin{aligned} \text{RV}(X, Y) &= \frac{\text{tr} S(X)S(Y)}{\text{tr} S^2(X) \text{tr} S^2(Y)} = \frac{\text{tr}(X'XY'Y)}{(\text{tr}(X'X)^2 \text{tr}(Y'Y)^2)} \\ &= \frac{\text{tr}(XY'YX')}{(\text{tr}(XX')^2 \text{tr}(YY')^2)}. \end{aligned}$$

Ahora, escribiendo $S_{11} = XX'$, $S_{22} = YY'$, $S_{12} = XY'$, $S_{21} = YX'$, resulta:

$$\text{RV}(X, Y) = \frac{\text{tr}(S_{12}S_{21})}{\text{tr} S_{11}^2 \text{tr} S_{22}^2} = \frac{\langle S_{21}, S_{21} \rangle}{\|S_{11}\| \|S_{22}\|} = \frac{\|S_{21}\|^2}{\|S_{11}\| \|S_{22}\|}.$$

De acuerdo a la fórmula (2), vemos que el coeficiente $\text{RV}(X, Y)$ puede usarse como una medida de la cercanía entre las configuraciones $C(X)$ y $C(Y)$, y

además se puede observar que este coeficiente varía entre 0 y 1. Conviene observar que el coeficiente RV es fácilmente calculable a partir de las tablas de datos.

Seguidamente haremos la presentación de diferentes métodos del análisis lineal multivariado, y trataremos a la vez de establecer con detalle la naturaleza teórica común de todos ellos.

3 Análisis en Componentes Principales (ACP) de X

Sea $X_{p \times n}$ la matriz de datos correspondiente a la medición de p variables sobre una muestra de n individuos. En ACP clásico se trata de encontrar t nuevas variables, con $t < p$, que sean combinaciones lineales de las p variables originales, no correlacionadas y de varianza máxima. El problema anterior se puede replantear de la manera siguiente:

Encontrar t nuevas variables combinaciones lineales de las variables originales, esto es, buscar una transformación lineal L de los datos originales, de modo que la distancia entre las configuraciones de X y $Y = L'X$ sea mínima, siempre que las variables sean no correlacionadas.

Sean $C(X)$ y $C(Y)$ las configuraciones correspondientes a X y $Y = L'X$ respectivamente. De acuerdo a (2), se tiene que

$$d(C(X), C(Y)) = \sqrt{2} \sqrt{1 - RV(X, Y)}.$$

Luego, para que la distancia entre ambas configuraciones sea mínima, basta que $RV(X, Y)$ sea máximo. Es decir se trata de maximizar la función $f(L) = \max RV(X, L'X)$, sujeto a la restricción de que las nuevas variables no estén correlacionadas, o equivalentemente, $(L'X)(L'X)' = L'S_{11}L = D$, donde D es una matriz diagonal de dimensión $t \times t$, con t no excediendo el número de valores propios no nulos de S_{11} .

Resumiendo, el problema a resolver es:

$$\begin{cases} \max_L \frac{\text{tr}(S_{11}LL'S_{11})}{\text{tr}(S_{11}^2) \text{tr}(L'S_{11}L)^2} \\ \text{con } (L'X)(L'X)' = D \end{cases}$$

Usando multiplicadores de Lagrange, el problema se transforma en maximizar la función Φ definida por

$$\Phi(L) = \text{tr}(S_{11}LL'S_{11}) - \sum_i \lambda_i [L'S_{11}L]_{ii},$$

bajo la restricción $(L'X)(L'X)' = D = (\delta_i)$, D fijo. Aquí λ_i ($i = 1, \dots, t$) son los multiplicadores de Lagrange.

Conviene aquí, hacer la siguiente observación: Debido a que $L'S_{11}L = D = (\delta_i)$ y si fijamos D , la expresión $\frac{\text{tr}(S_{11}^2)}{\text{tr}(L'S_{11}L)^2}$ es independiente de L , y por esa razón el denominador se desprecia a la hora de efectuar la maximización.

Seguidamente derivaremos respecto a L la función $\Phi(L)$. Sin embargo antes de entrar propiamente a la derivación, se harán algunos detalles previos:

$$\begin{aligned} \text{tr}(S_{11}LL'S_{11}) &= \sum_{i,k} \left(\sum_t (S_{11})_{it} L_{tk} \right)^2 = \sum_i \lambda_i (L'S_{11}L)_{ii} \\ &= \sum_i \lambda_i \sum_k L'_{ik} (S_{11}L)_{ki} = \sum_i \sum_k L'_{ik} \sum_t (S_{11})_{kt} L_{ti} \\ &= \sum_i \sum_k \sum_t \lambda_i L_{ki} L_{ti} (S_{11})_{kt}. \end{aligned}$$

Calculando la entrada uv de $\frac{\partial}{\partial L} \text{tr}(S_{11}LL'S_{11})$, se obtiene:

$$\begin{aligned} \sum_{i,k} \frac{\partial}{\partial L_{uv}} \left(\sum_t (S_{11})_{it} L_{tk} \right)^2 &= \sum_{i,k} \left(2 \sum_t (S_{11})_{it} L_{tk} \right) \frac{\partial}{\partial L_{uv}} \sum_t (S_{11})_{it} L_{tk} \\ &= 2 \sum_i \sum_t (S_{11})_{it} L_{tv} (S_{11})_{iv} \\ &= 2 \sum_i (S_{11}L)_{iv} (S_{11})_{iu}, \end{aligned}$$

puesto que los únicos términos no nulos corresponden a los casos $t = u$ y $k = v$. Nótese además que debido a la simetría de S_{11} se cumple que

$$2 \sum_i (S_{11}L)_{iv} (S_{11})_{iu} = 2 \sum_i (S_{11})_{ui} (S_{11}L)_{iv} = 2(S_{11}^2 L)_{uv}.$$

Calculando ahora la entrada uv de la matriz, se obtiene:

$$\frac{\partial}{\partial L} \sum_i \lambda_i (L'S_{11}L)_{ii} = \frac{\partial}{\partial L} \sum_i \sum_k \sum_t \lambda_i L_{ki} L_{ti} (S_{11})_{kt},$$

de donde:

$$\begin{aligned} \frac{\partial}{\partial L_{uv}} \left(\sum_i \lambda_i (L'S_{11}L)_{ii} \right) &= \sum_t \lambda_v L_{tv} (S_{11})_{ut} + \sum_k \lambda_v L_{kv} (S_{11})_{ku} \\ &= \lambda_v \left(\sum_t (S_{11})_{ut} L_{tv} + \sum_k (S_{11})_{uk} L_{kv} \right) \\ &= \lambda_v \{ (S_{11}L)_{uv} + (S_{11}L)_{uv} \} = 2(S_{11}L\Delta)_{uv}, \end{aligned}$$

con $\Delta = \text{Diag}(\lambda_i)$ una matriz diagonal de tamaño $t \times t$.

