



**Mg. Diana Atar<sup>1</sup>**

Lic. en Estadística. Profesora Titular de Metodología de la Ciencia y Trabajo de Integración Final, Facultad de Humanidades (UdeMM)

# El análisis de los datos en las ciencias sociales: elementos para la reflexión metodológica

*Una vez descartado lo imposible,  
lo que queda, por improbable  
que parezca, debe ser la verdad.*

Arthur Conan Doyle  
El mastín de los Baskerville

## Consideraciones Preliminares

En la actualidad son muchas las encuestas que se llevan a cabo, tanto en el ámbito público como en el privado, y que intentan dar cuenta de múltiples aspectos de la realidad social, como por ejemplo, investigaciones de mercado, relevamientos demográficos y epidemiológicos, ratings de audiencia, encuestas de opinión pública sobre temas variados tales como la percepción social de la inseguridad o de la inflación, la valoración de los medios de comunicación, la calidad académica de nuestras instituciones educativas, el impacto de la tecnología en la vida cotidiana, y la lista podría seguir aumentando.

Sin embargo, y en general, los análisis que se realizan a partir de los datos recogidos son metodológicamente limitados y, en consecuencia, los resultados obtenidos son de proyección

y utilidad también limitadas. En consecuencia, podemos decir que estos estudios sociales están siendo poco aprovechados en su riqueza informativa debido al instrumental metodológico empleado habitualmente.

*"(...), los decisores de cada área social demandan información específica que a su vez genere otros flujos de información dirigidos de nuevo a los decisores. Esta nueva información generada se configura utilizando métodos y programas estadísticos cada vez más potentes y sofisticados que permiten profundizar más exhaustivamente en los datos recabados en el comienzo de la investigación."* (Iglesias Antelo, S. y Aranzazu Sulé Alonso, M.; 2003; p. 7)

La medición de los conceptos que se plantean en muchas investigaciones psicológicas y sociales –nivel de violencia, grado de bienestar, comportamiento del consumidor, calidad educativa, etc., se evalúan a partir de un conjunto de indicadores que permite contar con una descripción, a veces precisa y otras no tanto, del fenómeno que se está estudiando.

Estos indicadores no sólo brindan conocimiento sino que además pueden

ser un instrumento útil para tomar de decisiones, y por ello, es necesario un esfuerzo de análisis, tanto en la forma en que se recopilan los datos como en la aplicación de métodos y técnicas estadísticas.

La tarea de realizar encuestas no es sencilla: comprende el empleo de recursos humanos, materiales y económicos, y es un proceso que va desde el diseño del cuestionario hasta el procesamiento y la impresión final de los resultados. Tal trabajo merece un mayor aprovechamiento de los datos, para lo cual podrían aplicarse técnicas más sofisticadas que las utilizadas habitualmente.

Uno de los problemas centrales que presentan los estudios de los fenómenos sociales, es el de la medición. La noción clásica de medición implica la asignación de números a objetos de acuerdo con reglas. Una concepción más moderna considera a la medición como el proceso de vincular conceptos abstractos con indicadores empíricos. Desde el punto de vista teórico, el interés se centra en el concepto subyacente no observable que está representado por la respuesta observable, mientras que desde el punto

de vista empírico, el foco está en esta última.

*“Cuando formulamos una pregunta en un cuestionario, las respuestas son observables. Ahora bien, frecuentemente el interés del investigador se enfoca en propiedades que no son observables. Siguiendo la terminología de la tradición lazarsfeldiana, llamaré a esos dos planos el manifiesto y el latente. El problema crítico de la teoría de la medición es establecer la relación entre esos dos planos; cómo medir propiedades latentes a través de observaciones realizadas mediante variables manifiestas (...)”* (Mora y Araujo, M.; 2005; p.436)

Medir, por ejemplo, las actitudes de una persona hacia los contenidos de un programa de televisión es una tarea más difícil que la de medir el grado de conocimientos que esa persona posee sobre los contenidos de ese programa de televisión, ya que esto último puede evaluarse tomando en cuenta la cantidad de información correcta que posee de dicho programa.

En cambio, el concepto “actitud”, como otras variables psicológicas, es imposible de observar en forma directa.

*“Para medir una actitud, debemos inferirla de la forma en que responde un individuo (por una expresión verbal o comportamiento manifiesto) a cierto estímulo. El término construcción hipotética describe una variable que no se observa directamente, pero que se puede medir por medios indirectos, como la expresión verbal o el comportamiento manifiesto.”* (Zikmund, W.; 1998; p. 346).

Entonces, el problema de medir una actitud consiste en poder establecer la relación entre lo no observable (latente) y lo observable (manifiesto).

*“Podemos concebir que esa relación es de carácter probabilístico, en el*

*sentido de que no existe una determinación completa entre el estado real en el que se encuentra una persona en una variable actitudinal en un momento dado y su respuesta manifiesta a una pregunta en ese mismo momento.”* (Mora y Araujo, M.; 2005; p. 437)

Por otra parte, la actitud –que podemos pensarla “como una disposición duradera a responder constantemente de una manera determinada a diversos aspectos del mundo, entre los que se encuentran personas, acontecimientos y objetos” (Zikmund, W.; 1998; p. 346), –comprende tres elementos: afectivo, cognoscitivo y de comportamiento.

El afectivo refleja lo que siente una persona hacia un objeto y es de carácter emocional. La componente cognoscitiva refleja el grado de conocimiento y de conciencia de la persona sobre el objeto. Y el elemento de comportamiento muestra la intención o la predisposición hacia la acción.

Supongamos que una persona declara: *“Me parece muy interesante la iniciativa de emitir cursos de formación técnica por televisión, porque la capacitación es un recurso importante para la inserción laboral. Por eso, apoyo la creación de un canal exclusivamente educativo”*.

En esta declaración se reflejan los tres componentes mencionados: el afectivo (“*me parece muy interesante*”), el grado de conocimiento y conciencia de la persona sobre el contenido (“*la capacitación es un recurso importante para la inserción laboral*”) y la predisposición a la acción (“*apoyo la creación de un canal exclusivamente educativo*”).

Es obvio que esta declaración refleja la actitud frente a un canal de televisión. Esta persona bien puede ignorar los temas tratados en los distintos programas que se emiten, o la grilla horaria de los mismos, pero aún así,

puede tener una actitud favorable ante la iniciativa de un canal exclusivamente educativo.

Todo proceso de medición impone que los conceptos se transformen en operativos; es decir, formular una descripción que de cuenta del significado del concepto, mediante la especificación de las operaciones necesarias para medirlo.

