

El análisis multidimensional de datos como herramienta metodológica en ciencias sociales.

Nora Moscoloni.

Cita:

Nora Moscoloni (2004). *El análisis multidimensional de datos como herramienta metodológica en ciencias sociales. VI Jornadas de Sociología. Facultad de Ciencias Sociales, Universidad de Buenos Aires, Buenos Aires.*

Dirección estable: <https://www.aacademica.org/000-045/435>

EL ANÁLISIS MULTIDIMENSIONAL DE DATOS COMO HERRAMIENTA METODOLÓGICA EN CIENCIAS SOCIALES

Nora Moscoloni

Programa Interdisciplinario de Análisis de Datos

(IRICE - Universidad Nacional de Rosario)

nmoscolo@sedede.unr.edu.ar

Acerca del Análisis Multidimensional de Datos

En el campo de las ciencias sociales, el Análisis Multidimensional de Datos se revela como una opción de gran interés para el procesamiento de la información que en la generalidad de los casos es rica en categorías y no en continuos, de naturaleza ambigua, con grandes dificultades de diseño.

Para la descripción de estos objetos complejos, la condición de multidimensionalidad así como la posibilidad de consideración de lo contradictorio es sin duda favorable. Las representaciones en planos factoriales permiten la observación de los opuestos en diferentes dimensiones, sin llegar sin embargo a pretender comprender la totalidad. Se habla de un mayor o menor porcentaje de variación explicada, de configuraciones multidimensionales más o menos estables, etc.

El practicante de análisis de datos no trata de adecuar los datos a sus modelos, sino que busca las estructuras a partir de los datos.

Una ventaja importante de estos métodos es que la interpretación de los resultados se hace en el terreno de lo real, en el sentido de no abstracto. A pesar

de los complicados cálculos que se realizan con los paquetes de programas, en la etapa final de interpretación se produce lo que se llama el “retorno a la realidad”, es decir los resultados se expresan en unidades de medida coherentes con el objeto de estudio.

Estas características hacen que esta corriente capte el interés de investigadores con enfoques no deterministas y permitan un mayor reconocimiento de los objetos en su complejidad.

Los algoritmos desarrollados para estos tipos de análisis se adaptan a diferentes niveles de complejidad de la información: datos numéricos, textuales, simbólicos. Es decir que el dato puede ser algo más que un único valor numérico resultado de la asignación de una medida o código a una unidad de análisis: puede ser una palabra, un conocimiento, una posibilidad, una conjunción de valores, en fin, un objeto simbólico.

Se ve así que las características disciplinares de la estadística se transforman no sólo en cuanto a los métodos sino también al objeto de estudio, que pasan de ser las poblaciones univariadas y sus entrecruzamientos, a la complejidad de una población multivariada donde los individuos no pierden totalmente su significación. Sin pretender ubicar al análisis de datos en alguna de las posiciones intermedias entre los enfoques cualitativo versus cuantitativo de las ciencias sociales, es cierto que es visto con mayor simpatía por los cultores del primer enfoque. En efecto, las características habitualmente citadas de la metodología cualitativa: ‘proceso interactivo’, ‘volver sobre los datos’, ‘analizarlos y replantear el proceso’, ‘búsqueda de tendencias, tipologías, regularidades o patrones’, ‘necesidad de

traducir los datos en categorías', 'análisis exploratorio'; se corresponde con el vocabulario que puede encontrarse en los textos del *analyse des données*.

Dos familias de técnicas

En el ámbito del AMD existen dos familias de técnicas que permiten realizar reducciones sin perder de vista la estructura fundamental de los datos.

Los *métodos factoriales* basados en el álgebra lineal que producen las representaciones gráficas sobre las cuales las cercanías entre los puntos líneas y los puntos columnas traducen las asociaciones estadísticas entre líneas y columnas. A este grupo pertenecen, entre otros, el análisis de componentes principales, el análisis de correspondencias binario y el análisis de correspondencias múltiples. Este último, por trabajar con variables nominales, es el más adecuado para el procesamiento de encuestas.

Los *métodos de clasificación*, que realizan los agrupamientos en clases de las líneas o de las columnas. Con estos agrupamientos, llamados también tipologías o clusters de individuos con características semejantes se puede obtener una visión macroscópica de la información.