Finalmente se obtiene la fórmula

$$\frac{\partial}{\partial L} \Phi(L) = 2(S_{11}^2 L - S_{11} L \Delta).$$

Así, una solución para $\frac{\partial}{\partial L} \Phi(L) = 0$ se obtiene para $S_{11} L = L \Delta$. Eligiendo L como la matriz cuyas columnas son t vectores propios de S_{11} normalizados de modo que satisfagan $L' S_{11} L = D$, obtenemos:

$$\text{tr}(S_{11} L L' S_{11}) = \text{tr}(L' S_{11}^2 L) = \text{tr}(L' S_{11} L \Delta) = \sum_{i=1}^t \alpha_i \lambda_i,$$

y se obtiene por tanto

$$\begin{aligned} \text{RV}(X, L' X) &= \frac{\sum_{i=1}^t \alpha_i \lambda_i}{\{\sum_{i=1}^p \lambda_i^2\}^{1/2} \{\sum_{i=1}^t \alpha_i^2\}^{1/2}} \\ &= \frac{\{\sum_{i=1}^t \lambda_i^2\}^{1/2}}{\{\sum_{i=1}^p \lambda_i^2\}^{1/2}} \cdot \frac{\sum_{i=1}^t \alpha_i \lambda_i}{\{\sum_{i=1}^t \lambda_i^2\}^{1/2} \{\sum_{i=1}^t \alpha_i^2\}^{1/2}}. \end{aligned}$$

El segundo cociente es el coseno del ángulo entre los vectores $(\alpha_1, \dots, \alpha_t)$ y $(\lambda_1, \dots, \lambda_t)$, por lo que una elección óptima para α_i es $\alpha_i = \lambda_i$, y en tal caso

$$\text{RV}(X_1 L' X) = \left(\sum_{i=1}^t \lambda_i / \sum_{i=1}^p \lambda_i^2 \right)^{1/2}.$$

Claramente el máximo global se obtiene eligiendo los λ_i como los t valores propios mayores, esto es $\lambda_1 \geq \lambda_2 \dots \geq \lambda_t$.

4 Componentes Principales de Y con respecto a X

En esta sección se aborda el problema siguiente: dadas dos tablas de datos $X_{p \times n}$ y $Y_{q \times n}$ sobre una misma muestra de individuos, encontrar combinaciones lineales de las variables definidas en la tabla Y ; es decir encontrar la matriz M' de manera que la distancia entre $C(X)$ y $C(M'Y)$ sea mínima.

El interés práctico de la solución del problema anterior radica en el hecho de que en algunas situaciones se quiere estudiar las variables de X , pero sin embargo resulta mucho más fácil la observación de las variables de Y .

Las nuevas variables obtenidas de Y bajo la transformación M' se denominan *Componentes Principales* de Y con respecto a X . Se trata aquí de maximizar respecto a M , la expresión:

$$\text{RV}(X, M'Y) = \frac{\text{tr}(XY' M M' Y X')}{\text{tr}(X X')^2 \text{tr}(M' Y Y' M)^2},$$

sujeto a la restricción de que las variables de $M'Y$ sean no correlacionadas. En otras palabras se trata de maximizar

$$RV(X, M'Y) = \frac{\text{tr}(S_{12}MM'S_{21})}{\text{tr}S_{11}^2 \text{tr}(M'S_{22}M)^2},$$

sujeto a $M'YY'M = M'S_{22}M = D = \text{Diag}(\alpha_i)$, y fijando D . Utilizando multiplicadores de Lagrange (λ_i), la función a maximizar resulta ser:

$$\Phi(M) = \text{tr}(S_{12}MM'S_{21}) - \sum_i \lambda_i (M'S_{22}M)_{ii}.$$

Derivando con respecto a M tenemos que la entrada uv de la matriz

$$\frac{\partial}{\partial M} \sum_i \lambda_i (M'S_{22}M)_{ii}$$

es:

$$\begin{aligned} \frac{\partial}{\partial M_{uv}} \sum_i \lambda_i (M'S_{22}M)_{ii} &= \frac{\partial}{\partial M_{uv}} \sum_{i,k,t} \lambda_i M_{ki} (S_{22})_{kt} M_{ti} \\ &= \sum_t \lambda_v (S_{22})_{ut} M_{tv} + \sum_k \lambda_v M_{kv} (S_{22})_{ku} \\ &= [(S_{22}M)_{uv} + (S_{22}M)_{vu}] \lambda_v. \end{aligned}$$

Luego, $\frac{\partial}{\partial M} \sum_i \lambda_i (M'S_{22}M)_{ii} = 2S_{22}M\Delta$, con $\Delta = \text{Diag}(\lambda_i)$. La entrada uv de la matriz $\frac{\partial}{\partial M} \text{tr}(S_{12}MM'S_{21})$ es:

$$\begin{aligned} \frac{\partial}{\partial M_{uv}} \text{tr}(S_{12}MM'S_{21}) &= \frac{\partial}{\partial M_{uv}} \sum_{i,k} \left(\sum_t (S_{12})_{it} M_{tk} \right)^2 \\ &= 2 \sum_{ik} \left(\sum_t (S_{12})_{it} M_{tk} \right) \sum_t \frac{\partial}{\partial M_{uv}} (S_{12})_{it} M_{tk} \\ &= 2 \sum_{i,t} (S_{12})_{it} M_{tv} (S_{12})_{iu} \\ &= 2 \sum_{i,t} (S_{21})_{ui} (S_{12})_{it} M_{tv} \\ &= 2 \sum_t (S_{21}S_{12})_{ut} M_{tv} = 2(S_{21}S_{12}M)_{uv}. \end{aligned}$$