Entonces, luego de haber definido conceptual y operacionalmente las variables de la investigación y de haber diseñado el instrumento de medición acorde a los conceptos establecidos, la etapa del análisis de los datos es fundamental, no solamente para obtener la mejor y más amplia información posible, sino también porque las técnicas que se utilicen para analizar los datos recogidos, deben ser las adecuadas al tipo de variable que se está estudiando.

*“La complejidad de la realidad (...) y el hecho de que en su conocimiento confluyan disciplinas científicas de origen diverso hacen que el contenido de los métodos multivariados se proyecte como un cuerpo de conocimientos de naturaleza plural e interdisciplinaria. Esta característica se acentúa aún más si consideramos el método operativo o las técnicas en las que se inspira para alcanzar los fines que persigue. Todo ello se refleja en la amplitud de categorías conceptuales (organización, marketing, inversión y financiación, recursos humanos e investigación), en la diversidad de materias científicas (psicología, sociología, estadística, economía, derecho, cibernética...), y en la variedad de técnicas de investigación operativa (programación, simulación, teoría de juegos,...) que se encuentran inmersas en el mundo del análisis multivariable.*

*La finalidad primordial de cualquier investigación a emprender es la de*

ayudar a los responsables en la toma de decisiones dentro de sus áreas respectivas. Actualmente, las necesidades de información de los decisores sociales para la planificación, ejecución de acciones o el control de resultados son cada vez mayores". (Iglesias Antelo, S. y Aranzazu Sulé Alonso, M.; 2003; ps. 6-7).

En ese sentido, entonces, planteamos alternativas metodológicas de análisis multivariante para el tratamiento de los datos de las encuestas que se llevan a cabo en los diversos estudios psicológicos y sociales, enfocándonos en métodos que intenten superar los análisis más frecuentes y algo acotados –porcentajes y medidas descriptivas.

Vale señalar que los modelos que se proponen *no son excluyentes* de los mencionados más arriba, no son mejores ni tampoco los reemplazan, sino que agregan información y amplían el conocimiento que podemos obtener.

## Los Modelos Probabilísticos

"Debido a que existe una incertidumbre considerable al tomar decisiones, resulta importante que todos los riesgos implícitos conocidos se evalúen en forma científica. Ayuda en esta evaluación, la teoría de la probabilidad, a la que frecuentemente se denomina ciencia de la incertidumbre. El empleo de tal teoría probabilística permite a quien toma decisiones, analizar –con información limitada– los riesgos y minimizar el azar inherente (...)". (Mason, R., Lind, D. y Marchal, W.; 2000; ps. 146-147)

Las investigaciones sociales, en general, se realizan con datos provenientes de una muestra y no de la población.

"Como es sabido, una muestra es un subconjunto de una población. Y

aunque la información que maneja el investigador se reduce a una, o en el mejor de los casos, a varias muestras, sus objetivos suelen consistir en extraer conclusiones y conocer el modelo que mejor explica la realidad en la población. Para ello empleará la inferencia estadística; proceso que, a partir de una muestra, permite inferir valores sobre características poblacionales en términos de probabilidad". (Iglesias Antelo, S. y Aranzazu Sulé Alonso, M.; 2003; p. 11).

Cabe destacar, entonces, qué entendemos por *probabilidad*: es una medida de la posibilidad de ocurrencia de un fenómeno, y como medida, tiene un valor numérico. Tal como señala Wayne Daniel:

"Podemos definir (...) la probabilidad como un número, de 0 a 1, que le asignamos a un fenómeno para indicar su posibilidad de ocurrir." (Wayne, D.; 1993; p. 33)<sup>1</sup>

Por lo tanto, los modelos probabilísticos permiten conocer cuál es la probabilidad de que determinado fenómeno se produzca.

De allí, la importancia que le asignamos como herramienta para el análisis de conceptos psicológicos y sociales, tales como actitud, percepción, valoración, motivación, etc., ya que estos modelos son útiles tanto para clasificar a los individuos de una población, como para predecir su "probable" comportamiento.

Y, ¿en qué radica esta importancia de predecir un comportamiento?; ¿cuál es la utilidad de contar con pronósticos sobre situaciones futuras?

"Los juicios sobre el futuro encierran dos tipos diferentes de aproximación

a la realidad: anticipaciones e intervenciones. Por un lado, se formulan pronósticos con el propósito de tomar decisiones, buscando anticipar los acontecimientos más probables que podrán ocurrir para adaptar las decisiones a esos acontecimientos. Si puedo anticipar la ocurrencia de un terremoto, me iré de la zona sísmica antes de que ocurra; si se anticipa tiempo tormentoso, me abstendré de volar hasta que mejore el tiempo; si anticipo la política económica, puedo jugar a la Bolsa con ventaja. Por otro lado, se formulan pronósticos con el propósito de intervenir en la realidad para modificar el curso de los acontecimientos. Nos gustaría impedir que los terremotos ocurran; en muchos lugares se procura producir lluvias o nevadas; si me conviene que las acciones bajen, y soy suficientemente poderoso, puedo tratar de producir una corrida." (Mora y Araujo, M.; 2005; p. 125)

## El Análisis Multivariado<sup>2</sup>

"En todos los casos podemos diferenciar un enfoque taxonómico, descriptivo, tendiente a lo unidimensional, y un enfoque multivariable, o multidimensional, tendiente a penetrar en la complejidad, el cual puede incluir nexos causales entre las variables. A veces parece que todo el saber –al menos el saber sobre lo social– no es más que una gran «estructura de sistemas taxonómicos»; de hecho la estadística es taxonómica. Pero el analista puede buscar relaciones que van más allá de lo descriptivo, conexiones causales, una realidad latente cuya taxonomía es, por así de-

1 Puede decirse que "0" equivale a la probabilidad de un fenómeno imposible y "1" a la probabilidad de un fenómeno cierto.

2 Según la bibliografía, este tipo de análisis se denomina multivariado, multivariante o multivariable.

cir, más compleja. Cuando pensamos la realidad la pensamos dinámica; cuando buscamos actuar sobre ella necesariamente la concebimos dinámica; el investigador tiene entonces que esforzarse para dinamizar las herramientas analíticas. Las taxonomías no son dinámicas, pero el análisis estadístico puede ser puesto al servicio de análisis dinámicos de la realidad.” (Mora y Araujo, M.; 2005; p. 428)

La mayoría de los fenómenos –sean sociales o naturales– comprende una multiplicidad de variables. Analizar dichos fenómenos requiere encontrar una combinación de variables que explique, de la forma más apropiada, la ocurrencia de los mismos.

