









INTRODUCCIÓN A LOS MODELOS ESTRUCTURALES  
EN INVESTIGACIÓN SOCIAL



ANTONIO ALAMINOS CHICA  
FRANCISCO FRANCÉS GARCÍA  
CLEMENTE PENALVA VERDÚ  
ÓSCAR SANTACREU FERNÁNDEZ

INTRODUCCIÓN A LOS  
MODELOS ESTRUCTURALES  
EN INVESTIGACIÓN SOCIAL



**PYDLOS**  
**ediciones**

© de la presente edición: Universidad de Cuenca

**INTRODUCCIÓN A LOS MODELOS ESTRUCTURALES  
EN INVESTIGACIÓN SOCIAL**

Antonio Alaminos Chica  
Francisco José Francés García  
Clemente Penalva García  
Óscar Antonio Santacreu Fernández

ISBN: 978-9978-14-314-8  
Derecho de autor: CUE-2349

Diseño Portada: Óscar Santacreu  
Diagramación: Patricia Barbero  
Corrección de estilo: María Eugenia Estrella  
Impresión: Editorial Don Bosco-Centro Gráfico Salesiano - Telf.: 2831745  
Tiraje: 300  
Impreso en Ecuador - *Printed in Ecuador*

2015

Este libro ha sido debidamente examinado y valorado por evaluadores ajenos a PYDLOS EDICIONES, con el fin de garantizar la calidad científica del mismo.

Edición corregida, revisada y ampliada del texto “El análisis de la realidad social”, para servir de referencia durante el Curso de formación de Posgrado: “Investigación Aplicada en Ciencias Sociales: Técnicas de producción de datos y análisis”, actividad académica organizada por el Grupo de investigación PYDLOS del Departamento de Investigación “Espacio y Población”, en coordinación con las Facultades de Ciencias Económicas y Administrativas, Jurisprudencia, Psicología, Filosofía Letras y Ciencias de la Educación, y con aval de la DIUC de la Universidad de Cuenca.

# ÍNDICE

1. PRESENTACIÓN .....	9
2. INTRODUCCIÓN: REGULARIDADES Y ESTRUCTURAS .....	11
3. ELEMENTOS DE LOS MODELOS ESTRUCTURALES .....	19
3.1. CONCEPTOS TEÓRICOS Y VARIABLES .....	19
3.2. MEDICIÓN .....	27
3.3. LOS NIVELES DE MEDICIÓN .....	30
3.4. ERROR DE MEDICIÓN .....	34
3.5. CIFRADOS Y TRANSFORMACIONES .....	38
3.6. EXPLICACIÓN Y CAUSALIDAD .....	40
3.7. LA DETERMINACIÓN TEÓRICA DEL ORDEN EXPLICATIVO .....	46
4. MODELOS ESTRUCTURALES CON VARIABLES OBSERVADAS .....	49
4.1. EL SIGNO DE LAS RELACIONES .....	54
4.2. CONSTRUCCIÓN DE MODELOS ESTRUCTURALES .....	55
5. EXPRESIÓN MATEMÁTICA .....	59
5.1. NOTACIÓN DE SISTEMAS ESTRUCTURALES .....	59
5.2. SISTEMAS DE ECUACIONES .....	60
5.3. PRESUNCIONES .....	62
5.4. TRANSFORMACIONES .....	63
5.5. PARÁMETROS TEÓRICOS Y ESTIMADOS EMPÍRICOS .....	65
5.6. PRIMERA REGLA DE DESCOMPOSICIÓN .....	66
5.7. SEGUNDA REGLA DE DESCOMPOSICIÓN .....	67
6. TIPOLOGÍA DE SISTEMAS .....	69
6.1. SISTEMAS SUPRESORES O DE REFUERZO .....	69
6.2. MODELOS RECURSIVOS Y NO RECURSIVOS .....	72

7. EL PROBLEMA DE LA IDENTIFICACIÓN .....	77
7.1. LA DETERMINACIÓN DEL ESTADO .....	79
7.2. IDENTIFICACIÓN DEL SISTEMA .....	81
7.3. CONDICIONES DE ORDEN .....	82
7.4. CONDICIONES DE RANGO .....	82
7.5. LOS PROCEDIMIENTOS DE RESTRICCIÓN .....	84
8. TESTADO DE MODELOS. LA BONDAD DE AJUSTE .....	89
9. MODELOS ESTRUCTURALES DE MEDICIÓN .....	99
9.1. LA EXPLICACIÓN ENTRE LATENTES .....	111
10. MODELOS ESTRUCTURALES CON LISREL .....	115
10.1. INTRODUCCIÓN: LISREL Y SIMPLIS .....	115
10.2. GENERACIÓN DE UN ARCHIVO SIMPLIS .....	116
10.3. CONVENCIONES DEL LENGUAJE SIMPLIS .....	119
10.4. TESTADO Y AJUSTE DE MODELOS CON LISREL .....	122
10.5. MODELOS DE MEDICIÓN .....	126
10.6. MODELOS ESTRUCTURALES CON VARIABLES LATENTES .....	129
11. EJEMPLOS DE MODELOS .....	133
12. BIBLIOGRAFÍA .....	147

# 1. PRESENTACIÓN

Durante años, el Programa Interdisciplinario de Población y Desarrollo Local Sustentable (PYDLOS) de la Universidad de Cuenca (Ecuador), y el grupo de OBETS, del Instituto de Desarrollo Social y Paz de la Universidad de Alicante (España) vienen manteniendo una estrecha colaboración en docencia e investigación, fruto de la cual es producto esta serie de monografías.

Esta estrecha colaboración ha sido posible gracias a la intensa actividad desarrollada por el Director de PYDLOS, el profesor Dr. Alejandro Guillén, gracias a cuyo apoyo se han propuesto y plasmado diversas líneas de cooperación. Este libro es una prueba más de ello. Estos textos se han preparado para servir de referencia durante el Curso de Formación de Posgrado: “Investigación Aplicada en Ciencias Sociales: Técnicas de producción de datos y análisis”, impartido en PYDLOS-Universidad de Cuenca en los años 2014 y 2015.

Los cinco volúmenes se estructuran de acuerdo a los contenidos de este curso de postgrado.

- I: *Medición. Encuestas y cuestionarios.*
- II: *Del cuestionario a la base de datos. Socioestadística y análisis de datos.*
- III: *Estadística multivariable para el análisis de encuestas de opinión pública.*
- IV: *La perspectiva cualitativa. Análisis de datos textuales.*
- V: *Análisis Espacial Socioeconómico.*

Este curso, que se imparte por segundo año consecutivo, ha sido posible gracias a la cofinanciación de las dos universidades, la Universidad de Cuenca y la Universidad de Alicante. En el caso de la Universidad de Alicante, la actuación es posible gracias a la Resolución del 14 de abril de 2014 del Vicerrectorado de Relaciones Internacionales de la Universidad de Alicante por la

que se aprueba la ayuda a este proyecto, dentro de la Convocatoria “Cooperación Universitaria al desarrollo” del año 2014.

Esta actividad académica ha sido organizada por el Grupo de investigación PYDLOS del Departamento de Investigación "Espacio y Población", en coordinación con las Facultades de Ciencias Económicas y Administrativas, Jurisprudencia, Psicología, Filosofía Letras y Ciencias de la Educación, y con aval de la DIUC de la Universidad de Cuenca. Este esfuerzo transversal de la Universidad de Cuenca tiene como principal objetivo fomentar el fortalecimiento de las Ciencias Sociales y se realiza en el marco del proyecto de investigación “Construcción conceptual y medición del Buen Vivir”.

Estamos convencidos de que estos eventos constituyen un aporte significativo para la formación de investigadores y docentes de nuestras universidades y para el robustecimiento de los vínculos entre ambas.

LOS AUTORES

Sant Vicent del Raspeig (Alicante)

julio 2015

## 2. INTRODUCCIÓN: REGULARIDADES Y ESTRUCTURAS

La sociología existe en la medida en que existe regularidad y estructura. Esto no implica o excluye la existencia del cambio y transformación. Toda sociedad es histórica, y en ese sentido, cambiante. No obstante, la posibilidad de conocer y más aún, de explicar, exige de una pauta o regularidad. Así, el cambio podrá ser objeto de estudio siempre que no sea de tipo aleatorio. Es importante advertir, asimismo, que esta regularidad no es de tipo “normativo”, impuesta por unas “normas sociales”. Debe diferenciarse entre diferentes fuentes de regularidad social, señalando aquella que puede ser objeto de la investigación sociológica.

El problema radica en que el sistema normativo puede hacer *comprensibles* las regularidades del comportamiento individual, pero no *explicar* tales comportamientos ni tampoco las regularidades del agregado y ambos fenómenos son sustancialmente diferentes. Así, Von Hayek ha propuesto distinguir tajantemente entre “el sistema de reglas de conducta que gobierna el comportamiento de los miembros individuales del grupo [...] por un lado, y por otro, el orden o pauta de acciones que resulta de aquel para el grupo como totalidad”<sup>1</sup>. De este modo, la norma interpretada individualmente no responde de los comportamientos agregados. Es decir, la regularidad normativa (o ética) permite *comprender* acontecimientos *singulares*, *acciones*, pero no *explicar* regularidades típicas del agregado<sup>2</sup>, pues éstas no dependen *sólo* de las acciones. Del análisis, concluye E. Lamo<sup>3</sup>, se debe distinguir radicalmente el orden normativo del social. No cabe en todo caso señalar una situación determinística, en la medida que existen posibilidades de cambio en los dos

---

1. Von Hayek, “Notes on the Evolution of the Systems of Rules of Conduct”, *Studies in Philosophy, Politics and Economics*, Citado en E. Lamo de Espinosa, *op. cit.*, p. 8.

2. K. Popper, *Rationality and the Status of the Rationality Principle (1967)*, citado por J.C. Zapatero en “K. Popper y la metodología de las ciencias sociales”, en *Cuadernos Económicos de ICE*, 3-4 (1977), p. 103.

3. E. Lamo de Espinosa, *La sociedad reflexiva*, *op. cit.*, pp. 88-91.

niveles individual y agregado. Es importante la distinción entre la acción normativa (basada en el debe hacer/ debe ser) y la de tipo probabilístico que se postula para la regularidad no normativa. Las regularidades sociales, en términos probabilísticos es lo que se denominaría “ley sociológica”. Así, “(la sociología) necesita que haya hechos sociales; pero si además quiere ser ciencia *nomotética* necesita que haya leyes sociales”<sup>4</sup>. No debe en todo caso confundirse la noción de ley con la de teoría. El rasgo diferencial entre leyes y teorías se encuentra en que las leyes hacen referencia a las características “empíricas” de los fenómenos, o sea, a aspectos observables y no a “conceptos teóricos” o abstractos.

Tal y como señala M. Navarro, la aproximación desde el constructivismo afirma la naturaleza no determinista en la constitución de cualquier regularidad social al mismo tiempo que destaca el papel contingente e histórico de las construcciones sociales. “El construccionismo social que inicia la sociología weberiana parte, pues, de acciones sociales individuales entrelazadas, más o menos permanentes o transitorias. Cualquier relación social puede quebrarse en todo momento y por ello, sólo podemos afirmar la regularidad en términos de probabilidad de conducta y no de normas: «la relación social *consiste* sólo y exclusivamente –aunque se trate de formaciones sociales como ‘Estado’, ‘iglesia’, ‘corporación’, ‘matrimonio’, etc.– en la probabilidad de que una forma determinada de conducta social, de carácter recíproco por su sentido, haya existido, existe o pueda existir»<sup>5</sup>.

Las regularidades que estamos considerando serían en todo caso, agregadas y probabilísticas. Wilfredo Pareto, ilustre antecedente de la sociología matemática, destacaba los problemas en la detección de regularidades: las distorsiones que se aprecian en las diferentes regularidades son producto de los errores y limitaciones propias del ser humano. En ese sentido, afirmaba que:

“(7) Hablando propiamente no puede haber excepciones a las leyes económicas y sociológicas, en la misma forma que las otras leyes científicas. Una uniformidad no uniforme no tiene sentido. Pero las leyes científicas no tienen una existencia objetiva. La imperfección de nuestro espíritu no nos permite considerar los fenómenos en su conjunto y estamos obligados a estudiarlos separadamente. En consecuencia, en lugar de uniformidades generales que están y que quedarán siempre ignoradas, estamos obligados a considerar un número infinito de uniformidades parciales, que crecen, se superponen y se oponen de mil maneras. Cuando consideramos una de esas uniformidades,

4. *Ibid.*, p. 95.

5. Max Weber, *Economía y sociedad*, México, Fondo de Cultura Económica, 1969, p.22, citada por M. Navarro, “Apuntes para una teoría de la cultura económica”, en VV. AA. *Escritos de Teoría Sociológica en homenaje a Luis Rodríguez Zúñiga*, Madrid, CIS, 1992, p.792.

y que sus efectos son modificados u ocultos por los efectos de otras uniformidades, que no tenemos la intención de considerar, decimos de ordinario, pero la expresión es impropia, que la uniformidad o la ley considerada sufre de excepciones. Si es admitida esta forma de hablar, las leyes físicas, y aún las matemáticas, comportan excepciones, lo mismo que las leyes económicas. (...)

(8) Una ley o una uniformidad no es verdadera sino bajo ciertas condiciones, que nos sirven precisamente para indicar cuáles son los fenómenos que queremos destacar del conjunto. Por ejemplo, las leyes químicas que dependen de la afinidad son diferentes, según que la temperatura se mantenga en ciertos límites, o los sobrepase. Hasta una cierta temperatura los cuerpos no se combinan; más allá de esa temperatura se combinan, pero si aumenta todavía más allá de cierto límite se disocian.

(9) Esas condiciones son unas implícitas y otras explícitas. No se debe hacer entrar entre las primeras más que las que son entendidas fácilmente por todos y sin el menor equívoco; si no eso sería un jeroglífico y no un teorema científico. No hay proposición que no se pueda certificar como verdadera bajo ciertas condiciones a determinar. Las condiciones de un fenómeno son parte integrante de ese fenómeno, y no pueden ser separadas.

(10) No conocemos, ni podremos jamás conocer, un fenómeno concreto en todos sus detalles; siempre hay un residuo. (...).

(11) Puesto que no conocemos enteramente ningún fenómeno concreto, nuestras teorías de esos fenómenos no son más que aproximadas. (...)"<sup>6</sup>.

Destaca en la noción de Pareto que el fenómeno es indisociable de sus condiciones de realización pero también que un mismo proceso puede generar regularidades distintas, incluso contradictorias. En ese sentido, M. Navarro destaca cómo la progresión de los esquemas de decisión instrumental adquieren una mayor complejidad conforme se difunden como culturalmente adecuados para la toma de decisiones. "El hecho sociológico básico que hay que resaltar y que guía toda la concatenación de fenómenos que se han expuesto hasta este momento, es la aparición en nuestras sociedades de conductas económicas definidas, que han tenido, a su vez, influencia en la aparición de un complejo de conductas no económicas, todas ellas con un carácter masivo y que se hacen más centrales, en la vida de los hombres, que tienen un carácter abstracto y complejo, que requieren un mayor número de conocimientos y que ofrecen una previsible perspectiva de aumentar el número de complejidad y de generalizarse socialmente"<sup>7</sup>

6. W. Pareto, *Manual de economía política*, Genève, Librairie Droz, 1964

7. M. Navarro, "Apuntes para una teoría de la cultura económica", en VV. AA. *Escritos de Teoría Sociológica en homenaje a Luis Rodríguez Zúñiga*, Madrid, CIS, 1992, p. 786.

La noción misma de “regularidad” puede aparecer denominada como “ley social” en un sentido “blando”. Otros autores como J. Elster prefieren matizar sustituyendo la palabra “ley” por “mecanismo” “(...) debilidad de la teoría más famosa de la explicación científica, la propuesta por Carl Hempel. Él sostiene que la explicación equivale a la deducción lógica del acontecimiento a explicar, con leyes generales y declaraciones de las condiciones iniciales como las premisas. Una objeción es que las leyes generales pueden expresar correlación pero no causa. Otra es que las leyes, aunque sean genuinamente causales, pueden ser anticipadas por otros mecanismos. Es por eso que aquí he puesto el acento en los mecanismos y no en las leyes. Esto no es un profundo desacuerdo filosófico. Un mecanismo causal tiene un número finito de eslabones. Cada eslabón se debe describir mediante una ley general y en ese sentido por una ‘caja negra’ acerca de cuyos engranajes internos permanecemos en la ignorancia. Pero para los fines prácticos (los fines del científico social en acción) es importante el lugar del acento dinámico de la explicación científica: el impulso a producir explicaciones cada vez más finas”<sup>8</sup>. Esta matización se corresponde en un sentido global con los planteamientos de Pareto y responde a la dinámica entre generalización y detalle en la explicación para alcanzar una mejor determinación de las regularidades.