Por lo tanto, $\frac{\partial}{\partial M} \text{tr}(S_{12}MM'S_{21}) = 2S_{21}S_{12}M$, resultando entonces que

$$\frac{\partial}{\partial M} \Phi(M) = 2(S_{21}S_{12}M - S_{22}M\Delta).$$

Así, el máximo se alcanza cuando $S_{21}S_{12}M - S_{22}M\Delta = 0$. Se trata entonces de encontrar una matriz M que satisfaga las ecuaciones

$$M'S_{22}M = D \quad (3)$$

$$S_{21}S_{12}M - S_{22}M\Delta = 0 \quad (4)$$

Sea v_j la columna j -ésima de M . Luego se tiene que

$$\begin{aligned} S_{21}S_{12}v_j &= \lambda_j S_{22}v_j \\ (S_{21}S_{12} - \lambda_j S_{22})v_j &= 0, \end{aligned}$$

con lo cual resulta que v_j es un vector propio de la matriz $S_{21}S_{12}$ respecto de la matriz S_{22} , asociado al valor propio λ_j , normalizado de modo que satisfaga (3).

Sean $\lambda_1, \dots, \lambda_t$, t valores propios distintos de cero de la matriz $S_{21}S_{12}$ respecto a S_{22} , y v_1, \dots, v_t los correspondientes vectores propios elegidos de modo que satisfaga las ecuaciones

$$\begin{aligned} v_j'S_{22}v_j &= \alpha_j \\ v_j'S_{22}v_i &= 0, \quad i \neq j. \end{aligned}$$

Definiendo M como la matriz cuyas columnas son los v_j , se satisfacen las condiciones deseadas. Además:

$$\begin{aligned} \text{tr}(S_{12}MM'S_{21}) &= \text{tr}(S_{21}S_{12}MM') = \text{tr}(S_{22}M\Delta M') \\ &= \text{tr}(M'S_{22}M\Delta) = \sum_{i=1}^t \lambda_i \alpha_i. \end{aligned}$$

Por lo tanto, se obtiene:

$$RV(X, M'Y) = \frac{\sum_{i=1}^t \lambda_i \alpha_i}{\text{tr} S_{11} \sum_{i=1}^t \lambda_i^2}^{1/2},$$

expresión que alcanza su máximo cuando $\lambda_i = \alpha_i$. Así, para esta elección de M se tiene que el máximo absoluto de $RV(X, M'Y)$ es

$$|RV(X, M'Y)| = \left(\sum_{i=1}^t \lambda_i^2 / \text{tr} S_{11} \right)^{1/2},$$

con $\lambda_1 \geq \lambda_2 \dots \geq \lambda_t$.

5 Regresión multivariada

La regresión multivariada expresa un conjunto de p variables, medidas sobre n individuos, como una combinación lineal de q variables medidas sobre los mismos n individuos, con $n \geq p + q$.

Matricialmente el modelo se plantea como:

$$X = M'Y + U, \quad \text{donde:}$$

X : Es una matriz $p \times n$, que representa las respuestas de n individuos a p variables.

Y : Es una matriz $q \times n$, que contiene q variables independientes, medida sobre los mismos n individuos.

U : Es una matriz de errores, de orden $p \times n$, cuyas filas se suponen no correlacionadas y que siguen una distribución normal de media 0 y varianza común Σ , $U \sim N_p(0, \Sigma)$

M' : Es la matriz de parámetros a estimar y es de tamaño $p \times q$.

Si suponemos que Y es de rango completo, esto el rango de Y es igual a q , entonces existe la inversa $(YY')^{-1}$ y podemos definir el proyector:

$$P : \mathbb{R}^n \longrightarrow W^\perp,$$

con $P = I - Y'(YY')^{-1}Y$, donde $W = \text{gen}\{Y^1, \dots, Y^q\}$ es el subespacio generado por los vectores fila de Y . Se tiene entonces que la proyección de X es

$$X = X(I - P) = XY'(YY')^{-1}Y,$$

y la matriz de errores se estima como

$$U = (X - M'Y) = XP$$

$$M' = XY'(YY')^{-1} = S_{12}S_{22}^{-1}.$$

El modelo anterior lo podemos enfocar desde el punto de vista del RV-coeficiente como sigue:

Encontrar una matriz M , de modo que las configuraciones $C(X)$, $C(M'Y)$ sean lo más cercanas posible, que coincida con el problema de encontrar las componentes principales de Y respecto a la matriz X , para el caso particular en que $S_{22} = YY'$ es invertible.

Se busca una matriz M que satisfaga el sistema de ecuaciones

$$M'S_{22}M = D \quad (5)$$

$$S_{21}S_{12}M - S_{22}M\Delta = 0. \quad (6)$$

Sea H la matriz que diagonaliza ortogonalmente a la matriz simétrica

$$S_{12}S_{22}^{-1}S_{21}.$$

Es decir, H verifica

$$HS_{12}S_{22}^{-1}S_{21}H' = \Lambda = \text{Diag}(\alpha_i).$$

Definamos $M' = HS_{12}S_{22}^{-1}$ y hagamos $D = \Delta = \text{Diag}(\alpha_i)$. Verifiquemos que satisfacen las ecuaciones (5) y (6). En efecto:

$$\begin{aligned}M'S_{22}M &= HS_{12}S_{22}^{-1}S_{22}S_{22}^{-1}S_{21}H' \\ &= HS_{12}S_{22}^{-1}S_{21}H' = \Delta,\end{aligned}$$

y

$$\begin{aligned}S_{21}S_{12}M - S_{22}M\Delta &= S_{21}S_{12}S_{22}^{-1}S_{21}H' - S_{22}S_{22}^{-1}S_{21}H'\Delta \\ &= S_{21}H'HS_{12}S_{22}^{-1}S_{21}H' - S_{21}H'\Delta \\ &= S_{21}H'\Delta - S_{21}H'\Delta = 0.\end{aligned}$$

Como el operador $H : \mathbb{R}^p \mapsto \mathbb{R}^p$ es una rotación, y el RV-coeficiente es independiente bajo rotaciones, podemos elegir $M' = S_{12}S_{22}^{-1}$, que es precisamente la matriz de parámetros en la regresión multivariada clásica.