“(…) dicha combinación debe representar las interrelaciones que existen entre esas variables o explicar el comportamiento de alguna otra –ya sea con fines predictivos o clasificatorios– de la mejor manera posible.” (Iglesias Antelo, S. y Aranzazu Sulé Alonso, M.; 2003; p. 10)

Sin embargo, los análisis de los resultados obtenidos de las encuestas sobre temas sociales quedan limitados a los de tipo descriptivo y univariado<sup>3</sup> (distribuciones porcentuales, o medias aritméticas de algunas variables cuantitativas como la edad).

Este tipo de análisis, si bien sirve para conocer el comportamiento en forma separada de las variables, no nos permite establecer asociaciones (o interrelaciones) entre ellas, de modo de comprender más profundamente el fenómeno que estamos investigando. En consecuencia, el análisis univariado parece acotado como forma de

abordaje de la realidad social, que es sumamente compleja.

Como ya dijimos, muchos de los conceptos que aparecen en estudios sociales comprenden múltiples dimensiones; por lo tanto, merecen un tratamiento multidimensional; esto es, la aplicación de métodos de análisis que tengan en cuenta no sólo la intervención de más de dos variables en la ocurrencia del fenómeno sino también las relaciones entre las variables involucradas.

Por ese motivo, proponemos la aplicación de modelos de Análisis Multivariado, que permiten considerar el efecto de más de una variable, vincular las variables entre sí, observar correlaciones –si las hubiere– y suponer causalidades.

“Multivariate analytical techniques are being widely applied in industry, government, and university-related research centers. Moreover, few fields of study or research have failed to integrate multivariate techniques into their analytical toolbox. (...) Applications-oriented books are of crucial interest to behavioral scientists and business or government managers of all backgrounds who have to expand their knowledge of multivariate analysis to gain a better understanding of the complex phenomena in their work environment. Any researcher who examines only two-variable relationships and avoids multivariate analysis is ignoring powerful tools that can provide potentially useful information.” (Hair Jr., J., Anderson, R., Tatham, R. y Black, W.; 1995; p. 3.)<sup>4</sup>

“Las técnicas de análisis multivariado están siendo ampliamente aplicadas en la industria, en el gobierno y en los centros universitarios relacionados con la investigación. Además, pocos campos de estudio e investigación han dejado de integrar las técnicas multivariadas a su bagaje de herramientas de análisis. (...)”

Para obtener entonces, en lo posible, información potencialmente útil, proponemos recurrir al Análisis Multivariado.

“Se entiende por Análisis Multivariante a la rama de la estadística y del análisis de datos, que estudia, interpreta y elabora el material estadístico sobre la base de un conjunto de  $n > 1$  variables, que pueden ser de tipo cuantitativo, cualitativo o una mezcla de ambos. La información del Análisis Multivariante es, por lo tanto, de carácter multidimensional. (...) El Análisis Multivariante cumple también con los objetivos descriptivo e inductivo, pero trabajando simultáneamente con varias variables en lugar de una sola. Es una metodología más complicada, pero también más potente, que utiliza extensamente los métodos del álgebra lineal, cálculo numérico, geometría lineal y otras clases de geometrías (ultramétrica, no euclídea).” (Cuadras, C. M.; 1981; ps. 3-4)

Cuadras explica que su aplicación se extiende a campos disciplinarios tan diversos como la Psicología, la Biometría, la Taxonomía, la Política, la Diagnóstico, la Psicofisiología, la Lingüística y la Meteorología.

Hair, por su parte, dice que resulta de utilidad para el análisis psicológico, sociológico y otros tipos de datos de comportamiento<sup>5</sup>.

Los libros orientados a las aplicaciones son de interés crucial para los científicos del comportamiento y para los administradores de negocios o del gobierno de todas las formaciones, que tienen que expandir su conocimiento del análisis multivariado para obtener una mejor comprensión de la complejidad de los fenómenos de sus contextos de trabajo. **Cualquier investigador que examina solamente relaciones bivariadas y evita el análisis multivariado está ignorando poderosas herramientas que pueden proveerle información potencialmente útil.** Traducción y destacado de la autora.

5 Ver Hair et al., op. cit., p. 3.

3 Análisis univariado es aquel que trata las variables por separado, sin establecer relación entre ellas.



En tanto, Iglesias Antelo y Aránzazu Sulé Alonso nos dicen que:

*“Sin lugar a dudas hemos entrado en una nueva era que modifica el rumbo de la investigación. Ésta tiene que permanecer alerta para enfrentarse a los profundos y rápidos cambios que están aconteciendo. Con total seguridad su futuro vendrá marcado por una interrelación mayor que nunca entre las distintas disciplinas –psicología, sociología, computación, estadística, economía, administración, recursos humanos...– para dar respuesta a las necesidades de información, tanto de los decisores de las organizaciones, como de los usuarios en general.”* (Iglesias Antelo, S. y Aranzazu Sulé Alonso, M.; 2003; p. 5-6).

Por lo expuesto hasta aquí, el tipo de análisis que proponemos serviría para enriquecer el estudio de los datos de las distintas encuestas existentes: podemos extraer de ellos mucho más que “porcentajes”, para ir hacia una interrelación de datos que permitiría, entre otras cosas:

- Analizar el impacto de un conjunto de variables explicativas sobre la probabilidad de ocurrencia de un hecho determinado (por ejemplo: evaluar la probabilidad de que un sector con determinadas características, valore de forma positiva o negativa un medio de comunicación determinado o un programa de televisión, con relación a variables como edad, condición laboral, hábitos de consumo de medios de información, nivel de instrucción, entre otras).
- Distribuir en grupos a los individuos de una población, conforme un conjunto de características determinadas, tales como actitud hacia los medios de comunicación, conocimientos sobre temas de comunicación social, opinión sobre las empresas y organismos de comunicación (radios, diarios, televisión, etc.).

Por ejemplo, si queremos estudiar la *valoración pública de los medios de comunicación e información*, podríamos analizarlo desde el concepto de la *representación social*. Éste es un constructo teórico intermedio entre lo *psicológico* y lo *social*, que integra conceptos cognitivos como actitud, estereotipo, imagen, creencia, etc., de una manera que no es una mera suma de partes. Por otro lado, si bien la encuesta no es el único instrumento para analizar este tipo de fenómenos, es una práctica difundida debido a que permite el manejo de una gran cantidad de observaciones.

Por lo tanto, si contamos con esos datos, entendemos que resultaría fructífero analizarlos de la manera más completa posible. Por ejemplo, utilizando la técnica multivariada, que nos permite relacionar variables como *actitud, estereotipo, imagen*, etc. En suma, que nos permitiría *predecir comportamientos y asociar características*.

