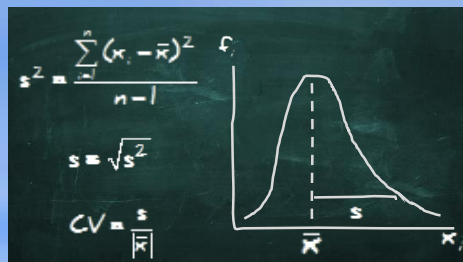


METODOLOGÍA DE LA INVESTIGACIÓN SOCIAL CUANTITATIVA

Pedro López-Roldán
Sandra Fachelli



METODOLOGÍA DE LA INVESTIGACIÓN SOCIAL CUANTITATIVA

Pedro López-Roldán
Sandra Fachelli




Bellaterra (Cerdanyola del Vallès) | Barcelona
Dipòsit Digital de Documents
Universitat Autònoma de Barcelona

UAB





Este libro digital se publica bajo licencia *Creative Commons*, cualquier persona es libre de copiar, distribuir o comunicar públicamente la obra, de acuerdo con las siguientes condiciones:

-  *Reconocimiento.* Debe reconocer adecuadamente la autoría, proporcionar un enlace a la licencia e indicar si se han realizado cambios. Puede hacerlo de cualquier manera razonable, pero no de una manera que sugiera que tiene el apoyo del licenciador o lo recibe por el uso que hace.
-  *No Comercial.* No puede utilizar el material para una finalidad comercial.
-  *Sin obra derivada.* Si remezcla, transforma o crea a partir del material, no puede difundir el material modificado.

No hay restricciones adicionales. No puede aplicar términos legales o medidas tecnológicas que legalmente restrinjan realizar aquello que la licencia permite.

Pedro López-Roldán

Centre d'Estudis Sociològics sobre la Vida Quotidiana i el Treball (<http://quit.uab.cat>)

Institut d'Estudis del Treball (<http://iet.uab.cat/>)

Departament de Sociologia. Universitat Autònoma de Barcelona

pedro.lopez.rolan@uab.cat

Sandra Fachelli

Departament de Sociologia i Anàlisi de les Organitzacions

Universitat de Barcelona

Grup de Recerca en Educació i Treball (<http://grupsderecerca.uab.cat/gret>)

Departament de Sociologia. Universitat Autònoma de Barcelona

sandra.fachelli@ub.edu

Edició digital: <http://ddd.uab.cat/record/129382>

1ª edición, febrero de 2015

Edifici B · Campus de la UAB · 08193 Bellaterra
(Cerdanyola del Vallés) · Barcelona · España
Tel. +34 93 581 1676

Índice general

PRESENTACIÓN

PARTE I. METODOLOGÍA

- I.1. FUNDAMENTOS METODOLÓGICOS
- I.2. EL PROCESO DE INVESTIGACIÓN
- I.3. PERSPECTIVAS METODOLÓGICAS Y DISEÑOS MIXTOS
- I.4. CLASIFICACIÓN DE LAS TÉCNICAS DE INVESTIGACIÓN

PARTE II. PRODUCCIÓN

- II.1. LA MEDICIÓN DE LOS FENÓMENOS SOCIALES
- II.2. FUENTES DE DATOS
- II.3. EL MÉTODO DE LA ENCUESTA SOCIAL
- II.4. EL DISEÑO DE LA MUESTRA
- II.5. LA INVESTIGACIÓN EXPERIMENTAL

PARTE III. ANÁLISIS

- III.1. SOFTWARE PARA EL ANÁLISIS DE DATOS: SPSS, R Y SPAD
- III.2. PREPARACIÓN DE LOS DATOS PARA EL ANÁLISIS
- III.3. ANÁLISIS DESCRIPTIVO DE DATOS CON UNA VARIABLE
- III.4. FUNDAMENTOS DE ESTADÍSTICA INFERENCIAL
- III.5. CLASIFICACIÓN DE LAS TÉCNICAS DE ANÁLISIS DE DATOS
- III.6. ANÁLISIS DE TABLAS DE CONTINGENCIA
- III.7. ANÁLISIS LOG-LINEAL
- III.8. ANÁLISIS DE VARIANZA
- III.9. ANÁLISIS DE REGRESIÓN
- III.10. ANÁLISIS DE REGRESIÓN LOGÍSTICA
- III.11. ANÁLISIS FACTORIAL
- III.12. ANÁLISIS DE CLASIFICACIÓN

Metodología de la Investigación Social Cuantitativa

Pedro López-Roldán
Sandra Fachelli

PARTE III. ANÁLISIS

Capítulo III.5 Clasificación de las técnicas de análisis de datos

Bellaterra (Cerdanyola del Vallès) | Barcelona
Dipòsit Digital de Documents
Universitat Autònoma de Barcelona

UAB



Cómo citar este capítulo:

López-Roldán, P.; Fachelli, S. (2015). Clasificación de las técnicas de análisis de datos. En P. López-Roldán y S. Fachelli, *Metodología de la Investigación Social Cuantitativa*. Bellaterra (Cerdanyola del Vallès): Dipòsit Digital de Documents, Universitat Autònoma de Barcelona. 1^a edición. Edición digital: <http://ddd.uab.cat/record/131468>

Capítulo redactado en abril de 2015

Índice de contenidos

1. CLASIFICACIÓN DE LAS TÉCNICAS DE ANÁLISIS BIVARIABLE.....	6
2. EL ANÁLISIS MULTIVARIABLE: CARACTERÍSTICAS Y CLASIFICACIÓN	10
2.1. Definición y caracterización del análisis multivariable	11
2.2. Criterios y clasificación de las técnicas de análisis multivariable.....	13
3. BREVE RESEÑA DE ALGUNAS TÉCNICAS DE ANÁLISIS MULTIVARIABLE.....	15
3.1. Técnicas de análisis multivariable de interdependencia.....	16
3.1.1. <i>El análisis de tablas de contingencia multidimensionales</i>	16
3.1.2. <i>El análisis log-lineal</i>	16
3.1.3. <i>El análisis factorial exploratorio</i>	17
3.1.4. <i>El análisis de clasificación</i>	17
3.1.5. <i>El análisis de escalonamiento multidimensional</i>	18
3.2. Técnicas de análisis multivariable de dependencia.....	18
3.2.1. <i>El análisis de regresión múltiple</i>	18
3.2.2. <i>El análisis de regresión logística</i>	19
3.2.3. <i>El análisis de la varianza multifactorial y multivariable</i>	19
3.2.4. <i>El análisis de covarianza</i>	20
3.2.5. <i>El análisis discriminante</i>	20
3.2.6. <i>El análisis del camino</i>	20
3.2.7. <i>El análisis factorial confirmatorio</i>	21
3.2.8. <i>Modelo de ecuaciones estructurales</i>	21
3.2.9. <i>El análisis de segmentación</i>	22
4. BIBLIOGRAFÍA	22

Clasificación de las técnicas de análisis de datos

En los capítulos precedentes hemos tenido ocasión de ver algunas de las técnicas de análisis de datos más básicas y fundamentales, presentes en todo ejercicio de estudio de los datos cuantitativos, pero que tratan con una sola variable. Con ellas podemos responder a preguntas de investigación más sencillas y nos posibilitan acondicionar nuestros datos para un análisis más elaborado que ahora iniciamos. Se busca responder a cuestiones o hipótesis de investigación más complejas e interesantes donde se plantean relaciones entre varias variables, para así reflejar la naturaleza multidimensional de los fenómenos sociales y de los modelos analíticos que dan cuenta de ellos. En la sociedad de la información actual, además, disponemos cada vez más de inmensos volúmenes de datos (*Big Data*) que requieren soluciones de procesamiento mediante el tratamiento simultáneo de diversidad de variables y muchos casos. En este sentido ya hemos apuntado anteriormente que las técnicas de análisis de datos estadísticos se pueden clasificar en tres grupos principales atendiendo a la complejidad que resulta de considerar el **número de variables implicadas**, así se pueden distinguir las técnicas univariadas, bivariadas y multivariadas.

