

EL ANÁLISIS MULTIVARIADO: DEFINICIÓN, CRITERIOS Y CLASIFICACIÓN

Carlos Lozares Colina
Pedro López Roldán
Departament de Sociologia
Universitat Autònoma de Barcelona

Resumen

El artículo, que se presenta como introducción al número monográfico, ofrece un panorama del conjunto de los métodos y las técnicas multivariadas. El esfuerzo no es sencillo dada la multiplicidad de los mismos y la diversidad de criterios de clasificación que se pueden adoptar. Por ello se insiste en dar a los métodos multivariados una identidad que vaya más allá de una definición simplista como sería caracterizarlos únicamente por el número de variables. Se hace intervenir, tanto en su definición como en su clasificación, otros criterios como la naturaleza de las variables en el proceso explicativo, métrica de las mismas, número, etc.

Resum

L'article, que es presenta com a introducció al monogràfic, ofereix un panorama del conjunt dels mètodes i tècniques multivariades. L'esforç no és senzill donada la multiplicitat que presenten i la diversitat de criteris de classificació que es poden adoptar. Per això s'insisteix en donar als mètodes multivariades una identitat que vagi més enllà d'una definició simplista, com seria si es caracteritzessin únicament pel nombre de variables. Es fa intervenir, tant a la seva definició com a la seva classificació, altres criteris com el tipus de variables en el procés explicatiu, la mètrica corresponent, el nombre, etc.

Abstract

This article, as an introduction to this monographic volume, offers an overview of the methods and multivariate techniques. It is not easy due to the multiplicity and diversity of classification criteria that can be adopted. Therefore, we insist on giving to multivariate methods an identity which goes further than a simplistic definition as the latter would identify them only by the

number of variables. Other criteria are introduced into the definition as well as into classification, such as the nature of the variables in the explanatory process, its metrical, number, etc.

Este número de PAPERS está dedicado a diversos métodos y técnicas de análisis multivariado que comienzan a ser habituales en la investigación sociológica y politológica, aunque en otras disciplinas se hayan desarrollado con más asiduidad e intensidad. La sociología, que ha generado sistemas o paradigmas explicativos y/o comprensivos importantes sobre los fenómenos sociales, está, sin embargo, «habituada» y posiblemente resignada a vivir de prestado en cuanto a los métodos y las técnicas de observación y análisis se refiere. Se reconoce el papel importante de los mismos cuando se afirma que todo método y técnica de investigación ha de ser inherente al propio objeto investigado pero, al mismo tiempo, pueden quedar marginados e impartidos por especialistas ajenos a la misma sociología programándose autónomamente y con poca o nula interacción con el objeto investigado, lo que aumenta su condición de prestados.

La dedicación de un número en PAPERS a este contenido es novedosa, deseando que marque el principio de una presencia más regular de tal problemática en la revista. En este número no podrá recogerse el amplio abanico de métodos y técnicas de análisis multivariado que se han desarrollado en su aplicación al ámbito de las Ciencias Sociales, basta recorrer la clasificación que proponemos más adelante para llegar a comprender la diversidad de los mismos.

Nos ha parecido de interés, previo al desarrollo de algunos de los métodos y técnicas en los diferentes artículos, dar una panorámica general del análisis multivariado. Iniciando dicho panorama con un intento de definición, se continúa con la exposición detallada de los diferentes criterios utilizados para la clasificación de tales métodos y técnicas, con una clasificación lo más abaricable posible de los mismos y con una breve reseña de los más importantes.

1. DEFINICIÓN

Es habitual en la investigación empírica sociológica el tratamiento bivariado. La puesta en juego de más de dos variables, simplemente por la introducción de una tercera variable de control, adquiere una cierta dificultad de análisis que se incrementa si el número de variables o la complejidad de los modelos aumenta. El cálculo que conlleva el tratamiento

conjunto de un número elevado de variables ha sido otra de las principales limitaciones por las que el investigador en sociología no ha desarrollado hasta hace poco tiempo estos tipos de análisis. Fue durante las décadas de los años treinta y cuarenta cuando empezaron a emplearse las técnicas de análisis multivariado. Sin embargo, el desarrollo experimentado por la informática en los últimos años y la facilidad y posibilidad de cálculo que conlleva han favorecido la extensión del uso de las técnicas estadísticas de análisis multivariado en numerosas disciplinas y también en la sociología, muy en particular a partir de la difusión, en los años sesenta y setenta, de los paquetes estadísticos para el análisis de datos (SPSS, BMDP, SAS, SPAD, etc.).

Las ventajas que el multivariado presenta sobre el bivariado son puestas de manifiesto por Cattell¹:

- economía en el almacenamiento de datos,
- mayor consistencia en la inferencia estadística,
- desarrollo de conceptos teóricos más adecuados,
- mayor precisión y perspectiva conceptual.

Sin embargo, hay que insistir en que se trata de instrumentos que facilitan:

- el análisis y síntesis de los datos,
- la operativización y validación de los conceptos,
- la mejora en la obtención y calidad subsiguiente de los propios datos.

Pero en nada sustituyen la labor imprescindible de contrucción del objeto de investigación, de diseño de la misma, del análisis y de la interpretación de los resultados.