Otro aspecto interesante procede desde el ámbito de la generalización y lo local. En este contexto reaparece el debate sobre la utilidad local de las generalizaciones. Nuevamente la dicotomía entre sociología como ciencia o tecnología. P. von Morpurgo<sup>9</sup>, reflexiona cómo “El carácter de las ciencias sociales está íntimamente vinculado a su reivindicación de universalidad. Si se puede determinar la regularidad de algunos fenómenos, y por consiguiente establecer leyes que tengan un alcance casi universal, esa reivindicación se puede aceptar. No obstante, con una vasta aplicación de este principio se corre el riesgo de ignorar la diversidad local. En efecto, en estos últimos cincuenta años, las ciencias sociales y sus repercusiones han cobrado mayor fuerza gracias a la mejora de sus métodos e instrumentos de investigación aplicables a escala reducida. Es evidente que, en materia de elaboración de políticas, los resultados y recomendaciones adecuadamente detallados y relacionados con un entorno específico son más pertinentes que las vastas generalizaciones”. No parece, sin embargo, que ambas tareas sean excluyentes, sino más bien complementarias.

La característica distintiva en la sociología aplicada es el empleo de métodos formales para la construcción de modelos en el análisis de los fenómenos

---

8. J. Elster, *Tuercas y tornillos* Barcelona, Gedisa. 1996 p. 16

9. P. von Morpurgo, *op. cit.*,

sociales. Así, T. Fararo en la introducción de su manual afirma cómo<sup>10</sup> “este libro pretende ser una contribución a acelerar la formalización de las teorías así como a restituir la capacidad explicativa de la teoría mediante el empleo de modelos”. Los dos conceptos clave son formalización y representación mediante modelos. La actividad de la sociología matemática no se dirige a una labor de formalización *per se* de la teoría sociológica, sino que contribuye a ello mediante el desarrollo de modelos formalizados, generalmente de carácter matemático. La observación es importante, dado que la función del análisis lógico se circunscribe a la elaboración de los modelos. Es evidente que el empleo y desarrollo de modelos dentro de una teoría ayuda a su formalización y potencia la consistencia interna. Esta tarea constituye una aproximación metodológica bastante definida a la realidad social.

En ese sentido, cabe recordar que una de las características del método científico es, según Hanson, la búsqueda de un modelo en el que insertar los datos: “En una disciplina de búsqueda en crecimiento, la investigación se dirige, no a reordenar viejos hechos y explicaciones en modelos formales más elegantes, sino más bien al descubrimiento de nuevos esquemas de explicación”<sup>11</sup>. A este respecto, la misma opinión sostiene Allais: “Cuando se analizan los fenómenos sociales sobre todo los económicos, se revela la existencia de regularidades tan sorprendentes como las que encontramos en las ciencias físicas (...). Toda ciencia se basa en modelos, ya sean descriptivos o explicativos o estén destinados a hacer pronóstico o a tomar decisiones”<sup>12</sup>. En ese sentido, la noción de representación es uno de los núcleos importantes en la actividad de la sociología aplicada.

El estudio de modelos estructurales lleva asociado habitualmente el penetrar en el futuro, ya sea a efectos de control metodológico (como parte de un diagnóstico del ajuste), o como finalidad del modelado. Por ello, el modelar implica el descubrir pautas regulares y sistemáticas en los datos, de modo que nos permitan construir un modelo matemático que explique las relaciones conocidas entre los fenómenos sociales. A partir de ello, se intentará predecir el comportamiento futuro más probable en el fenómeno social.

No obstante, en este escenario de modelado, para que el futuro sea predecible debe de ser semejante al pasado en lo que se refiere a los factores que

---

10. T. Fararo, *Mathematical Sociology: an introduction to fundamentals*. New York, Wiley, 1973. p.15

11. Citado en T.A. Sebeok “One, Two, Three...Uberty” en U. Eco y T.A. Sebeok (eds.) *El signo de los tres*. Dupin, Holmes, Peirce. Barcelona, Lumen. 1989. p.74

12. M. Allais, “La pasión por la investigación” en M. Szenberg (ed) *Grandes economistas de Hoy*, Madrid, Debate, 1994

generan el comportamiento de la serie. Es decir, las reglas que rigen el cambio deben mantenerse estables en el futuro, dado que el "mecanismo" modelado se pone a funcionar para generar los valores futuros de la variable. Por lo tanto, cualquier cambio brusco en el medio donde se desenvuelve el "mecanismo", o en los elementos que componen el "mecanismo" podrán inducir a error. Podemos pensar que la eficacia del modelo dependerá, por lo tanto, de la "profundidad" en que estén operando sus reglas en el sistema social, por un lado, y del contexto "evolucionario" o "revolucionario" en el que se desarrolla el fenómeno social en estudio.

Un modelado elaborado sobre una realidad social epidérmica (epifenoménica), determinará un "mecanismo" con predicciones excesivamente contingentes. La eficacia del modelado de una estructura de covarianzas dependerá más de la reflexión sociológica que le concede un ámbito de existencia, determinando los "mecanismos" a revelar, que de las técnicas instrumentadas para acceder a ellos. En otro sentido, el proceso o fenómeno social en estudio puede desenvolverse en un entorno social en equilibrio dinámico o en profunda y rápida transformación. La posible existencia de discontinuidades (una guerra civil, una revolución, un golpe de estado, etc.) matiza las probabilidades futuras de realización del fenómeno social modelado. De este modo, un conjunto relacionado de variables o indicadores se modela según sus propias características, pero su futuro se evalúa por sus circunstancias o contexto.

La investigación sociológica recorrió un camino importante desde los diseños correlacionales, el análisis de regresión y el "análisis de senderos". A principios de la década de los 70 quedaba claro que desde diferentes disciplinas se estaba considerando un problema parecido. Así, el análisis de senderos en sociología, los modelos de ecuaciones estructurales en econometría o los modelos de análisis factorial en psicología respondían a un mismo intento de modelado. Esta aproximación se consolidaría en una estrategia general de modelado denominado modelos estructurales de covarianzas. Los modelos estructurales se emplean tanto en la medición como en la explicación. Ambas dimensiones pueden combinarse, de modo que simultáneamente a la realización de una medición (por ejemplo, de variables latentes) se puede estructurar un diseño explicativo. Los denominados modelos causales son un área de actividad dentro de este campo. Es importante contextualizar el concepto de explicación (o causalidad) a un doble nivel, epistemológico y operativo.

Esta dinámica se vería muy influenciada por el impacto "operativo" de los nuevos métodos de computación, que ayudarían a generar modelos matemáticos de elevada sofisticación, como los desarrollados en la Escuela de Wisconsin. El motor del proceso no fue un salto cualitativo en los desarrollos de la

sociología que exigiese de expresión matemática, sino un impacto tecnológico que operando sobre los procedimientos (que no sobre la teoría) intento arrastrar a ésta tras sí, enganchada en el atractivo de los modelos matemáticos.

Como podremos evaluar más adelante, la investigación primero nombra y después numera. Cuando existen números sin nombre (como en el análisis factorial exploratorio o modelos estructurales de medición) la desorientación está servida y el resultado queda disponible a las veleidades argumentales del investigador. De este modo, los desarrollos de la computación (numerología) sólo podía potenciar los procedimientos e indirectamente, el método “positivista” que se articula sobre dichos procedimientos. Indirectamente, en la medida que los nuevos procedimientos no se vertebran claramente sobre él; de hecho, en términos funcionalistas, los procedimientos son denominados métodos, pero muy raramente se encontrará una articulación explícita entre investigación y teoría.

La tendencia iniciada en los años 70 iniciaba una autonomización del método (en los términos anteriores), que conduciría no ya al contraste de teorías (realmente difícil) sino de “resultados”. Es el refugio en el metodologismo que Coser (1975) denunciará y que Hope (1981<sup>a</sup>), recuperando a Lakatos, afirmaría como proceso ritual: “El gran error de la construcción de modelos reside en concentrarse en cuestiones de ajuste y el examen de las desviaciones del modelo, con la exclusión de la investigación del significado del modelo en sí mismo, de manera que éste no constituye ya un instrumento de progreso, sino más bien el centro ritual de un proceso degenerativo”. No eran, en general, ideas sociológicas servidas con técnicas estadísticas, sino desarrollos estadísticos ilustrados con una sociología simplificada. La sociología matemática entendida en ese modo entró en crisis, evidentemente. Ya desde principios de los años 70 (Golthorpe, Hope, Bertaux y Kreckel en el coloquio de Constanza) y especialmente en su segunda mitad, surgen fuertes críticas que cuestionan el ateoricismo en que se mueven estos metodólogos y la no significatividad de los resultados obtenidos mediante modelos matemáticos muy elaborados sobre ideas sociológicas rudimentarias.

Destaca especialmente el discurso de Coser (1975) como presidente de la Asociación Americana de Sociología, donde crítica duramente la sustitución de ideas por números. Como ya comentábamos, la eclosión de modelos producidos por la primera “sociología matemática” era esencialmente una explosión inducida desde el desarrollo de otros ámbitos (estadística, computación, etc.) y que en esas condiciones, tendía a carecer de intereses propios y contenido.

*"Se trata, sencillamente, de una sana preocupación ante el trance por el que atraviesa el hombre moderno. (Quede aquí definido el hombre moderno como toda persona nacida después del edicto de Nietzsche "Dios ha muerto", y antes del éxito pop "I Wanna Hold Your Hand".) Tal "trance" puede enunciarse de una manera o de otra, si bien ciertos filósofos del lenguaje prefieren reducirlo a una ecuación matemática, fácil no ya de resolver sino de llevar en la cartera."*

W. Allen "Mi discurso a los graduados"

### 3. ELEMENTOS DE LOS MODELOS ESTRUCTURALES

#### 3.1. CONCEPTOS TEÓRICOS Y VARIABLES

La sociedad no viene definida exclusivamente por aquello que es visible culturalmente. Existen conceptos que no pueden ser percibidos directamente por el investigador social. Así, por ejemplo, la noción Weberiana del proceso de racionalización. Dicho proceso entendido como una secuencia histórica se aprecia indirectamente en multitud de indicadores, pero sin embargo rehúye a una observación directa. Del mismo modo, conceptos como ideología política, religiosidad, racismo, o los referidos a valores, actitudes, etc., expresan realidades no directamente medibles.

En ese sentido, la realidad que investiga el sociólogo presenta diferentes niveles de profundidad, desde una opinión epidérmica a una actitud racista o xenófoba.

*"Existen también numerosos cuerpos artificiales cuyas estructuras no son perceptibles; tal es el vino; es un cuerpo producido por el Arte, y, sin embargo, la fuerza que lo hace fermentar no es perceptible; su existencia se conoce sólo por su acción, y esta fuerza (quwwa) es la forma (sura) y la estructura (sigha) del vino; desempeña, respecto al vino, el mismo papel que el filo respecto a la espada, puesto que el vino ejerce su acción gracias a esa fuerza. (...) Así también los medicamentos, como la teriaca y otros, que son compuestos (murakkaba) por el arte de la Medicina, obran en el cuerpo merced a fuerzas que implica su composición; esas fuerzas no son perceptibles; los sentidos perciben sólo las acciones que resultan de ellas. Por tanto, un medicamento no llega a ser tal sino por dos cosas: las mezclas complejas (ajlat) de que se compone y la fuerza por la cual ejerce su acción. Las mezclas constituyen su materia, y la fuerza por la cual obra constituye su forma."*

*Al-Farabi. "Ihsa' al-'Ulum"*

Un análisis de la sociedad debe considerar la existencia de procesos, estructuras y dimensionalidades latentes, con capacidad explicativa y que en algunos

casos puede carecer incluso de nombre. Este es un caso muy frecuente en el análisis dimensional, donde la estructura de los datos revela variables latentes que poseen un contenido claro y carecen de un nombre preciso.

Cada vez es más evidente para quien se dedica a la investigación social que disfrutamos de un catálogo incompleto y culturalmente sesgado de las variables que puedan definir la realidad social. El descubrimiento de variables latentes ayuda a simplificar, integrar y des-trivializar la sociedad. Las variables son elementos que describen la realidad sustituyéndola, y también afirman cómo se construye explicándola.

Una de las primeras cuestiones que deben plantearse cuando nos referimos al concepto de variable es precisamente el de su existencia. Al emplear una variable estamos reemplazando la realidad por algo que la sustituye. En principio, en los manuales de técnicas de investigación, el problema principal es la dificultad para traducir de un modo válido y fiable dicha realidad en variables e indicadores. No obstante, previamente a ese esfuerzo técnico para efectuar una medición correcta ha existido la decisión de que esa variable es importante.

Quizás uno de los temas más interesantes al tratar con variables es precisamente la doble selección que efectúa todo investigador: primero con respecto a qué realidad le interesa y, segundo, qué variables la van a reemplazar en el análisis. Pensemos en un investigador que, para analizar un fenómeno, decida un conjunto de indicadores absolutamente anómalo. Diríamos que su análisis está mal especificado. Y esa puede considerarse la clave del desarrollo de la investigación social: la especificación correcta de la realidad. No solamente qué realidad es importante, dado que eso lo decidirá la financiación o los intereses personales del científico, sino qué variables van a traducirla, a sustituirla para que la describamos, a reemplazarla para que la expliquemos.

Un investigador decide qué variables o conjuntos de variables son relevantes para definir y explicar el fenómeno social que es objeto de su atención. Hablar de variables es, en definitiva, decidir qué partes de la realidad social existen y qué otras no.

En la mayor parte de los casos, las variables que un investigador considera sociológicamente relevantes han sido heredadas de la tradición en la investigación de ese tema, o propuestas por la teoría sociológica. En ese sentido, es necesario explicitar que existe una decisión previa tomada consciente o inconscientemente por el investigador con respecto a qué partes de la realidad social va a conceder existencia.

Precisamente la existencia de variables latentes, no percibibles directamente por el investigador, así como la dificultad para etiquetarlas conceptualmen-

te, es un indicador de que gran parte de la realidad social escapa a las definiciones que previamente han sido definidas.

Esta selección de aquello que va a describir la realidad social es especialmente importante cuando se trata de variables explicativas. Si las variables en general desagregan parcialmente el mundo, la atribución a una variable de propiedades explicativas implica decisiones sobre la reconstrucción misma de la realidad social. Esta percepción de la potencia de reconstrucción del mundo social es prioritaria en el desarrollo de un análisis de la realidad. Si consideramos que una variable explica, es basándonos en la percepción de que discrimina la realidad. Si la variable género explica una diferencia de opinión, afirmamos que hombres y mujeres muestran una actitud diferente ante la realidad social. Toda capacidad explicativa o descriptiva es esencialmente la afirmación de una diferencia, de que algo es distinto. O no. Así, una variable de control que no explica un fenómeno social pierde su naturaleza de control, ya no explica el mundo, existe algo que debería diferenciar y no diferencia. Precisamente, una cuestión fundamental es reconocer la dependencia de la naturaleza de control atribuible a una variable de las realidades que se le asocian.

La capacidad de definición del mundo no es el único atributo de una variable. Su capacidad más argumentada es su potencia operativa de medición. Como ya sabemos, la vía para cuantificar un concepto teórico es identificar su definición teórica con una definición que lo haga operativo, medible. Para aproximarse a la realidad social, un concepto teórico se apoya sobre variables. ¿Qué es una variable? Una variable es un sistema de categorías o factor que representa estados o valores que puede adoptar una característica sociológicamente relevante (como la edad, los ingresos, el sexo, etc.), y que son mutuamente exclusivas y totalmente inclusivas. El primer elemento en la definición de variable nos dice que ésta es un conjunto de categorías. Por ejemplo, la variable sexo es un conjunto de categorías cuyos elementos son hombres y mujeres. La segunda parte de la definición dice que para que un conjunto de categorías sea una variable, estas deben de ser mutuamente excluyentes y totalmente inclusivas. Esto quiere decir que cada caso sólo podrá ser asignado a una sola categoría, y que deberá haber alguna categoría donde asignar cada caso.