Esta es una primera diferenciación importante que tenemos presente en este manual. Por el momento hemos visto las técnicas relacionadas con el análisis de una sola variable introduciendo en particular la diferenciación entre el aspecto descriptivo y el aspecto inferencial cuando la información se obtiene por muestreo estadístico. A partir del próximo capítulo veremos las técnicas bivariadas y multivariadas. Esta distinción es útil para diferenciar instrumentos de análisis que responden a modelos e hipótesis de investigación donde se relacionan dos variables o bien más de dos, pudiendo llegar a tratar simultáneamente decenas de ellas.

No obstante las técnicas bivariadas no son más que una versión particular del caso más general multivariado, y en este sentido no veremos capítulos específicos que las diferencien sino que se tratará de formulaciones más básicas bivariadas que se ampliarán como técnicas multivariadas.

En la clasificación de las técnicas de análisis de datos, además de la cantidad, es definitorio el tipo de relación que se establece entre las variables, donde se realiza una

distinción entre técnicas de **análisis de interdependencia y de dependencia**. En el primer caso se disponen de un modelo de análisis que plantea de forma más o menos precisa la existencia de una serie de interrelaciones entre un conjunto de variables pero donde no se diferencia su papel al tratarlas de forma simétrica, podemos decir que todas ellas se consideran como independientes e interesa encontrar o constatar su patrón o estructura de relaciones mutuas y simultáneas. En el segundo caso sí que se establece la diferenciación entre variable dependiente e independiente, una(s) actúan de variables explicadas y otra(s) de variables explicativas según un modelo de análisis explicativo que conceptualiza y explicita hipótesis sobre el vínculo de dependencia.

Existe un tercer aspecto fundamental para caracterizar las técnicas de análisis de datos: el **nivel de medición** de las variables, su métrica. En función de si las variables son cuantitativas (métricas, de tipo discreto o continuo) o cualitativas (no métricas, de tipo nominal u ordinal) cabe establecer este tercer criterio clasificatorio que las diferencia y define.

1. Clasificación de las técnicas de análisis bivariable

Con este criterio de la escala de medición vamos a establecer una clasificación inicial básica de las técnicas de análisis de datos que relacionan dos variables. Veremos que estas técnicas serán el fundamento de muchos conceptos y técnicas multivariantes que se irán introduciendo. El esquema clasificatorio es sencillo. Tratándose de dos variables, nos podemos encontrar con que o ambas son cualitativas, o son cuantitativas las dos o bien analizamos la relación entre una cualitativa y una cuantitativa (Tabla III.5.1).

Tabla III.5.1. Clasificación de las técnicas de análisis de datos



La Tabla III.5.1 se presenta resumida la clasificación de las técnicas de análisis de datos básicas, las univariantes que hemos visto en los capítulos III.3 y III.4, las bivariables que comentaremos brevemente a continuación y que serán objeto de tratamiento en

los capítulos III.6, III.7 y III.8, y las multivariantes de las que daremos una reseña al final de este capítulo y que serán desarrolladas en el resto de los capítulos del libro.

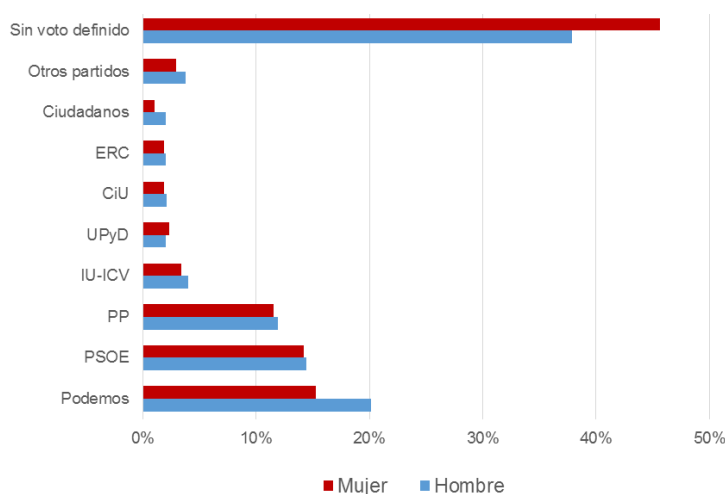
Cuando analizamos la relación entre dos variables cualitativas la técnica de análisis de datos característica es el **análisis de tablas de contingencia**. Una tabla de contingencia en una tabla de frecuencias que resulta del cruce de las categorías o valores de ambas variables y nos permite estudiar la distribución de una variable, por ejemplo, la intención de voto en la próximas elecciones generales (considerada habitualmente como variable de interés a explicar o variable dependiente), según los valores de una segunda variable, por ejemplo, el sexo. La pregunta que nos formulamos es si existen diferencias en el comportamiento intencional de voto por los distintos partidos políticos entre varones y mujeres, y en qué medida, o si por el contrario son similares. En la Tabla III.5.2 la tabla de contingencia que cruza ambas variables y se representa en un gráfico de barras en el Gráfico III.5.1¹.

Tabla III.5.2. Intención de voto en supuestas elecciones generales según sexo

		Sexo		
		Hombre	Mujer	Total
Intención de voto en supuestas elecciones generales	Podemos	20,1%	15,3%	17,6%
	PSOE	14,4%	14,2%	14,3%
	PP	11,9%	11,5%	11,7%
	IU-ICV	4,0%	3,4%	3,7%
	UPyD	2,0%	2,3%	2,1%
	CiU	2,1%	1,9%	2,0%
	ERC	2,0%	1,9%	1,9%
	Ciudadanos	2,0%	1,0%	1,5%
	Otros partidos	3,8%	2,9%	3,3%
	Sin voto definido	37,9%	45,6%	41,9%
	Total	100,0%	100,0%	100,0%

Fuente: Centro de Investigaciones Sociológicas, Estudio 3041 de 2014

Gráfico III.5.1. Intención de voto en supuestas elecciones generales según sexo

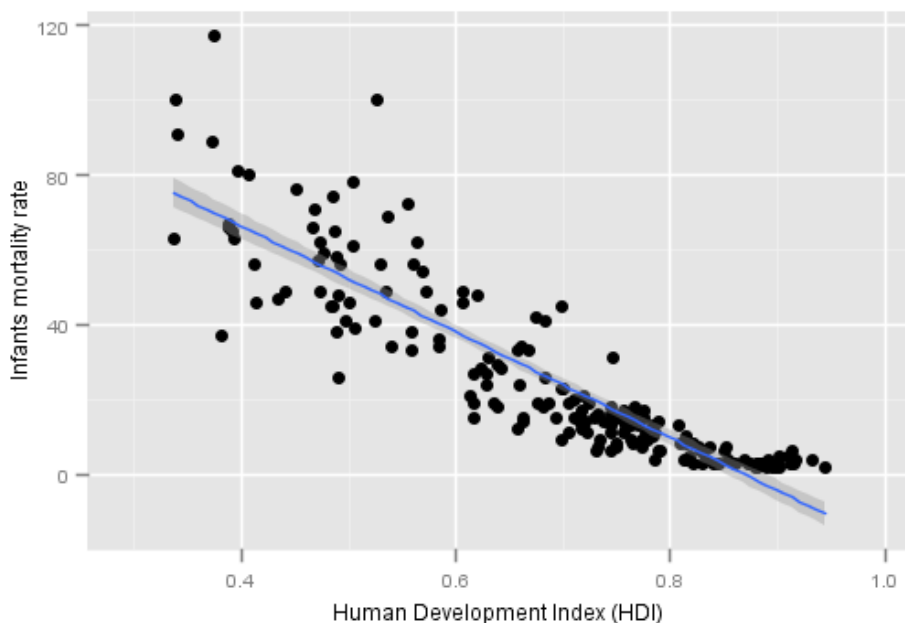


¹ Los datos corresponden a la matriz CIS3041.sav.