1. El *Análisis de Datos* se asimila en algunas tradiciones, como en la francesa, al *Análisis Multivariable*. En otras como la inglesa el concepto es más limitativo, refiriéndose a veces exclusivamente a los análisis univariados o bivariados. Desde la perspectiva aquí adoptada se considera el *Análisis Multivariable* inserto en el *Análisis de Datos* y éste, a su vez, inscrito en el *Análisis Estadístico* o de los *Métodos Estadísticos*. El *Análisis de Datos* puede diferenciarse de una Estadística más clásica, descriptiva o inferencial, por la utilización de métodos algebraicos y la algoritmación de los procesos, por la

1. Cattell, R.B. (ed). *Handbook of Multivariate Experimental Psychology*. Rand McNally, Chicago, 1966.

ausencia de modelos apriorísticos a testar y por el número de dimensiones/variables puestas en juego. La distinción hecha entre Análisis de Datos y Estadística clásica no significa su incompatibilidad. Los métodos y técnicas del Análisis de Datos han de incorporar la Estadística inferencial para validar sus resultados. En cuanto al término Análisis Multivariable será equivalente a los términos multivariado y multivariante pues dependen más bien de la disciplina en la que han nacido y se han desarrollado. La acepción que se dará al Análisis Multivariable es en una primera y elemental aproximación la de todo método o técnica que *comporte un nivel de análisis que exceda o vaya más allá del bivariado*. Se trata de una definición muy general por lo que se proponen una serie de condiciones más específicas, a fin de asegurar más convenientemente los límites conceptuales del término. De entrada y a partir del contenido elemental adoptado puede ya verse la adecuación y el interés de tales métodos y técnicas en una disciplina como la sociología en la que el campo de aplicación, mirado desde el espacio de los atributos o desde el universo poblacional, suele implicar una visión multidimensional de la realidad social.

2. El Análisis Multivariable, y será una de las condiciones adicionales a la definición precedente, supone también una clara distinción, ruptura o desagregación, entre lo que se denomina universo de individuos, entidades, etc., sujetos de la atribución conceptual, y el espacio de las variables. Ello conlleva y posibilita que los *datos al ser expresables en forma matricial puedan ser interpretados en una doble lectura*: desde las entidades y desde las variables. Kendall² va en esta dirección cuando propone como definición de análisis multivariable, a partir de n individuos y de p variables, la de ser una *rama de la estadística que se interesa en el estudio de la relación entre series de variables dependientes de los individuos que las sustentan*. Se entiende por dependencia la que vinculan determinadas variables a los individuos de los que se predicán.

3. Este soporte matricial necesario pasa a ser la condición de otra característica definitoria del análisis multivariable. Se trata de una característica polar dada la naturaleza dual del dato. Por un lado, *el dato es un resultado algebraico*, por tanto, sometido y sometible a la lógica del lenguaje matemático utilizado que lo ha *leído*, medido, tratado y analizado, y, por otro, es también el *resultado de una construcción de naturaleza científica*, en cuanto a inserción y pertenencia del mismo a una determinada teoría más o menos elaborada, a unos presupuestos e hipótesis y, sobre todo, al haber sido objeto

2. Kendall M. G. *Multivariate Analysis*. Griffin, Londres, 1975, p. 1.

de una validación interna y externa en diversos grados. Esta dualidad descrita, intrínseca al dato, es extensible a la matriz.

3.1. Sobre la primera de las polaridades mencionadas las técnicas tratadas son multivariantes en el sentido de que posibilitan y ofrecen una *flexibilidad en cuanto al lenguaje algebraico*, y en particular en cuanto a las *diversas «métricas»* requeribles en función del tipo de estructura algebraica utilizada en la operativización de los conceptos. La métrica definida es también múltiple y diversa, adaptable a la naturaleza de la operativización de los conceptos. Las matrices pueden ser, según los métodos, de diferente naturaleza. Así puede tratarse de matrices lógicas, matrices nominales y con ellas de contingencia, de frecuencias absolutas, relativas o relativas condicionales, matrices de rangos y matrices cuantitativas. El definir las exclusivamente como técnicas estadísticas es de hecho restrictivo: más ampliamente el álgebra matricial y vectorial, así como las teorías de categorización, suelen estar en la base de los métodos utilizados. Así pues, y dentro de estas consideraciones, *la exigencia de una métrica dada aparece como uno de los requisitos importantes.*

3.2. En cuanto a la segunda, esto es, las exigencias ligadas a su inserción teórica, a la corroboración de hipótesis, etc., el conjunto de las técnicas multivariantes posibilita *diversos niveles de análisis* correlativos a dichas exigencias. De ahí también la conveniencia de su asignación como multivariante atendiendo a sus virtualidades múltiples. Diversos análisis y orientaciones metodológicas pueden plantearse a partir de estas técnicas que enunciaremos sin detallar:

3.2.1. En el orden de la economía y de la parsimonia exigible a todo método se resalta el *efecto simplificador y reductor* que muestran y procuran. De ahí también otro de los resultados de estas técnicas en cuanto que permite un tratamiento *global*, de todas las variables y de todas las unidades independientemente o de forma combinada entre variables e individuos, o *parcial*, tomando subconjuntos de variables o agrupaciones de individuos. Estas técnicas permiten *tratamientos simultáneos* de todas las variables o unidades puestas en juego o un *tratamiento escalonado* en la intervención de alguna(s) unidad(es) o variable(s). Se dan técnicas que permiten la introducción de determinadas variables o unidades que sin intervenir en el cálculo algebraico o algorítmico solamente tienen una función descriptiva o de localización en un conjunto.