Estas dos familias de métodos pueden utilizarse complementariamente con los mismos datos logrando así una síntesis explicativa más accesible al usuario no experimentado.

Principios básicos de las técnicas de Análisis Multidimensional de Datos

Cada dimensión de una matriz de datos numéricos permite definir las distancias (o proximidades) entre los elementos de la otra dimensión. Así el conjunto de las columnas (que pueden ser variables, atributos) permite definir con la ayuda de

fórmulas apropiadas las distancias entre líneas (que pueden ser los individuos, las observaciones). De la misma forma, el conjunto de líneas permite calcular las distancias entre columnas.

Se obtienen así matrices de distancias que se asocian a representaciones geométricas de las similitudes existentes entre las filas y las columnas de las matrices de datos a describir.

El problema es comprender intuitivamente estas representaciones gráficas perdiendo lo menos posible de información. Este proceso podría ser visto desde el análisis exploratorio como una sucesión de construcciones y de-construcciones de una imagen. Tal como en un rompecabezas al que le faltaran piezas y al que en sucesivos armados nuestra intuición y conocimiento del problema permitiera ir agregándolas.

Etapas en un AMD clásico

La complementariedad entre las técnicas factoriales y de clasificación se plasma en la práctica de procesamiento como el encadenamiento de una serie de etapas que pueden describirse como sigue:

Etapas 1: Análisis factorial

Elección de una batería homogénea de variables activas: a pesar de que una característica importante del AMD es la utilización de un gran número de variables, ellas no pueden ser incluidas indiscriminadamente en el análisis esperando que como una caja negra el programa resuelva los algoritmos. Las variables activas son aquéllas que intervienen en la determinación de los ejes del análisis factorial. Para que el análisis tenga coherencia será necesario que el

conjunto de las variables activas sea homogéneo. En el lenguaje metodológico ello implicará que se trate en realidad de indicadores que midan el mismo tipo de dimensiones y variables teóricas. La selección de variables activas implicará asimismo una contraparte: la selección de variables ilustrativas. Éstas últimas no son menos importantes, su inclusión en esta categoría se debe a que constituyen un conjunto de diferente 'especie' que las activas, ya sea porque son indicadores de otra dimensión o variable teórica, porque juegan en diferente sentido que las activas o porque pueden distorsionar el análisis.

Descripción gráfica de la población: la proximidad entre individuos es función de la semejanza de las respuestas a las variables activas. En la misma matriz de datos será equivalente analizar los perfiles filas en el espacio de las columnas que los perfiles columnas en el espacio de las filas. Esta correspondencia que se establece entre ambos espacios es la base de la interpretación gráfica. La visualización de los puntos individuos en el espacio de las variables deberá corresponder naturalmente al valor de sus coordenadas (valores de las variables) en ese espacio. Si el conjunto de valores relativos de un individuo (perfil) es parecido al conjunto de valores relativos o perfil de otro, entonces ambos individuos se representarán como dos puntos cercanos en el espacio de las variables. Esos espacios multidimensionales son reducidos a través del análisis factorial y los puntos podrán ser representados en el nuevo espacio determinado por los ejes factoriales.

Posicionamiento de los elementos ilustrativos sobre los planos factoriales: una vez construido el nuevo espacio factorial puede ser visualizado a través del cruce entre dos ejes factoriales, los cuales serán todos perpendiculares entre sí (o sea

estadísticamente independientes). Dos ejes factoriales determinarán un plano factorial de máxima inercia sobre el cual podrán luego proyectarse las variables ilustrativas.

Un ejemplo de esto se puede buscar en el análisis de una encuesta de opinión. Si elegimos como variables activas a los indicadores de opinión, podremos observar en el plano factorial la estructura de respuestas a favor, indiferentes o en contra de un determinado tema. Las variables contextuales que indicamos como ilustrativas estarán en cambio respondiendo a la pregunta: ¿quiénes son, qué características de contexto tienen los que contestan de una forma u otra?

Etapa 2: Clasificación mixta sobre un subespacio factorial

Las distancias entre los puntos individuos son calculadas en el espacio de los primeros ejes factoriales con la distancia euclidiana usual.