Referencias

- [1] Anderson, T. W.: *Introduction to Multivariate Statistical Analysis*. Wiley & Sons, New York 1958.
- [2] Robert P., Escoufier Y.: *A Unifying tool for linear multivariate statistical methods: The RV-coefficient*. Applied Statistical, Vol. 25, Num. 3, pp. 256-265, 1976.
- [3] Rao, C. R. : *The used and interpretation of principal components analysis in applied research*. Sankhyà, Vol. A, Num. 26, pp. 329-358, 1965.

La métrica inducida por un Análisis en Componentes Principales

Eduardo Piza*

1 Introducción

En Clasificación Automática con frecuencia se trabaja con colecciones de objetos caracterizados por atributos cuantitativos o cualitativos ordinales o atributos dicotómicos indicadores de presencia-ausencia de características específicas. En tales situaciones es recomendable realizar algunos estudios preliminares de orden descriptivo sobre los atributos, que permitan visualizar mejor los datos en general, las relaciones entre los atributos caracterizantes, así como posibles estructuras o sesgos en la matriz de correlación entre los mismos.

Una metodología ampliamente difundida como análisis descriptivo previo a un estudio de Clasificación Automática es el Análisis en Componentes Principales (ACP). Es costumbre de los investigadores realizar un ACP de sus datos sobre el espacio de los individuos (algunos paquetes de software lo permiten, sobre todo los paquetes franceses) y representar en el primer plano principal la nube de "individuos" (en este caso los "individuos" son los objetos a clasificar). Con ello se obtiene la mejor representación euclídea plana de los objetos, dentro de un cierto porcentaje de inercia explicada, como se ilustra en la Figura 1. En el primer plano principal se reflejan las cercanías y lejanías entre los objetos con mayor o menor grado de confiabilidad, dependiendo de la bondad de representación de cada objeto en el plano. Esto permite entonces establecer una agrupación preliminar por particiones de los objetos. El ACP previo al estudio de Clasificación Automática se realiza generalmente por una de las siguientes dos razones, o ambas:

- (a) Emplear las proyecciones de los "individuos" sobre cada uno de los ejes principales (o sobre los primeros ejes) para establecer una medida de disimilitud a emplear (distancia euclídea clásica de las proyecciones) en los estudios de Clasificación Automática.

*Profesor de la Escuela de Matemáticas de la Universidad de Costa Rica.

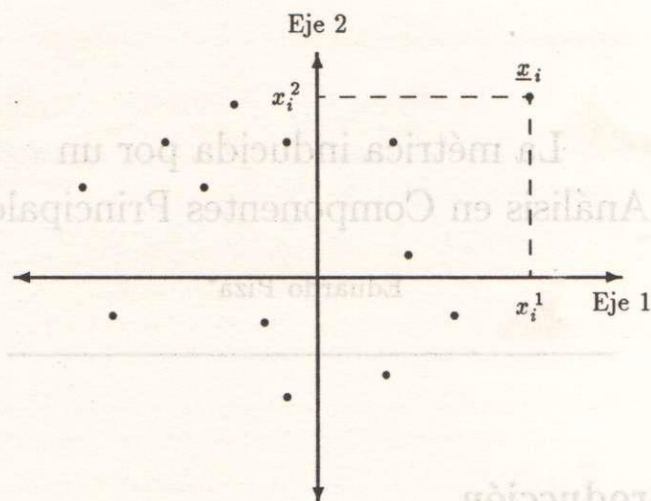


Figura 1: Representación típica del primer plano principal del ACP.

- (b) Emplear las agrupaciones sugeridas por el ACP como criterio de coherencia que ayuda a evaluar la validez de los resultados que se obtengan posteriormente en los análisis de Clasificación Automática, utilizando otras medidas de disimilitud clásicas.

Sin embargo al realizar un ACP previo con sus datos el investigador debe escoger adecuadamente la *métrica* a emplear en el espacio euclídeo de los "individuos". Dependiendo de la métrica escogida variarán sus resultados en el primer plano principal. Existe una costumbre bastante generalizada en nuestro país de establecer la disimilitud entre objetos a clasificar de acuerdo a las *distancias relativas* de los mismos en un mapa típico del primer plano principal de un ACP [5]. Lo anterior implica en la práctica que para cada escogencia de una métrica particular en el ACP se induce una nueva métrica euclídea, la cual es utilizada en los estudios de Clasificación Automática. Cabe entonces preguntarse:

1. ¿Cuál es la forma explícita de la métrica inducida por un ACP? ¿Qué relación tiene con la métrica del ACP?
2. ¿Cómo escoger la métrica en el estudio de ACP, para que la métrica inducida no difiera mucho de la misma?

El presente estudio contesta a estas dos preguntas. Básicamente se obtiene que la "mejor" métrica euclídea para realizar un ACP con fines posteriores de Clasificación Automática es la métrica de Mahalanobis, y no se recomienda el empleo de ninguna otra pues se puede llegar a métricas inducidas sin ningún sentido práctico para fines de Clasificación Automática.

2 Los datos disponibles

Se dispone de una colección $\Omega = \{\omega_1, \dots, \omega_n\}$ de n objetos a clasificar y una *representación euclídea* de los objetos, esto es, una aplicación

$$\begin{aligned} g: \Omega &\longmapsto E = \mathbf{R}^p \\ \omega_i &\longmapsto g(\omega_i) := \underline{x}_i \end{aligned}$$

donde E denota el espacio euclídeo p -dimensional y $p = \dim E$ es la cantidad de atributos (cuantitativos o cualitativos con escala ordinal o dicotómicos 0-1) que son tomados en consideración para representar los individuos. El i -ésimo individuo ω_i se representa en E por \underline{x}_i , donde:

$$\underline{x}_i = \begin{pmatrix} x_i^1 \\ x_i^2 \\ \vdots \\ x_i^p \end{pmatrix} \in E \quad ; \quad i = 1, 2, \dots, n.$$

La matriz de datos es entonces $X_{p \times n} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)$. Supondremos además que $n > p$ y todos los atributos han sido previamente centrados, de manera que X es una matriz centrada por filas.