3.2.2. *Efectos de consistencia y de validación* en la construcción y definición de las variables. En este sentido muchas de las técnicas, aunque aparezcan

como exploratorias, permiten y facilitan, por un lado, la construcción de tipologías más allá del núcleo inicial adoptado extendiéndose a otras variables introducidas en el análisis y, por otro, la validación interna de determinadas variables. En este sentido estos métodos tienen una función instrumental en cuanto que pueden mejorar la calidad del proceso de operativización y de los datos en general, pudiendo ser así un procedimiento heurístico y reiterativo para la construcción de variables. Aparece pues de esta forma una virtualidad inferencial y pragmática.

3.2.3. Dichas técnicas llenan también otros objetivos metodológicos:

a) Así, pueden servir para procurar:

- *la descripción multivariable de la problemática tratada,*
- *la categorización de las entidades e individuos que componen el universo que se elige,*
- *la agrupación, asociación y dependencia de las variables elegidas así como su dimensionalización.*

Como efecto de la complementariedad entre las dos últimas ópticas, se puede llegar a la construcción de tipologías que van más allá de la simple taxonomización del universo. Desde esta perspectiva son soporte importante de los llamados análisis empíricos y, en general, inductivos, sirviendo como instrumentos iniciales exploratorios para cualquier tratamiento más «exigente».

b) Por fin los métodos y técnicas multivariados se prestan en tanto que soporte algebraico/estadístico a un tipo de análisis de *tipo confirmatorio o más general de naturaleza de estructura causal*. Es aquí donde algunas de las técnicas muestran una gran virtualidad por el camino de la dependencia o interdependencia de variables a partir de hipótesis previamente establecidas.

En ambos objetivos puede introducirse el factor tiempo y dar origen además a los diversos tipos de análisis temporales que también entran en el ámbito de los llamados análisis multivariados.

Huelga decir la ventaja que este tipo de análisis puede tener sobre los análisis habitualmente utilizados en sociología como son los univariados o bivariados. Con todo, es necesario insistir en que estas técnicas no han de ser utilizadas como un salto evasivo para evitar todo lo que puede suponer:

- un exigente planteamiento teórico,
- una operativización cuidada de los conceptos,

- la metrización adecuada de las variables así como su validación interna,
- la elaboración verificable de las hipótesis o de otros niveles en la construcción del objeto de investigación
- y, por fin y sobre todo, la comprensión cercana y casi masticable del conjunto de las variables introducidas en la problemática tratada.

En este último sentido, somos de la opinión de que el empleo de métodos y técnicas de análisis multivariado es imposible sin una fase previa fundamental de análisis minucioso y atento de los resultados obtenidos en la investigación a los niveles menos complejos, univariados y bivariados.

2. CRITERIOS DE CLASIFICACIÓN DE LOS MÉTODOS MULTIVARIABLES

Todo intento de clasificación se encuentra con la dificultad proveniente de la variedad existente de métodos/técnicas y del hecho de que diversos criterios operen simultáneamente, algunos de ellos en parte superponibles, lo que introduce además un grado mayor de complejidad. Así pues parece conveniente introducir primero los distintos criterios habitualmente utilizados como principios clasificatorios para, a continuación, presentar la propuesta de clasificación.

Pueden ser diferenciados dos grandes grupos de criterios. El primer grupo va más en la dirección de tener en cuenta los diferentes objetivos analíticos que puedan presentarse en la investigación, sea del lado de las formas de construir el objeto como de los modelos de hipótesis elaborados. El segundo en función de criterios más bien algebraicos y/o técnicos. Evidentemente entre ambos se da un grado de correlación.

2.1. CRITERIOS SEGÚN EL MODELO DE ANÁLISIS

2.1.1. Se puede adoptar un criterio que tenga en cuenta la finalidad o nivel analítico pretendido con el método o técnica adoptada. Se trata de un criterio de clasificación que depende del *diseño de la investigación*. En este sentido se puede hablar de:

a) Métodos/técnicas más bien exploratorios como las Escalas Multidimensionales, Análisis de Componentes Principales, Análisis de Correspondencias, etc.

b) Métodos/técnicas más vinculados al análisis explicativo causal como los de Estructura Causal, Análisis del Camino (*Path Analysis*), Análisis Factoriales Confirmativos, Regresión y Varianza Múltiples, etc.

La base inicial de la clasificación de *Evrard y Le Maire*³ va en esta dirección.

2.1.2. Dentro de este grupo puede incluirse también un criterio clasificatorio que tenga en cuenta las diversas formas de construcción en que puede presentarse el objeto de investigación construido. En este sentido unas técnicas son más adecuadas o pertinentes que otras a determinadas formas de construcción del objeto.

a) Formas de construcción que impliquen niveles relativamente elementales de categorización/clasificación o de contrastación comparativa:

- desde las más empírico-pragmáticas a otras de elaboración, al menos elemental, de modelos comparativos;
- desde propósitos exclusivamente descriptivos a otros con construcciones de categorías poblacionales previas.