En este sentido, siguiendo a Hair et al.<sup>6</sup>, pueden establecerse dos tipos de Análisis Multivariado:

1. *Análisis de dependencia*: se aplica cuando una o más variables dependientes van a ser explicadas por un conjunto de variables independientes. Éstas son técnicas de *predicción*.

2. *Análisis de interdependencia*: se aplica cuando no se hace distinción entre variable dependiente o independiente, sino que se consideran las interrelaciones entre las variables y, en definitiva, la estructura subyacente. Son técnicas de *clasificación*.

Existen muchas técnicas de Análisis Multivariado. Lamentablemente, no todas ellas pueden aplicarse en todos los casos: dependemos de los datos que tenemos: cuáles son las variables y cuáles sus niveles de medición.

6 Ver Hair et al., op. cit., ps. 17/21.

## Las Variables disponibles y sus Niveles de Medición

*“La materia prima del análisis multivariable son los datos. Estos datos son los valores que toman las variables, las cuales a su vez son magnitudes que representan distintos conceptos o atributos de individuos u objetos. La precisión de tal representación dependerá directamente de la escala de medida elegida, entendiéndose por escala los valores numéricos asignados a cada variable según ciertas reglas. Éste es el significado más básico del término (...). Y es que escala también puede adoptar otros sentidos, entre ellos el de instrumento de medida compuesto por una serie de ítems que se utiliza para cuantificar características de las personas a partir de sus respuestas a los ítems (...).*

*La tipología de escalas de medida más clásica en estadística multivariable es la propuesta por Stevens (1946), que distingue cuatro básicas: la nominal, la ordinal, la de intervalo y la de razón, cada una de ellas más precisa que la anterior. Las dos primeras son escalas no métricas o cualitativas, puesto que reconocen en cada encuestado una determinada cualidad o propiedad, mientras que las dos últimas son escalas métricas o cuantitativas, capaces de reflejar diferencias de grado o cantidad”.* (Iglesias Antelo, S. y Aranzazu Sulé Alonso, M.; 2003; p. 8)

Vale la reiteración: la elección del tipo de técnica a aplicar va a estar condicionada por el tipo de variables con las que se cuente y qué tipo de nivel de medición presentan esas variables. *Sólo cuando sabemos con qué tipo de variables contamos estamos en condiciones de definir correctamente los métodos de análisis estadístico que pueden aplicarse.*

Llegado este punto, debemos señalar que muchas de las variables tratadas en los estudios sociales son *cualitativas, nominales u ordinales*. Esto nos lleva a pensar cuidadosamente qué técnicas de análisis pueden utilizarse y para qué.

### Tratamiento de las Variables

Tal como dijimos más arriba, las variables cualitativas pueden ser de *nivel nominal* o de *nivel ordinal*.

A su vez, las de nivel nominal pueden ser *dicotómicas* (cuando lo que se mide puede ser ubicado en una de dos categorías excluyentes, por ejemplo, la variable *Sexo*) o *politómica* (cuando hay más de dos categorías posibles, por ejemplo, la variable *Estado Civil*).

El tratamiento estadístico de este tipo de datos cualitativos, lleva a realizar transformaciones en las variables, del tipo *Escala de Likert* o variables *dummy* –simuladas–, a fin de adjudicar un valor numérico a la observación y de esta forma, poder emplear algunas técnicas multivariadas particulares. En el caso de la Escala de Likert –método propuesto por Rensis Likert en la década de 1920 para la *medición de actitudes*–, la técnica consiste en que los encuestados califican qué tan de acuerdo están con las afirmaciones construidas. Se utiliza una escala con opciones que varían desde una actitud o valoración *muy positiva* hasta una *muy negativa* hacia un objeto. Se asigna valores a cada respuesta alternativa, las cuales se suman para obtener un índice. En general, los encuestados evalúan las afirmaciones en una escala de 5 puntos.

A modo de ejemplo, supongamos el siguiente enunciado, que podría aplicarse para medir, entre otros, la actitud hacia la línea editorial de un periódico.

*“Al comprar un diario, la línea editorial del mismo es un factor tenido en cuenta por el lector”.*

- Muy en desacuerdo (1)
- En desacuerdo (2)
- Ni de acuerdo ni en desacuerdo (3)
- De acuerdo (4)
- Muy de acuerdo (5)

En cuanto a las variables simuladas, éstas no son numéricas pero se transforman en variables numéricas al asignar el valor 1 o 0 a una persona, dependiendo de si ésta posee o no una determinada cualidad o atributo<sup>7</sup>. Siguiendo el mismo tipo de ejemplo que el anterior:

*“Compro el mismo diario todos los días”*

- |     |     |
|-----|-----|
| SI  | NO  |
| (1) | (0) |

### Clasificación y Predicción

*Clasificar* es uno de los temas centrales de la actividad investigadora. La clasificación de una persona o un objeto dentro de un determinado grupo a partir de un conjunto de mediciones u observaciones, comprende cierto grado de incertidumbre que puede cuantificarse mediante el empleo de algún método probabilístico.

La clasificación de personas u objetos se basa, principalmente, en la presencia o la ausencia de ciertas características, y para poder llevarla a cabo, es necesario establecer una regla de decisión que permita clasificar esas personas y objetos de acuerdo a ciertas funciones, llamadas *discriminantes*. La regla óptima será aquella que minimice la probabilidad de una mala clasificación.

En los modelos de variables cuantitativas, el problema reside en evaluar magnitudes tales como *¿cuántos lectores de diarios compran el diario A?*, o *¿qué porcentaje de lectores compra el diario todos los días?* Sin embargo, hay otro aspecto previo: *¿el lector del diario A compra el diario todos los días?*, *¿el lector del diario A lee además otro diario?* En tales situaciones intervienen otro tipo de variables que hacen de indicadores de la presencia o ausencia de *procesos cualitativos*.

En el primer caso, tendremos como respuesta valores numéricos que indican cantidades; en el segundo caso, las respuestas posibles podrán ser “*si*” o “*no*”, o bien del tipo de “*nunca*”, “*a veces*”, “*siempre*”.

En cuanto a la *predicción*, frecuentemente los investigadores realizan estudios para explicar y pronosticar un atributo en función de otras variables, sean éstas cualitativas o cuantitativas.

*“Attributes, of course, are not described numerically. In models with a dependent attribute, we seek to predict the category or the attribute by means of a rule assigning an element to one or another category depending on given values and categories of the explanatory variables and attributes.”* (Tryfos, P.; 1998; p.338)<sup>8</sup>

La importancia de la predicción, de la elaboración de pronósticos, radica en que contribuye a disminuir la incertidumbre y atenuar las dudas de quienes son responsables de tomar decisiones.