Esta etapa marca una diferencia importante con los otros enfoques de análisis multivariado: a pesar de que los programas computacionales estadísticos clásicos incluyen técnicas de cluster o clasificación, en ellos los cálculos se realizan con los valores 'crudos' de las variables, o sea los verdaderos valores de cada individuo en cada una de las variables medidas en su escala original. El enfoque de la escuela francesa se caracteriza por la complementariedad de las técnicas, lo cual indica que la clasificación se realiza tomando en cuenta no los valores crudos de las variables sino los valores de cada punto individuo respectivamente en cada eje factorial. De esta manera se suaviza la posible distorsión ocasionada por escalas de medidas diferentes y se trabaja con el bloque más importante de variación que contiene las oposiciones fundamentales encontradas en los datos.

Etapa 3: Descripción automática de las clases

Una vez que se clasificaron los individuos de acuerdo con sus cercanías en el espacio factorial, o sea de acuerdo con la similitud de sus respectivos perfiles, se tratará de describir estas clases o grupos de individuos de acuerdo con sus características en común, pero 'retornando a la realidad' de sus verdaderos valores en las variables de origen. Así se calculan los porcentajes de individuos de la clase que poseen una categoría (una determinada opinión por ejemplo) y se los compara con los porcentajes de individuos que la poseen en la población, de esta manera se puede determinar la eventual concentración de esa categoría de opinión en la clase en cuestión. Así, pueden asignarse categorías características de ciertas clases facilitando la interpretación de cada uno de los grupos.

Etapa 4: Posicionamiento de las clases en el plano factorial

Visualizar las posiciones relativas de las clases en el espacio y poner en evidencia trayectorias ocultas por la discontinuidad de las clases:

La posición de cada individuo identificado por el número de su clase permite representar la densidad y la dispersión de las clases en el plano factorial.

La correspondencia entre los espacios permite calcular las coordenadas factoriales tanto de los puntos variables como de los puntos individuos. Podrá observarse que ambas configuraciones son visualmente equivalentes. Ello permite aprovechar esta posibilidad de representación proyectando los puntos individuos en los primeros ejes factoriales e identificando cada punto con el número (y eventualmente un color) de la clase a la cual pertenece. Esta representación permitirá observar la relativa concentración que presenta cada clase, las posibles

superposiciones de clases cuando lo que se está representando a veces son sólo manifestaciones indicadoras de la continuidad de una variable teórica.

Si tenemos en cuenta que las clases construidas en AMD clásico son mutuamente excluyentes, es decir que un individuo pertenece sólo a una clase, se verá que la posibilidad de observar la vecindad de ciertas clases en el espacio factorial proveerá de gran riqueza interpretativa.

Análisis textual

Aproximación informática al texto

El texto ha sido objeto de estudio de distintas disciplinas con objetivos, métodos y perspectivas muy diferentes: la lingüística, el análisis de discurso, el análisis de contenido, la búsqueda documentaria, la inteligencia artificial.

En la actualidad las facilidades de disponer de los textos en computadora han aumentado considerablemente, no siempre es necesario introducirlos manualmente a través del teclado, los métodos de scaneado (operación que realiza el reconocimiento de los diferentes caracteres) han realizado progresos interesantes, así como la posibilidad de obtención de textos directamente desde internet, han puesto a disposición de los investigadores una abrumadora masa de textos en los cuales es posible aplicar métodos lexicométricos.

La computadora ha permitido disponer de la potencia de la informática al servicio del texto, de su lectura, de su exploración, para satisfacer la curiosidad del investigador, permitiéndole moverse libremente en el interior del corpus, encontrar sistemáticamente las repeticiones, seguir la fluctuación de las significaciones, conocer instantáneamente las características del locutor o de la frase, identificar

las asociaciones semánticas fuertes, las influencias susceptibles de explicar la formación del discurso, seleccionar las expresiones típicas, raras o especiales, esas pequeñas frases que a veces dan vida a un informe. Leer, seleccionar pero también anotar, asociar a cada frase o cada unidad de significación el o los temas que los caracterizan para así referir la variedad de un texto a una temática general.