En estos ámbitos pueden entrar las Técnicas de Taxonomización, de Segmentación, Escalas Multidimensionales, Perfiles, Coeficientes de Contrastes, Anova, Análisis de Clasificación (*Cluster Analysis*), etc.

b) Formas del objeto construido que tiendan a la búsqueda de las dimensiones latentes o de los factores más discriminantes de la realidad social considerada, esto es, estructuras dimensionales a través de las relaciones de interdependencia entre las variables. Los objetivos pueden ir desde los más descriptivos y exploratorios hasta otros situados en los linderos de lo explicativo. Estos métodos/técnicas se prestan también a la validación de variables operativizadas y a la eliminación de las redundancias. Se trata de métodos como el Análisis Discriminante, el Análisis de Componentes Principales, el Análisis de Correspondencias, Modelos Log-lineal, etc.

c) Formas de construcción del objeto que impliquen modelos más elaborados, como las tipologías concretas, donde no se da una simple estratificación sino que, además de serlo, los tipos extraídos se sitúan y adquieren identidad *en y a partir* de la estructura dimensional del campo de atributos. Esta forma de construcción del objeto sociológico nos parece sumamente efi-

3. Evrard, Y.; Le Maire, P. *Information et Décision en Marketing*. Dalloz, París, 1976.

caz y atractiva en la investigación sociológica. Llena además objetivos de gran interés como:

- la validación de las elaboraciones tipológicas,
- la de ofrecer criterios decisivos para su misma construcción, bien sea empíricamente, haciendo emerger determinadas realidades no bautizadas conceptualmente recibiendo así su identidad social, bien sea deductivamente tratando de aseverar y contrastar determinados conceptos no confirmados realmente. Técnicas de Análisis Discriminante, de Análisis de Correspondencias Múltiples, de Varianza Múltiple, de complementariedad entre el Análisis de Componentes Principales, de Cluster y Modelos Log-lineal pueden perfectamente ir en esta dirección.

d) Formas de construcción del objeto que comporten hipótesis proposicionales simples o sistemas de hipótesis, esto es, modelos ajustados, en general de dependencia lineal, con objetivos explicativos-causales. Tendríamos en este apartado numerosos métodos y técnicas que llenan dicho objetivo: Anova, Anovas Factoriales, Análisis de Regresión Múltiple, Modelos Logit, Análisis Factoriales Confirmatorios, Análisis del Camino, Correlaciones Canónicas y en general los modelos de Ecuaciones Estructurales.

Una de las clasificaciones dada por Bisquerra⁴ se acerca en buena medida al criterio aquí propuesto.

2.1.3. Otro de los criterios clasificatorios dentro de este grupo, parcialmente superponible a los anteriores está marcado por un punto de vista más pragmático que el de su finalidad analítica, como lo han sido los 2.1.1 y 2.1.2. Es el criterio adoptado por Prieto⁵ y puede parecer limitativo a no ser que se entienda como predictivo lo explicativo/causal. Se trata pues de un criterio que clasifica los métodos según:

a) Una orientación predictiva, y nos encontraríamos con técnicas y métodos como el Análisis de Regresión, de Covarianza, Varianza, etc.

b) Una orientación reductiva, y nos encontraríamos con métodos y técnicas como el Análisis de Componentes Principales, Correlación Canónica, Cluster, etc.

4. Bisquerra, R. *Introducción conceptual al Análisis Multivariable. Un enfoque informático con los paquetes SPSSX, BMDP, LISREL y SPAD*. PPU, Barcelona, 1989, pp. 20-21.

5. Prieto, G. «Análisis Multivariable», en A. de la Orden Hoz, *Investigación educativa*. Anaya, Madrid, 1985, pp. 26-28.

2.2. CRITERIOS ALGEBRAICOS Y TÉCNICO-INSTRUMENTALES

2.2.1. Como condición y exigencia previa a toda finalidad analítica, por ejemplo exploratoria o explicativa, se puede establecer un criterio que diferencie técnicas que:

a) Implican una relación de dependencia entre variables, así las de Regresión, Logit, Análisis del Camino, Varianza, Correlación Canónica, Factorial Confirmatorio, etc.

b) Implican una relación de interdependencia entre las variables como Cluster, Correspondencias, Componentes, Escalas Multidimensionales, etc.

Este criterio aparece en la mayor parte de las clasificaciones: Dillon y Goldstein⁶, Kinnear y Taylor⁷, Sheth⁸.

2.2.2. Otros criterios clasificatorios tienen en cuenta la cantidad de variables puestas en juego:

- en las dependientes, ya que se parte de la idea de que entre las independientes se dan más de dos,
- en las interdependientes que se suponen más de dos variables.

Este criterio suele aparecer en muchas de las clasificaciones existentes: Sheth, Kinnear y Taylor, Evrard y Le Maire, Bisquerra, Tabachnick y Fidell⁹, Harris¹⁰.

2.2.3. Por fin y dentro de este grupo de criterios aparece una conjugación de los mismos como son la métrica de las variables, la naturaleza cuantitativa o cualitativa de las mismas, aunque estos criterios se encuentren en escalones más bajos de la clasificación.

6. Dillon, W. R.; Goldstein, M. *Multivariate Analysis. Methods and Applications*. John Wiley, Nueva York, 1984.

7. Kinnear, P.R.; Taylor, J.R. «Multivariate Methods in Marketing Research. A Further Attempt at Classification», en *Journal of Marketing*, 35, octubre, 1971, pp. 56-59.