<sup>8</sup> Los atributos, por supuesto, no están descritos numéricamente. En modelos con una variable cualitativa dependiente, buscamos predecir la categoría o el atributo por medio de una regla que asigne un elemento a una u otra categoría dependiendo de los valores o categorías dadas de las variables y atributos explicativos.” Traducción de la autora.

<sup>7</sup> Ver Gujarati, D.; 1981; ps.285-286

“En las ciencias sociales, el procedimiento más habitual para generar juicios prospectivos es apoyarse en regularidades empíricas conocidas y proyectarlas hacia delante, suponer que las mismas pautas conocidas en el pasado se replicarán a sí mismas en el futuro. Sin duda, las regularidades empíricas no son fundamento explicativo suficiente; pero aunque no expliquen en un sentido estricto, funcionan para reducir incertidumbre sobre los acontecimientos futuros. Que la ocurrencia de un cierto evento no sea estrictamente explicable no excluye que pueda estimarse una alta probabilidad de ocurrencia; para la toma de decisiones muchas veces esto es más importante que una buena explicación”. (Mora y Araujo, M.; 2005; p. 128).

Sin embargo, hemos observado que gran parte del trabajo relacionado con las encuestas sobre temas sociales enfoca, principalmente, el análisis del tipo descriptivo para una o dos variables. Trataremos, entonces, de aportar herramientas para iniciar un análisis más fértil.

**Los Modelos de Análisis**

Presentamos a continuación, tres modelos de análisis para la clasificación y predicción:

1. Modelo de probabilidad lineal.
2. Modelo de probabilidad no lineal: Logit
3. Modelo de probabilidad no lineal: Probit.

La utilidad de estos modelos radica en que permiten la modelización de variables cualitativas (que son las que encontramos en gran cantidad en los estudios sociales).

Dentro de estos modelos, puede ocurrir que la variable dependiente o explicada tome únicamente dos valo-

res alternativos (0 y 1). El modelo, entonces, se denomina de *respuesta dicotómica*. En los casos en que la variable dependiente puede asumir tres o más valores, el modelo es de *respuesta múltiple* o multinomial.

De esta manera, por ejemplo, se puede estimar, en función de un conjunto de variables explicativas como el nivel educativo alcanzado, el hábito de lectura de diarios, el empleo de Internet como fuente de información, etc., la probabilidad de que un individuo:

- Opine a favor o en contra de un medio de comunicación determinado (respuesta dicotómica).
- Apoye, se oponga o le resulta indiferente un medio de comunicación determinado (respuesta múltiple).

En suma, la utilidad de los modelos, en sentido estricto, es proporcionar, con los datos disponibles, la mayor cantidad de información posible. Por lo tanto, no decimos *qué hacer* con la información que se extraiga de los datos, sino *qué información se puede suministrar a aquel que la necesite*.

A continuación, trataremos algunos aspectos generales de los modelos de respuesta dicotómica. Luego, se presenta la aplicación de éstos en los modelos de probabilidad propuestos.

**Modelos de Respuesta Dicotómica**

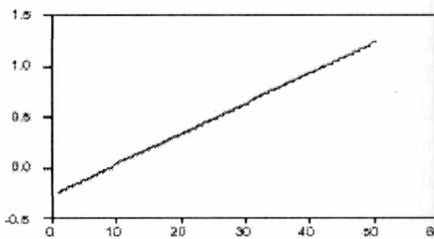
Tomemos *n* elementos (personas u objetos), cada uno de ellos con un conjunto de variables (características) y sean dos categorías mutuamente excluyentes.

El resultado de la aplicación del modelo de respuesta dicotómica es un valor numérico que indica la probabilidad de que un elemento, de la muestra o de la población, con un conjunto de características determinadas pertenezca a alguna de las dos categorías excluyentes.

Es decir, que la probabilidad obtenida es función de las variables explicativas y que sirven para discriminar entre las categorías y determinar la pertenencia de un elemento a una u otra<sup>9</sup>.

**1. Modelo de Probabilidad Lineal**

Este modelo permite, mediante una función lineal, determinar la combinación de variables independientes cuyas variaciones inciden sobre la variable dependiente; es decir, que permiten estimar la probabilidad de que una persona con determinadas características pertenezca o no a cada una de las dos categorías excluyentes. Su representación gráfica es una recta<sup>10</sup>.



El modelo es el siguiente:

$$P = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$$

Siendo

- P = Probabilidad
- b<sub>i</sub> = Coeficientes estimados del modelo
- X<sub>i</sub> = Variable explicativa

9 Sean *n* elementos, cada uno de ellos con un conjunto de *k* variables y sean C<sub>1</sub> y C<sub>2</sub> dos categorías mutuamente excluyentes. El resultado de la aplicación del modelo es un valor numérico P<sub>i</sub> = P(C<sub>i</sub>/x) que indica la probabilidad de que un elemento, de la muestra o de la población, con un conjunto de características x = (X<sub>1</sub>, X<sub>2</sub>, ..., X<sub>k</sub>) pertenezca a la categoría C<sub>i</sub>. P<sub>i</sub> es una función de las variables X<sub>1</sub>, X<sub>2</sub>, ..., X<sub>k</sub>, llamadas explicativas.

10 El gráfico corresponde a una ecuación bivarriada (una variable independiente y una variable dependiente).

Veamos una aplicación del modelo en el siguiente ejemplo<sup>11</sup>:

**Clasificación en dos categorías**     **C1:** valoración positiva de un medio de comunicación determinado  
**C2:** valoración negativa de un medio de comunicación determinado

**Objetivo:** estudiar la proporción de personas con estudios universitarios completos, en función de la edad y su condición laboral, que tienen una valoración positiva de un medio de comunicación determinado

**Variables explicativas:**  $X_1$ : edad  
 $X_2$ : condición laboral (Desocupado = 0; Ocupado = 1)

**Variable dependiente:** valoración de un medio de comunicación determinado  
P estimada por  $f/n$  (frecuencia relativa)

**El modelo es:**  $P = b_0 + b_1X_1 + b_2X_2$

Supongamos 4 conjuntos distintos de estas características:

Conjunto N°	Edad	Condición Laboral	Universitarios	Valoración positiva	Frecuencia relativa	Probabilidad estimada
	$X_1$	$X_2$	n	f	$fr = f/n$	
1	25	1	36	10	0,2778	0,2908
2	35	0	29	8	0,2759	0,2411
3	50	0	23	5	0,2174	0,2522
4	65	1	12	4	0,3333	0,3203

$$P = 0,2153 + 0,0007X_1 + 0,0571X_2$$

- La probabilidad de que una persona con estudios universitarios completos, desocupado y de 40 años de edad valore positivamente es igual a

$$P = 0,2153 + 0,0007(40) + 0,0571(0) = 0,2433 \text{ ó } 24,33\%$$

**La interpretación de estos coeficientes es la siguiente:** si se produce un incremento de una unidad en la variable explicativa  $X_k$ , la probabilidad sufrirá una variación igual a  $b_k$ . El signo de  $b$ , sea éste positivo (+) o negativo (-), indica si se trata de un incremento o bien de una reducción.