Este trabajo artesanal de lectura, selección, anotación, desglose, archivo, puede ser en parte automatizado. La computadora simplemente es más rápida que la tijera, que la lapicera, que el ojo, aún cuando se trate de recorrer rápidamente el texto buscando una palabra que no quiere aparecer.

Asimismo ciertas estructuras muy difícilmente perceptibles mediante la lectura pueden revelarse a través del cruce entre las palabras y las características que identifican a un texto del cual ellas provienen, por ejemplo, atributos del producto, de la marca o del soporte, cuando se citan términos publicitarios o identidad del entrevistado en el caso de encuestas. La informática permite producir instantáneamente estas tablas cruzadas o tablas de contingencia (asimismo llamadas tablas léxicas) pero también analizarlas automáticamente haciendo aparecer las correspondencias. El retorno al texto, guiado por las palabras más notables, dará cuerpo a las estructuras armadas por el análisis.

La construcción de indicadores léxicos y del discurso (tamaño de un corpus, intensidad de un campo léxico, banalidad) pueden ayudar a la selección y a la interpretación de los contenidos y en ciertos casos llegan a medir los efectos de discurso o de las “intensidades semánticas”.

Análisis de Datos Textuales

La idea de tratar textos con métodos estadísticos no es nueva. Pueden citarse las investigaciones lexicométricas del Laboratorio de Lexicología Política de la Escuela Normal Superior de Saint Cloud en Francia, así como los trabajos de P. Guiraud y de Ch. Muller, citados por Maingueneau (1989). En los primeros textos se aplicaban los métodos estadísticos elaborados para tratamiento de variables continuas, sin embargo con los métodos de AMD se obtienen resultados más interesantes.

El Análisis de Datos Textuales es una derivación de los métodos de AMD es decir métodos exploratorios, que son llamados asimismo métodos estadísticos lexicométricos.

Las primeras aplicaciones en este enfoque fueron realizadas por Jean Paul Benzécri (1976), quien desarrolló el análisis de correspondencias.

Posteriormente, los aportes de Ludovic Lebart¹, se dirigieron a resolver los problemas de tratamiento de preguntas abiertas en las encuestas sobre condiciones de vida utilizando métodos más automáticos que la post - codificación manual que es el procedimiento alternativo habitual.

El Análisis de Datos Textuales consiste en aplicar el análisis de correspondencias y los métodos de clasificación, a tablas específicas creadas a partir de datos textuales. Estos métodos se completan con otros propios del dominio textual como los glosarios de palabras, las concordancias y la selección del vocabulario más específico de cada texto, para así proveer una herramienta comparativa de los mismos.

¹ A partir de su trabajo en el CREDOC (Centre de recherche pour l'étude et l'observation des conditions de vie)

El ámbito en el cual es más fácil aplicar estos métodos son las preguntas abiertas de las encuestas y en general a numerosos textos cortos, como los textos políticos o las entrevistas en profundidad. Si bien el campo de aplicación es bastante amplio, es deseable contar con textos que presten cierto grado de homogeneidad y de exhaustividad en el tema a estudiar.

El principio fundamental de estos métodos es el análisis a través de la comparación. Se busca comparar entre sí el discurso de los individuos que han contestado a una encuesta, o de grupos de individuos generalmente formados a partir de la respuesta a una pregunta cerrada. Por ejemplo: el lenguaje de los hombres con el de las mujeres, el lenguaje de los jóvenes con el de los mayores. En un ámbito literario permite comparar textos a los efectos de identificar estilos. La comparación implica llegar eventualmente a clasificar a los individuos en clases homogéneas en cuanto al vocabulario empleado aunque también puede interesar clasificar palabras.

Estos métodos pueden resumir los textos mediante las palabras más características y mediante las respuestas o frases más características. En este sentido sería otra forma de construir un verbatim.

Un objetivo importante es conectar las repuestas abiertas con toda la información proporcionada por las respuestas cerradas o las variables categóricas relativas a características contextuales de los individuos.

Los métodos estadísticos lexicométricos se proponen como sistemáticos, en el sentido de que cuentan la presencia de las palabras sin una selección a priori, sin embargo cabe tener en cuenta que no se trata de un simple recuento de palabras, sino que, como en todo AMD, su objetivo es descubrir una estructura subyacente.