Por lo tanto, a cada caso le corresponde una y sólo una categoría, por ejemplo en la variable "edad" según los años que posea. Es decir, las categorías entre sí son excluyentes. También podemos apreciar que todos los casos encuentran una categoría que cuadre con sus características, o lo que es lo mismo, que el conjunto de categorías incluye a todos los casos. Las variables son una de las claves del conocimiento científico de la realidad, ya sea social o física. Lo esencial es seleccionar aquellas variables que representen lo más

ajustadamente posible conceptos teóricos. Frecuentemente se emplea indistintamente la denominación de concepto teórico y variable para nombrar lo que hemos definido como variable. Esto no es correcto, dado que ambos conceptos implican grados diferentes de abstracción.

*"Needleman no era un hombre fácil de comprender. Su reticencia era tenida por frialdad, pero poseía una gran capacidad de compasión: testigo casual de una horrible catástrofe minera, no pudo concluir una segunda ración de tarta de manzana. Su silencio, por otra parte, enervaba a la gente, pero es que Needleman consideraba el lenguaje oral como un medio de comunicación defectuoso y prefería sostener sus conversaciones, hasta las más íntimas, mediante banderas de señales. (...)*

*Como siempre, cuando murió, Needleman tenía entre manos varias cosas a la vez. Desarrollaba una ética, basada en su teoría de que "el comportamiento bueno y justo no sólo es más moral, sino que puede hacerse por teléfono". Andaba igualmente por la mitad de un nuevo ensayo sobre semántica, donde demostraba (según insistía con particular vehemencia) que la estructura de la frase es innata pero el relincho es adquirido. Y en fin, otro libro más sobre el Holocausto. Éste con figuras recortables. A Needleman le obsesionaba el problema del mal y argüía con singular elocuencia que el auténtico mal es sólo posible cuando quien lo perpetra se llama Blackie o Pete"*

*W. Allen "Recordando a Needleman"*

Los conceptos teóricos rara vez representan un fenómeno social unidimensional. La dimensionalidad se refiere al número de cualidades distintas que son inherentes a un concepto teórico. Por ejemplo, "status social", "alienación", o "prejuicio" son conceptos teóricos que engloban aspectos bastante diferentes. Así, en el concepto "prejuicio" se incorporan aspectos tales como "tendencia a discriminar", "presencia de sentimientos negativos", o "el contenido de estereotipo" que pueda existir en el prejuicio. En general, la caracterización multidimensional de los conceptos teóricos contribuye al desarrollo de la ciencia. Esto, porque al presentar diferentes lecturas de un concepto teórico se favorece la formulación de hipótesis a la vez que se indican posibles vías de acceso a su medición. En el caso de los prejuicios contra los gitanos, se podría medir éste mediante las creencias despectivas que pudiesen existir, o por el rehuir a relacionarse con personas de esta raza, lo que ofrece un mayor repertorio de estrategias de medición.

Un aspecto reconocido por la mayoría de los científicos sociales es el hecho de que es necesaria una formulación más estricta de los conceptos teóricos. En palabras de Blalock: "la cuidadosa reelaboración de las teorías verbales constituye una de las tareas más desafiantes con que nos enfrentamos. La principal tarea de esta empresa consistirá en clasificar conceptos, eliminar o consolidar variables, traducir las teorías verbales existentes a un lenguaje común,

investigar la bibliografía a la búsqueda de proposiciones, y detectar las presunciones que conectan las principales proposiciones con el trabajo teórico serio”. Se precisa, pues, de una tarea previa de formalización que permita a la encuesta muestral tener una mayor eficacia teórica de la que posee actualmente. Este esfuerzo desborda, con mucho, las tareas estrictamente metodológicas. Tal y como lo expresa Lazarsfeld (1951), “¿Cómo nos ponemos a formar tales categorías, en primer lugar? ¿Por qué escoger ciertos elementos de la situación, y no otros? ¿Por qué combinarlos precisamente en estas categorías? Se puede argüir acertadamente que no podemos redactar un conjunto de instrucciones manuales para categorizar los fenómenos sociales: tales instrucciones no serían más que un programa general para desarrollar la teoría social. No podemos escribir un manual sobre cómo formar fecundos conceptos teóricos de la misma manera que escribimos manuales sobre cómo seleccionar muestras o redactar cuestionarios”. Este es un reconocimiento a la evidencia de que el desarrollo de la investigación social sólo puede ser una consecuencia de un desarrollo previo de la teoría sociológica.

La investigación social es siempre investigación de algo, y ese algo debe de estar definido, como ya hemos indicado, desde una matriz teórica previa que identifique los límites lógicos y consecuencias teóricas de aquello que se evidencie empíricamente. Sin embargo, existe un desarrollo técnico de la investigación social que no se relaciona directamente con una eficacia teórica, dado que la teoría sociológica ha padecido siempre de una escasa formalización de conceptos, planteando más imágenes globales que detalles particulares, lo que constituye un obstáculo a su cuantificación. La investigación social es posible actualmente porque el investigador participa de la sociedad que estudia, de modo que posee un conocimiento, en gran parte intuitivo, de las características que son relevantes. Cicourel (1982) lo expresa claramente: “la falta de una teoría social desarrollada obliga a todos los investigadores en sociología a emplear conceptos vulgares que reflejan los conocimientos comunes a los sociólogos y a los miembros “medios” de la comunidad o sociedad. Suponiendo desde el principio que el sociólogo y sus sujetos constituyen una cultura común que cada uno entiende más o menos de la misma manera, los sentidos “obvios” de las preguntas operativizadas del cuestionario en que se basan los indicadores incorporarán propiedades sólo vagamente definidas en la teoría social, pero cuya importancia para el proyecto de investigación se da por supuesta”. Hay que aceptar, en parte, la crítica de Cicourel en el sentido de que se debe mejorar la definición teórica de los conceptos sociológicos, mostrando aquellos aspectos que puedan orientar una cuantificación más precisa. No obstante, el compartir una cultura común es una ventaja indudable y un punto de partida obvio para la cuantificación de conceptos teóricos.

Es decir, está claro que se debe desarrollar la medición de los conceptos teóricos desde la teoría que los genera, pero también es cierto que no nos encontramos en un camino equivocado, sino sólo al comienzo de la vía correcta. Utilizar como punto de partida un conocimiento vulgar, en tanto que socialmente compartido, es lo que permite que la investigación social recoja actualmente una parte, mayor o menor, de la realidad de los fenómenos sociales. Ciertamente este conocimiento no posee una gran trascendencia teórica, pero sí la tiene práctica, tal y como demuestran los gastos cada vez más crecientes en investigación de mercados. Recordemos que el origen de la encuesta no se dio en el ámbito universitario, ni en el de las instituciones científicas, sino en la vida pública moderna, en la necesidad urgente de conocimientos prácticos, demandados por un sistema económico competitivo y un régimen político democrático.

Se hace necesario un nuevo estilo de teoría sociológica capaz de crear nuevas estrategias de medición. Para que esto sea posible hay que hacer operativos los conceptos; es decir expresarlos de modo que sean mensurables.

Una definición operativa consiste en la descripción detallada de los procedimientos utilizados en la investigación para asignar las unidades de análisis a las diferentes categorías de una variable. Esto incluye, en una encuesta, las preguntas efectuadas, qué categorías se ofrecen como respuestas (o en todo caso las reglas empleadas para codificar, si la pregunta es abierta), las instrucciones completas para recoger los datos, etc. Formalmente debe recurrirse a definiciones preexistentes del mismo concepto generadas por otros científicos, de modo que las conclusiones sean equiparables. No obstante, está abierta la vía a generar nuevas definiciones operativas siempre y cuando tengan en consideración las ya existentes.

En ese sentido, la interpretación de un concepto teórico debe considerar como una de las características más destacables de la investigación científica el estar basada en el empirismo.

Los conceptos teóricos, con los que pensamos el mundo, se plasman desde la perspectiva empírica en un conjunto de variables o indicadores, no necesariamente percibibles directamente. Un modelo de covarianzas plantea, esencialmente, que existe un sistema de relaciones de interdependencia entre un conjunto de variables. En ese sentido los conceptos importantes para el análisis estructural están articulados entre sí. La noción de explicación conlleva la distinción de tipos de variables, así como de las relaciones entre ellas. La complejidad que adopten las relaciones nos lleva a la idea de sistemas y tipos de sistemas. Los modos que empleemos para hablar de dichos sistemas de relaciones nos conduce a las notaciones alternativas para poder expresarlos.

Una de las primeras consecuencias del análisis estructural es el refinamiento conceptual de las variables. Estas poseerán una cualidad explicativa especial dependiendo de su posición estructural. El análisis no se limita a las variables manifiestas sino que incluye las variables latentes. De este modo, un modelo puede combinar simultáneamente variables manifiestas y variables latentes, que vendrán definidas por un conjunto de variables identificadoras de la variable latente. Por una parte, en función a su visibilidad encontramos las variables manifiestas, las variables latentes y las variables indicadores que actúan de forma instrumental para determinar los efectos de las variables latentes. Por otra parte, si atendemos al *criterio explicativo*, la principal distinción atañe a las variables que se intenta explicar y aquellas otras que no. Tendremos por lo tanto variables que explican y variables que son explicadas. Estos dos conjuntos de variables no son disjuntos. Esto implica que tendremos variables que explican y a su vez son explicadas.

Las variables exógenas son variables que sólo aportan explicación, pero que ellas mismas no son explicadas dentro del modelo que se propone. En el orden estructural implican factores de explicación no explicada. Las variables endógenas intervinientes son aquellas que explican y son explicadas dentro del modelo explicativo que estamos desarrollando. Por último, las variables endógenas no intervinientes son aquellas que vienen explicadas dentro del modelo pero no contribuyen a explicar ninguna otra variable.

Realmente, un modelo estructural consiste en narrar una historia donde se explica algo. Nuestro repertorio de personajes está limitado a los que consideramos dentro del modelo, sus acciones se explicarán entre sí. Habrá personajes (variables) que actúan dentro la trama, pero no sabemos qué les dirige o explica. Habrá personajes cuyos actos están explicados por las acciones de otros y que a su vez influyen en unos terceros. Por último, estarán aquellos que, dentro del ámbito de la historia que intentamos contar, son explicados al máximo en sus actuaciones pero cuyo impacto en los demás no es tenido en cuenta, por esta vez, en la historia que se está contando.

En ese sentido, un modelo estructural no deja de ser una narración donde se intenta explicar el porqué de algo con un número limitado de personajes (variables exógenas, endógenas intervinientes o endógenas). La explicación nunca es completa, ni del conjunto ni de cada personaje endógeno, por lo que restan residuales entre la explicación y los comportamientos que se detectan (varianza de las variables). La lógica explicativa indica que a más residuales peor explicación, y por tanto, a menos residuales, es decir, cuanto mejor se puedan predecir los cambios en el comportamiento de los actores-variables, mejor modelo. Podemos pues diferenciar entre descripción, desarrollada me-

diante la introducción de variables intervinientes que refinan el modelo y lo hace más detallado, y explicación, que dependerá del grado de ajuste con la realidad de las relaciones que encontramos.

Una noción importante es la de variable común. Esta variable común (A) puede explicar dos variables relacionadas entre sí (BC) sin disminuir su relación. En este caso la variable común valida indirectamente la relación entre B y C. Pero también una variable común a otras puede explicarlas simultáneamente. Consideremos un ejemplo intuitivo: dos personas que se conocen y se relacionan (B y C). Ambas son inicialmente amigas de otra tercera persona (A) que los presentó, y que casi siempre está cuando los otros dos se encuentran. Puede plantearse una relación de orden entre los dos primeros, pero lo evidente es que el conocido común explica por qué se reúnen. El número de reuniones entre ambos puede ser elevada (B,C), pero cuando se tiene en cuenta cuántas reuniones se realizan sin la presencia del amigo común (A), este número deviene insignificante. Es decir, la reunión (covariación) entre B y C es dependiente de A, y si se controla por ella, la relación entre B y C puede devenir “espuria”. Una relación espuria se revela cuando la asociación entre dos variables viene explicada por otra común a las dos. Así, la noción de causa común es aquella variable explicativa que siendo previa a otras dos que covarían, explica o no la covariación entre éstas. Como veremos, la noción de causa común es muy importante para falsar las hipótesis estructurales.

En resumen, introduciendo el máximo de causas comunes (sean exógenas o endógenas) incrementamos la capacidad explicativa del modelo y su robustez a la falsación, mientras que la introducción de variables endógenas intervinientes no comunes refina el análisis, haciéndolo más descriptivo.

Una vez que poseemos un lenguaje para expresar ideas y explicaciones, el siguiente paso es contar una historia e intentar compararla con los datos que aporta la realidad (tal y como ha sido seleccionada y definida). Esta fase se denomina especificación (contar una historia) y ajuste (comparar lo que decimos con los datos de que disponemos). Evidentemente, ninguna explicación es completa ni definitiva. No obstante, la existencia de una realidad (un proceso, una historia, acontecimientos) a explicar se impone inexcusable. En ese sentido, la tarea del modelado, que pudiera parecer absurda, adquiere su verdadero sentido, deviene reveladora en la medida que exige una explicación al mundo, tanto de lo que creemos saber como de lo que no sabemos.

Pasemos pues a introducirnos en el vocabulario y notaciones que nos forzará a reconstruir lo que incluso damos por evidente.

### 3.2. MEDICIÓN

El proceso de medición ha sido definido de modos muy diversos. En el nivel más simple de medición, la regla consiste en adjudicar cifras<sup>13</sup> que sirven como etiquetas de categorías mutuamente exclusivas. En un ámbito más complejo, la regla genera cifras que pueden ser tratadas como números auténticos, en un sentido matemático, y sobre los que es posible efectuar todas las operaciones propias de los números reales (R). Es un error pensar que una vez reconocida la necesidad de medir puede generarse una asignación numérica como si todos los procedimientos fuesen aceptables. Es necesario mantener una correspondencia entre las propiedades del sistema de cifrado y el de los fenómenos sociales que se van a representar. Tal y como lo expresara Ackoff, la medición es "una forma de obtener símbolos para representar las propiedades de los objetos, sucesos o estados; símbolos que tienen entre sí la misma importante relación que las cosas que representan". Así pues, en la medición lo que nos interesa es la correspondencia entre ciertas entidades empíricas y un modelo formal (generalmente un sistema numérico, aun cuando no necesariamente) en términos de las relaciones que, presuntamente, existen entre los elementos de cada sistema.

Torgerson desarrolla acertadamente un último aspecto importante: "La medición atañe a las propiedades de los objetos, no a los mismos objetos. Así, en nuestro uso del término, no es mensurable un palo, aunque sí podrían serlo su longitud, peso, diámetro y dureza... Medir una propiedad implica, pues, atribuir números a sistemas para representarla. Y, para ello, ha de prevalecer un isomorfismo, es decir una relación o correspondencia exacta, entre ciertas características del sistema numérico implicado y las relaciones entre diversas cantidades de la propiedad por medir. La esencia de este procedimiento es la atribución de números de tal manera que se refleje esta correspondencia precisa entre dichas características de los números y las correspondientes relaciones entre las cantidades". Aquí encontramos varias ideas clave: en primer lugar, no se miden los individuos sino determinadas características que se consideran relevantes (actitudes, comportamientos, opiniones, etc.). Es decir, la unidad de análisis es el individuo, pero éste no es considerado en su integridad; se cuantificarán aquellas características del individuo que son teórica y sociológicamente relevantes: las que se han definido como variables. Segundo, debe de existir, como ya sabemos, una relación "exacta" entre las propiedades relacionales del sistema de cifrado que se utilice y las

---

13. Una cifra puede entenderse como guarismo, y por lo tanto notación simple, o como número con sus propiedades cuantificadoras y relacionales correspondientes. Así, la atribución de cifras a las diferentes categorías de una variable puede poseer o no propiedades numéricas.

características relacionales de las cantidades del concepto que se pretende medir. Aceptar y desarrollar esa evidencia es esencial a las posibilidades de medición que puedan tener los conceptos teóricos y, por lo tanto, al proceder científico de la sociología. Por ejemplo, la característica sexo no admite ser sumada, ni restada, ni dividida, ni multiplicada; esta evidencia implica que el sistema de cifrado que se emplee para medir dicha característica no puede admitir dichas operaciones. La razón es simple, las operaciones que efectuemos sobre los "cifras" que representan las características del concepto que se pretende medir no pueden violentar la realidad. Si a la categoría "Hombre" le asignamos un 1 y a la categoría "Mujer" un 2, el tratamiento o tratamientos posibles para ese 1 o 2 no pueden ignorar aquello que representan, un sistema de clasificación. Podemos dividir 1 entre 2 y nos resultará un valor, 0,5, pero, ¿qué significado posible tiene esa cantidad en términos de aquello que pretendemos medir? 0,5 en el concepto teórico sexo no significa nada. Del mismo modo, la asignación de representaciones numéricas incorrectas a un concepto teórico implica que los hallazgos o conclusiones a que se llegue mediante el análisis son incorrectos y carentes de significado.