8. Sheth, J. N. «The Multivariate Revolution in Marketing Research», en *Journal of Marketing*, 35, enero, 1971, pp.13-19.

9. Tabachnick, B.G.; Fidell, L.S: *Using Multivariate Statistics*. Harper & Row, Nueva York, 1983.

10. Harris, R.J. *A Primer of Multivariate Statistics*. Academic, Orlando (Florida), 1985.

3. CLASIFICACIÓN

Se trata de ofrecer un esquema sintético pero al mismo tiempo que contenga el mayor número posible de métodos/técnicas teniendo en cuenta la complejidad de los criterios adoptados (ver Gráfico 1).

A. Tomando como cuerpo de análisis el *campo de las variables* con índices de similitud tales como los de varianza, covarianza y correlación para variables cuantitativas o índices de asociación o proporciones en las cualitativas:

A.a. En el caso de que se *explícite una relación de dependencia* según modelos relacionales lineales, pueden darse diferentes supuestos:

De *una sola variable métrica (cuantitativa) explicada* o dependiente y sea(n) también *métrica(s) la(s) independiente(s)*, las técnicas de análisis pertinentes son las de:

Regresión simple

Regresión múltiple

Idénticamente si se da *una sola variable métrica explicada* o dependiente pero *la(s) independiente(s) no es (son) métrica(s)*, tenemos las técnicas de:

Anovas

Análisis de varianza simple

Análisis de varianza múltiple

Anova factorial

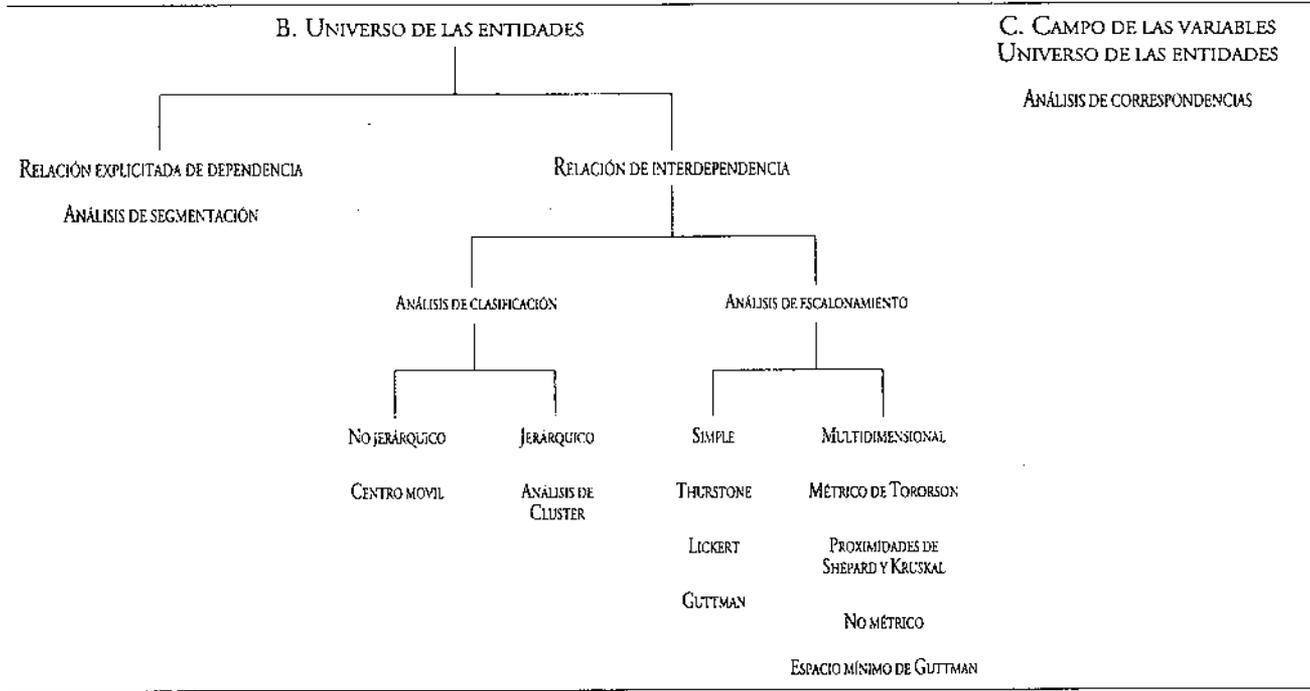
Igualmente con *una sola variable métrica dependiente* o explicada pero *las independientes no todas son métricas*, tenemos como técnica apropiada:

Anacova

Considerando también *una sola variable explicada* o dependiente, pero ahora *no métrica (o cualitativa)*, siendo las *independientes métricas*, la técnica adecuada será la del:

Análisis discriminante

Igualmente con *una sola* (posiblemente en el interior de un modelo más complejo) *variable no métrica (cualitativa) explicada* o dependiente con *in-*



dependiente(s) no métrica(s), con lógicas de modelos causales para variables cualitativas, se pueden utilizar técnicas como:

Diferencias entre proporciones
Modelos logit

En el caso de que existan *dos o más variables métricas dependientes* o explicadas y la(s) *independiente(s) es (son) métrica(s)*, según modelos de lógica causal, se tienen las técnicas:

Análisis factorial confirmatorio
Análisis del camino (o path analysis)
Análisis de estructuras de covarianza
Modelos de Ecuaciones Estructurales