De esta información se desprende un perfil similar entre varones y mujeres excepto en la indefinición del voto, que es superior en las mujeres, y en el voto a **Podemos** donde predomina claramente el voto masculino.

Cuando analizamos la relación entre dos variables cuantitativas la técnica de análisis que las trata es el **análisis de regresión lineal**. En este caso tras constatar que existe un grado de correlación lineal entre las variables, la técnica expresa a una de ellas, la variable dependiente, en función de otra variable que se considera como independiente o factor explicativo. El vínculo que se establece es una relación funcional, y se expresa a través de la función lineal, es decir, mediante la ecuación de una recta. Así, por ejemplo, considerando los países del mundo podemos plantear un modelo de relación para explicar en qué medida la tasa de mortalidad infantil está en función del nivel de desarrollo del país considerando el IDH (Índice de Desarrollo Humano). Así esperamos obtener que a medida que el índice de desarrollo es mayor la tasa de mortalidad infantil será menor, es decir, se dará una relación inversa. Así lo podemos comprobar en el gráfico de dispersión del Gráfico III.5.2 y en los coeficientes de la Tabla III.5.3. Cada punto del gráfico es un país y éstos se disponen de manera que los que tiene alto IDH tiene bajos valores de mortalidad, y al viceversa.

Gráfico III.5.2. Mortalidad infantil según el IDH



Se configura así una nube de puntos que se puede ajustar a la forma de una línea recta, y cuyos coeficientes permiten interpretar que por cada punto más de desarrollo la tasa de mortalidad se reduce en -141,270, teniendo el modelo una capacidad explicativa del 79,8%².

² Los datos corresponden a la matriz **IDH2014.sav**.

Tabla III.5.3. Regresión Mortalidad infantil según el IDH

Modelo lineal	Coefficientes de regresión
Constante	122,779
Human Development Index (HDI)	-141,270
Capacidad explicativa	Coefficientes
Correlación	-0,893
Coefficiente de determinación	0,798

Fuente: United Nations, IDH 2014

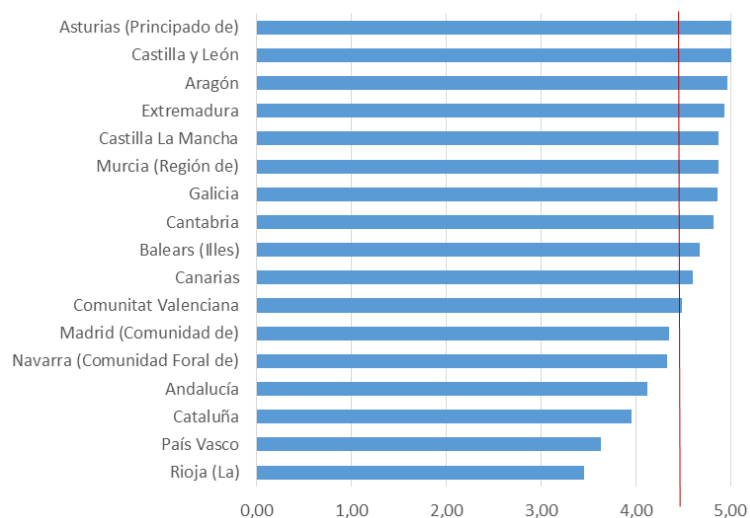
Cuando analizamos la relación entre una variable cuantitativa y una variable cualitativa la técnica de análisis principal que se considera es la **comparación de medias** y, de forma más general, el **análisis de varianza**. En este caso se trata de ver hasta qué punto la distribución de los valores de la variable cuantitativa, que es considerada como la dependiente, cambia, es diferente entre los diferentes grupos definidos por la variable independiente. Por ejemplo, nos podemos preguntar en qué medida el posicionamiento ideológico, tratado como una variable cuantitativa en una escala de 1 a 10 (siendo 1 izquierda y 10 derecha), difiere entre las personas de las distintas comunidades autónomas y, por tanto, determinar así en qué grado el lugar de residencia es una variable explicativa de las diferencias ideológicas. Para ello podemos comparar las distribuciones por comunidad autónoma calculando la media de cada una como se presenta en la Tabla III.5.4 y el Gráfico III.5.3

Tabla III.5.4. Ideología según Comunidad Autónoma

Comunidad autónoma	Escala de autoubicación ideológica (1=izq-10=der)		
	Media	n	Desviación típica
Rioja (La)	3,46	13	1,613
País Vasco	3,63	87	1,622
Cataluña	3,96	338	1,786
Andalucía	4,13	335	1,741
Navarra (Comunidad Foral de)	4,33	27	1,732
Madrid (Comunidad de)	4,35	253	2,136
Comunitat Valenciana	4,48	229	1,791
Canarias	4,60	75	1,993
Baleares (Illes)	4,67	55	2,373
Cantabria	4,82	22	1,651
Galicia	4,86	123	2,086
Murcia (Región de)	4,87	46	1,529
Castilla La Mancha	4,88	91	2,299
Extremadura	4,94	48	2,409
Aragón	4,96	57	1,812
Castilla y León	5,04	111	1,926
Asturias (Principado de)	5,06	48	1,731
Total	4,41	1958	1,944

Fuente: Centro de Investigaciones Sociológicas, Estudio 3041 de 2014

Gráfico III.5.3. Media de la escala ideológica según Comunidad Autónoma



2. El análisis multivariable: características y clasificación

En este apartado³ daremos una visión panorámica del conjunto de procedimientos que de manera genérica se denominan como métodos y técnicas de análisis **multivariable**, también nombrados como de análisis **multivariante** o **multivariado**⁴.

Se trata de un conjunto de procedimientos de análisis estadístico donde se suele considerar un número elevado de variables, de hecho, como venimos diciendo, cuando hay más de dos variables en el análisis se identifica como multivariable. Esta es una definición que en primera instancia nos orienta sobre una característica importante de estas técnicas. Sin embargo, podemos ser más precisos y contemplar otros rasgos importantes. A esta tarea de definición nos dedicaremos en este apartado y se completará con la especificación de varios criterios clasificatorios de las diversas técnicas que nos conducirán a proponer una clasificación de las mismas.

En la investigación social empírica que se realiza en la actualidad es habitual encontrar el tratamiento multivariable de los datos estadísticos. La complejidad que conlleva el tratamiento conjunto de un número elevado de variables y las dificultades de asimilación de los conceptos y del lenguaje propio de las técnicas que las tratan ha sido una de las razones por las que el investigador/a en sociología, y en las ciencias sociales en general, no haya hecho un uso generalizado de este tipo de análisis hasta hace relativamente poco. La introducción y generalización del uso de los softwares estadísticos marca un punto de inflexión (Alvira, 1990) en el camino hacia “normalidad” de estos instrumentos en la investigación social aplicada para dar cuenta

³ Este apartado es reflejo de la revisión de publicaciones anteriores (Lozares y López-Roldán, 1991, 2000).

⁴ La expresión multivariante, al igual que bivariante o univariante, alude sobre todo al tipo de distribución estadística de los datos; multivariado tendría en cuenta la presencia de varios grados de diversidad, mientras que multivariable destaca la cantidad de variables que intervienen. En inglés la expresión utilizada es *multivariate* y en francés *multivariée*.

de la complejidad de los modelos de análisis que intentan reflejar una visión multidimensional de la realidad social⁵.