*"Me hallaba yo acostado en mi casa de la playa, pero no podía dormir pensando en que se me antojaba una pechuga de pollo que había en la nevera. Esperé a que mi mujer se quedase transpuesta, y fui de puntillas a la cocina. Eran las cuatro y cuarto en punto. Estoy completamente seguro, porque el reloj de la cocina no funciona desde hace veintiún años y marca siempre esa hora."*

W. Allen "La amenaza O.V.N.I."

El problema de la medición supone una dificultad diferente según las características del fenómeno social que se desea cuantificar. Para la encuesta, Sudman y Bradburn (1974) distinguen a grandes rasgos entre dos tipos de respuestas: de comportamiento y psicológicas. Las primeras son, o pueden ser, verificables (ingresos, edad, voto, etc.) mientras que las segundas sólo pueden ser inferidas indirectamente. La actitud de un individuo hacia un partido político es una referencia mucho más vaga que el voto en las últimas elecciones. Son más fáciles de cuantificar las variables de comportamiento, dado que los individuos mantienen una relación objetiva con ellas, que aquellas otras referidas a estados emocionales o subjetivos. Cicourel (1964) nos recuerda claramente el problema "Al suponer que las variables estructurales o de actitud son cuantificables automáticamente, obligamos a los conceptos a tomar la apariencia de precisión, de manera que puedan dividirse en dicotomías, tricotomías, series ordinales, intervalos y distancias métricas. Pero el concepto no es *per se* cuantitativo; sólo llega a serlo cuando lo situamos dentro de

cierta textura teórica que origine explícitamente dicotomías significativas, tricotomías, relaciones ordinales e intervalos que se suponen iguales y distancias con rasgos métricos. La noción de "variable" puede significar una colección no aditiva de elementos que caracterizan cierto rasgo del mundo del actor, definido culturalmente. La "variable" no constituiría un continuo unitario, diferenciable, ni aun una dicotomía, forzosamente, a menos que la teoría lo exija y lo justifique específicamente. Toda comprensión de las operaciones de cifrado, en cuanto relacionadas con la estructuración de los cuestionarios y de los programas de entrevista, debe tener en cuenta lo que ofrece el conocimiento vulgar del mundo que compartimos con el entrevistado, nuestra teoría sociológica y lo impuesto por los recursos de medida". Las operaciones que puedan efectuarse sobre los resultados de una medición deben de mantener una relación estricta con la realidad del fenómeno social que se pretende medir. Por ello, Torgeson habla de una relación de isomorfismo<sup>14</sup> entre el sistema numérico y el sistema empírico que es objeto de medición. Asimismo, también se isonómia en lo que se refiere a las leyes combinatorias, combinaciones y relaciones que son compatibles en ambos conjuntos.

*"El mundo material –siguió Dupin– es abundante en analogías exactas con el inmaterial, y eso da un tinte de verdad a ese dogma retórico de que la metáfora o el símil pueden servir para fortalecer un argumento o para embellecer una descripción".*

*E.A. Poe, La carta robada*

En ese sentido, medir supone la concurrencia de tres elementos: un dominio teórico a modo de código, un sistema relacional empírico y un sistema relacional numérico que le es particular. Esto es, una regla que ordene y exprese los datos en números en función del sistema empírico y del sistema teórico. Así, actuaremos sobre el sistema numérico y observaremos los resultados, intentando comprender mejor el sistema empírico. Esto implica que si el sistema numérico refleja correctamente el sistema empírico (isomorfismo) podremos tener la oportunidad de profundizar en el conocimiento de los fenómenos sociales. Si la medición es mala y no existe una correspondencia correcta entre sistemas, habremos perdido la referencia empírica de la investigación y los resultados carecerán de sentido. No obstante, este isomorfismo entre el sistema numérico y el empírico requiere un control teórico. Como ya se ha indicado, no hay medición sin teoría previa. De hecho, cada medición constituye por sí misma una teoría en pequeña escala. Como afirma

14. Para que exista isomorfismo entre los dos sistemas debe existir una aplicación biyectiva entre ambos sistemas (conjuntos). Esto implica que las relaciones deben ser aplicaciones inyectivas y sobreyectivas (exhaustivas).

Coombs "Nuestras conclusiones, incluso en el nivel de medición, son ya una consecuencia de la teoría. Un modelo de medición es realmente una teoría sobre el comportamiento, ciertamente en un nivel de miniatura, pero no obstante teoría". Más formalmente, Blalock denomina a las teorías que arrojan la operativización de los conceptos teóricos "teorías auxiliares". En general, la operativización y medición precisan de un respaldo teórico que dé cuenta de aquellos procedimientos que se efectúan en el proceso de medición.

*"Fantasma. Persona que se ha desvanecido hasta hacerse impalpable por muerte, ausencia o cambio de costumbres".*

*James Joyce "Ulyses"*

### 3.3. LOS NIVELES DE MEDICIÓN

Existen muchos niveles de medición, en la medida que cada uno de ellos refleja un conjunto de propiedades en el sistema de cifrado adoptado. Así, los economistas acostumbran a utilizar con frecuencia un nivel denominado cardinal. Para la sociología, en general, se admite que la medición de los fenómenos sociales puede efectuarse en cuatro formas sin perder información sustancial con respecto a las propiedades relacionales del sistema empírico.

- a) nominal
- b) ordinal
- c) intervalos
- d) razón o proporciones

Cada uno de estos grados de medición se corresponde con diferentes fenómenos sociales, debiéndose elegir aquél que sea más adecuado.

**La medición nominal.** Un nivel de medición nominal implica básicamente determinar la ausencia o presencia de una característica, también denominada cualidad. En otras palabras, se trata de clasificar a los individuos en virtud de una determinada característica. El nivel de medición nominal ofrece apenas la información más indispensable, ya que los "números" sirven sólo como "nombres" para identificar las propiedades que se estudian. A todos los miembros de una clase se les asigna el mismo número o nombre, pero jamás se da el mismo a dos clases distintas. Para utilizar un nivel de medición nominal basta que seamos capaces de distinguir si dos elementos son "iguales" o si son "distintos" con respecto a determinadas propiedades. Carece de sentido añadir números e incluso ordenarlos. Así los valores de "varón" y "hembra", constituyen un tipo de medición nominal, según el cual se clasifican los individuos por su género.

En qué grado la medición nominal puede discriminar entre los individuos depende de la intención del investigador. Por ejemplo, en España el número

del documento nacional de identidad no posee capacidad numérica ni refleja ningún orden. Es una medición nominal que identifica a cada ciudadano, es decir, posee capacidad de discriminar hasta un nivel individual. El grado de clasificación que puede desarrollar una medición nominal depende del investigador, que decide el tipo de agregación que necesita.

Estadísticamente, la medida de tendencia central que corresponde con un nivel de medición nominal es la moda. Existe una predisposición a considerar la medición nominal como inferior a los otros tipos de medición. Es este un error importante, pues las limitaciones que pueda encontrar una medición nominal en el tratamiento estadístico es una cuestión secundaria ante la evidencia de que muchos fenómenos sociales son eminentemente nominales, y como tales hay que tratarlos.

**La medición ordinal.** Un nivel de medición que además de clasificar ordena, es el denominado ordinal. La medición ordinal requiere que los objetos se puedan ordenar por categorías, de acuerdo a algún atributo; señala, pues, una jerarquía pero sin postular una unidad fija de medida por lo que carece de sentido añadir grados o hablar, por ejemplo, de que uno posee el doble que otro de una determinada característica. Esta medición implica determinar un orden de intensidad para una cualidad, es decir, que el fenómeno en estudio está ordenado, por ejemplo, "mucho", "bastante", "poco", "nada". Este gradiente sólo indica la relación de una respuesta con otra; "mucho" es más que "bastante", o "poco" más que "nada", pero no expresa qué distancia puede existir entre las diferentes respuestas. No tiene sentido preguntarse cuántas veces "mucho" es más que "nada". La medida de tendencia central que es factible utilizar con el nivel de medición ordinal es la mediana, además de la moda. Como vemos, conforme la medición es más refinada incorpora las características de las anteriores. Así, al existir un orden existe, a su vez, clasificación.

**La medición por intervalos.** El nivel de medición por intervalos implica un continuo compuesto por intervalos igualmente espaciados; es decir, que las distancias numéricamente iguales en el instrumento de medición representan distancias iguales en la propiedad que se mide. Aquí se introduce una unidad constante de medida. Tiene pues sentido el uso de la adición o sustracción entre los valores, pero sin poder afirmar aún que un objeto posea una cierta característica en una cuantía doble a otro, dado que no se postula la existencia de un cero natural. En otras palabras, en este continuo no existe un valor que signifique ausencia de la cualidad, es decir, un 0 absoluto. Por ello, podemos apreciar las diferencias entre dos mediciones, pero no sabemos cuánta cualidad implica esa diferencia. Un coeficiente de inteligencia de 110 es 10 puntos mayor que uno de 100, y esa distancia de 10 es igual a la existente entre

una diferencia entre 115 y 105, dado que los intervalos son equivalentes entre sí. Al ser los intervalos iguales, son adecuados para poner en relación una medición respecto de otra. La medida estadística de tendencia central para un nivel de medición por intervalos es la media (conjuntamente con la moda y la mediana).

El grado de refinamiento que alcance el nivel de medición por intervalos debe estar ajustado con la realidad social del fenómeno que se está considerando. Por ejemplo, no por plantear que el 1 es extrema izquierda y 10.000 extrema derecha, vamos a conseguir una cuantificación más exacta del fenómeno. Mucho más prudente sería entre 1 y 7 ó 1 a 10. Como ya indicamos, la formulación de una medición debe de ajustarse sobre las características de aquello que se desea cuantificar.

**La medición de razón o proporciones.** El nivel de medición que permite un mayor grado de operaciones estadísticas es el de razón o proporciones, consistente en un continuo compuesto por intervalos igualmente espaciados donde existe el valor 0, que representa la ausencia de cualidad. Quiere esto decir que el 0 posee un significado. Cero grados centígrados no representan ausencia de temperatura. Indica a lo sumo el punto térmico de transición entre el estado líquido y el sólido en el agua, pero no ausencia absoluta de "calor". La escala Kelvin sí posee un nivel de medición de razón o proporción. Cero grados Kelvin (k) indica ausencia de calor, no hay movimiento molecular. Al existir un cero absoluto, puede decirse en propiedad que 5000 K es el doble de temperatura que 2500 k. Evidentemente en la escala Kelvin no existen signos positivos o negativos al no haber la posibilidad de estar bajo cero Kelvin. Este nivel de medición permite realizar todas las operaciones aritméticas con los valores obtenidos al incorporar las características de los niveles anteriores, añadiendo a su vez la existencia de un cero como origen natural. Al existir una referencia absoluta, es posible evaluar los incrementos que supone cada unidad sobre la anterior. Aquellos fenómenos sociales cuyo nivel natural de medición es el de proporciones son factibles de ser tratados con el mayor número de procedimientos estadísticos. Esto anterior ha sido especialmente cierto durante bastante tiempo, si bien en la actualidad se van desarrollando nuevos procedimientos que potenciarán el análisis de los otros niveles de medición considerados. Es una cuestión de tiempo el que los sociólogos (gracias al trabajo de los profesionales de la matemática y la estadística) pierdan su complejo por los niveles de medición que su objeto de estudio les permite emplear más frecuentemente: nominal u ordinal.

Como ya sabemos, no debe de primar la consideración estadística sobre la lógica de la realidad social. El nivel de medición que se adopte debe de ser el más adecuado al concepto teórico que se pretende cuantificar, aun

cuando quepa la opción de emplear otro con mayores posibilidades estadísticas. Así, los ingresos familiares pueden considerarse como una variable de razón. Sin embargo, si el concepto teórico que deseamos medir es el de "estatus social" resulta evidente que unos ingresos mensuales de 0 no reflejan un estatus social de 0, por lo que fijarle el nivel de medición en "razón" no es lo más apropiado. Tampoco sería correcto el nivel de medición interval, dado que no existe igualdad de intervalos. Un incremento de 0 a 1.000.000 supone un incremento de estatus mucho más sustancial que pasar de 100.000.000 a 101.000.000. Por todo esto, si los ingresos se van a emplear para medir "estatus social" deberían de hacerlo en un nivel ordinal, que es el que mejor refleja la naturaleza social del estatus. Esto no obsta para que los ingresos, cuando miden otro concepto teórico diferente al de estatus social, puedan adoptar un nivel de medición interval o de razón.

Este ejemplo anterior recoge la evidencia de que la realidad social requiere el nivel de medición que le corresponda adecuadamente; ello implica que, aún cuando exista la posibilidad de un mayor tratamiento estadístico, el nivel de medición debe de ajustarse a la naturaleza del fenómeno que representa. La clave de una buena cuantificación no consiste en emplear aquel nivel de medición que permite más operaciones matemáticas, sino el que es más apropiado a las características de aquello que se desea medir.

*"Chimista, que era autodidacto y que sabía de todo, le explicó algunas cosas que el muchacho no entendía bien. Luego, para embromarle, le hacía preguntas capciosas de pronto.*

*—A ver, Iturriza, ¿cuántos kilogramos de distancia hay desde Manila a Hong-Kong? —le preguntaba.*

*—No sé... ¿Habrán quinientos? ¿Habrán ochocientos?*

*—No seas tonto... Kilogramos es una medida de peso.*

*—Tiene usted razón. A ver, pregúnteme usted otra cosa.*

*—Si un chino come con los palillos cuatro granos de arroz por segundo y cada cubo tiene una cabida de cinco libras, ¿cuánto tardará en comerse el cubo?*

*— Pues tardará... a ver... a ver...; dice usted que...*

*—No, no hay manera de averiguarlo. Es un problema mal planteado, porque no se dice lo que pesa cada grano de arroz.*

*—Tiene usted razón. Es verdad.*

*—Es lo primero que hay que ver en todo problema. Si está bien planteado, porque si no lo está, todo esfuerzo es cosa perdida.*

*—Hazle caso a éste —le decía yo a Iturriza—, que sabe lo que dice.*

*—Ya lo veo —contestaba él.*

*Pío Baroja. La estrella del capitán Chimista*

### 3.4. ERROR DE MEDICIÓN

Un proceso de medición difícilmente puede obtener la cuantificación precisa de un fenómeno social sin que incorpore algún grado de error. En una encuesta, como ya sabemos, se recogen las respuestas a diferentes preguntas, como por ejemplo la edad del entrevistado. Lo que obtenemos son, no obstante, respuestas, es decir, valores observados de la variable. La cuestión es qué relación existe entre lo que se recoge y los valores reales de la variable que han constituido la base de referencia para la respuesta. En el caso de la edad, lo habitual es redondear en el último cumpleaños, despreciando los meses. Tendremos pues, un valor observado (en años) y un valor real de la variable (años y meses). La diferencia entre el valor real de la variable y el valor observado se define como un error de medición. En el caso de la variable "edad" la situación es bastante simple, donde la respuesta (R) del entrevistado, o valor observado ( $X_R$ ) es una función lineal del valor real de la variable ( $\tau_R$ ) y el error de medición ( $\varepsilon_R$ ). Esta combinación lineal se expresaría para un individuo como

$$X_R = \tau_R + \varepsilon_R$$

de donde el error de medición es definido como

$$\varepsilon_R = X_R - \tau_R$$

En el caso general, considerando la respuesta de todos los entrevistados en esa medición (es decir, preguntas) la función lineal sería

$$X = \tau + \varepsilon$$

Esto anterior implica que en un proceso de medición podemos considerar dos factores que explican la variación que se recoge al aplicar una definición operativa.

(1) La primera fuente de variación corresponde con las diferencias reales que se pretenden medir ( $\tau_R$ ). Si se obtiene la autodefinición ideológica de los individuos se apreciarán diferencias entre ellos, donde unos son más de izquierdas y otros más de derechas. Esta variación da cuenta de aquello que queremos medir, y constituye lo que es la medición en sí.