Igualmente con *dos o más variables métricas dependientes* pero la(s) *independiente(s) no es (son) métrica(s)*, las técnicas de análisis apropiadas son:

Análisis multivariable de varianza (manova)

Igualmente con *dos o más variables métricas dependientes* pero las *independientes algunas son métricas y otras no métricas*, las técnicas pertinentes son las de:

Análisis de covarianza múltiple

Si las *dos o más variables dependientes no son métricas* pero son *métricas las independientes*, la técnica adecuada es la del:

Análisis discriminante múltiple

Igualmente si las *dos o más variables dependientes no son métricas* y las *independientes tampoco* tenemos como técnica:

Análisis logit múltiple

A. b. En el caso en que *no se explicita la relación de dependencia (relación de interdependencia)* no se dará una distinción neta entre variables dependientes y otras independientes.

En el supuesto de *variable no métrica* y se den (o no) *modelos de asociación* entre ellas tendremos:

Análisis Log-lineal

En los supuestos de *interdependencia* entre las *variables métricas* pero donde *todas* y cada una de las variables son *explicadas por todas* sin ningún modelo previo de relación tenemos:

Análisis de componentes principales

En el caso de la introducción de *variables temporales* y sin detallar las diferentes modalidades de análisis diacrónicos se tienen los:

Análisis de series temporales

B. Centrando el análisis en el *universo de las entidades (unidades, individuos, objetos)* a través de índices de distancias.

B.a. Si se supone una *parcial distinción o jerarquía en la utilización de las variables* (o simplemente no se tienen en cuenta todas las variables), esto es, la existencia de diferencias entre las variables o al menos una distinción en el orden de introducirlas en el análisis como criterio de categorización de los individuos, tenemos como técnicas pertinentes:

Análisis de segmentación

B.b. Si se parte de una *relación de interdependencia entre las variables* puestas en juego a la hora del cálculo de las distancias y al mismo tiempo *se fija el número de los conglomerados* o agrupamientos se obtienen las técnicas *no jerárquicas*, concretamente la de:

Centro móvil

En el mismo supuesto de *interdependencia* pero en un proceso ascendente o descendente de establecer sucesivas agrupaciones de individuos según algoritmos e índices de distancia dados, tenemos *las técnicas jerárquicas*, concretamente:

Análisis de Cluster

Dentro del mismo supuesto de *interdependencia* y con la finalidad de descubrir estructuras subyacentes podemos incorporar las técnicas de *escalonamiento sencillo* como las:

Escalas sencillas de Thurstone, Lickert y Guttman

o de *escalonamiento multidimensional* de preferencias, proximidades o de percepciones como el:

Escalonamiento multidimensional métrico de Tororson

Análisis de proximidades de Shepard y Kruskal

Escalonamiento multidimensional no métrico

Análisis del espacio mínimo de Guttman

C. Si se trata equivalentemente el *campo de las variables y el universo de las entidades* por su proyección en un único espacio métrico y sin previa distinción entre variables (explicativas o explicadas) según un modelo de *interdependencia* tenemos como técnica pertinente el:

Análisis de correspondencias

4. BREVE RESEÑA DE ALGUNOS MÉTODOS

Se dará cuenta de algunos de los métodos y técnicas que más utilidad pueden tener para la investigación en sociología, sin entrar por consiguiente en todos los enumerados. Por otro lado la explicación será necesariamente breve dado el objetivo mismo de esta introducción.

4.1. REGRESIÓN SIMPLE Y MÚLTIPLE

Dado un modelo en el que se supone la existencia de una variable dependiente (VD) cuantitativa como función de otra variable independiente (VI) -regresión simple- o de otras VI -regresión múltiple-, el análisis de regresión trata de encontrar, según los datos, una función lineal que se ajuste, en determinadas condiciones como modelo matemático, a dicha dependencia. Por ejemplo si se trata de analizar la dependencia que puede suponerse lineal entre la variación, como VD, de los *gastos* en función de las VI, como los *ingresos*, *número de hijos*, etc. El modelo establece que la variación de la VD está dada por las variaciones de las VI afectadas por un coeficiente que se trata de calcular más otra parte que queda inexplicada que se denomina error del modelo. El ajuste, o cálculo de los coeficientes, se realiza a partir de la técnica estadística de mínimos cuadrados por el que se minimiza la parte inexplicada por las VI. Cada coeficiente o parámetro expresará la importancia relativa en la explicación de esta varianza total.

4.2. ANÁLISIS DE VARIANZA SIMPLE (ANOVA) Y MÚLTIPLE (MANOVA)

Se parte de la existencia de una variable dependiente de naturaleza cuantitativa y una o varias independientes cualitativas, denominadas factores. Por ejemplo si se pretenden analizar los *gastos* en función de la *categoría socio-profesional*, *nivel educativo*, etc. Por el análisis de varianza se trata, bajo un modelo lineal en el que la variación, VD, se desglosa como una suma de contribuciones de cada una de las categorías de los factores y bajo determinadas condiciones que ha de cumplir el modelo, de diferenciar, en la varianza total del sistema, la parte debida a la varianza inter explicada por los factores y la parte intra. El modelo nos da noticia sobre el grado de significación de la dependencia.