3 Algunos resultados teóricos en ACP

En $E = \mathbf{R}^p$ podemos representar a los objetos mediante los vectores columna de la tabla de datos centrada X :

$$X = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n) \quad ; \quad \text{donde } \underline{x}_i \in E \text{ es la representación del } i\text{-ésimo objeto.}$$

En $F = \mathbf{R}^n$ podemos representar a los atributos mediante los vectores fila de la tabla de datos centrada X :

$$X = \begin{pmatrix} \underline{x}^1 \\ \underline{x}^2 \\ \vdots \\ \underline{x}^p \end{pmatrix} \quad ; \quad \text{donde } \underline{x}^j \in F \text{ es la representación del } j\text{-ésimo atributo centrado.}$$

Sea M una métrica definida sobre E . Luego tenemos el siguiente esquema de dualidad:

$$\begin{array}{ccc} E = \mathbf{R}^p & \xleftarrow{X} & F^* \\ M \downarrow \uparrow V & & W \downarrow \uparrow D_r \\ E^* & \xrightarrow{X^t} & F = \mathbf{R}^n \end{array}$$

donde:

- D_p es la métrica de los objetos en el espacio F de los atributos:

$$D_p \doteq \text{diagonal}(p_1, p_2, \dots, p_n) \quad ; \quad \text{donde } p_i \text{ es el peso del objeto } \omega_i, \text{ con } \sum_{i=1}^n p_i = 1.$$

- E^* y F^* son los espacios duales algebraicos de E y F respectivamente.
- X^t es la traspuesta de X .
- $V = X \circ D_p \circ X^t$ es la forma cuadrática de inercia o matriz de variancias y covariancias de los p atributos. Es simétrica y semidefinida positiva.
- $W = X^t \circ M \circ X$.

El ACP sobre el espacio de los "individuos" básicamente es el estudio de los valores y vectores propios del operador $V \circ M$:

$$E \xrightarrow{M} E^* \xrightarrow{V} E$$

Sean $\lambda_1, \lambda_2, \dots, \lambda_p$ los p valores propios de $V \circ M$ ordenados en orden decreciente contando posibles repeticiones (todos los λ_i son reales y positivos pues $V \circ M$ es simétrica y definida positiva):

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$$

Sean $\underline{u}_1, \underline{u}_2, \dots, \underline{u}_p$ los vectores propios correspondientes, escogidos de manera que $\{\underline{u}_i : i = 1, \dots, p\}$ es base M -ortonormal de E , esto es:

$$\underline{u}_i^t M \underline{u}_j = \delta_{ij} \quad ; \quad \forall i, \forall j.$$

Luego, los ejes principales del ACP están dados por $\Delta_{u_1}, \Delta_{u_2}, \dots, \Delta_{u_p}$, donde Δ_{u_i} denota la recta engendrada por \underline{u}_i . Cada eje aporta un porcentaje de explicación de la inercia, el cual es igual a $(100 \lambda_i / \sum_{k=1}^p \lambda_k)\%$ para el eje Δ_{u_i} . Los objetos se proyectan en los diversos ejes principales, obteniéndose

$$\underline{x}_i = \sum_{k=1}^p x_i^k \underline{u}_k \quad ; \quad x_i^k = \underline{x}_i^t M \underline{u}_k$$

donde x_i^k es la coordenada de \underline{x}_i en el eje Δ_{u_k} .

El primer plano principal es $\{\Delta_{u_1}, \Delta_{u_2}\}$ y dentro de todos los posibles planos de E es el que mejor representa a los objetos, con un porcentaje de inercia explicada igual a $100 [(\lambda_1 + \lambda_2) / \sum_{k=1}^p \lambda_k]\%$. Mayores referencias sobre esta metodología pueden encontrarse en [1] y [6].

4 Estudio de la métrica inducida por un ACP

Se desea analizar con detalle la métrica inducida por un estudio previo de ACP en el espacio de los "individuos" (objetos). Esta métrica se define como

$$d_M(\underline{x}_i, \underline{x}_j) = \sum_{k=1}^p \lambda_k (x_i^k - x_j^k)^2$$

donde:

- M es la métrica empleada en el ACP.
- λ_k es la inercia explicada por el k -ésimo eje principal Δ_{u_k} del ACP.
- x_i^k y x_j^k son la coordenadas de los objetos \underline{x}_i y \underline{x}_j proyectados sobre el k -ésimo eje principal Δ_{u_k} .

La métrica d_M define a la vez una medida de disimilitud $d_\Omega^{(M)}$ en el conjunto de objetos Ω :

$$d_\Omega^{(M)}(\omega_i, \omega_j) := d_M(\underline{x}_i, \underline{x}_j).$$

Obsérvese que la disimilitud entre los objetos ω_i y ω_j se mide ponderando adecuadamente la distancia entre las proyecciones de los objetos en los ejes principales Δ_{u_k} . El factor de ponderación, λ_k , indica la importancia relativa del eje Δ_{u_k} en el cómputo global de la disimilitud. Esta manera de medir disimilitudes tiene mucho sentido desde el punto de vista geométrico, como se observa en la Figura 2.

Aunque las coordenadas de los objetos en el eje principal Δ_{u_1} se encuentran en la Figura 2 a la misma distancia que las correspondientes en el eje principal Δ_{u_2} , hay una separación mayor entre \underline{x}_3 y \underline{x}_4 que entre \underline{x}_3 y \underline{x}_1 debido a que el primer eje principal Δ_{u_1} aporta el doble de información ($\lambda_1 = 2$) que el segundo eje principal Δ_{u_2} ($\lambda_2 = 1$). Al ponderar cada eje principal Δ_{u_k} por su porcentaje de inercia respectivo se está en realidad cambiando la escala, de manera que los cuatro objetos representados en la Figura 2 luzcan como se muestra en la Figura 3, en la cual hemos re-escalado los ejes de acuerdo a los valores propios λ_1 y λ_2 .