(2) Sin embargo, parte de la variación que se detecta en las respuestas que se obtienen es inherente al hecho de emplear una determinada definición operativa o un instrumento de medición y no otro ( $\varepsilon_R$ ). Por ejemplo, la elección de aplicar una escala de autoubicación ideológica que recorra de 1 a 7 posiciones o de 1 a 10 puede afectar a las respuestas. En esta variación interviene así mismo, la introducción de errores de tipo aleatorio o fortuito, que cabe esperar tiendan a compensarse unos con otros.

Como podemos apreciar, lo que se denomina error de medición está constituido por dos fuentes de variación diferentes. Estos errores adquieren dos formas: sistemáticos y aleatorios.

### *Tipos de error*

Un error sistemático es aquel que introduce un sesgo que se produce en un mismo sentido. Es el caso ya citado de la edad, donde se menciona siempre el último cumpleaños, aun cuando éste sucediese hace once meses. En este caso se produce un error sistemático que reduce en meses todas las edades, por lo que el sesgo está orientado en la misma dirección. Con carácter más general, puede aparecer un aspecto reactivo del instrumento de medición. Por ejemplo, en el caso de las entrevistas "cara a cara", se produce un sesgo por efecto de la interacción. Pueden aparecer otros, tales como una pregunta cuya redacción está sesgada, o que se responda de acuerdo a lo que es deseable socialmente. Este tipo de error es sistemático, dado que los sesgos están orientados en un mismo sentido.

El error aleatorio es un tipo de error que no está ligado al concepto que se quiere medir, y que aparece puntualmente sesgado en cualquier dirección. Por ejemplo, un entrevistado que no presta atención, o que está cansado. La redacción de una pregunta puede generar errores aleatorios en la medida que sea ambigua y genere diferentes interpretaciones, especialmente en las preguntas abiertas.

Estos tipos de error, sistemático y aleatorio, mantienen una relación estrecha con los dos conceptos clave para evaluar un instrumento de medición: la fiabilidad y la validez.

### *Fiabilidad y validez de una medición*

**Fiabilidad.** El concepto de fiabilidad expresa la estabilidad o consistencia de una definición operativa, es decir, muestra la consistencia y reproductibilidad de una medición. La reproductibilidad viene dada cuando una misma pregunta repetida en diferentes aplicaciones produce mediciones semejantes. La consistencia significa que una pregunta, planteada en otros términos dentro de un mismo cuestionario, debe de ofrecer la misma medición. La fiabilidad examina, en definitiva, en qué grado la variación observada en una medición  $X$  ( $\sigma_x^2$ ) se debe a posibles errores. Por ello, la fiabilidad<sup>15</sup> de  $X$ , que nota-

---

15. Se puede demostrar, y constituye una formulación alternativa de la fiabilidad, que ésta es así mismo:

$$P_x = 1 - (\sigma_\varepsilon^2 / \sigma_x^2)$$

remos como  $P_x$ , es simplemente la razón entre la variación real de la variable ( $\sigma^2_\tau$ ) y la variación observada:

$$P_x = \sigma^2_\tau / \sigma^2_x$$

Lógicamente, la fiabilidad de una medición, expresada en estos términos, es un coeficiente que oscila entre 0 y 1, donde 1 es muy fiable ( $\sigma^2_\tau = \sigma^2_x$ ) y 0 nada fiable. Despejando en la fórmula anterior tendríamos:

$$\sigma^2_\tau = \sigma^2_x P_x$$

Es decir, la variación real de la variable X es igual a la variación observada multiplicada por el coeficiente de fiabilidad.

En lo que se refiere a los errores, el error sistemático es transparente a los test de fiabilidad, dado que este tipo de error se reproduce en términos semejantes en todas las mediciones. Generalmente, la medición de la fiabilidad se plantea mediante dos estrategias principales, las mediciones de estabilidad y las mediciones de equivalencia. Las mediciones de fiabilidad mediante la evaluación de estabilidad se basan en correlacionar las mediciones a lo largo del tiempo. El procedimiento más conocido es el denominado "test-retest" que traduciremos por "doble test". Se asume que la medición que produzca un ítem o escala (x) estará correlacionado consigo misma en las mediciones que ofrezca en diferentes tiempos ( $t_1, t_2$ ), debido a que recoge la variación real de la variable:

$$X_{t1} = \tau + \varepsilon_{t1} \quad X_{t2} = \tau + \varepsilon_{t2}$$

Si consideramos que  $\sigma^2\varepsilon_{t1} = \sigma^2\varepsilon_{t2}$  y que los errores no están asociados  $C(\varepsilon_{t1}, \varepsilon_{t2}) = 0$ , entonces:

$$P_{X_{t1}X_{t2}} = P_x$$

**Validez.** La validez, considera la congruencia o bondad del ajuste entre una definición operativa y el concepto que pretende medir. Se trata, pues, de analizar si el instrumento de medición mide realmente lo que quiere medir. Podemos pensar procedimientos para comprobar la validez desde dos perspectivas diferentes: la validez teórica y la validez empírica. La correlación entre  $\tau$  y  $x$ ,  $P_{tx}$ , es la denominada validez teórica, y mide la correlación de la variación en las respuestas observadas (x) con el valor real de la variable ( $\tau$ ). Por otro lado, si la correlación es entre (x) y otro ítem observado del mismo concepto (y),  $P_{xy}$  se denomina validez empírica. Evidentemente, la comprobación de la validez empírica requiere más de una variable observada. Lord (1968) relaciona los conceptos de validez y fiabilidad demostrando que:

$$P_{xy} \leq P_{tx} = \sqrt{P_x}$$

Es decir, la validez empírica ( $P_{xy}$ ) de una medida (x), con relación a otra segunda medida (y), no puede exceder la validez teórica ( $P_{tx}$ ) o la raíz cuadrada de su fiabilidad ( $\sqrt{P_x}$ ). De ello se deduce que ninguna medida puede ser válida sin ser también, a la vez, fiable. Sin embargo, una medida puede ser fiable sin ser válida. Supongamos un termómetro de los utilizados para medir la temperatura corporal, que se encuentra obturado a 37°C. Si un individuo que cree sentirse enfermo lo emplea, marcará 37°C. Si lo empleara repetidas veces, marcaría siempre 37°C por más fiebre que tuviese. El termómetro es fiable, pues ofrece siempre una medida consistente, pero no es válido para medir la fiebre. Solamente sería válido en este caso para indicar a qué altura se ha obturado el termómetro. Por consiguiente, se trata de un instrumento de medición que no mide lo que debe medir (la temperatura corporal), sino que mide otro concepto diferente (dónde se obturó el instrumento). Como conclusión, un instrumento de medición puede ser fiable y no tener validez. Óptimamente, deberá de poseer las dos cualidades de fiabilidad y validez.

En general, la validez protege contra los posibles errores sistemáticos, dado que evalúa la relación entre lo que se quiere medir y lo que se mide realmente. Sin embargo, los errores que pudieran suceder de forma aleatoria no son detectables mediante las pruebas de validez, dado que estos errores no se repiten regularmente. El error aleatorio es, no obstante, detectable en las pruebas de fiabilidad. Tanto la fiabilidad con la validez son dos aspectos esenciales en la investigación que deben de considerarse especialmente cada vez que se genera un nuevo instrumento de medición.

### *Correlación epistémica*

La relación entre un concepto teórico y la medición que se obtiene es denominada "correlación epistémica". La correlación epistémica indica la relación teórica entre el componente correcto de una medición y el concepto que desea representar.

*"Un hombre del Reino de Chu cruzaba un río cuando se le cayó la espada al agua. Al momento hizo una marca en el costado de la barquilla.*

*—Aquí fue donde cayó mi espada —dijo.*

*Cuando la barca atracó se sumergió en el agua para buscar su espada, bajo el punto que había marcado, pero como la barca se había movido y la espada no, el método para localizar la espada resultó ineficaz."*

*Discursos de Llui Bu-wei<sup>16</sup>*

---

16. Obra atribuida a Llui Bu-wei (?-235 antes de Nuestra Era), una parte de la cual fue escrita por sus protegidos.

La correlación epistémica no es cuantificable, sino que se determina lógicamente, y expresa el contenido de validez de una medición. Así, los ingresos mensuales poseen una correlación epistémica alta con el concepto teórico "status social", mientras que el número de esposas, en occidente, mantiene una correlación epistémica baja con ese mismo concepto teórico. Tal y como pone de relieve el ejemplo, la correlación epistémica, así como la validez de un instrumento de medición, no es universalizable. No se pueden generar operativizaciones cross-culturales estándar, dado que no serían correctas para alguna realidad social que se pretenda medir. Esto último es algo a tener muy en consideración al replicar investigaciones en los estudios cross-culturales. Lo fundamental, en cada caso, es que los fenómenos sociales que se pretende medir se correspondan del modo más preciso con el nivel de medición adoptado y se consiga, así mismo, una correlación epistémica alta entre el concepto teórico y la realidad social que representa.

### 3.5. CIFRADOS Y TRANSFORMACIONES

Como sabemos, un sistema de cifrado con propiedades numéricas intenta mantener una relación de isomorfismo con el sistema empírico, así como ser isonómico en las operaciones que son permitidas en el sistema de cifrado con las operaciones que son factibles en el sistema empírico. La idea que existe detrás de estas restricciones es que al simular un comportamiento del sistema empírico lo hacemos operando sobre el sistema de cifrado, y éste no debe ir más allá en sus desarrollos y combinaciones de lo que lo puede hacer el sistema empírico. Así, operando en el sistema de cifrado "operamos" sobre el sistema empírico.

En ese sentido, respetados los criterios de isomorfismo y de isonomía entre ambos sistemas, puede adoptarse cualquier sistema de cifrado. Es decir, no existe ninguna naturalidad en la relación entre sistemas de cifrado determinados y sistemas empíricos.

*"El mundo material –siguió Dupin– es abundante en analogías exactas con el inmaterial, y eso da un tinte de verdad a ese dogma retórico de que la metáfora o el símil pueden servir para fortalecer un argumento o para embellecer una descripción".*

*E.A. Poe, La carta robada*

Una cuestión diferente proviene de lo que se denominan transformaciones. Éstas no suponen sustituir un sistema de cifrado por otro que posea las mismas propiedades, sino que transforma el sistema de cifrado de tal modo que

permita un análisis más fácil de la estructura presente en los datos (en definitiva, esa estructura existe y se revelaría a la luz de un sistema de cifrado y no a otro, permaneciendo independientemente de cómo se reexpresa el sistema de cifrado). Es decir, se pueden considerar las transformaciones en el sistema de cifrado como aceptables en tanto que la operación matemática ayude al sistema de cifrado a revelar estructuras que están distorsionadas por la rigidez de la representación numérica. Así, transformar un sistema de cifrado tomando logaritmos ofrece un sistema nuevo de cifrado logarítmico donde el cero no existe y lo que antes era multiplicación " $a*b$ " deviene en adición  $\log(a*b) = \log(a)+\log(b)$ ; es decir un proceso multiplicativo se convierte en aditivo. Puede, en algunas circunstancias, resultar más operativo matemáticamente ajustar un modelo aditivo a uno multiplicativo. ¿Es la realidad aditiva o multiplicativa? Dependerá del sistema de cifrado, pues una misma relación alcanzará valores equivalentes expresadas en un sistema u otro.

La utilidad de transformar no es estética, sino hacer el análisis más fácil, permitir ajustar modelos más simples y por tanto mejorar nuestra capacidad para entender los datos. No obstante, en esa operación pueden perderse aspectos importantes y siempre debe mantenerse la posibilidad de restituir los datos a su sistema de cifrado original, de modo que los datos tengan una interpretación inmediata en términos accesibles y facilite discernir los posibles absurdos a que puedan conducir unas operaciones efectuadas mediante transformaciones inapropiadas.

Porque no sólo interesan los beneficios matemáticos, algunas veces pueden interesar más los beneficios teóricos. Así, no es lo mismo transformar los ingresos expresados en número de euros, que en log de euros o en euros al cuadrado. ¿Cuál expresaría mejor la realidad asimétrica del fenómeno? Simetrizar una distribución puede contribuir a ocultar una realidad aunque facilite el ajuste matemático de un modelo; quizás la mejor transformación sea tomar potencias elevando al cuadrado acentuando el efecto no lineal de ganar más dinero. Así pues, no sólo deben interesarnos las transformaciones que simplifican el análisis, sino también aquellas que lo expresan más adecuadamente. En el sentido de alejarse de la realidad, las transformaciones afectan a diferentes propiedades (magnitud, distancia y orden). En ese sentido, el sacrificio de la realidad a su analista está asociado al grado en que el cifrado nuevo afecte a la realidad. Es evidente que sólo afectar a la cantidad del valor (lineales) y no a la distancia o al orden es menos grave que aquellas transformaciones que afectan al orden de los casos.

### 3.6. EXPLICACIÓN Y CAUSALIDAD

*"Los hombres aprenden de otros hombres lo que saben de sí mismos, del mundo en el que deben vivir y del mundo en el que desearían vivir"*

*Harlan Ellison, Dangerous Visions*

Existe una dificultad evidente en establecer la noción de causalidad. En general, se consideran necesarias tres condiciones para que se pueda plantear la relación de causalidad entre dos variables (Selltiz et al., 1959). La primera es la necesaria covariación entre las variables que se suponen ligadas estructuralmente. La segunda impone la existencia de una asimetría u orden estructural (temporal) entre la variable causa y la variable efecto. La tercera de las condiciones es algo más problemática, en la medida que exige que no existan otros factores o variables comunes que puedan explicar la covariación de las variables que se suponen relacionadas estructuralmente. De suceder esto, la desaparición de esa causa común supondría la desaparición de la supuesta relación entre causa y efecto. Es evidente que la explicación estructural es fuertemente dependiente de la teoría. Existe un número potencialmente infinito de variables (reconocidas como tal o no) que pueden estar influenciando sobre las dos simultáneamente. No existe ningún test estadístico que pueda revelar la existencia de estas variables. En ese sentido, la construcción de un modelo estructural supone el cierre formal de un sistema (abierto en la realidad). Para ello se selecciona un conjunto finito de variables y se proponen y evalúan relaciones entre ellas. Es muy importante tener presente la contingencia del modelo.

Es especialmente importante recordar esta provisionalidad porque, en evitación de la parálisis investigadora, la recomendación tal y como señala Asher (1983) consiste en operar sobre la base de que todas las variables están consideradas. En palabras de Blalock (1964), "no importa cuan elaborada sea una teoría, siempre es necesario establecer algunas presunciones. Particularmente, debemos asumir hasta cierto punto que los factores externos que afectan al modelo son estadísticamente despreciables. La aleatorización puede ayudar a resolver este problema, pero la plausibilidad de este tipo de presunción simplificadora es siempre una cuestión de grado." Como tendremos ocasión de comprobar, las presunciones a las que alude Blalock se refieren preferentemente al error.

En términos generales, hablaremos de causalidad cuando los cambios en una variable (llamémosle X) producen (tarde o temprano) cambios en otras variables (llamémosle Y). No obstante, es preciso establecer una serie de aclaraciones de las condiciones en que se plantea ese "producir cambios" en otra variable.

Estamos hablando de cambios medios. Estos cambios, al estar referidos a grupos de individuos (o diferentes unidades de análisis, pero siempre conjuntos de ellas) nos llevan a hablar básicamente de promedios o tendencias. Es decir, que la relación observada lo es para segmentos mayores o menores de población; esto implica que, al ser promedios, puede que dicho cambio detectado no exista realmente (es decir, que ninguna unidad de análisis se vea afectada realmente de esa forma); debe entenderse entonces que, no sólo no es obligatorio o necesario el cambio a todas las unidades de análisis (es bastante probable que existan excepciones individuales o de grupos específicos de individuos), sino que además podría suceder perfectamente que los coeficientes de cambio que se determinan empíricamente no existan realmente en la magnitud estimada. Una relación causal, en este sentido, es probabilística y no obliga a todos (a diferencia de las leyes en una democracia, que sí obligan a todos). Cuestión diferente es el tratamiento o etiqueta con que se identifiquen las excepciones al promedio. Las excepciones en términos de distribución pueden ser casos atípicos o casos extremos.