No obstante, algunos de estos procedimientos tienen sus antecedentes a finales del siglo XIX y principios del XX, siendo durante las décadas de los años treinta y cuarenta cuando comenzaron a emplearse las técnicas de análisis multivariable. Aun así, no es hasta la década de los ochenta con el desarrollo experimentado por la informática aplicada al análisis estadístico y la facilidad y posibilidades de cálculo que conlleva su implementación en macro y microordenadores, que se produce la extensión del uso de las técnicas estadísticas de análisis multivariable en numerosas disciplinas y también en la sociología.

Las ventajas que presenta el análisis multivariable sobre el bivariado eran puestas de manifiesto por Cattell (1966):

- Economía en el almacenamiento de datos.
- Mayor consistencia en la inferencia estadística.
- Desarrollo de conceptos teóricos más adecuados.
- Mayor precisión y perspectiva conceptual.

Hay que insistir sin embargo, que se trata de instrumentos que facilitan:

- El análisis y la síntesis de los datos.
- La operativización y la validación de los conceptos.
- La mejora en la obtención y la calidad derivada de los propios datos.

Pero también hay que decir que en nada sustituyen la tarea imprescindible de construcción del objeto de investigación y su fundamentación teórica, del diseño del mismo, del análisis y de la interpretación de los resultados guiados por referentes teóricos-conceptuales.

2.1. Definición y caracterización del análisis multivariable

El **análisis multivariable** es un conjunto de métodos y técnicas destinadas al análisis estadístico de las relaciones entre tres o más variables que son tratadas simultáneamente. Se trata de una definición básica e incluyente en la línea de otros autores como Hair et al. (2011) o Lévy y Varela (2003: 6), y que supone la extensión y generalización de los conceptos y las técnicas del análisis univariable y bivariado. Partiendo de esta primera acepción general se propone una definición ampliada a partir de una caracterización que introduce una serie de condiciones más específicas para asegurar más convenientemente los límites conceptuales del término.

Así se contemplan dos aspectos. En primer lugar, el análisis multivariable lo situamos en una perspectiva atributiva que supone una clara distinción, ruptura o desagregación entre lo que denominamos el **universo de los individuos** o poblacional (unidades,

⁵ Recordemos para los más jóvenes que el primer ordenador personal (los IBM PC) aparecen en el año 1981. En esa época se trabajaba con terminales de grandes ordenadores en instituciones y centros de investigación que podían acceder a este tipo de recursos. No es hasta 1984 que aparece el SPSS/PC por ejemplo, y es por entonces que empieza lentamente a extenderse su uso y la enseñanza en las universidades tanto de las técnicas como del software estadístico que hace posible su utilización.

casos,...), sujetos de la atribución conceptual, y el **espacio de las variables** (o espacio de atributos) presentes en una matriz de datos atributiva. Esta distinción conlleva y posibilita que los datos, al ser expresables en forma matricial, puedan ser interpretadas en una doble lectura: desde las unidades y desde las variables. Kendall (1975) se orienta 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, entendiendo por dependencia la vinculación de las variables a los individuos de los que se predicen.

En segundo lugar, el soporte matricial introduce una segunda 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 que lo lee, mide, trata y analiza, y, por otra parte, también es el resultado de una construcción de naturaleza científica, en tanto en cuanto el dato se inscribe a una determinada teoría más o menos elaborada, a unos presupuestos e hipótesis y, sobre todo, al haber sido objeto de una validación interna y externa en diversos grados. Esta dualidad, intrínseca al dato concreto, es extensible a la matriz de datos.

En relación a la primera de las polaridades mencionadas de la dualidad del dato, las técnicas son multivariables en el sentido que posibilitan y ofrecen una flexibilidad en cuanto al lenguaje algebraico, y en particular en cuanto a las **diversas métricas** exigibles en función de la estructura algebraica utilizada cuando operativizan los conceptos. La métrica definida es múltiple y diversa, adaptable a la naturaleza de la operativización. Las matrices pueden ser, según los métodos, de diferente naturaleza: lógicas, nominales, de frecuencias absolutas, relativas o condicionales, matrices de rangos o matrices cuantitativas. Definir las técnicas de análisis multivariable exclusivamente como técnicas estadísticas es de hecho restrictivo: más ampliamente el álgebra matricial y vectorial suele ser el fundamento de los métodos utilizados⁶.

En relación a la segunda polaridad, es decir, la relativa a las exigencias ligadas a la inserción de las técnicas en la instancia teórica, de la corroboración de hipótesis, etc., el conjunto de las técnicas de análisis multivariable posibilita **varios niveles de análisis** vinculados a estas exigencias. Son análisis y orientaciones metodológicas que pueden plantearse a partir de estas técnicas:

- En el orden de la **parsimonia** exigible a todo método, resalta el efecto simplificador y reductor que procuran, permitiendo también tratamientos simultáneos de todas las variables o unidades que intervienen o un tratamiento escalonado de alguna(s) unidad(es) o variable(s). En algunos casos incluso las unidades o variables son introducidas sin intervenir en los cálculos con una finalidad descriptiva o de localización en el conjunto.
- También hay que considerar los efectos de consistencia y de validación en la **construcción y definición de las variables**. Mediante la construcción de tipologías o la obtención de variables métricas que subyacen en la combinación de un conjunto inicial de información, disponemos de métodos que tienen una función instrumental destinada a mejorar la calidad del proceso de operativización y de los

⁶ En el anexo del libro se recoge un apartado dedicado al tratamiento matemático de la información para el análisis estadístico donde se introducen los conceptos de vector y espacio vectorial, las operaciones con vectores y matrices, etc., que están en la base de buena parte de las técnicas de análisis multivariable.

datos en general, pero que se constituyen en un procedimiento heurístico para la construcción de variables.

- Las técnicas de análisis multivariable permiten satisfacer un objetivo metodológico **descriptivo** de la problemática tratada donde se busca dar cuenta de la realidad investigada especificando los elementos y propiedades fundamentales que lo caracterizan y estructuran como resultado de una visión conjunta de la información seleccionada y medida. Metodológicamente pueden responder tanto a necesidades y dinámicas de exploración e inducción como de deducción y validación.
- Cuando el objetivo del método es dar cuenta del porqué del fenómeno identificamos al procedimiento como **explicativo**, lo que nos exige un modelo previamente establecido que explicita y fundamenta los factores explicativos que determinan el comportamiento de una realidad social medida a través de una o más variables que dependen de aquéllos. Con este planteamiento las técnicas de análisis multivariable se orientan hacia un tipo de análisis deductivo y confirmatorio y en general de estructura causal.
- La descripción y la explicación pueden llevar asociada una finalidad también **predictiva**, en este caso, en función del modelo de análisis operativizado, es posible realizar afirmaciones basadas en los datos que establecen un comportamiento previsible, una tendencia probable del fenómeno estudiado en un tiempo futuro próximo.
- Podríamos introducir adicionalmente el factor tiempo y dar lugar a los diversos **análisis temporales** que también se incluyen dentro del ámbito del análisis multivariable. Y junto al tiempo el **espacio**, ya sea para una caracterización de fenómenos que tienen una clara expresión espacial y mapeable, como para dar cuenta de fenómenos que presentan niveles distintos de agregación territorial.

Las posibilidades y bondades de este tipo de análisis, más allá de los procedimientos más sencillos univariados y bivariados, no nos debe hacer perder de vista que su utilización no puede eludir todo lo que supone un exigente planteamiento teórico, la construcción de un objeto de investigación que haga verificables las hipótesis, una operativización cuidadosa de los conceptos, la calidad de la producción de los datos en todas sus fases, la utilización de una métrica adecuada de las variables así como su validación interna.