La nueva métrica d_M inducida por el ACP depende completamente de la métrica M utilizada en el análisis. Una vez obtenida d_M , podemos realizar otro ACP con dicha métrica para obtener otra nueva métrica inducida, y así hasta el infinito. Para fijar ideas, sea $Q_0 = M$ la métrica inicial y sea Q_1 la matriz de la métrica inducida $d_M = d_{Q_0}$. En general definimos

$$Q_{n+1} := \text{matriz de la métrica } d_{Q_n} \quad ; \quad n = 1, 2, \dots$$

Obtenemos entonces una secuencia de métricas $(Q_n)_{n \in \mathbb{N}}$ en E que provienen de una secuencia de aplicaciones del APC, como lo muestra el esquema de dualidad de la Figura 4.

4 Estudio de la métrica inducida por un ACP

Se desea analizar con detalle la métrica inducida por un estudio previo de ACP en el espacio de los "individuos" (objetos). Esta métrica nos como

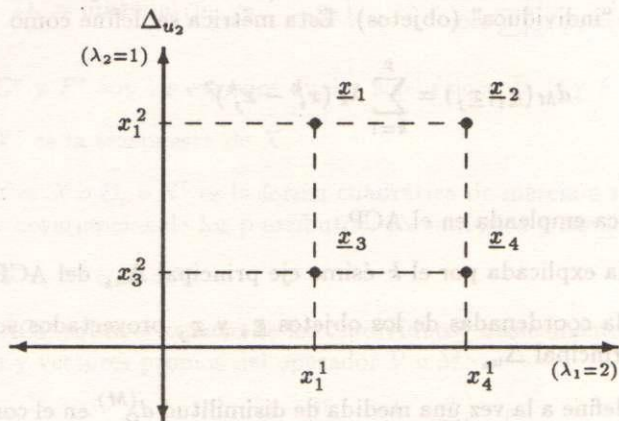


Figura 2: Ilustración de la disimilitud inducida por d_M en el primer plano principal del ACP.

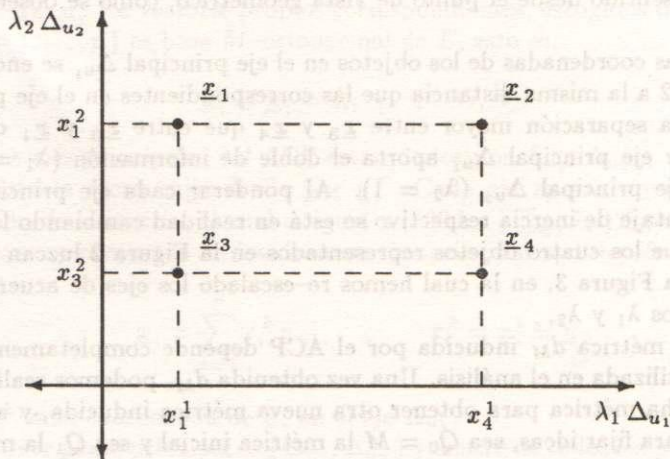


Figura 3: Mismo gráfico del primer plano principal del APC que en la Figura 2, pero escalando los ejes de acuerdo a los valores propios λ_1 y λ_2 . Se observa que x_1 y x_3 están más próximos entre sí que x_3 y x_4 .

Sobre la escogencia apropiada de la métrica inicial $Q_0 = M$ para que este proceso iterativo tenga algún límite con sentido presentamos dos resultados, dependiendo del tipo de operador V .

4.1 Caso en que V es no singular

Supóngase que V es invertible. Entonces se tiene el siguiente resultado:

Teorema 1 *La métrica d_M inducida por el ACP tiene matriz $Q_1 = MVM$. Además una condición necesaria y suficiente para que la secuencia de métricas $(Q_m)_{m \in \mathbb{N}}$ sea convergente a una métrica de E es que M sea la métrica de Mahalanobis V^{-1} .*

Demostración: Trabajando matricialmente se tiene que:

$$\begin{aligned}
 d_M(\underline{x}_i, \underline{x}_j) &= \sum_{k=1}^p \lambda_k (\underline{x}_i^k - \underline{x}_j^k)^2 \\
 &= \sum_{k=1}^p \lambda_k (\underline{x}_i^t M \underline{u}_k - \underline{x}_j^t M \underline{u}_k)^2 \\
 &= \sum_{k=1}^p \lambda_k (\underline{x}_i - \underline{x}_j)^t M \underline{u}_k \underline{u}_k^t M (\underline{x}_i - \underline{x}_j) \\
 &= (\underline{x}_i - \underline{x}_j)^t M \sum_{k=1}^p \lambda_k \underline{u}_k \underline{u}_k^t M (\underline{x}_i - \underline{x}_j) \\
 &= (\underline{x}_i - \underline{x}_j)^t M \sum_{k=1}^p VM \underline{u}_k \underline{u}_k^t M (\underline{x}_i - \underline{x}_j) \\
 &= (\underline{x}_i - \underline{x}_j)^t MVMUU^t M (\underline{x}_i - \underline{x}_j)
 \end{aligned}$$

donde $U = (\underline{u}_1, \underline{u}_2, \dots, \underline{u}_p)$. Como los vectores propios \underline{u}_k son M -ortonormados, entonces

$$\underline{u}_i^t M \underline{u}_j = \delta_{ij} \quad ; \quad \forall i, \forall j$$

de donde $U^t M U = I_p$. Luego se obtiene que $M = (U^t)^{-1} U^{-1} = (U U^t)^{-1}$, lo cual lleva a la simplificación

$$MVMUU^t M = MVM,$$

de donde $d_M(\underline{x}_i, \underline{x}_j) = (\underline{x}_i - \underline{x}_j)^t MVM (\underline{x}_i - \underline{x}_j)$. Para demostrar la segunda parte, obsérvese que

$$\begin{aligned}
 Q_1 &= MVM = M(VM) \\
 Q_2 &= d_{Q_1} = Q_1 V Q_1 = MVMVMVM = M(VM)^3 \\
 Q_3 &= d_{Q_2} = M(VM)^3 VM(VM)^3 = M(VM)^7
 \end{aligned}$$