En el diagnóstico previo a cualquier análisis estructural es imprescindible estudiar estos casos que se encuentran fuera de la norma de variación o de covariación, es decir, que se comportan de modo distinto al patrón detectado en el conjunto de la población para la que poseemos datos. Existen razones para ello. En primer lugar, validar los datos descartando errores de medición o grabación. Segundo, explicarnos por qué esos casos o unidades de análisis son distintos. Podría suceder que dichos casos con un comportamiento extraño sean simplemente una cuestión de visibilidad. El riesgo evidente es que la producción de los datos o de las mediciones haya excluido (o se hayan excluido) segmentos de sociedad con un comportamiento equivalente. En ese sentido, el patrón detectado es simplemente consecuencia de la homogeneidad producida por la tarea de recoger información, lo que ha hecho más visible a una parte de la sociedad e invisible a otras. Es urgente explicar los casos atípicos y extremos respondiendo a la pregunta ¿son unidades de análisis en condiciones individuales particulares o representan un caso de visibilidad de los segmentos sociales invisibilizados en el proceso de traducción de la realidad social en datos e información?

En cualquiera de los tres casos, error de grabación, casos individuales o segmentos invisibles, debe tomarse una decisión antes de proceder al análisis. El motivo es evidente; los casos extremos ejercen una influencia notable sobre los ajustes. En el caso de errores de grabación la solución es quizás más evidente, bastando con corregirlos o suprimirlos. En las otras situaciones, una vez explicados los casos atípicos o extremos, retirarlos del análisis sólo parece legítimo si son casos particulares y además hemos aprendido lo que significan

en relación al resto de los datos. La decisión en caso de que se determine mediante el análisis que dichos casos representan situaciones de visibilidad excepcional de segmentos (y condiciones) sociales ocultos sólo podría ser redefinir de qué se está hablando y de qué no se está hablando. Sería inaceptable la ocultación de los casos atípicos o extremos y mantener la ficción de investigar la sociedad.

No debe olvidarse que la transformación de la sociedad en información/datos implica una traducción de ésta que aporta, a su vez, una fuente de variabilidad. En ese sentido, la variación, la covariación o los cambios medios determinados entre dos variables pueden verse condicionados por el proceso de obtención de los datos. De este modo, la traducción de la realidad con unas escalas y no otras, con unas variables y no otras, con un método y no otro, puede producir variabilidad en los coeficientes sintéticos y por lo tanto en los mismos cambios medios.

En esta revisión del significado de los cambios medios y sus condiciones de existencia e interpretación (validación, visibilidad/invisibilidad, casos particulares, traducción) nos falta considerar aquellos que, siendo reales y habiendo superado todas las consideraciones anteriores, dependen de condiciones teóricas para ser definidos como parte estructural en un proceso explicativo. La relación empírica debe ir acompañada de una relación lógica que la haga verosímil. Se trata de un caso especial de relaciones espurias, donde una estrecha relación estadística no responde a ninguna explicación lógica (no existe contigüidad en el espacio teórico entre las variables o conceptos relacionados). En ese sentido, la noción de causalidad es evidentemente teórica, en la medida que la relación no es por sí sola suficiente para definir causalidad.

La noción de causa es una noción asimétrica. Implica una ordinalidad temporal, una secuencia ordenada de covariaciones. La covariación entre dos variables no prueba por sí sola la presencia de causación. Como ya hemos advertido, un coeficiente de covariación expresa la fuerza con que está asociada la variación de dos variables, pero no afirma nada acerca de una hipotética dirección estructural entre ambas. Un ejemplo de ello es cuando se presentan correlaciones expresas, es decir, asociaciones entre variables que no están relacionadas teóricamente. Cuando no existe una relación teórica, la asimetría tiende a ser arbitraria. En ese sentido, las covariaciones espurias sin contigüidad explicativa acostumbran a estar, a su vez, desordenadas lógicamente. La estructuralidad u orden de una asociación es un constructo teórico, que primero considera la asociación entre un conjunto de variables y no otras y segundo establece un orden entre ellas. En resumen, el concepto de correlación habla de coordinación en la variabilidad de dos variables, mien-

tras que el de causa implica subordinación entre ellas (es decir, varían conjuntamente, pero una de ellas dependiendo de la otra, de algún modo a sus órdenes).

Esto es importante, en la medida que la noción de estructuralidad implica que el ajuste empírico (determinación de una relación mediante estimación) es condición necesaria pero no suficiente para definir causalidad. Por el contrario, la contigüidad lógica en un contexto cultural puede llegar, erróneamente, a aceptarse teóricamente como condición suficiente de causalidad, incluso en ausencia de evidencia empírica. Precisamente, esa es parte de la crítica de la sociología matemática con respecto a la sociología exclusivamente teórica: su tendencia a satisfacerse con la contigüidad lógica como condición suficiente para explicar el mundo. Recordemos que el análisis estructural intenta explicar la sociedad, pero en todo momento tiene el freno de lo que puede saber y en qué condiciones. Este freno a reconstruir el mundo sobre condiciones suficientes (propio de las ideologías) y atender a las urgencias de las condiciones necesarias es uno de los elementos centrales de la sociología matemática y, evidentemente, del análisis estructural.

Cuando en un análisis estructural se afirma que "X" es causa de "Y" no se afirma que "X" sea única causa de "Y". Existirán, probablemente, fenómenos sociales que dependerán de una sola y exclusiva causa. Sin embargo, es difícil que en un sistema tan interconectado como es el sistema social este caso sea el más frecuente. Ello es cierto incluso para coeficientes de correlación iguales a 1, donde la relación sea perfecta, en la medida que pueden existir otras variables que contribuyan a explicar la variabilidad de la dependiente, coordinada con la otra variable explicativa.

*“Cuando bajé hacia las naves, acompañado de mis padres, mi orgullo de guerrero había sido desplazado en mi ánimo por una intolerable sensación de hastío, de vacío interior, de descontento de mí mismo. Y cuando los timoneles hubieron alejado las naves de la playa con sus fuertes pértigas, y se enderezaron los mástiles entre las filas de remeros, supe que habían terminado las horas de alardes, de excesos, de regalos, que preceden las partidas de soldados hacia los campos de batalla. Había pasado el tiempo de las guirnaldas, las coronas de laurel, el vino en cada casa, la envidia de los canijos, y el favor de las mujeres. Ahora, serían las dianas, el lodo, el pan llovido, la arrogancia de los jefes, la sangre derramada por error, la gangrena que huele a almíbares infectos. No estaba tan seguro ya de que mi valor acrecería la grandeza y la dicha de los acaienos de largas cabelleras. Un soldado viejo que iba a la guerra por oficio, sin más entusiasmo que el trasquilador de ovejas que camina hacia el establo, andaba contando ya, a quien quisiera escucharlo, que Elena de Esparta vivía muy gustosa en Troya, y que cuando se refocilaba en el lecho de Paris sus estertores de gozo encen-*

*dían las mejillas de las vírgenes que moraban en el palacio de Príamo. Se decía que toda la historia del doloroso cautiverio de la hija de Leda, ofendida y humillada por los troyanos, era mera propaganda de guerra, alentada por Agamenón, con el asentimiento de Menelao. En realidad, detrás de la empresa que se escudaba con tan elevados propósitos, había muchos negocios que en nada beneficiarían a los combatientes de poco más o menos. Se trataba sobre todo –afirmaba el viejo soldado– de vender más alfarería, más telas, más vasos con escenas de carreras de carros, y de abrirse nuevos caminos hacia las gentes asiáticas, amantes de trueques, acabándose de una vez con la competencia troyana. La nave, demasiado cargada de harina y de hombres, bogaba despacio. Contemplé largamente las casas de mi pueblo, a las que el sol daba de frente. Tenía ganas de llorar. Me quité el casco y oculté mis ojos tras de las crines enhiestas de la cimera que tanto trabajo me hubiera costado redondear a semejanza de las cimeras magníficas de quienes podían encargar sus equipos de guerra a los artesanos de gran estilo, y que, por cierto, viajaban en la nave más velera y de mayor eslora."*

*A. Carpentier. "Semejante a la noche"*

Como casi siempre, será desde la interpretación que ofrezca la teoría como se resolverán las cuestiones de dependencia. Recordando nuevamente el aspecto central, en un modelo estructural se presume implícitamente que las causas no son únicas, y precisamente la importancia del residual (perturbación) en una relación indicará la importancia, en cantidad y calidad, de las variables que no están presentes en el modelo.

Los modelos sociológicos son porosos a la realidad que los rodea. Esto nos conduce al concepto de "Control". La noción de control es posiblemente la más alejada de la práctica de la investigación sociológica. La idea central es la necesidad de controlar el máximo de factores que puedan estar influenciando el proceso o fenómeno en observación. De este modo, es factible atribuir de forma única la explicación de la variabilidad observada. En ciencias naturales nos conduce directamente a los diseños experimentales. En ciencias sociales la psicología ha sido, probablemente, la disciplina que más se plantea el empleo de este tipo de diseños. En sociología, estos diseños están evidentemente vedados. La investigación sociológica se centra sobre todo en los diseños correlacionales y en todo caso, en los diseños cuasi-experimentales. En ese sentido, hablar de control es hablar de errores en la investigación. No obstante, eliminar todos los sesgos es prácticamente imposible, dado que el lenguaje mismo estructura nuestra percepción del mundo. Los sesgos alcanzan incluso a la selección de los problemas para investigar o las mismas estrategias utilizadas.

En el caso de la investigación experimental, existen diversos procedimientos empleados en el transcurso de la investigación destinados a minimizar los errores, como la asignación aleatoria a grupos experimentales. Es importante recordar que el tipo de diseño de investigación empleado afecta a la aplicación de técnicas de análisis. Por ejemplo, los problemas del análisis de varianza con dos factores o más, empleando datos de encuestas, y por ello con casos desiguales en cada celdilla. Este ideal de control del proceso de investigación es una referencia importante en la actividad denominada ciencia. El empleo de los procedimientos de control elimina las explicaciones confusas de los sucesos en estudio, y es uno de los rasgos principales por el cual el conocimiento científico difiere del casual. Mientras que el investigador puede diseñar estudios para recoger información y testar determinadas explicaciones o respuestas, los hallazgos están frecuentemente abiertos a diferentes interpretaciones. La idea de control pretende emplear procedimientos que efectivamente descarten aquellas explicaciones que realmente no son coherentes con la información. En el caso de la investigación sociológica, el control es muy limitado y se reduce, en el caso de la investigación cuantitativa, al control estadístico.

Los datos que provienen de diseños no experimentales aportan evidencias menos concluyentes que aquellos otros que se obtienen de diseños experimentales. Uno de los motivos por los que la evidencia es más sólida en términos experimentales que en no experimentales viene dado por el hecho de que en diseños experimentales existe un control sobre las fuentes de variabilidad. Es decir, al mutilar o aislar el sistema de relaciones, la atribución a las variables de control de la variabilidad apreciada en las variables dependientes es neta, dado que se intenta que no existan otras fuentes de variabilidad que influyan. Sin embargo, en investigación sociológica, los sistemas estructurales se ajustan sobre datos que provienen de contextos permeables a la realidad que les rodea (variables no incluidas en el análisis, variables desconocidas). Esto implica que la permeabilidad de la información y su sensibilidad a factores no controlados obliga a un escrutinio minucioso de la variación no explicada. Dentro del método científico, el método correlacional es una alternativa al método experimental, siempre y cuando se acepten y reconozcan sus limitaciones.

Como resumen, para que se pueda afirmar la noción de causalidad, es decir, que las variaciones en una variable (X) causan las variaciones en otra (Y), deben de estar presentes tres condiciones:

(1) La variable que se considera causa (X) debe preceder temporalmente/lógicamente a la variable efecto (Y). (criterio de asimetría)

(2) Las variables que se consideran causa (X) y efecto (Y) deben de covariar. (coordinación estadística)

(3) No deben existir explicaciones alternativas que, de un modo satisfactorio, den cuenta teórica de la covariación observada entre X y Y. (criterio de control desde la perspectiva correlacional).

### 3.7. LA DETERMINACIÓN TEÓRICA DEL ORDEN EXPLICATIVO

Como hemos mencionado anteriormente, consideradas dos variables, éstas pueden mostrar una coordinación estadística de tal forma que covarían. Sobre dicha coordinación observada puede articularse una relación teórica de subordinación entre ellas, donde una característica correspondiente a un caso (sea entrevistado, país, en definitiva, unidad de análisis) cambia de valor en función a los valores que adopte otra característica diferente. Por ejemplo, horas de ver televisión en función de ingresos. Probablemente existe un umbral de ingresos donde otros tipos de ocio reemplazan las horas de televisión, por lo tanto, de existir correlación estadística, las horas de televisión están subordinadas a los ingresos o renta disponible del entrevistado.

Parece evidente que el orden de subordinación que se establezca entre dos variables se "construye" sobre la existencia de covariación entre los valores de dos variables. De no existir covariación estadística, la subordinación (relación estructural) entre variables pierde fundamento empírico. Analíticamente, podemos diferenciar cuatro posibilidades en ese orden estructural para el caso de dos variables<sup>17</sup>:

$$(1) \quad y_1 \longrightarrow y_2$$

En este caso, "y<sub>1</sub>" puede influir en "y<sub>2</sub>" pero no lo contrario.

$$(2) \quad y_1 \longleftarrow y_2$$

En este segundo caso, "y<sub>2</sub>" es la que influye en la variabilidad de "y<sub>1</sub>", siendo imposible, teóricamente lo contrario.

$$(3) \quad y_1 \longleftrightarrow y_2$$

Aquí las variables "y<sub>1</sub>" e "y<sub>2</sub>" se influyen mutuamente, se retroalimentan en su variación. Por ejemplo, los presupuestos de defensa de USA e URSS durante la guerra fría.

17. Hemos notado la relación mediante una flecha, es decir, mediante un grafo orientado. Cuando la relación de subordinación es unidireccional implicará un solo grafo orientado, cuando es una relación de interdependencia serán dos grafos, cada uno de ellos indicando una dirección opuesta. En el caso de coordinación estadística, sin relación de dependencia, se expresa mediante un grafo con dos cabezas de flecha.

$$(4) \quad y_1 \longleftrightarrow y_2$$

" $y_1$ " e " $y_2$ " presentan en este caso una coordinación estadística sobre la que no presumimos ningún orden teórico. Básicamente, existe una covariación no explicada.

En el momento de construir la secuencia de relación entre variables sobre la base del criterio de subordinación lógica pueden surgir problemas importantes. Algunos de ellos sólo se resolverán desde la aceptación de una teoría previa, especialmente en el caso de variables que acostumbran a ser de naturaleza exógena (como las variables Religión e Ideología política). No obstante, si es posible establecer unas orientaciones metodológicas para la determinación teórica del orden estructural.

La noción de orden estructural se establece habitualmente sobre la noción de "tiempo". Esencialmente, se resumen en la afirmación "lo que sucede después no puede causar lo que sucedió antes". Si algo cambia, lo hace en función de un cambio previo en la variable de la que depende. Si instrumentamos la determinación del orden estructural en función a qué variable cambia primero, es útil operativizar esa secuencia temporal.

Es útil pensar que las variables tienen una fecha de "inicio" (o de fabricación) y otra de "término". Así, la fecha de inicio de una variable es el momento antes del cual no puede existir valor para esa variable. La fecha de término de una variable es el momento tras el que no existen cambios de valor posible. Por ejemplo, la variable "participar en la última Guerra Civil española" tiene una fecha de inicio en 1936, y de terminación en 1939. Antes de 1936 no existía valor ni variable, después de 1939 el valor que posea cada caso es fijo (sí o no). Siguiendo este criterio de inicio y término de una variable (cuando esto es posible) es factible establecer una serie de reglas para determinar el orden estructural entre variables (allí donde la teoría o la incertidumbre no prescriban un orden de subordinación en la covariación).

a) Consideremos que  $y_2$  varía en función de  $y_1$  cuando la variable  $y_2$  tiene una fecha de inicio posterior a la fecha de terminación de  $y_1$ . Por ejemplo, la variable "actitud hacia la entrada de España en la OTAN" y "participación en la última Guerra Civil española". La opinión de los españoles acerca de la entrada de España en la OTAN variará en función a su participación en la Guerra Civil, dado que la fecha de inicio (entrada de España en la OTAN) es posterior a la de terminación de "participación en la última Guerra Civil española"<sup>18</sup>.

18. La anticipación de una variable (es decir, pensar sobre su comportamiento futuro) puede hacer variar otra que le antecede temporalmente. Sin embargo, hay que considerar que la variable "anticipación de  $y_2$ " es una variable distinta y anterior a la variable " $y_2$ ".