En nuestro ejemplo, todos los coeficientes son de signo positivo. Pero veamos qué ocurre si modificamos la edad: tomemos un individuo con estudios universitarios completos, desocupado y de 25 años:

$$P = 0,2153 + 0,0007(25) + 0,0571(0) = 0,2328 \text{ ó } 23,28\%$$

La diferencia que se produce en la probabilidad es muy pequeña (de 24% a 23%), modificando la edad.

11 Si para cada conjunto de características  $x$ , hay  $f$  elementos que pertenecen a la categoría de interés dentro de una muestra aleatoria de  $n$  elementos que poseen el set de característi-

cas  $x$ , se puede utilizar  $f/n$  como estimador de  $P$  y estimar los parámetros  $b_0, b_1, b_2, \dots, b_k$  por regresión, mediante el *Método de Mínimos Cuadrados*. Esto es una técnica de estimación

de los coeficientes  $b_i$ , que minimiza la suma de los cuadrados de los errores (diferencias) entre los valores observados de la variable dependiente y los valores estimados o predichos.



Sin embargo, en el caso que se modifique la condición laboral, es decir, un individuo con estudios universitarios completos, de 40 años y ocupado, la probabilidad de que la valoración sea positiva será entonces:

$$P = 0,2153 + 0,0007(40) + 0,0571(1) = 0,3004 \text{ ó } 30,04\%$$

La diferencia es ahora mayor (de 24% a 30%); y lo mismo ocurre si el individuo es ocupado y de 25 años.

$$P = 0,2153 + 0,0007(25) + 0,0571(1) = 0,2899 \text{ ó } 28,99\%$$

Estas diferencias en los valores de la probabilidad con respecto al primer ejemplo, se deben a que las variaciones provocadas por la variable edad son muy pequeñas porque el coeficiente que la acompaña es muy cercano a cero (0,0007); en cambio, el coeficiente que corresponde a la condición laboral es de mayor peso (0,0571); por lo tanto, el incremento en los niveles de probabilidad es mayor.

Esto muestra que al querer pronosticar una respuesta o clasificar una persona, basándonos en un conjunto de características, no todas ellas tienen la misma ponderación; y esto, muchas veces, es contrario a nuestras intuiciones.

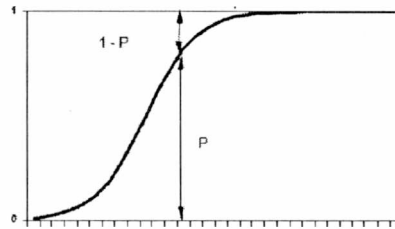
Entonces, al momento de definir una estrategia publicitaria, formular un discurso, o reorientar una línea editorial, teniendo en cuenta esta evidencia empírica, el "decisor" debería considerar que la edad de aquellos a quienes dirige su campaña, o la edad de aquellos a quienes destinará su discurso, *no es relevante. En cambio, sí lo es, la condición laboral.*

Cabe aclarar que el modelo hasta aquí presentado –de probabilidad lineal–, puede presentar algunos inconvenientes relativos a supuestos teóricos de la Estadística Inferencial; pero no es objetivo de este artículo desarrollar estos temas<sup>12</sup>.

### Modelos de Probabilidad no Lineal

Para superar los inconvenientes que podrían plantearse con la aplicación del modelo de probabilidad lineal, se utilizan otros modelos de funciones no lineales, tales como el **logit** (basado en la función de distribución logística acumulada) y el **probit** (que utiliza la distribución normal estándar acumulada).

Ambas distribuciones son simétricas con respecto a cero y sus gráficas son curvas; es decir, no lineales. Tal el caso de la distribución logística en forma de "S":



Siendo  $P$  = la probabilidad de que el elemento pertenezca a una categoría determinada; y  $1 - P$  = la probabilidad complementaria o probabilidad de que el elemento no pertenezca a esa categoría.

Se observa, además, que los valores de probabilidad variarán entre 0 y 1.

El *objetivo* de estos modelos es determinar cómo se relaciona la probabilidad de pertenecer a determinada categoría con los valores de las variables explicativas  $X_i$ <sup>13</sup>.

Como la relación es no lineal, el efecto sobre la probabilidad  $P$  de incrementar una unidad de  $X$ , no puede interpretarse como en el modelo lineal<sup>14</sup>. Sin embargo, podemos interpretar el valor de  $Y$  como un puntaje a favor de la categoría de interés: a medida que aumenta  $Y$ , aumenta la probabilidad  $P$ , como se mostrará a continuación.

### 2. El Modelo Logit

Este modelo analiza la relación que existe entre una variable dependiente dicotómica y varias variables independientes, cuantitativas (métricas) o

12 El modelo de probabilidad lineal, podría presentar los siguientes problemas (conf. Gujarati, D., op. cit., Capítulo 10):

- 1) La estimación de la probabilidad puede tomar valores fuera del rango 0 a 1.
- 2) No normalidad de la perturbación o error aleatorio.
- 3) Heteroscedasticidad

13 Para predecir estas probabilidades, efectuamos dos pasos: *primero* estimamos los coeficientes  $b_i$ . Así obtenemos  $Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$ . Luego, las probabilidades quedan determinadas por la aplicación de la función de probabilidad acumulada elegida, sea ésta la *logística* o la *normal estándar*.

14 Para conocer el efecto de cada variable explicativa sobre la probabilidad debe calcularse la derivada  $dP/dX$ .

cualitativas (no métricas). El propósito de este tipo de análisis es estimar la probabilidad de que se produzca un determinado suceso (“con valoración positiva”) si las variables independientes (“edad”, “condición laboral”) asumen determinados valores. El modelo es el siguiente:

$$P = \frac{e^Y}{1 + e^Y}$$

Siendo

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$$

Resulta obvio que  $P$  necesariamente caerá en el intervalo 0 a 1, dado que  $e^Y$  es siempre positivo<sup>15</sup>.