2.2. Criterios y clasificación de las técnicas de análisis multivariable

Todo intento de clasificación se encuentra con la dificultad que proviene de la variedad existente de métodos/técnicas de análisis multivariable y que varios criterios operen simultáneamente. Algunos de ellos incluso se superponen en parte, lo que introduce un grado más de dificultad. Así pues, introduciremos primero los diferentes criterios habitualmente utilizados como principios clasificatorios para, a continuación, presentar la propuesta de clasificación.

Un primer criterio de clasificación depende del **modelo de análisis** de la investigación y se establece teniendo en cuenta la finalidad analítica explicativa⁷. En este sentido se puede hablar de dos tipos de técnicas:

⁷ La base inicial de la clasificación de Evrard y Le Maire (1976), por ejemplo, va en esta dirección.

- Técnicas de análisis multivariable explicativo-causales que buscan formalizar un modelo de relaciones que conlleven hipótesis proposicionales simples o sistemas de hipótesis de dependencia lineal.
- Técnicas exploratorias y de estructuración donde se incluyen modelos diversos: los que implican niveles relativamente elementales de categorización o de contrastación comparativa; los que buscan las dimensiones latentes o de los factores más discriminantes de la realidad social considerada estableciendo estructuras de interrelación entre las variables; o los que buscan construir tipologías concretas como formas de tipificación de la realidad como forma una explicación estructural de los fenómenos

Un segundo criterio es algebraico y técnico-instrumental, como condición previa a toda finalidad analítica, que establece la elección de las técnicas según:

- Impliquen una relación de **dependencia** entre las variables, donde se establece la distinción entre variables dependientes o explicadas y variables independientes o explicativas.
- Impliquen una relación de **interdependencia** o la simple correlación/asociación entre las variables, donde todas las variables tienen la misma consideración (si se quiere de variables independientes).

Un posible criterio adicional técnico sería el **número de variables** que intervienen en el análisis:

- Cuando se trata de técnicas de dependencia, se valora cuántas variables dependientes intervienen, asumiendo que se considerará más de una variable independiente.
- Cuando se trata de técnicas de interdependencia se considera que se relacionan más de dos variables.

Por último, se pueden conjugar los dos criterios anteriores con la **métrica** de las variables: si son variables métricas o cuantitativas (numéricas, medidas como discretas o continuas) o no son métricas o cualitativas (categóricas, medidas a nivel nominal u ordinal).

En la Tabla III.5.5 se ofrece una clasificación donde se conjugan varios criterios sucesivos que sintetizan y contienen el mayor número de técnicas de análisis multivariable teniendo en cuenta la complejidad de los criterios adoptados. De hecho la propuesta clasificatoria tiene en cuenta los tres últimos criterios comentados: análisis de relaciones de dependencia o de interdependencia, una variable dependiente o dos o más, y métricas de las variables. La tabla se completa con otras técnicas de análisis cuya caracterización es difícil de situar en el esquema propuesto.

Tabla III.5.5. Clasificación de las técnicas de análisis multivariable

Análisis de relaciones de interdependencia $V \leftrightarrow V$		Vs cualitativas	Análisis de Tablas de Contingencia Multidimensionales Análisis Log-Lineal Análisis de Correspondencias Análisis de Clasificación Análisis de Escalamiento Multidimensional No Métrico Análisis de Clases Latentes	
		Vs cuantitativas	Análisis Factorial Exploratorio. Análisis de Componentes Principales Análisis de Clasificación Análisis de Escalamiento Multidimensional Métrico	
Análisis de relaciones de dependencia $VI \rightarrow VD$	1 variable dependiente	VD cuantitativa	VIs cuantitativas	Análisis de Regresión Lineal Múltiple Análisis de Regresión No Lineal Múltiple
			VIs cualitativas y cuantitativas	Análisis de Covarianza
			VI cualitativas	Análisis de Series Temporales Análisis de Supervivencia
	2 ó más variables dependientes	VDs cuantitativas	VIs cuantitativas	Análisis de Varianza Múltiple Análisis de Segmentación Análisis Conjunto
			VIs cualitativas y cuantitativas	Análisis Discriminante
			VI cualitativas	Análisis de Regresión Logística Análisis de Regresión Ordinal Análisis de Regresión Probit Análisis de Regresión Lineal Múltiple con variables ficticias
	2 ó más variables dependientes	VDs cuantitativas	VIs cuantitativas	Análisis Log-lineal Logit Análisis de Segmentación Análisis Conjunta
			VIs cualitativas y cuantitativas	Análisis de Modelos de Ecuaciones Estructurales
		VDs cualitativas	VIs cuantitativas	Análisis de Covarianza Múltiple Análisis de Correlación Canónica
			VIs cualitativas	Análisis Multivariable de Varianza
Otras técnicas de análisis		VIs cuantitativas	Análisis Discriminante Múltiple	
		VIs cualitativas	Análisis Log-lineal Logit Múltiple	
		Análisis de Redes Sociales Minería de Datos Análisis de Redes Neuronales Simulación Social Análisis de Decisión Multicriterio		

Notación: **V**: variable **VD**: variable dependiente **VI**: variable independiente.

Variables **cualitativas** (no métricas o categóricas): nominales u ordinales.

Variables **cuantitativas** (métricas o numéricas): de intervalo o de razón (discretas o continuas).

Variable ficticia o *dummy*: variable binaria codificada con 0 i 1.

3. Breve reseña de algunas técnicas de análisis multivariable

En este apartado presentaremos de forma muy resumida y ejemplificada las características del conjunto de procedimientos que hemos presentado en la clasificación anterior para así tener una noción general de los objetivos y aplicaciones de cada técnica analítica. En esta tercera parte del libro, a partir del próximo capítulo, se dará cuenta de buena parte de estos procedimientos de análisis multivariable.

Dividiremos la exposición en dos apartados para diferenciar claramente las técnicas de de análisis de interdependencia de las de dependencia.

3.1. Técnicas de análisis multivariable de interdependencia

3.1.1. *El análisis de tablas de contingencia multidimensionales*

El análisis de tablas de contingencia lo hemos presentado más arriba en su versión bivariable. El análisis multivariable implica introducir terceras y cuartas variables en la tabla, habitualmente no más, pues el cruce entre ellas aumenta el número de casillas con la consecuente insuficiencia de frecuencias en ellas y la complejidad de la lectura de la información. El análisis trivariable por ejemplo implica introducir una tercera variable que se considera como variable de control para ver la relación entre la tres y en particular para observar en qué medida una relación bivariable se modifica a cada nivel de la tercera. De hecho del análisis multidimensional consiste en multiplicar el análisis bidimensional con tantas tablas como valores tenga la tercera variable. Así por ejemplo podríamos analizar la relación entre el nivel de ingresos y el sexo y preguntarnos en qué medida su desigual distribución depende de la categoría ocupacional que se tiene. Si tuviéramos dos niveles ocupacionales (alto y bajo) compararíamos la relación entre ingresos y sexo en una primera tabla entre los que tienen nivel alto y en una segunda para los que tienen nivel bajo, y las compararíamos entre sí. Diversos resultados son posibles, uno podría ser que la relación desapareciera por que la verdadera razón de ingresar diferente es tener una ocupación mejor remunerada, y cómo los varones suelen estar ocupados en mejores ocupaciones ganan más en consecuencia.