**b)** Podemos considerar que  $y_2$  depende de  $y_1$  cuando la variable  $y_1$  se vincula a un paso, etapa o momento anterior de  $y_2$ , dentro de una secuencia temporal bien conocida. Un ejemplo de secuencia en sociología es el ciclo vital, donde se describe una trayectoria vital que es seguida por la inmensa mayoría de los que viven en una sociedad. Así, la secuencia vendría a ser, esquemáticamente,

- (1) características de la familia donde se crece
- (2) educación escolar
- (3) primer trabajo a tiempo completo
- (4) primer matrimonio
- (5) nacimiento de hijos
- (6) disolución de matrimonio por divorcio o muerte.

Y así continuadamente, concretándose tanto como el investigador desee.

De hecho, el ciclo vital de un individuo se trenza, por así decirlo, con el de sus padres y sus propios hijos o su entorno social. De este modo prácticamente todo aquello que acontece en el ciclo vital de un individuo, se podría hacer depender, teóricamente, del entorno familiar en que éste desarrolla su personalidad y sus costumbres, y así sucesivamente.

**c)** Otro criterio es cuando una de las variables posee valores o atributos que se pueden considerar definitivos para cada individuo, mientras que la otra registra la posibilidad de que un individuo cambie de valor o atributo. Por ejemplo, una variable en la que difícilmente se cambia de valor es la variable género, y de hecho esta variable se emplea habitualmente para explicar la variación de otras variables sociológicas.

**d)** Otra regla, que más bien es una orientación, afirma que si los valores o atributos que cada individuo o caso tiene registrado en una variable  $y_1$  son relativamente estables, lentos de cambiar, mientras que los valores que registra en otra variable  $y_2$  son volátiles o poco estables,  $y_2$  dependerá normalmente de  $y_1$ . Así por ejemplo, si las variables son "preferencias religiosas" y "opinión sobre el gobierno", probablemente la segunda dependerá de la primera. Es difícil infravalorar la importancia que posee una correcta especificación del modelo, tanto por sus consecuencias teóricas como por la prima de ajuste que se obtendrá al postular relaciones con probabilidad de ocurrir. Estas orientaciones anteriores ayudan a establecer con mayor consistencia el orden estructural, en el sentido de postular una dirección estructural, bidireccionalidad o simple coordinación estadística sin determinación de orden.

## 4. MODELOS ESTRUCTURALES CON VARIABLES OBSERVADAS

Los modelos estructurales, en tanto que explicación narrativamente compleja de la realidad, aspiran a explicar sistemas de relaciones donde interviene un número importante de variables. La intervención de diferentes variables contribuye, en la práctica, a clarificar las relaciones existentes entre ellas. Recordemos el ejemplo de las relaciones espurias: se denomina relación espuria a aquella covariación existente entre dos variables que es consecuencia de que ambas dependen de otra variable que es causa común de ellas y que da cuenta de la covariación. Ésta es una posibilidad existente que debe ser evaluada en detalle, y en principio constituye una sospecha que pesa sobre toda covariación bivariable. Dado que la determinación de la condición de relación espuria entre dos variables (donde su covariación observada viene inducida por su dependencia común de una tercera variable) se afirma teóricamente al definir la tercera variable como causa común, la introducción de variables comunes en los modelos incrementa su validez.

*"Los hechos son obstinados, dice un proverbio inglés. Este proverbio inglés nos viene a menudo a la memoria, especialmente cuando algún escritor se despacha, trinando como un ruiseñor, sobre la grandeza del "principio de la nacionalidad" en sus diversos sentidos y correlaciones, a tiempo que este "principio" se aplica con tanto acierto como acertadas y oportunas fueron las exclamaciones de un célebre héroe de un cuento popular que, a la vista de una procesión fúnebre, le deseó: "Ojalá tengáis siempre un muerto que llevar." Hechos exactos, hechos indiscutibles: he aquí lo particularmente insoporrible para esta clase de escritores y lo verdaderamente necesario, si uno desea orientarse con seriedad en el complejo y difícil problema, a menudo enredado con toda premeditación. Pero ¿cómo reunir los hechos? ¿Cómo establecer su nexo e interdependencia? En el terreno de los fenómenos sociales no existe procedimiento más difundido y más inconsistente que tomarse de los pequeños hechos "aislados", jugando a los ejemplos. Escoger los ejemplos, en general, es bastante fácil,*

*pero resulta que, o no significan nada o son negativos, puesto que el fondo reside en el ambiente histórico concreto de cada caso. Los hechos, considerados en su conjunto, en su mutua correlación intrínseca, no sólo son "obstinados" sino absolutamente demostrativos. En cambio, los pequeños hechos tomados en forma aislada y sin relación intrínseca, fragmentaria y arbitrariamente se transforman en un juguete o en algo peor. Por ejemplo, si un escritor con fama de persona seria, deseoso de que se lo siga considerando como tal, toma el caso del yugo mongólico y lo expone como ejemplo para aclarar ciertos acontecimientos ocurridos en la Europa del siglo XX, ¿podrá considerarse su proceder sólo como un juego, o más correctamente como charlatanismo político? El yugo mongol es un hecho histórico indudablemente ligado con el problema nacional. También en la Europa del siglo XX se observa una serie de hechos cuya conexión con este problema es asimismo obvia. Sin embargo, pocas personas habrá –del tipo que los franceses tildan de "payasos nacionalistas"– susceptibles de pretender seriedad y al mismo tiempo obrar con "hechos" como los del yugo mongol para ilustrar lo que sucede en la Europa del siglo XX.*

*La conclusión es clara: hay que tratar de establecer una base de hechos exactos e indiscutibles sobre la cual apoyarse para comparar cualesquiera de esas argumentaciones "generales" y "ejemplares" que en la actualidad se usan en forma abusiva en algunos países. Para que esa base sea verdadera, es necesario no tratar hechos aislados, sino todo el conjunto de los hechos que atañen al problema en discusión, sin una sola excepción, puesto que de otra manera es inevitable que surja la sospecha, muy legítima, de que los hechos se eligieron o adoptaron arbitrariamente y que, en lugar de una correlación objetiva y una interdependencia de los fenómenos históricos en su conjunto, nos sirven un mejunje "subjetivo" para justificar posiblemente un asunto sucio. Eso ocurre más a menudo de lo que se cree.*

*Partiendo de estas premisas, hemos resuelto comenzar con estadísticas, conscientes de la gran antipatía que suelen provocar en algunos lectores y escritores, quienes prefieren la "noble mentira" a las "bajas verdades"; por su afición a pasar, bajo la bandera de meditaciones "generales", contrabando político sobre internacionalismo, cosmopolitismo, nacionalismo, patriotismo, etcétera."*

V.I. Lenin "Estadística y sociología", Bolchevik, n° 2, 1935

Si bien es interesante determinar las relaciones no explicativas, es evidente que la finalidad última de los modelos estructurales es determinar relaciones explicativas. El término empleado para nombrar la relación entre variables es el de *efecto*, dado que se postulan relaciones de causa-efecto. Según el tipo de relación entre variables, es decir, según su posición en el sistema, el efecto (relación) se denominará de un modo u otro. Debe recordarse que todo efecto responde a la presunción de una relación con contenido teórico y que responde a una hipótesis.

Estableceremos dos tipos generales de relación, la que se produce en los dos sentidos estableciendo una dinámica de retroalimentación, y la que se produce en un sentido único. De este modo tendremos:

Causación unidireccional:

- efecto directo
- efecto indirecto
- efecto directo e indirecto
- efecto condicional

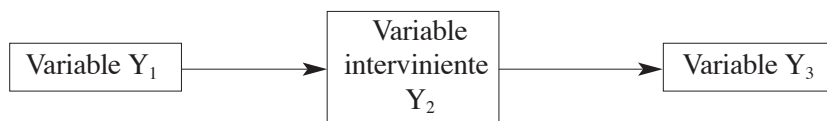
Causación bidireccional:

- efecto recíproco directo
- efecto recíproco indirecto

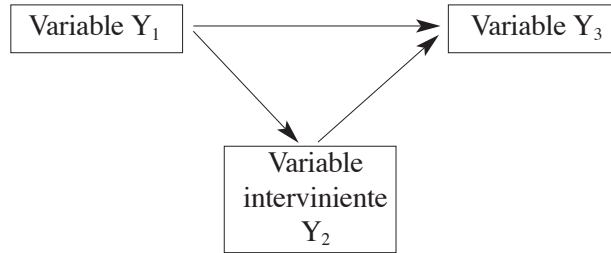
En primer lugar, un efecto directo indica una relación no mediada entre dos variables. En ese sentido expresa que, de existir variables que medien entre ellas dos, éstas carecen de entidad o significación teórica para ser explicitadas. Esto no siempre es así, y el recurso a los efectos directos permite ocultar, incluso de modo no voluntario, relaciones importantes. En ese sentido, recordemos que el entimema es una forma no correcta de razonamiento donde se da por obvia la segunda premisa. Los efectos directos deben evaluarse cuidadosamente, dado que muy posiblemente se den por evidentes variables mediadoras que deberían ser explicitadas para una mejor comprensión del proceso en estudio. Los efectos directos responden al siguiente esquema.



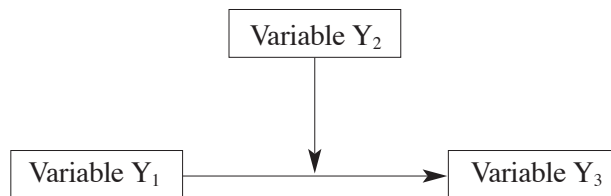
Un efecto indirecto se produce cuando una variable causa influye en otra variable a través de una tercera variable que actúa como variable mediadora. Esta tercera variable que convierte lo que sería un efecto directo en uno indirecto se denomina, como ya se advirtió, variable interviniente. Una de las ventajas de la introducción de variables intervinientes es que desvelan con una mayor nitidez la secuencia que sigue el mecanismo estructural.



Esta secuencia muestra un efecto indirecto de la variable que se postula como causa, sobre la variable que se postula como efecto. Pueden presentarse conjuntamente efectos directos e indirectos entre una variable causa y otra efecto. Esta posibilidad se recoge en el diagrama siguiente.



Hemos podido apreciar cómo la variable interviniente convierte una relación directa en indirecta. Existe otro tipo de variable que puede mediar en otra forma sobre el efecto existente entre dos variables. Es la denominada variable condicional. Las variables condicionales determinan la intensidad de los efectos estructurales. Así, en el diagrama siguiente podemos apreciar cómo la variable condicional  $Y_2$  no orienta su grafo hacia otra variable, sino que lo hace en dirección a otro grafo.



Por ejemplo, podemos afirmar que el grado en que se conozcan las normas que rigen las interacciones de los miembros de un grupo tendrá como efecto el grado de integración en dicho grupo. Sin embargo, aún con un alto conocimiento de las normas, la integración en el grupo se verá determinada por el interés que tenga el individuo en pertenecer a él. Un estudio puede tener un conocimiento completo de las normas de un grupo de “punkies” y no por ello estar integrado en uno de ellos. En cierto modo, las variables condicionales están siempre presentes si bien no se suelen explicitar, excepto cuando su intervención es especialmente relevante para la relación estudiada.

Un tratamiento aparte requiere las relaciones condicionales bidireccionales. En algunos planteamientos teóricos no está clara la distinción entre variable causa y variable efecto, en la medida que ambas se afectan mutuamente. Este tipo de relación se denomina relación recíproca y es aquella en la que dos variables se influyen mutuamente. Es decir, la teoría prevé que una variable produce variación en otra, y ésta segunda en la primera.

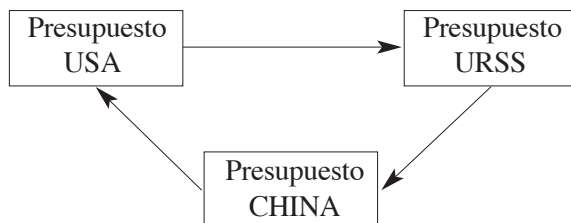


Un tipo de fenómeno modelado con frecuencia de este modo son los conflictos sociales, por ejemplo estudiantes y policías, oposición y represión, etc. Así, las variables se afectan una a la otra secuencialmente en el tiempo. Esta retroalimentación está asociada a sistemas dinámicos, donde se producen espirales de calentamiento o enfriamiento según los signos de la relación. Un efecto recíproco implica la presencia de ecuaciones simultáneas. Se trata, por lo tanto, de acciones y reacciones entre variables. Dada la variabilidad en los posibles ritmos de alternancia, pueden aparecer problemas específicos de medición, al detectar o no la presencia de sincronía. Se trata, en definitiva, de diagnosticar el posible retardo entre la evolución de las dos variables.

En otros casos dicho problema no aparece, como por ejemplo al considerar el efecto recíproco entre URSS y USA de los presupuestos de defensa durante la guerra fría, dado que la unidad temporal año detecta bien la variabilidad existente. Así, en el modelado de la influencia de los presupuestos de defensa de USA en la antigua URSS y viceversa, debe considerarse que el conocimiento de los presupuestos de un año en USA influían en el siguiente en URSS y así sucesivamente.



Estos efectos recíprocos pueden establecerse directamente, en cuyo caso trataremos generalmente con dos variables. Otra posibilidad viene dada por la presencia de efectos recíprocos indirectos, donde pueden estar involucradas más de dos variables; para el caso de tres variables se establecería una dinámica circular.



En este caso de efectos recíprocos indirectos, las variables se afectan entre sí en una dinámica circular. Este tipo de relación es característica, en la medida que refleja claramente dinámicas de crecimiento o decrecimiento en un sis-

tema. Dan forma por sí solas a unas tipologías específicas de modelos, así como a las técnicas para determinar los parámetros<sup>19</sup>.

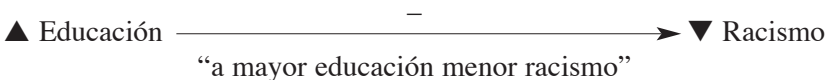
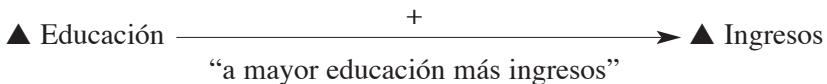
Un aspecto distinto al de los efectos es el de la covariación. Como sabemos, un efecto es una covariación expresada en términos de causalidad. Cuando nos referimos a la covariación en los diagramas, estos son usualmente simbolizados mediante líneas con puntos de flecha señalando en ambas direcciones. Dado que no está especificada una subordinación entre variables, son denominados efectos conjuntos.



#### 4.1. EL SIGNO DE LAS RELACIONES

Los efectos pueden tomar signos dependiendo de la relación en que se mueva la variabilidad entre las variables. Si los valores en una variable efecto tienden a crecer cuando los valores en la variable causa tienden a crecer se establece un signo positivo, dado que la coordinación estadística entre ambas variables se mueve en el mismo sentido. Por el contrario, cuando una de ellas decrece en el caso de que la otra crezca el signo es negativo, debido a que los valores en las dos variables se mueven en sentido distinto. Una cuestión interesante es, dado que los sistemas estructurales concatenan varios efectos estructurales con diferentes signos, determinar cuál es la relación entre una variable y otra. Por ejemplo, supongamos una variable causa, 15 variables intervinientes mediando y una variable final. ¿Cómo podremos saber si la coordinación entre ambas es directa o inversa?

En los modelos estructurales los senderos tienen signo. Un sendero es una serie de variables conectadas entre sí mediante grafos (efectos), siempre que el orden de los efectos se desplace en el mismo sentido. Es decir, no aparezcan mediando efectos recíprocos.



19. Los efectos recíprocos se formulan mediante sistemas de ecuaciones simultáneas, que a su vez son el alma de las simulaciones basadas en retroalimentaciones.

Como regla para determinar el signo final de un sendero, es decir, en qué direcciones se mueven la primera y la última variable del sendero, se deben multiplicar el signo de sus relaciones. Un sendero será positivo a menos que contenga un número impar de signos negativos. Si recordamos la regla de multiplicación de signos es evidente,

$$(+ * + = + ; + * - = - ; - * - = +)$$

						signo del sendero		
a)	$y_1$	$\xrightarrow{+}$	$y_2$	$\xrightarrow{+}$	$y_3$	$\xrightarrow{+}$	$y_4$	+
b)	$y_1$	$\xrightarrow{+}$	$y_2$	$\xrightarrow{-}$	$y_3$	$\xrightarrow{-}$	$y_4$	+
c)	$y_1$	$\xrightarrow{+}$	$y_2$	$\xrightarrow{-}$	$y_3$	$\xrightarrow{+}$	$y_4$	-
d)	$y_1$	$\xrightarrow{-}$	$y_2$	$\xrightarrow{-}$	$y_3$	$\xrightarrow{-}$	$y_4$	-

Las relaciones que hemos considerado hasta el momento se establecen para un conjunto de variables, dando forma a sistemas de variables interconectadas denominados modelos estructurales. La noción de sistema es central en la investigación social actual. Ello viene dado por su gran utilidad, al permitir y exigir explicitar las variables que se consideran importantes, así como la forma en que se relacionan entre sí. Para ello debe superarse la idea que afirma "todo está relacionado con todo", explicitando aquellos nudos de covariación que son especialmente significativos para comprender y explicarnos la sociedad en que vivimos.