Son exhaustivos, porque trabajan a partir del texto de todas las respuestas y por lo tanto se dice que son métodos que permiten una mayor 'objetividad', aunque en realidad lo que sucede es que posibilitan retrasar la intervención del investigador hasta una fase más tardía del trabajo.

El programa de análisis textual más utilizado en este campo ha sido el SPAD.T (Bécue, 1991), que se complementaba con el SPAD.N (numérico). En la actualidad ambos programas se agrupan en un mismo sistema (Lebart, 2000).

Existen algunos indicadores léxicos cuyo objetivo es caracterizar un texto:

Tamaño del corpus y de las respuestas: se evalúa en número de palabras, corresponde a la longitud del texto total o del de cada respuesta. Este indicador es muy útil para diferenciar los comportamientos de las respuestas según las cuestiones o la identidad de los entrevistados. El tamaño puede ser restringido a las palabras de un diccionario especializado (los términos del placer, de la insatisfacción...) e indicar entonces una intensidad semántica según que las palabras de los campos considerados sean numerosas o raras. Se tratará entonces de un indicador de contenido.

Riqueza del vocabulario: la riqueza del léxico es igual al número de palabras diferentes que lo componen. Más la riqueza es grande, más el vocabulario es abundante y la información producida será más rica.

Banalidad de la respuesta: la banalidad se aplica a una respuesta y es igual a la ocurrencia media de las palabras que la componen. Más la respuesta está formada de palabras que aparecen muy frecuentemente (de ocurrencia elevada) más banal es ella.

Balance lexical: hacer un balance léxico es establecer globalmente, o según determinadas modalidades de una respuesta cerrada, las características léxicas de las respuestas aportadas. Tamaño, riqueza, intensidad de determinado campo lexical.

Se trata de buscar las asociaciones entre una pregunta y la otra y no ya al interior de una misma pregunta. Las técnicas estadísticas utilizadas son: frecuencias de encadenamiento/asociación y análisis factorial de correspondencias.

Aplicaciones de AMD

Procesamiento de encuestas

Las técnicas de AMD permiten un análisis más exhaustivo de los datos de una encuesta. Es cierto que es necesario un entrenamiento especial para su puesta en marcha, ya que no son tan simples de utilizar como las tradicionales. Sin embargo, la rapidez en los cálculos y la percepción casi inmediata de la estructura de los mismos, permiten encontrar relaciones que pasarían inadvertidas si el usuario está inmerso en una montaña de cifras, tal como resulta en el caso de los laboriosos análisis de cuadros cruzando variables de a dos por vez.

La utilización de técnicas de AMD modifica las primeras fases del tratamiento de datos de encuesta, ya que no se trata de complementos refinados que intervienen luego de los métodos tradicionales; por el contrario el AMD cambia radicalmente el encadenamiento de etapas y define una metodología y unos conceptos diferentes. Con el AMD se puede probar la coherencia global de los datos de una encuesta de una forma rápida y sistemática, da un panorama general, permite criticar la información y orientar los pasos siguientes, elegir las recodificaciones de variables

que sean necesarias, agrupamientos, cambios de escala, así como descubrir si una variable ha sido mal definida. Finalmente y como presentación simple de resultados, seleccionar los cruces de variables que más convengan.

Estas operaciones interviniendo al comienzo del procesamiento permiten monitorear los análisis subsiguientes corrigiendo las direcciones de análisis si fuera necesario. En este sentido se encuadran dentro de los procedimientos generales del Análisis Exploratorio.

Respuestas a preguntas abiertas en encuestas

El conjunto de las respuestas abiertas a una pregunta de encuesta o entrevista constituye lo que llamamos, siguiendo a los lingüistas, el “corpus estudiado”. Sin embargo, éste es un corpus particular en el sentido del tratamiento de análisis textual que se hace del mismo.

Las respuestas a preguntas abiertas, también llamadas respuestas libres, son elementos de información muy específicos, que pueden desconcertar a la vez a los estadísticos y a los especialistas de estudios textuales. Los primeros pueden decepcionarse por el carácter impreciso y multiforme de estas respuestas, los segundos por su carácter artificial, y su fuerte redundancia global.