Tal como dijimos,  $Y$  puede interpretarse como un puntaje de la categoría de interés, que es función lineal de las variables explicativas  $X$ . A mayor valor de  $Y$ , mayor será la probabilidad  $P$ <sup>16</sup>.

A continuación, el siguiente ejemplo:

**Clasificación en dos categorías** C1: valoración positiva de un medio de comunicación determinado  
C2: valoración negativa de un medio de comunicación determinado

**Objetivo:** estudiar la proporción de personas con estudios universitarios completos, en función de la edad y su condición laboral, que tienen valoración positiva de un medio de comunicación determinado.

**Variabes explicativas:**  $X_1$ : edad  
 $X_2$ : condición laboral (Desocupado = 0; Ocupado = 1)

**Variable dependiente:** valoración de un medio de comunicación determinado  
 $P$  estimada por distribución logística

La estimación se realiza en dos pasos:

**1er. Paso:** Se ajusta linealmente  $Y = \ln(f/(n-f))$ , obteniéndose

$$Y^* = b_0 + b_1 X_1 + b_2 X_2$$

**2do. Paso:** Se estima la probabilidad como

$$P = e^Y / (1 + e^Y)$$

Supongamos 4 conjuntos distintos de estas características:

Conjunto Nº	Edad $X_1$	Condición Laboral $X_2$	Universitarios $n$	Con valoración positiva $f$	Frecuencia relativa $fr = f/n$	$f/(n-f)$	$\ln(f/(n-f))$
1	25	1	36	10	0,2778	0,3846	-0,9555
2	35	0	29	8	0,2759	0,3810	-0,9651
3	50	0	23	5	0,2174	0,2778	-1,2809
4	65	1	12	4	0,3333	0,5000	-0,6931

<sup>15</sup> La constante  $e = 2,71828\dots$

<sup>16</sup> Además,  $e^Y = \frac{P}{1-P}$

Aplicando logaritmo natural, se obtiene  
 $\log \frac{P}{1-P} = Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$

Tomando  $f/n$  como estimador de  $P$ , se obtiene

$$\log \frac{(f/n)}{1 - (f/n)} = \log \frac{f}{n-f}$$

Cabe aclarar que, si  $f$  es un valor dentro del intervalo  $(0;n)$  para todos los valores de la variable explicativa, los coeficientes pueden calcularse por regresión. Si  $f$  es un valor igual

a 0 o  $n$ , no puede calcularse  $Y$  utilizando regresión. Los parámetros  $b_i$  se estimarán mediante el Método de Máxima Verosimilitud (el estimador de máxima verosimilitud de un parámetro para una muestra dada, es el valor del parámetro que hace máxima la probabilidad de la muestra).

Y*	P estimada
-0,8874	0,2916
-1,1467	0,2411
-1,0993	0,2499
-0,7612	0,3184

- La probabilidad de que una persona con estudios universitarios completos, desocupado y de 40 años de edad valore de forma positiva un medio de comunicación determinado es igual a

**1er. Paso:**  $Y^* = -1,2571 + 0,0032X_1 + 0,2908X_2$   
 $Y^* = -1,2571 + 0,0032(40) + 0,2908(0) = -1,130$

**2do. Paso:**  $Exp(-1,130)/[1+Exp(-1,130)] = 0.2442$  ó 24,42%

Veamos, como lo hicimos en el modelo anterior, qué ocurre si modificamos la condición laboral; es decir, tomemos una persona con estudios universitarios completos, de 40 años y ocupado.

En el primer paso, obtenemos:

$$Y^* = -1,2571 + 0,0032(40) + 0,2908(1) = -0,8383$$

Luego, en el segundo paso, se obtiene la probabilidad igual a:

$$Exp(-0,8383)/[1+Exp(-0,8383)] = 0.3019$$
 ó 30,19%

Se ve que al aumentar el valor de  $Y^*$ , aumenta la probabilidad de que la valoración sea positiva, pasando del 24% al 30%. Además, el valor de  $Y^*$ , se ha visto incrementado por la variable *condición laboral*.

### Comparación entre los resultados del Modelo de Probabilidad Lineal y del Modelo Logit

Siguiendo el mismo ejemplo (mismas variables, mismos datos, mismas

frecuencias), se presenta el siguiente cuadro comparativo de los resultados de la aplicación de ambos modelos.

Conjunto Nº	Edad X <sub>1</sub>	Condición Laboral X <sub>2</sub>	Universitarios n	Valoración positiva f	Probabilidad estimada	
					Probabilidad Lineal	Logit
1	25	1	36	10	0,2908	0,2916
2	35	0	29	8	0,2411	0,2411
3	50	0	23	5	0,2522	0,2499
4	65	1	12	4	0,3203	0,3184

Se puede observar que los resultados –en ambos modelos– son muy similares. La aplicación de uno u otro modelo dependerá de –entre otros

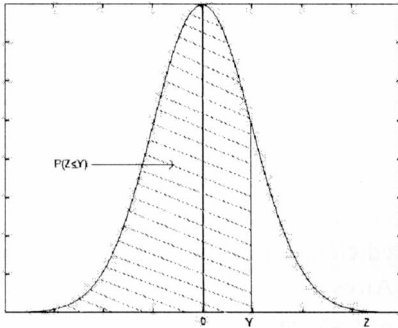
factores– si el rango de variación de las variables explicativas conduce a valores de probabilidad fuera del intervalo [0;1].

### 3. El Modelo Probit

En este modelo, la probabilidad que un elemento pertenezca a una determinada categoría está dada por la probabilidad de que una variable  $Z$  (variable normal estándar) sea menor o igual a un puntaje  $Y$ ; por lo tanto, se calcula hallando el área bajo la curva de la distribución normal estándar a izquierda del valor  $Y$ <sup>17</sup>, siendo  $Y$  una función de las variables explicativas  $X$ , tal como en el modelo anterior.

<sup>17</sup> El cálculo de  $P$  es sencillo y se puede realizar mediante el empleo de la Tabla de Distribución Normal Estándar.

Se puede graficar de la siguiente manera:



Entonces, para cada conjunto de variables explicativas  $X_i$ , la probabilidad estará dada por:

$$P(Z \leq Y)$$

donde

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$$

Siendo  $Y$ , una función lineal de las variables explicativas  $X_i$ .

Los coeficientes  $b_i$ , se estiman por el método de máxima verosimilitud<sup>18</sup>.

### Modelos de Respuesta Múltiple

Por motivos de extensión del presente artículo, no expondremos en detalle, los modelos de respuesta múltiple. Solamente diremos que, los elementos pueden clasificarse en más de dos categorías, por ejemplo,  $m$  categorías,  $C_1, C_2, \dots, C_m$ , en función de las variables explicativas  $X_1, X_2, \dots, X_k$ . Tal el caso de clasificar a una persona en tres categorías como “mira programas de actualidad política regularmente”, “mira programas de actualidad política ocasionalmente” y “no mira programas de actualidad política”.