4.3. ANÁLISIS DE COVARIANZA (ANACOVA)

Se trata de una extensión de los análisis de Anova o Manova en el sentido de que en el modelo que postula la dependencia hipotética entre la variable dependiente, de naturaleza cuantitativa, y los factores o variables independientes, de tipo cualitativo, se introducen variables control, de tipo cuantitativo. Se trata de un método aplicable a cualquier Anova o Manova. De alguna manera es una combinación de Anova/Manova con regresión.

4.4. ANÁLISIS DISCRIMINANTE SIMPLE Y MÚLTIPLE

Se parte de un modelo en el que una o varias VD de naturaleza cualitativa, nominales u ordinales, se hace o hacen depender de varias VI, de tipo cuantitativo. Es decir, se trata de encontrar la combinación lineal —función discriminante— de las variables independientes que dé cuenta de la máxima discriminación entre dos o más grupos dados por una variable cualitativa, o entre un cruce de categorías, si intervienen varias variables cualitativas. Por ejemplo tomando como VD, los *votos según partidos y/o sexo de los votantes* y como VI, *ingresos, número de hijos, años de residencia en el municipio, edad*, etc.

4.5. DIFERENCIAS DE PROPORCIONES PARA MODELOS CON MÁS DE DOS VARIABLES

El método utiliza el índice de diferencia de proporciones (de VD) entre dos valores de la supuesta variable independiente para establecer modelos de dependencia sobre el peso que la variación entre dos valores de las VI procura a la variación de las proporciones entre las VD. El modelo que se establece es de orden lineal y permite una representación gráfica de los coeficientes

de dependencia. El modelo permite extenderse a más de dos variables. Las variables son de naturaleza cualitativa.

4.6. MODELOS LOG-LINEALES Y LOG-LINEALES MÚLTIPLES

Sin entrar, pues no es el objetivo buscado en esta introducción, en la distinción existente entre diversos tipos de análisis logarítmicos lineales, se trata de establecer el peso que las distintas frecuencias marginales de diversas variables de tipo cualitativo tienen en las diferentes frecuencias, de la tabla de contingencia proveniente de dichas variables cruzadas. El modelo de dependencia que se establece es lineal a partir de expresiones logarítmicas de dichas frecuencias. Es posible modelizar diferentes tipos de dependencia: modelos saturados, modelos de dependencia, modelos jerárquicos. Por ejemplo si se desea analizar la relación existente entre *nivel de conocimiento del catalán, nivel socioeconómico, lugar de nacimiento, ingresos*, etc. Se trata de una extensión del análisis bivariado de tablas de contingencia.

4.7. ANÁLISIS DEL CAMINO

Se parte de un modelo de ecuaciones estructurales en forma de relación lineal entre las variables que pueden ser múltiples, endógenas o variables dependientes, y exógenas o variables independientes, todas ellas observables. El modelo es recursivo, esto es, marcado unidireccionalmente. Se trata de hallar los coeficientes de relación o coeficientes *path* que den la dependencia y determinen las ecuaciones estructurales. Estos coeficientes son de regresión estandarizados y son, como en la regresión, interpretables como la relación entre la variación o cambio estandarizado de una variable dependiente en función del cambio estandarizado de una variable independiente.

4.8. ANÁLISIS RECURSIVO DE ESTRUCTURAS DE COVARIANZA

Como en el método precedente se parte también de un modelo de ecuaciones estructurales en forma de relación lineal entre las variables que pueden ser múltiples, endógenas o variables dependientes y exógenas o variables independientes, pero a diferencia del caso anterior no todas ellas son directamente observables, esto es, son latentes incluidas en el modelo. El modelo sigue siendo recursivo en el sentido de que se admite solamente una dirección de dependencia. Se trata también de hallar los coeficientes de dependencia del modelo.

4.9. ANÁLISIS NO RECURSIVO DE ESTRUCTURAS DE COVARIANZA

Las consideraciones precedentes son repetibles. La variación proviene de que el modelo es no recursivo, esto es, no se da unidirección en el proceso de dependencia y puede, por consiguiente, darse una relación recíproca entre dos variables.

4.10. ANÁLISIS FACTORIAL CONFIRMATORIO

Como modelo de análisis factorial su objetivo consiste en detectar e identificar, de un conjunto de variables cuantitativas observadas, otras inobservadas, latentes o factores, cuyo número se postula más reducido que el de las observadas. El modelo postula una dependencia que se establece entre las variables manifiestas (observadas, o iniciales) y las latentes (factores) lineal. El método confirmatorio supone que no todas las variables observadas contribuyen con el mismo peso a la emergencia de los factores comunes, dados por la parte común de las variables, sino que dichas variables contienen una parte específica no explicada por los factores. El modelo de hipótesis orienta el establecimiento de dichos pesos.

4.11. ANÁLISIS DE COMPONENTES PRINCIPALES O FACTORIAL EXPLORATORIO

Se trata también de un método factorial y, como tal, pretende objetivos idénticos al precedente: descubrir variables latentes, identificarlas y reducir el espacio vectorial dado por las variables cuantitativas primitivas, haciendo a su vez que los factores o componentes (variables latentes) sean entre sí independientes. Se trata también de establecer un modelo lineal de dependencia entre factores y variables iniciales. El criterio utilizado para determinar las variables latentes consiste en encontrar aquellos ejes que en el espacio inicial, de las variables observadas, tengan la máxima dispersión para las unidades habidas. Con métricas adecuadas puede trabajarse con variables cuantitativas.