3.1.2. *El análisis log-lineal*

El análisis log-lineal permite superar algunas limitaciones lo que se ha venido en llamar el análisis clásico de tablas de contingencia que acabamos de ver. La principal limitación radica en la imposibilidad de establecer conclusiones significativas estadísticamente de la relación entre las variables cuando se introducen terceras y sucesivas variables, si bien en ambas técnicas el número de variables que se pueden tratar simultáneamente es una limitación y no suele exceder de cinco. Otras técnicas como el análisis de correspondencias múltiples veremos que evita este aspecto restrictivo. En el caso del análisis log-lineal se trata de encontrar el mejor modelo y más sencillo (más parsimonioso) que se ajuste a los datos y exprese la relaciones de asociación entre las variables. Las asociaciones (o interacciones) pueden ser entre dos variables, entre tres o más, imponiendo en general un criterio de jerarquía: una relación entre tres variables presupone un vinculación entre parejas de ellas. Finalmente se obtiene un modelo y éste se cuantifica a través de unos parámetros que dan cuenta de la importancia de las relaciones entre las variables. Una de las aplicaciones más habituales en Sociología es el análisis de la movilidad social intergeneracional para obtener modelos que relacionan el origen social de los padres y madres con el destino social de los hijos e hijas (en términos ocupacionales, de clase social, de educación). Esta relación entre origen y destino se relaciona con las diferentes cohortes, diferenciando varones de mujeres y varones o comparando diferentes territorios entre sí. Estos análisis permitirían concluir

hasta qué punto las desigualdades sociales que expresa la estratificación social se han mantenido constantes en el tiempo, ha aumentado la reproducción social o la fluidez social.

3.1.3. *El análisis factorial exploratorio*

Dentro del análisis factorial exploratorio se incluyen procedimientos como el Análisis de Componentes Principales o el Análisis de Correspondencias. Como métodos factoriales tratan de descubrir variables latentes, identificarlas y reducir el espacio vectorial definido por el conjunto más numeroso de variables originales, haciendo a la vez que estos factores (variables latentes) sean independientes entre sí y estén ordenados progresivamente de mayor a menor importancia o poder explicativo. Se trata de análisis de interdependencia donde también se formaliza un modelo matemático lineal donde se muestra la relación entre las variables originales y los factores. El criterio utilizado para determinar las variables latentes consiste en encontrar aquellos ejes que en el espacio inicial tengan la máxima dispersión o variabilidad según las unidades observadas, es decir, son los factores que más discriminan a los individuos o a las unidades consideradas. Bajo este principio general, el Análisis de Componentes Principales trabaja a partir de variables cuantitativas mientras que el Análisis de Correspondencias lo hace con variables de naturaleza cualitativa. Como ejemplo de este tipo de análisis podríamos considerar un conjunto de indicadores de caracterización social de los municipios de una región (de actividad laboral, de distribución por sectores de actividad, de categorías profesionales, de niveles educativos, de movilidad espacial, demográficos, ...) y el análisis podría mostrar que el factor principal de diferenciación de estos municipios es una variable compleja asociada al concepto de clase social, en segundo lugar, nos podemos encontrar con un factor identificado por la dimensión demográfica de la edad, y un tercero marcado por el tipo de actividad y la dimensión rural-urbano, que resumirían un conjunto extenso de variables cuya estructura latente se expresa de forma sintética en estas tres dimensiones o factores.

3.1.4. *El análisis de clasificación*

Por Análisis de Clasificación (*cluster analysis*) identificamos una serie de procedimientos cuyo objetivo es clasificar un conjunto de individuos o unidades en grupos lo más homogéneos internamente y los más heterogéneos o diferentes entre ellos. Los procedimientos concretos tienen en cuenta la métrica utilizada y la definición de un criterio inicial de proximidad (de similitud o de distancia) entre las unidades; se trata entonces de comparar las unidades entre sí según este criterio de proximidad y unir en un mismo grupo las unidades más parecidas entre sí de acuerdo con el conjunto de variables que actúan de criterios clasificatorios. La forma de realizar estas comparaciones configuran los diferentes métodos clasificatorios, los cuales se pueden dividir en dos tipos. En primer lugar tenemos los métodos jerárquicos, donde el número de grupos no se establece a priori y se trata de que a partir de considerar cada unidad como un grupo en sí mismo se vayan uniendo progresivamente en grupos de manera ascendente o aglomerativa hasta tener un solo grupo, o bien, de forma inversa, descendente o por división, a partir de considerar todas las unidades en un solo grupo se trata de ir dividiéndolo sucesivamente. En este proceso ascendente o descendente

se tratará de decidir la partición o la clasificación en un número de grupos que son objeto de análisis y validación. Por su parte los procedimientos no jerárquicos parten de un número de grupos dado según hipótesis o análisis previos y se operan diferentes algoritmos de clasificación según el método. Así, por ejemplo, con un determinado procedimiento clasificatorio podríamos buscar la construcción de una tipología de zonas o de barrios, de municipios, de regiones, de países, etc. en función de un conjunto de variables o de indicadores sociales y económicos previamente seleccionados que actúan de criterios clasificatorios.

3.1.5. *El análisis de escalonamiento multidimensional*

El objetivo del análisis de escalonamiento múltiple (*Multidimensional Scaling*) consiste en descubrir la estructura implícita de 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 de dimensiones reducido (dos o tres como máximo) de tal forma que se posibilite expresar, y representar las proximidades o semejanzas entre objetos, individuos, situaciones, etc. Dado un conjunto de objetos a partir de los cuales se supone -hipotéticamente- o se observa -empíricamente- una serie de cercanías (similitudes o distancias), el método proporciona unas dimensiones a partir de las cuales se configura un espacio con una configuración de los objetos con distancias que son equivalentes o se corresponden, a partir de indicadores de la bondad del ajuste, con proximidades supuestas u observables. Se pueden distinguir métodos de escalonamiento multidimensional no métrico y métrico. En un estudio, por ejemplo, se podría analizar cómo se perciben las diferencias entre los partidos políticos (o de sus líderes) con datos de las valoraciones de semejanza o proximidad expresadas por personas encuestadas, mediante el escalonamiento multidimensional, se podrían identificar las dimensiones que describen las preferencias del electorado.

3.2. Técnicas de análisis multivariable de dependencia

3.2.1. *El análisis de regresión múltiple*

Dado un modelo donde se supone la existencia de una variable dependiente cuantitativa como función de otra variable independiente –regresión simple– o de otras variables independientes –regresión múltiple–, también cuantitativas, estos tipos de análisis trata de encontrar a partir de los datos una función lineal que se ajuste, bajo determinadas condiciones del modelo matemático, a esta dependencia. Mediante un modelo de regresión describimos la relación de dependencia entre las variables y podemos establecer predicciones basadas en él. Por ejemplo, si se trata de analizar la dependencia que puede suponerse lineal entre la variación de los gastos, como variable dependiente, en función de variables independientes como los ingresos, del número de hijos, etc. el modelo establece que la variación de la variable dependiente está dado por las variaciones de las variables independientes afectadas por coeficientes, uno para cada variable independiente, que se trata de calcular más otra parte que queda inexplicada y que se denomina error del modelo. El ajuste, o cálculo de los coeficientes, se realiza mediante, por ejemplo, la técnica estadística de mínimos cuadrados

ordinarios para la que se minimiza la parte inexplicada por las variables independientes. Cada coeficiente o parámetro (coeficiente de regresión) asociado a cada variable independiente expresará la importancia relativa en la explicación de la varianza total. Adicionalmente disponemos de tests estadísticos de ajuste para validar el modelo global y de significación de los coeficientes.

3.2.2. *El análisis de regresión logística*

La regresión logística sigue los principios generales de una regresión pero tratando variables dependientes cualitativas. Cuando la variable a explicar es dicotómica se habla de regresión logística binaria, cuando es politómica la regresión es multinomial. Si además las categorías están ordenadas y expresan latentemente la continuidad de una variable entonces la regresión es ordinal. Así por ejemplos podemos estar interesados en explicar el comportamiento electoral abstencionista y elegir una serie de variables independientes, cuantitativas o cualitativas, que determinen la acción de votar o no votar en unas elecciones: la edad, el sexo, la clase social, el nivel de estudios, el lugar de residencia, los ingresos, la cultura política, la valoración de la situación económica, etc. En lugar de explicar la participación electoral podríamos modelizar los perfiles diferenciados de cada partido político viendo los factores que más explican el voto a cada opción. En una regresión ordinal podríamos explicar los niveles de ingresos, el rendimiento académico o el grado de satisfacción con la prestación de un servicio en relación a un conjunto de variables independientes que deben ser objeto de elección en función de un modelo sustentado teóricamente que justifique su consideración. En todos estos casos se trata de explicar las razones de frecuencias entre las categorías a partir de un vínculo que formalmente se establece con las variables independientes a través de la función logística. Finalmente lo que obtenemos es una serie de coeficientes que estiman la importancia relativa de cada categoría o variable como factores explicativos teniendo en cuenta la presencia del resto de variables en el modelo multidimensional.