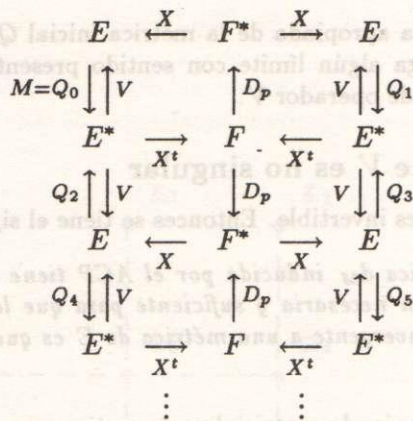


Figura 4: Secuencia de esquemas de dualidad correspondientes a la secuencia de métricas inducidas por un ACP con la métrica inicial $Q_0 = M$.

y en general

$$Q_n = M(VM)^{2^m-1} \quad ; \quad m \geq 1.$$

Si escogemos $M = V^{-1}$ obtenemos claramente que $Q_m = V^{-1}, \forall m \in \mathbb{N}$ y la secuencia de métricas converge trivialmente a la métrica de Mahalanobis. Considérese ahora la descomposición espectral de VM :

$$VM = R\Lambda R^t$$

donde

$$\Lambda = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{pmatrix}$$

es la matriz diagonal de los valores propios de VM (son todos no negativos) y R es la matriz ortogonal cuyas columnas son los vectores propios normalizados correspondientes. Luego, $(VM)^2 = R\Lambda R^t R\Lambda R^t = R\Lambda^2 R^t$, y en general $(VM)^k = R\Lambda^k R^t, \forall k \in \mathbb{N}$, de donde

$$Q_m = M R \Lambda^{2^m-1} R^t \quad ; \quad \forall m \in \mathbb{N}.$$

De lo anterior se deduce que la convergencia o divergencia de la secuencia de métricas $(Q_m)_{m \in \mathbb{N}}$ está determinada por la convergencia o divergencia de $(\Lambda^m)_{m \in \mathbb{N}}$, donde

$$\Lambda^m = \text{diagonal}(\lambda_1^m, \lambda_2^m, \dots, \lambda_p^m).$$

Esta última matriz converge a la matriz diagonal

$$D = \text{diagonal}(a_1, a_2, \dots, a_p),$$

con $a_i \in \{0, 1\}$, $i = 1, 2, \dots, p$ si y solo si $0 \leq \lambda_k \leq 1, \forall k$. Ahora bien, si alguno de los valores propios λ_k es menor que 1, entonces $a_k = 0$, y tendríamos que (Q_m) converge a una semimétrica (no invertible). Luego, debe cumplirse que todos los valores propios de VM son iguales a 1, de donde $VM = PI_pP^t = I_p$, y se concluye entonces que $M = V^{-1}$. ■

4.2 Caso en que V es singular

Cuando V es singular (no invertible), con $\text{rango}(V) = q < p$, el primer resultado del teorema precedente sigue siendo válido: MVM es la semimétrica inducida por un ACP a partir de la métrica M . Se trata sin embargo de una semimétrica, pues $\text{rango}(MVM) = q$.

Se analizará que sucede si se emplea esta semimétrica en un nuevo ACP. Más generalmente, se encontrará una fórmula explícita para la semimétrica inducida por un ACP realizado partiendo de una semimétrica M . Algunas restricciones naturales se impondrán a la semimétrica M .

Primeramente obsérvese que como $V = XD_pX^t$ entonces $\text{rango}(X) = q$, de donde solamente hay q atributos linealmente independientes. Sin pérdida de generalidad puede suponerse que los primeros q atributos son linealmente independientes (de otra forma basta reordenar inicialmente las filas de X). Entonces, particionando V por bloques,

$$V = \left(\begin{array}{c|c} V_{11} & V_{12} \\ \hline V_{21} & V_{22} \end{array} \right) \quad (1)$$

donde

- V_{11} es de tamaño $q \times q$, con $\text{rango}(V_{11}) = q$.
- V_{12} es de tamaño $q \times (p - q)$.
- $V_{21} = V_{12}^t$.
- V_{22} es de tamaño $(p - q) \times (p - q)$.

Los p atributos engendran un subespacio q -dimensional E_q de E . Se realiza un ACP en el subespacio E_q , tomando métricas M_{11} en E_q y luego se pasa al espacio E escribiendo

$$M = \left(\begin{array}{c|c} M_{11} & 0 \\ \hline 0 & 0 \end{array} \right) \quad (2)$$

donde 0 denota matrices nulas de tamaños apropiados. La semimétrica M así definida es la extensión trivial de M_{11} en E . Como en V las últimas $(p - q)$ filas

son linealmente dependientes de las primeras q , entonces existe una matriz K de tamaño $(p - q) \times q$ tal que

$$(V_{21} \mid V_{22}) = K(V_{11} \mid V_{12}).$$

Luego:

$$\begin{aligned} V_{21} &= KV_{11} \Rightarrow K = V_{21}V_{11}^{-1} \\ V_{22} &= KV_{12} \Rightarrow V_{22} = V_{21}V_{11}^{-1}V_{12} \end{aligned} \quad (3)$$

Sean $\lambda_1, \lambda_2, \dots, \lambda_q$ los q valores propios positivos en orden decreciente (contando posibles repeticiones) de $V_{11}M_{11}$:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q$$

Sean $\underline{u}_1, \dots, \underline{u}_q$ los correspondientes vectores propios M_{11} -ortonormalizados de $V_{11}M_{11}$. Escribiendo

$$\underline{v}_i = \begin{pmatrix} \underline{u}_i \\ K\underline{u}_i \end{pmatrix}; \quad i = 1, \dots, q$$

se obtiene entonces:

$$\begin{aligned} VM\underline{v}_i &= \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \begin{pmatrix} M_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \underline{u}_i \\ K\underline{u}_i \end{pmatrix} = \begin{pmatrix} \lambda_i \underline{u}_i \\ V_{21}M_{11}\underline{u}_i \end{pmatrix} \\ &= \begin{pmatrix} \lambda_i \underline{u}_i \\ KV_{11}M_{11}\underline{u}_i \end{pmatrix} = \begin{pmatrix} \lambda_i \underline{u}_i \\ \lambda_i K\underline{u}_i \end{pmatrix} = \lambda_i \underline{v}_i \end{aligned}$$