El efecto de la repetición y más generalmente el de la frecuencia con la cual se emplean las formas es muy particular. Las frecuencias léxicas observadas son por una parte artificiales, ya que la misma pregunta se hace a centenas o miles de personas, por lo tanto la yuxtaposición de las respuestas constituye un texto redundante por construcción, donde no son raros los estereotipos.

Sin embargo, las preguntas abiertas constituyen una prolongación indispensable de los cuestionarios cuando las encuestas van más allá de un simple sondeo

electoral, cuando se trata de explorar y profundizar un objeto complejo o poco conocido.

Luego de haber tomado conocimiento de las respuestas, será necesario analizar el contenido. La recodificación de las preguntas abiertas (cerrar a posteriori) es un trabajo que puede ser muy fastidioso a fuerza de repetitivo. Con el análisis de datos textuales es posible acelerarlo y además encontrar categorías semánticas imprevistas.

Entrevistas en profundidad, grupos focales

El desarrollo del análisis textual, como técnica de AMD, ha captado el interés de investigadores en ciencias sociales que utilizan como herramienta habitual la entrevista semiestructurada, en profundidad o grupos focales. La aplicación de las técnicas de AMD en este campo se ha visto muy difundida, sin embargo su utilización se distingue de otras herramientas informáticas de exploración más al uso de las técnicas cualitativas.

En estos casos las técnicas AMD no deberían utilizarse en todo su nivel de automatización, es importante complementar el procesamiento incluyendo un grado considerable de trabajo artesanal para evitar la supresión de cierta riqueza interpretativa, que es característica del trabajo cualitativo.

Investigación de mercados

En el área de investigación de mercados las técnicas de AMD encontraron un fecundo campo de aplicación, con diversas modalidades gráficas, que en la jerga del mercadeo se denominan 'mappings' a los que se agregan los mapas perceptuales.

Algunos de los problemas de marketing comercial o político que se pueden resolver con técnicas de AMD, se refieren a la segmentación del mercado en el marco de una campaña, búsqueda de taxonomías en los programas de audiencia de los medios, percepciones de preferencias en productos cualesquiera, estudios de estructuras de mercado, análisis de la publicidad de distintos productos y/o marcas.

En la actualidad ha tomado gran auge el análisis de las percepciones de los clientes no solamente con respecto a la marca de la campaña sino en relación con el conjunto de marcas. Se enfatiza el interés en desarrollar un 'mapa de producto', al cual se lo define como una representación gráfica de las formas en las cuales la gente percibe a los productos en términos de determinados atributos, así como una ayuda para entender sus preferencias. Esto se realiza a través de información que se recoge sobre niveles según algunos atributos y elecciones que se deben realizar entre productos, construyéndose así matrices de datos de preferencias. De esta manera se pueden encontrar asociaciones entre distintos atributos con varias marcas, productos o servicios y el correspondiente perfil del cliente (Verde, 1999).

Se intensifica el estudio de distintos modelos que son comercializados por diferentes empresas que se dedican al estudio del comportamiento del cliente, la identificación de los factores que componen la satisfacción, el grado de satisfacción de los clientes, de la compañía y/o de la competencia.

Un campo de especial importancia en el actual contexto de recesión económica, lo constituye el diseño de productos (entendiendo a éstos en su sentido más amplio, desde un clásico producto de consumo diario a un producto financiero, o un

servicio público) con el fin de ofrecer los que mejor puedan satisfacer las necesidades y expectativas de los consumidores. Por eso es de vital importancia incorporar la voz del cliente en el propio diseño del producto. La técnica más adecuada dentro del AMD que permite realizar esto es el Análisis Conjunto (trade-off) cuyo objetivo es determinar aquella combinación de características que hacen “ideal” un producto para un cierto segmento de consumidores (Aluja Banet, 1994). Otra aplicación se refiere a la obtención de un mapa interpretativo del lenguaje publicitario presente en un determinado mercado a través del estudio de los mensajes que se utilizan en publicidad mediante el análisis textual (Balbi, 1999).