3.2.3. *El análisis de la varianza multifactorial y multivariable*

El análisis de la varianza multifactorial (ANOVA multifactorial) considera una variable dependiente de naturaleza cuantitativa y diversas variables independientes cualitativas llamadas factores. Por ejemplo, si se pretende analizar los gastos en función de la categoría socioeconómica, el nivel educativo, el tipo de familia, etc. o bien si se quiere analizar la valoración que se hace de un producto anunciado en la televisión (mediante una escala cuantitativa) en función de la categoría socioeconómica y la edad (como variables cualitativas). Bajo un modelo lineal y con determinadas condiciones donde la variación de la variable dependiente se desglosa como una suma de contribuciones de cada una de las categorías de los factores y que debe satisfacer el modelo, mediante este análisis se trata de diferenciar, de la varianza total, la parte debida a la varianza entre o explicada por los factores y la parte intra o interna no explicada por factores. Además, el modelo proporciona información sobre la intensidad y el grado de significación de la dependencia. En esta dependencia se tienen en cuenta las diferencias observadas por los valores (grupos o niveles) definidos por cada una de las variables independientes y por las interacciones entre estas variables.

El análisis multivariable de varianza (MANOVA) consideraría varias variables dependientes. En el caso del ejemplo anterior sobre la valoración del producto podríamos considerar las valoraciones que se hacen según varios aspectos: valoración del contenido del mensaje, valoración de la imagen, valoración del sonido, etc. en función de las mismas variables independientes. Estos modelos de análisis de la varianza se pueden plantear desde modelos o diseños experimentales y no experimentales, según se dé una manipulación directa o no de las variables, con diseños adicionales particulares que aquí no consideramos.

3.2.4. *El análisis de covarianza*

El análisis de covarianza (ANCOVA) es una extensión del análisis de la varianza. Se considera el mismo modelo de relaciones entre las variables añadiendo un control estadístico de variables extrañas o perturbadoras. Se supone que hay variables intervinientes que influyen sobre la variable dependiente modificando los resultados de la dependencia con las variables independientes, por lo tanto, se trata de corregir este efecto (de la llamada variable concomitante o covariable, o de varias de ellas) considerando solamente los efectos atribuibles a las variables independientes. Si consideramos el modelo multifactorial anterior este podría completarse mediante la introducción de un factor de covariación, una variable que actúa como factor de comparación: si se quiere analizar la valoración que se hace de un producto anunciado en la televisión en función de la categoría socioeconómica y la edad, la introducción de la variable valoración la semana anterior de emisión nos permitiría ver si las diferencias se mantienen al margen del efecto que haya podido tener la semana de referencia.

3.2.5. *El análisis discriminante*

En este caso una o más variables cualitativas se consideran como variables dependientes en función de varias variables independientes de tipo cuantitativo. El análisis trata de encontrar la combinación (lineal) de las variables independientes (función discriminante) que dé cuenta de la máxima discriminación entre dos o más grupos definidos por la variable dependiente cualitativa, o los grupos que resultan del cruce de varias variables cualitativas. Así, por ejemplo, el partido político votado y el sexo podrían considerarse variables dependientes de los ingresos, del número de hijos, de los años de residencia en el municipio, de la edad, etc. El análisis discriminante se considera una técnica de clasificación donde una serie de grupos definidos previamente (el partido votado según el sexo) se analizan en función de las variables que los hacen más diferentes y, por tanto, explican el voto.

3.2.6. *El análisis del camino*

El análisis del camino (*path analysis*) parte de un modelo de ecuaciones estructurales donde se explicita en forma de relación lineal las dependencias entre las variables, todas ellas cuantitativas. Hay variables endógenas o dependientes y variables exógenas o independientes, todas ellas observables y bajo un modelo recursivo, es decir, marcado unidireccionalmente. El análisis trata de encontrar los coeficientes de relación

(coeficientes *path*) que den cuenta de la dependencia y determinen las ecuaciones estructurales en que se expresa un modelo teórico dado. Estos coeficientes son de regresión, y se interpretan igualmente como la relación entre la variación o cambio de una variable dependiente en función del cambio de una variable independiente, cuando ambas variables han sido previamente estandarizadas. Un posible modelo por ejemplo sería intentar explicar la ideología política en función de una serie de variables sociales, como la edad, los años de estudios del individuo, los años de estudios de los padres o los ingresos, las cuales al mismo tiempo se relacionan entre sí según un modelo explicitado de dependencia.

3.2.7. *El análisis factorial confirmatorio*

El análisis factorial pretende identificar, bajo un modelo predefinido de dependencia y a partir de un conjunto de variables cuantitativas originales observadas o manifiestas, las variables no observadas o latentes, llamadas factores, que en un número reducido subyacen a las originales. El análisis factorial confirmatorio se basa en un modelo previo donde se postula una dependencia entre las variables manifiestas iniciales y las latentes o factores, donde se supone que las variables originales contribuyen con un peso diferenciado en la emergencia de los factores comunes, dado por la parte común de las variables, y otra parte específica que se debe a las variables y que no es explicada por factores. De hecho, las variables independientes son los factores. El modelo previo de hipótesis, de relaciones causales, orienta el establecimiento de estos pesos y se somete a comprobación. Así por ejemplo, a partir de la realización de una serie de pruebas intelectuales de diferente naturaleza se intenta probar que algunas de estas miden un factor de habilidades verbales y lingüísticas, y otras un factor de habilidades de cálculo matemático.

3.2.8. *Modelo de ecuaciones estructurales*

Los modelos causales en general y los modelos de ecuaciones estructurales en particular tienen actualmente una importancia primordial dentro de los métodos empleados en la investigación sociológica. Representan de alguna manera la continuación, desarrollo y superación de los análisis de regresión, factoriales y del análisis del camino y, en general, de las que suponen relaciones de dependencia entre variables. Los métodos de ecuaciones estructurales suponen, por tanto, que la realidad social debe ser conceptualizada y tratada bajo modelos de causalidad. En sentido estricto se trata de explicitar relaciones causales entre variables cuantitativas, algunas de las cuales se declaran dependientes y otros independientes. La particularidad del modelo de ecuaciones estructurales es doble: las variables pueden ser manifiestas o latentes, en parte o todas ellas, y el modelo puede ser recursivo o no recursivo. En el caso de utilizar exclusivamente variables latentes debe ser complementado con los métodos factoriales confirmatorios o métricos que proporcionan las variables latentes a partir de variables observables. En este sentido se puede hablar de un modelo general de estructuras de covarianza (las ecuaciones con varianzas y covarianzas entre las variables son el núcleo algebraico de estos métodos) que comprende los llamados modelos factoriales confirmatorios o de medida con variables observables y no observables, y los modelos de ecuaciones estructurales con variables observables, o latentes o de ambos tipos a la vez. Un modelo confirmatorio, de entrada, puede darnos,

a partir de unas variables observables (por ejemplo, la categoría socioeconómica de la familia y de los alumnos/as, el nivel de estudios del padre y de la madre, el nivel de ingresos, etc.) unas variables latentes de entrada que explican las observables. Otro análisis confirmatorio, de salida, proporciona las aspiraciones del núcleo familiar, una relación entre variables observables (por ejemplo, las notas, las motivaciones y aspiraciones del alumnado según escalas de valoración de profesiones, etc.) y otros latentes, de salida, que dan cuenta o son causa de estas variables observables. Pues bien, el modelo de estructuras causales, cuando se trabaja sólo con variables latentes, establece relaciones causales entre ambos grupos de variables latentes, las de entrada y las de salida, ya sea de forma recursiva o no recursiva.