de donde $\lambda_1, \lambda_2, \dots, \lambda_q$ y $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_q$ son también los valores propios no nulos y vectores propios M -ortonormados correspondientes de VM en el espacio E . Considerando $U_{11} = (\underline{u}_1, \dots, \underline{u}_q)$ y

$$U = \begin{pmatrix} U_{11} & \mathbf{0} \\ KU_{11} & \mathbf{0} \end{pmatrix}$$

se obtiene que

$$\begin{aligned} UU^t M &= \begin{pmatrix} U_{11} & \mathbf{0} \\ KU_{11} & \mathbf{0} \end{pmatrix} \begin{pmatrix} U_{11}^t & U_{11}^t K^t \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} M_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} U_{11}U_{11}^t M_{11} & \mathbf{0} \\ KU_{11}U_{11}^t M_{11} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} I_q & \mathbf{0} \\ K & \mathbf{0} \end{pmatrix} \end{aligned}$$

Se tiene entonces el siguiente resultado:

Teorema 2 Cuando V es singular, $\text{rango}(V) = q$, en la forma (1) y $Q_0 = M$ es una semimétrica en la forma (2), la semimétrica inducida por un ACP realizado a partir de la semimétrica M es

$$Q_1 = \left(\begin{array}{c|c} M_{11}V_{11}M_{11} & 0 \\ \hline 0 & 0 \end{array} \right)$$

Además, una condición necesaria y suficiente para que el proceso de métricas $(Q_m)_{m \in \mathbb{N}}$ sea convergente a una semimétrica de rango q es que M sea una inversa generalizada de V .

Demostración: La primera parte es evidente, en vista de que un simple cálculo arroja

$$Q_1 = MVMUU^tM = \left(\begin{array}{c|c} M_{11}V_{11}M_{11} & 0 \\ \hline 0 & 0 \end{array} \right)$$

Ahora, Q_1 es una semimétrica del tipo (2), de donde al aplicar de nuevo el ACP con Q_1 se obtiene

$$Q_2 = \left(\begin{array}{c|c} M_{11}(V_{11}M_{11})^3 & 0 \\ \hline 0 & 0 \end{array} \right)$$

Inductivamente se obtiene en general que

$$Q_m = \left(\begin{array}{c|c} M_{11}(V_{11}M_{11})^{2^m-1} & 0 \\ \hline 0 & 0 \end{array} \right) ; \quad m \geq 1.$$

Supóngase que M es una inversa generalizada de V de rango q . Entonces $VMV = V$, lo cual lleva a la ecuación matricial

$$\left(\begin{array}{cc} V_{11}M_{11}V_{11} & V_{11}M_{11}V_{12} \\ V_{21}M_{11}V_{11} & V_{21}M_{11}V_{12} \end{array} \right) = \left(\begin{array}{cc} V_{11} & V_{12} \\ V_{21} & V_{22} \end{array} \right)$$

Luego, $V_{11}M_{11}V_{11} = V_{11}$, de donde $M_{11} = V_{11}^{-1}$. Obsérvese que $V_{21}M_{11}V_{12} = V_{21}V_{11}^{-1}V_{12} = V_{22}$ como se mostró en (3). Entonces la secuencia $(Q_m)_{m \in \mathbb{N}}$ es trivialmente convergente a la semimétrica de rango q

$$\left(\begin{array}{c|c} V_{11}^{-1} & 0 \\ \hline 0 & 0 \end{array} \right)$$

Recíprocamente, para que la secuencia (Q_m) sea convergente a una semimétrica de rango q es necesario y suficiente que la secuencia $(V_{11}M_{11})^{2^m-1}$ lo sea, lo cual sucede si y solo si $M_{11} = V_{11}^{-1}$, como se analizó en el teorema 1. En tal caso es evidente que la semimétrica M obtenida es inversa generalizada de V . ■

Paquete	Métrica empleada		Métrica inducida d_M
	M	VM	
SPSS/PC+	I_p	V	V
ANADA	I_p	V	V
	D_{1/σ^2}	C	CD_{1/σ^2}
CHADOC	I_p	V	V
	D_{1/σ^2}	C	CD_{1/σ^2}
	M	VM	MVM

Figura 5: Métricas implícitas y métricas d_M inducidas y los paquetes de software usuales. Aquí C indica la matriz de correlación entre las variables y D_{1/σ^2} la matriz diagonal cuyas entradas en la diagonal son los inversos de las variancias de los atributos.

5 Consideraciones finales

La única métrica inicial M que produce un proceso estable de métricas inducidas por el ACP es la métrica de Mahalanobis $M = V^{-1}$. En tal caso la métrica inducida es también la de Mahalanobis, por lo cual es del todo innecesario realizar el ACP.

Debe tenerse especial cuidado al emplear la métrica d_M inducida por un ACP. Los paquetes ordinarios de software que generalmente están a disposición de los usuarios realizan el ACP con limitadas o ninguna posibilidad de escogencia de la métrica M . En la Figura 5 se resumen las posibilidades de escogencia de la métrica M y la métrica inducida d_M en los paquetes SPSS [4], ANADA [3] y CHADOC [2].

Por lo general se obtiene una métrica inducida d_M que no refleja los intereses del investigador en estudios de clasificación automática.

Referencias

- [1] Caillez, F. y Pages, J. P. "Introduction a l'analyse des données"; Smash, Paris; 1976.
- [2] CHADOC "Manual de Referencia"; Département Informatique de l'IUT de Nice, Nice; 1986.
- [3] Guevara, R. "Nuevas opciones para el Análisis Estadístico Matemático: El paquete ANADA"; Rev. Ciencias Económicas, Vol. III, No. 1; 1983.

- [4] Norušis, M. J. "SPSS/PC+ for the IBM PC/XT/AT"; SPSS inc., Chicago; 1986.
- [5] Piza, E. "La clasificación automática jerárquica aglomerativa"; Rev. Ciencias Económicas, Vol. VII, No. 1, pp. 95-111; 1987.
- [6] Takeuchi, K. y Yanai, H. y Mukherjee, B. N. "The Foundations of Multivariate Analysis"; Wiley Eastern Limited, New Delhi; 1982.