<sup>18</sup> No se muestran los resultados del Modelo Probit porque, para la aplicación de este modelo, es necesario el uso de paquetes estadísticos como el SPSS for Windows, y en este trabajo se utilizó Microsoft Excel.

ca”, según las variables explicativas “edad” y “condición laboral”.

Al igual que en la clasificación de dos categorías, la probabilidad que un elemento pertenezca a la categoría  $i$ -ésima, dado un conjunto de características  $x$ , será

$$P_i = P(C_i/x).$$

Al ser la suma de las  $m$  probabilidades igual a 1, una probabilidad cualquiera puede expresarse como

$$P_m = 1 - \sum P_i$$

con  $i = 1, 2, \dots, m - 1$

En el caso del modelo de probabilidad lineal, éste queda expresado por

$$P_i = b_{0i} + b_{1i} X_1 + b_{2i} X_2 + \dots + b_{ki} X_k$$

$Y$  para el modelo Logit, será

$$P_i = \frac{e^{Y_i}}{1 + \sum e^{Y_j}}$$

con  $i = 1, 2, \dots, m - 1$

### Comentarios Finales

“Una técnica multivariada no puede decirnos lo que es, por ejemplo, la inteligencia, ni qué es el conocimiento, ni qué son ni cómo son las operaciones psicológicas; lo único que permite es distinguir las unidades de función que las operaciones psicológicas manifiestan en la conducta y ello, de manera aproximada y probable” (San Martín Castellanos, R., en Lévy Mangin, J. y Varela Mallou, J.; 2003, p. XXVII)

Como puede apreciarse, la información que se obtiene a partir de la aplicación de los modelos probabilísticos propuestos, es más amplia que la que surgiría de un tratamiento más simple de los datos. No arroja certezas, pero contribuye a disminuir el grado de incertidumbre en la evaluación de las actitudes, intereses, opiniones de las

personas (es decir, en las distintas manifestaciones observables) con respecto a un tema en particular.

Los “porcentajes” que se obtienen (en el primer ejemplo “el 27% de los universitarios ocupados de 25 años tienen una valoración positiva de un medio de comunicación determinado”), y que son el tipo de resultado habitual en el tratamiento de las encuestas, se limitan a esa muestra en particular. En cambio, mediante la aplicación de los modelos presentados, se puede estimar –es decir, predecir– valores para el conjunto de la población.

Este tipo de modelización permite, por un lado, la estimación de la probabilidad de que una persona pertenezca a un grupo o a otro (capacidad clasificatoria); y por otro, la identificación de las variables y su “peso” en la explicación del fenómeno (capacidad predictiva), basándose en características observables<sup>19</sup>.

Además, se evita una lectura ingenua<sup>20</sup> de los datos, del tipo “si de 36 universitarios ocupados de 25 años, 10 tienen una valoración positiva de un medio de comunicación determinado; entonces de 3600 universitarios ocupados de 25 años, 1000 tendrán una valoración positiva de un medio de comunicación determinado”. Ya que, como se vio en el ejemplo mencionado, siendo la probabilidad estimada 0,2908, habrá 1047 universitarios con valoración positiva.

<sup>19</sup> El “peso” de cada variable se establece después de aplicados los modelos; es decir, se puede corroborar o refutar una intuición a priori sobre la incidencia que tiene cada variable sobre la aparición de un fenómeno determinado. Analizando los resultados de los ejemplos desarrollados, surge que la variable Edad tiene un “peso” inferior a la variable Condición Laboral.

<sup>20</sup> Y errónea.



Las encuestas son, en general, ricas en datos y, entendemos, debería aplicarse a ellos la mayor cantidad de herramientas de análisis estadístico a nuestro alcance.

Cabe señalar que el tipo de análisis que se propone es más o menos habitual cuando se trata de establecer modelos predictivos (por ejemplo, en estudios de mercadotecnia). Sin embargo, en lo que respecta al análisis de las encuestas en otras áreas sociales, el mismo no se lleva a cabo o, al menos, no es el que se realiza con mayor frecuencia.

En ese sentido, entonces, se propone esta herramienta metodológica: no como una innovación, sino más bien como un aporte al manejo y al tratamiento de los datos, de forma tal que se tuviera en cuenta, por un lado, lo cualitativo y lo cuantitativo de las observaciones y, por el otro, la aplicación de métodos estadísticos de Análisis Multivariado, que permite establecer modelos predictivos y clasificar a los integrantes de la población, según un conjunto de características dadas.

Si bien no descartamos los habituales estudios descriptivos, pretendemos superar una limitación metodológica en el tratamiento de los datos y mostrar que las encuestas –técnica muy difundida en los estudios sociales– están siendo poco aprovechadas en su riqueza informativa.

### Bibliografía

Cuadras, C. M., (1981), *Métodos de Análisis Multivariante*, Editorial Universitaria de Barcelona.

Gujarati, D. (1981), *Econometría Básica*, Capítulo 13: Regresión de una variable dicótoma, McGraw Hill, Bogotá.

Hair Jr., J., Anderson, R., Tatham, R. y Black, W. (1995), *Multivariate Data Analysis*, Prentice Hall, 4th. Edition.

Iglesias Antelo, S. y Aranzazu Sulé Alonso, M., (2003). Capítulo 1, Introducción al análisis multivariable, en Lévy Mangin, J. P. y Varela Mallou, J. (directores), *Análisis Multivariable*

*para la Ciencias Sociales*, Pearson, Prentice Hall, Madrid.

Mason, R., Lind, D. y Marchal, W. (2000), *Estadística para Administración y Economía*, Alfaomega, 10ª edición, Colombia.

Mora y Araujo, M., (2005), *El poder de la conversación. Elementos para una teoría de la opinión pública*, 1ª edición, La Crujía Ediciones, Buenos Aires.

San Martín Castellanos, R., en Lévy Mangin, J.P. y Varela Mallou, J. (directores) (2003), *Análisis Multivariable para la Ciencias Sociales*, Pearson, Prentice Hall, Madrid.

Tryfos, P. (1998), *Methods for business analysis and forecasting: text and cases*, John Wiley & Sons, Inc.

Wayne, D. (1993), *Estadística con aplicaciones a las ciencias sociales y a la educación*, McGraw Hill, México.

Zikmund, W. (1998), *Investigación de Mercados*, 6ª edición, Prentice Hall, México.