3.2.9. El análisis de segmentación

El análisis de segmentación (*Tree Analysis*) toma un conjunto de variables, y determinados hipotéticamente dos grupos entre ellas, el grupo de variables dependientes y el grupo de variables independientes, se trata de ir segmentando o dividiendo el total de la población sucesivamente a partir de la elección de la variable independiente y de las agrupaciones de sus valores que maximicen la separación o las diferencias entre los grupos haciéndolos a su vez homogéneos en relación a la o las variables dependientes. Se trata de un procedimiento con fines clasificadores, pero a diferencia del análisis de clasificación se trata de grupos de individuos. Las variables dependientes pueden ser cuantitativas o cualitativas, mientras que las variables independientes se tratan a nivel nominal u ordinal. Esta técnica se podría utilizar para segmentar la opinión frente al aborto (a favor o en contra) a partir de considerar diversas variables independientes que configurarían los grupos característicos favorables o no al aborto: la ideología, la edad, el sexo, los ingresos, el estado civil, el nivel de estudios, la práctica religiosa, la actividad laboral, etc. Como resultado podría salir que la variable que tiene más poder explicativo es la práctica religiosa, configurándose dos grupos, los creyentes y los no creyentes, los cuales a su vez podrían dividirse según el efecto discriminador de una segunda variable: los creyentes se podrían dividir en practicantes y no practicantes, mientras que los no creyentes se podrían dividir en personas de ideología de izquierdas o de derechas, y así sucesivamente mientras hubiera divisiones y variables significativas.

4. Bibliografía

- Alvira, F. (1990). Comentario a la enseñanza de la Estadística en Sociología. *Estadística Española*, 31, 122, 435-437.
- Ato García, M.; López García, J. J. (1996). *Análisis estadístico para datos categóricos*. Madrid: Síntesis.
- Barton, A. H. (1985). Concepto de espacio de atributos en sociología. En *Metodología de las Ciencias Sociales. I. Conceptos e Índices*, editado por R. Boudon i P.F. Lazarsfeld. Barcelona: Laia, 195-219.
- Bertier, P.; Bourouche, J.-M. (1983). *Analyse des données multidimensionnelles*. Paris: PUF.
- Bennet, S.; Bower, D. (1976). *A introduction to multivariate techniques for social and behavioural sciences*. London: The Macmillan Press.

- Bishop, Y. M.; Fienberg, S. E.; Holland, P. W. (1989). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: Cambridge University Press.
- Bisquerra Alzina, R. (1989). *Introducción conceptual al análisis multivariable. Un enfoque informático con los paquetes SPSS-X, BMDP, LISREL y SPAD*. Barcelona: Promociones y Publicaciones Universitarias.
- Borràs, V. et al. (1991). «Bibliografía sobre anàlisi multivariant». *Papers. Revista de Sociologia* 37. P. 135-139.
- Bosque, J.; Moreno, A. (1994). *Prácticas de Análisis Exploratorio y Multivariante de Datos*. Barcelona: Oikos-Tau.
- Calvo Gómez, F. (1993). *Técnicas estadísticas multivariantes*. Bilbao: Deusto.
- Cattell, R. B. (1966). *Handbook of Multivariate Experimental Psychology*. Chicago: Rand McNally.
- Cuadras, C. M. (1996). *Métodos de análisis multivariante*. Barcelona: EUB.
- Díaz De Rada, V. (1999). *Técnicas de Análisis de Datos para Investigadores Sociales. Aplicaciones prácticas con SPSS para Windows*. Madrid: Editorial Ra-Ma.
- Etxeberría, J.; García, E.; Gil, J.; Rodríguez, G. (1995). *Análisis de datos y textos*. Madrid: RA-MA.
- García Ferrando, Manuel (1987) *Socioestadística. Introducción a la estadística en sociología*. 2a edició amp. Madrid: Alianza. Alianza Universidad Textos, 96.
- Grande Esteban, I.; Abascal Fernandez, E. (1989). *Métodos multivariantes para la investigación comercial*. Barcelona: Ariel.
- Hair, J. F. et al. (2011). *Multivariate Data Analysis*. New Jersey: Prentice Hall.
- Jöreskog, Karl G. (1993). *Modelado de ecuaciones con Lisrel*. Vitoria: Instituto Vasco de Estadística
- Kachigan, S. K. (1991). *Multivariate Statistical Analysis. A Conceptual Introduction*. 2a edició. New York: Radius Press.
- Kendall, M. G. (1975). *Multivariate Analysis*. London: Griffin.
- Kendall, Maurice G. (1978). The History of Statistical Method. En *International Encyclopedia of Statistics*, editado per W.H. Kruskal i J.M. Tuner. New York: Free Press, 1093-1101.
- Kinncar, P.R.; Taylor, J.R. (1971). «Multivariate Methods in Marketing Research. A Further Attempt at Classification». *Journal of Marketing* 35, 4, octubre, 56-59.
- Laforgue, H. (1981). *Analyse multivariante*. Montreal: Études Vivantes.
- Lebart, L.; Morineau, A.; Fenelon, J. P. (1985). *Tratamiento estadístico de datos. Métodos y programas*. Barcelona: Marcombo.
- Lefebvre, J. (1983). *Introduction aux analyses statistiques multidimensionnelles*. Paris: Masson.
- Lozares Colina, C.; López-Roldán, P. (1991). «El análisis multivariado: definición, criterios y clasificación». *Papers. Revista de Sociologia* 37. P. 9-29.
- Manly, B. F. J. (1986). *Multivariate Statistical Methods*. London: Chapman and Hall. P. 17-25
- Prieto, G. (1985). Análisis Multivariable. En *Investigación educativa*, editado por A. de la Orden Hoz. Madrid: Anaya, 26-28.
- Riba Lloret, M. D. (1989). *Una panoràmica de las técnicas estadísticas multivariantes*. Bellaterra: Universitat Autònoma de Barcelona.
- Sánchez Carrión, J. J. (1984). *Introducción a las técnicas de análisis multivariable aplicadas a las ciencias sociales*, Madrid: Centro de Investigaciones Sociológicas.
- Sierra Bravo, R. (1994). *Análisis Estadístico Multivariable: Teoría y Ejercicios*. Madrid: Paraninfo.

- Stevens, James (1986). *Applied Multivariate statistics for the social sciences*. Hillsday (New Jersey): Lawrence Erlbaum.
- Stigler, S. M. (1986). *The history of statistics: the measurement of uncertainty before 1900*. Cambridge: Belknap Press.
- Tabachnick, B. G.; Fidell, L. S. (1989). *Using Multivariate Statistics*. 2a. edició. New York: Harper Collins.
- Tacq, J. (1997). *Multivariate analysis techniques in social science research: from problem to analysis*. London: Sage Publications.
- Visauta Vinacua, B. (1999). *Anàlisis estadístic con SPSS para Windows: estadística multivariante*. Madrid: McGraw-Hill.
- Volle, M. (1978). *Analyse des données*. Paris: Economica.
- VV.AA. (1991). L'anàlisi multivariable de dades. Monogràfic de *Papers. Revista de Sociologia*, 37.
- VV.AA. (1996). La construcció de tipologies. Exemples. Monogràfic de *Papers. Revista de Sociologia*, 48.