4.12. SERIES TEMPORALES

En todos los métodos reseñados, como en los que seguirán, la variable tiempo puede intervenir explícitamente sea como cuantitativa que es, sea como discreta. Las posibilidades de análisis que ofrecen estos métodos subsisten pues bajo hipótesis o supuestos que contengan el tiempo como variable. Además se ofrecen modelos de tratamiento específico, en los que no se entrará, como modelos estacionarios/no estacionarios, cíclicos o estacionales, etc.

4.13. ANÁLISIS DE SEGMENTACIÓN

Con el análisis de segmentación (*Tree Analysis*) dado un grupo de variables, y determinados hipotéticamente dos grupos entre ellas, uno de VD y otro de VI (lo normal es trabajar con un grupo de una sola variable dependiente), se trata de ir segmentando o dividiendo la población sucesivamente a partir de la elección de la variable independiente (politémica) y de las agrupaciones de sus valores que maximicen la separación o distancia entre los diferentes grupos y den mayor homogeneidad a los grupos formados para la(s) variable(s) dependiente(s).

4.14. ANÁLISIS DE CLASIFICACIÓN

El método del análisis de clasificación o de *Cluster* se suele denominar también análisis tipológico. El objetivo de la misma consiste en clasificar a las unidades de la matriz original en grupos o *Clusters* lo más homogéneos posible dentro de sí (*intra*) y lo más heterogéneos entre ellos (*inter*). La definición de un criterio inicial de distancia se impone —normalmente la distancia cuadrática euclídea— con lo que el punto de partida es una matriz de distancias entre las unidades de la muestra o experimento. Los diferentes índices que se establecen para determinar las distancias entre grupos una vez constituidos dan lugar a diferentes técnicas más específicas: centroide, distancia mínima, máxima, *ward*, etc. Los métodos de clasificación se dividen en dos tipos:

- los *jerárquicos* en los que el número de grupos no se establece fijo a priori. Dentro de éstos se distinguen a su vez los *ascendentes* o aglomerativos (*clumping*) en los que se parte inicialmente de cada unidad como un grupo en sí que se va asociando con otros sucesivamente hasta tener al final un solo grupo, y los *descendentes* o divisorios (*clustering*) en los que se parte de todas las unidades como un grupo único y se va dividiendo en grupos sucesivos según máxima distancia.
- los *no jerárquicos* en los que se parte de un número dado de grupos, según hipótesis o cálculos previos.

4.15. ESCALONAMIENTO MULTIDIMENSIONAL

El objetivo del conjunto de métodos englobados como de escalonamiento multidimensional (*Multidimensional Scaling*) consiste en descubrir la estructura implícita en una matriz de datos, de las unidades (o de las variables). Se trata de construir un espacio, a partir de una métrica definida, pero con

un número reducido de dimensiones posibles (dos o tres a lo sumo) de tal forma que posibilite expresar, y representar, las proximidades (preferencias) o similitudes entre objetos, situaciones, individuos, problemas, etc. Dado un conjunto de objetos en el que, o sobre el que, se supone o se prevee (hipotéticamente), o se percibe (empíricamente), una serie de proximidades (similitudes) o distancias (disimilitudes) el método procura unas dimensiones en cuyo espacio se da una configuración tal de los objetos (situaciones, individuos, etc.) cuyas distancias son equivalentes o se corresponden, en el grado aceptado y a partir de una serie de indicadores de la bondad de ajuste, con las proximidades supuestas o percibidas. La ventaja interpretativa de tales métodos proviene de la posibilidad de representar gráficamente los resultados mencionados. Se dan diversas clasificaciones de tales métodos: métodos *no métricos* como los métodos de escalonamiento multidimensional no métrico de *Kruskal*, y métodos métricos como el escalonamiento multidimensional métrico de *Tororson*.

4.16. ANÁLISIS FACTORIAL DE CORRESPONDENCIAS

Se trata de un análisis factorial con la lógica del Análisis de Componentes Principales, es decir, de un modelo de interdependencia. La diferencia básica con él estriba en la métrica dada en los espacios vectoriales —el de las variables como dimensiones y las unidades como nube de puntos y, a la inversa, el de las unidades como dimensiones y de las variables como nube de puntos— que permite un tratamiento equivalente del campo de las variables y del universo de los individuos, proyectando, esto es, hallando las componentes de ambos espacios, en un espacio único. El método tiene como objeto detectar las dimensiones o factores no observados, y con ello la estructura latente, de este espacio único a partir de las relaciones que se establecen entre ambos. La métrica utilizada permite el tratamiento, tanto de variables cualitativas a través de matrices de datos constituidas por hipertablas de contingencia como de variables cuantitativas.

En el presente número de la revista se presentan cinco de las técnicas mencionadas: el *Análisis de Componentes Principales*, aplicado al análisis de datos secundarios para el cálculo de una muestra estratificada; el *Análisis de Cluster*, en una perspectiva general sobre su aplicación, interpretación y validación; el *Análisis de Diferencias entre Proporciones*, tratado con un ejemplo concreto; el *Análisis Lineal Logarítmico*, ejemplificándolo en un tratamiento concreto del abandono de los estudios universitarios, y el *Análisis de Correspondencias Múltiples*, utilizado para el tratamiento de datos textuales.