Los métodos estadísticos de las ciencias sociales

La traducción inglesa de *analyse des données* es *data analysis*, sin embargo esa expresión en la escuela anglosajona, remite al enfoque popularizado por John Tukey y ambos no tienen la misma filosofía.

Los métodos anglosajones consideran como una etapa previa al *análisis exploratorio*, el cual a través de métodos de examen y de visualización de una matriz de datos permiten formular las primeras hipótesis o esbozos de modelos probabilísticos, que son testeados luego por un *análisis confirmatorio* haciendo uso de las técnicas clásicas de la estadística matemática.

Otro de los métodos de amplia utilización en sociología es la regresión logística, que responde a una óptica metrológica más que clasificatoria como es el caso del AMD. En el primero, el objetivo es la búsqueda de un resultado, asociando los efectos a las causas, trabajando sobre variables descontextualizadas y supuestas como de comportamiento general. Desde este punto de vista, se enfocaría la

ciencia social como una disciplina que avanza por la acumulación de este tipo de resultados, Desroisières (1996)

Por el contrario, el análisis de datos a la francesa se presenta como un fin en sí mismo, alejándose de toda hipótesis de modelo probabilístico. Es sobre todo una técnica descriptiva y no se propone confirmar o inferir una teoría previamente formulada. Es más que nada un elemento de un conjunto de descripciones históricas de la complejidad y de las dimensiones de un universo social.

Las variables no cuentan tanto en cuanto tales, sino a través de las clases que ellas caracterizan. En realidad son las configuraciones singulares de estas clases y de sus propiedades, las que son el objeto del comentario del sociólogo.

La generalización eventual procede de una retórica diferente de la de las ciencias de la naturaleza. Es la yuxtaposición de configuraciones similares lo que provee un argumento.

Estas diferencias de uso de las distintas técnicas, reflejan la relativa fragmentación de una disciplina, la sociología, que extrae su legitimidad de diversos modelos de científicidad. Será sin duda más provechoso, explicitar esta mezcla en términos de inserción de su discurso en las prácticas sociales diferentes, más que buscar la primacía de uno u otro de estos modelos. La historia muestra que estos debates en apariencia epistemológicos, son en general estériles, ya que cada uno de estos modelos estadísticos tiene un uso social determinado.

Bibliografía

- Aluja Banet T (ed.) (1994): *Diseño del "producto ideal"*. Ponencias de la Jornada de Marketing y Estadística. Universidad Politécnica de Cataluña

- Balbi S(1999): *Lo studio dei Messaggi Pubblicitari con l'Analisi dei Dati Testuali*, Nápoles, Facoltà di Scienze Politiche, Università degli Studi di Napoli "Federicoll".
- Bécue M(1991): *Análisis de Datos Textuales*, París, CISIA
- Benzécri JP(1976) : *L'Analyse des données, T.I La taxonomie T.II L'Analyse des correspondances*. París, Dunod.
- Desrosières A (1996): *Les Apports Mutuels de la Methodologie Statistique et de la Sociologie* Comunicación en las Journées de Méthodologie Statistique organizadas por el INSEE. Disponible en Internet en: <http://www.upmf-grenoble.fr/adept/seminaires/desros/>
- Lebart L, Salem A (1994) : *Statistique Textuelle*, París, Dunod.
- Lebart L, Morineau A, Piron M (1995) : *Statistique Exploratoire Multidimensionnelle*, París, Dunod.
- Lébart, L., Morineau,A., y ot. (2000) : *Système SPAD, Versión 4.51*, París, ®CISIA-CERESTA
- Maingueneau D(1989): *Introducción a los métodos de Análisis del Discurso*, Buenos Aires, Hachette.
- Moscoloni N(1999): *El papel de la Estadística y el Análisis de Datos en la Interdisciplinariedad*, en "Comunicación: evaluación institucional y curriculum", Rosario, UNR Editora, pp 149-164
- Verde R (1999): *Techniche Multivariate per l'Analisi dei Dati di preferenza*, en Actas de la Jornada Internacional de AMD, 1era. Reunión de la Red Lamda, Proyecto Alfa de la Unión Europea, Rosario, UNR.