#### 4.2. CONSTRUCCIÓN DE MODELOS ESTRUCTURALES

No existe, evidentemente, ningún algoritmo que por sí solo genere modelos estructurales. Estos son el resultado de un análisis de la realidad y del establecimiento de unas hipótesis sobre ella. No obstante, desde un punto de vista instrumental sí es posible establecer algunas orientaciones sobre cómo organizar la tarea.

1. En primer lugar es importante determinar la lista de las variables que son importantes en el proceso estudiado. Este paso es esencial, en la medida que implica una definición de la realidad que se desea estudiar. No debe olvidarse que los modelos matemáticos requieren de variables operativizadas, es decir datos. En ese sentido, difícilmente existe libertad para utilizar todas las variables que podrían resultar interesantes. Esto es especialmente cierto en el caso de los datos provenientes de encuestas o secundarios. Sólo en el caso de datos

primarios y cuando el coste o el tema de investigación lo permite existe una mayor libertad de diseño.

2. Determinación del orden estructural que se postula en las variables. Una vez listadas las variables que operarán en el modelo estructural, es preciso establecer la secuencia en que se relacionan entre sí. Como se mencionó, se postulan relaciones asimétricas entre ellas, en función a qué variable explica y qué variable es explicada.

3. Especificación de las hipótesis estructurales. Es decir, establecer la cadena argumental explicativa del fenómeno social estudiado. En esta etapa se establece la potencia descriptiva de nuestro modelo explicativo.

4. Elaboración del diagrama estructural. A efectos prácticos, es útil establecer la secuencia mediante un grafo orientado que permita visualizar qué variables están conexas entre sí y qué variables están inconexas. En muchas ocasiones el grafo o diagrama estructural permite detectar incongruencias en la explicación que se pretende ofrecer. La visión conjunta del sistema ofrece una potencia importante para evaluar el modelo que se propone.

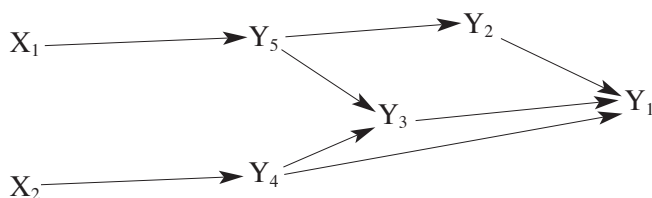
En términos prácticos, se procede escribiendo las variables con posiciones ordenadas indicando el orden estructural. Tras esta tarea se introducen las hipótesis como flechas entre las variables de acuerdo a los efectos directos. Es una convención que los efectos no especificados son cero (0). Una vez sobre el diagrama estructural es el momento de reflexionar si se han planteado todas las variables y relaciones que son pertinentes.

Resulta evidente que la formulación teórica es un proceso activo donde la articulación de los sistemas depende de la fase de la investigación, del empleo de datos secundarios o primarios, etc. En ese sentido, se desarrolla una reflexión sobre la coherencia lógica de las relaciones que se postulan entre las variables así como de las limitaciones de la explicación que se está ofreciendo. Una vez que se especifica una teoría estructural, ésta debe de ser contrastada con los datos para testar su eficacia empírica.

En la fase de diseño del modelo los principios rectores son esencialmente teóricos, dando cuerpo a las hipótesis estructurales. La determinación de las variables y su relación es una tarea previa al ajuste sobre los datos. En ese sentido resulta interesante, a la luz de la teoría, establecer los modelos y dejar que posteriormente las limitaciones del acceso a datos restrinjan el modelo: de ese modo se es más consciente de las variables que han podido quedar fuera (influyendo en el modelo desde fuera) así como de las limitaciones de la potencia explicativa del modelo.

Como sabemos, los datos simplemente determinan el grado de covariación. No obstante sabemos que la covariación no es una prueba de relación estructural, dado que ésta puede estar provocada por causas comunes a las variables de interés.

Como sabemos, el diagrama estructural está compuesto por las variables relacionadas mediante grafos orientados. Así, se disponen las variables y después se conectan entre sí aquellas para las que se proponga alguna relación teórica. El ejemplo siguiente muestra un grafo orientado, donde X y Y notan variables y las flechas relaciones.



5. Matriz de efectos. La matriz de efectos es esencialmente una matriz donde se expresan mediante ceros y unos la existencia o no de relación entre las diferentes variables. Normalmente, es una prueba más de comprobación de la completitud del diseño. Su planteamiento destaca sobre todo la ausencia de relaciones. En ese sentido, el diagrama estructural es útil para expresar lo que se quiere decir, mientras que la matriz de efectos destaca lo que no estamos diciendo. Así, en el diagrama se da cuerpo a la existencia de relación mediante el grafo, y no es fácil evaluar qué se está diciendo a su vez que no existe relación entre las variables donde no la hay. En la matriz de efectos destacan sobre todo los efectos que postulamos que no existen.

La matriz se construye listando todas las variables (tanto exógenas como endógenas) en la cabecera, y las variables efecto (o que son explicadas) en las filas.

	y <sub>5</sub>	y <sub>4</sub>	y <sub>3</sub>	y <sub>2</sub>	y <sub>1</sub>	x <sub>1</sub>	x <sub>2</sub> Variables exógenas al final
y <sub>5</sub>	-	0	0	0	0	1	0
y <sub>4</sub>	0	-	0	0	0	0	1
y <sub>3</sub>	1	1	-	0	0	0	0
y <sub>2</sub>	1	0	0	-	0	0	0
y <sub>1</sub>	0	1	1	1	-	0	0

El procedimiento a seguir es que cuando existe efecto directo entre dos variables se pone un 1. En el caso que no se postule efecto directo entre dos variables se anota un cero (0). Como hemos destacado, un aspecto importante es el de las relaciones que postulamos igual a cero, es decir, que no existen. En esa línea, el completar la matriz es una labor que puede ayudar a desarrollar hipótesis interactivamente, dando una mejor forma al modelo.

Estas orientaciones para el diseño de los modelos estructurales deben considerar también la necesidad de simplificación de las teorías estructurales. La noción de sistema es central en la investigación social actual. Ello viene dado por su gran utilidad, al permitir y exigir explicitar las variables que se consideran importantes, así como la forma en que se relacionan entre sí. Para ello debe matizarse la idea que afirma "todo está relacionado con todo", en la medida que algunas cosas están especialmente relacionadas, adoptando esa relación formas específicas arbitrarias. Hemos avanzado en el sentido de hacer operativa la noción de sistema, considerando analíticamente las unidades, variables y relaciones que lo componen.

## 5. EXPRESIÓN MATEMÁTICA

Un modelo teórico, una explicación en definitiva, puede encontrar diferentes formas de expresión; ya sea en la apariencia de un diagrama, adoptando una enunciación verbal o escrita, en todos los casos se trata del mismo modelo. Una forma nueva de representar el mismo modelo es mediante un sistema de ecuaciones. Para ello, deberemos adoptar una serie de convenciones para poder formular el modelo ecuacionalmente. No existe una notación universalmente aceptada, (evidentemente, no existe una notación natural) y la que empleamos no deja de ser una más de las existentes.

### 5.1. NOTACIÓN DE SISTEMAS ESTRUCTURALES

Las variables endógenas (dependientes) las notaremos mediante una  $Y$  con subíndice que expresa un número que la diferencia. Para el caso de las variables exógenas (independientes) emplearemos una  $X$  con subíndice.

Variable endógena  $y_i$

Variable exógena  $x_i$

En lo que se refiere a las relaciones o efectos, aquel que se postula entre variables endógenas lo notaremos  $\beta$  con dos subíndices ( $ij$ ) donde se identifican las variables que intervienen en dicha relación. El subíndice ( $i$ ) para la variable que recibe el efecto (y por tanto que es explicada) y el subíndice ( $j$ ) para la variable que explica.

$$y_j \xrightarrow{\beta_{ij}} y_i$$

Para la relación de una variable exógena sobre una endógena emplearemos una  $\gamma$  con la misma intencionalidad en los subíndices.

$$x_j \xrightarrow{\gamma_{ij}} y_i$$

Como ya se advirtió al hablar del contenido de las perturbaciones, éstas serán notadas  $\zeta$  con el subíndice de la variable correspondiente. Evidentemente, suponemos una perturbación por cada ecuación. Ya nos es posible especificar un sistema de ecuaciones lineales, donde habitualmente los efectos son aditivos. El sistema tendrá tantas ecuaciones como variables endógenas contenga, dado que cada variable endógena posee alguna previa que explica su variabilidad. En ese sentido, recordemos que la variación y covariación entre variables endógenas están, de algún modo, determinadas por la variación y covariación entre variables exógenas, lo que nos lleva a reconocer que la varianza y covarianza de las variables exógenas son fundamentales en todo modelo.

Necesitamos, por lo tanto, una forma de notación para las varianzas de cada variable y las covarianzas entre ellas. La cuantía de la  $x_i$  (la variación de la variable exógena) se nota  $\phi_{ii}$ . Cuando se trate de la covarianza entre dos variables exógenas  $x_i$  e  $x_j$  serán los subíndices los encargados de indicarlo:

$$\phi_{ij}$$

La varianza de las perturbaciones  $\zeta_i$  se nota como

$$\Psi_{ii}$$

y nuevamente cuando nos refiramos a la covarianza entre dos perturbaciones  $\zeta_i$  e  $\zeta_j$

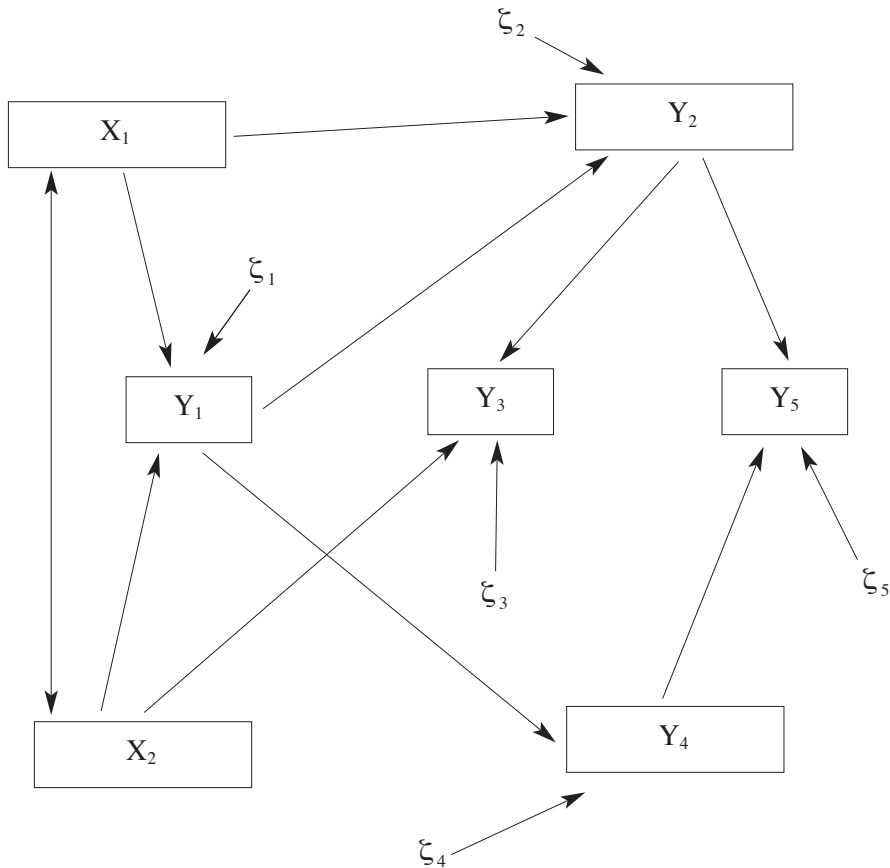
$$\Psi_{ij}$$

Una vez acordadas las convenciones de notación, podemos utilizarlas para construir ecuaciones.

## 5.2. SISTEMAS DE ECUACIONES

Antes de comenzar, debemos recordar que un modelo estructural no es simplemente un sistema de ecuaciones. Lo esencial es que dicho sistema represente el mecanismo estructural que ha producido los valores observados en las variables endógenas. En ese sentido, el diagrama estructural siguiente (ver página 61) expresaría una secuencia explicativa.

Sobre la base del sistema de notación que se ha introducido, las relaciones entre variables endógenas se expresarán mediante una Beta ( $\beta$ ) con los subíndices correspondientes a las variables que está relacionando. Recordemos que primero se posiciona el subíndice de la variable efecto (la que recibe el grafo) y seguidamente el subíndice correspondiente a la variable que se propone como causa de ella. Los efectos de exógenas sobre endógenas se nota-



rán con la letra griega Gamma ( $\gamma$ ) la constante de la ecuación con la letra Alfa ( $\alpha$ ) y las perturbaciones con Zeta ( $\zeta$ ). Cuando una variable no produce efecto sobre la endógena en cuestión, se pone directamente un cero. Evidentemente, tendremos tantas ecuaciones como número de endógenas.

$$\begin{aligned}
 Y_1 &= 0y_1 + 0y_2 + 0y_3 + 0y_4 + 0y_5 + \gamma_{11} X_1 + \gamma_{12} X_2 + \alpha_1 + \zeta_1 \\
 Y_2 &= \beta_{21}y_1 + 0y_2 + 0y_3 + 0y_4 + 0y_5 + \gamma_{21} X_1 + 0X_2 + \alpha_2 + \zeta_2 \\
 Y_3 &= \beta_{31}y_1 + \beta_{32}y_2 + 0y_3 + 0y_4 + 0y_5 + 0X_1 + \gamma_{32} X_2 + \alpha_3 + \zeta_3 \\
 Y_4 &= \beta_{41}y_1 + 0y_2 + 0y_3 + 0y_4 + 0y_5 + 0X_1 + 0X_2 + \alpha_4 + \zeta_4 \\
 Y_5 &= 0y_1 + \beta_{52}y_2 + \beta_{53}y_3 + \beta_{54}y_4 + 0y_5 + 0X_1 + 0X_2 + \alpha_5 + \zeta_5
 \end{aligned}$$

### 5.3. PRESUNCIONES

En el planteamiento de modelos estructurales son habitualmente necesarias un conjunto de presunciones que definan el marco de la especificación del sistema que se propone. Estas presunciones son testadas durante la fase de ajuste empírico del sistema de ecuaciones sobre los datos. Para un modelo expresado con las variables no transformadas, es decir tal y como se han registrado, encontraremos normalmente cuatro presunciones básicas.

La primera a considerar afirma que la media de los errores es cero para todas las ecuaciones. Lo que se afirma mediante esta presunción es que la ecuación estructural explica correctamente la variable endógena, en la medida que los efectos de las variables que no están en el modelo (y que son representadas por el error) tienden a cancelarse entre si.

$$\mu \zeta_i = 0 \text{ para todo } i \quad (1)$$

Una segunda presunción importante afirma que los errores de las diferentes ecuaciones no covarían con las variables exógenas. La razón principal por la que el error y las variables exógenas podrían covariar es que ambas tengan alguna causa previa que sea común, y en ese sentido la presunción indica que no existen dichas causas comunes omitidas a variables endógenas y exógenas.

$$\text{Cov}(\zeta_i, x_j) = 0 \text{ para todo } i, j \quad (2)$$

La tercera presunción afirma que los errores no covarían. La interpretación de dicha covariación, en el caso de producirse, sería esencialmente que en la fase de especificación se han olvidado variables que son causa común a las endógenas. No debe pensarse que habitualmente la varianza de un error sea cero, dado que esto implicaría que el error es cero o una constante, ambas cosas bastante improbables. La media de un error sí que puede ser cero, pero no su variación alrededor de la media.

$$\Psi_{ij} = 0 \text{ para todo } i \neq j \quad (3)$$

Por último, una cuarta presunción plantea la posibilidad de que las variables exógenas, es decir, aquellas que no son explicadas dentro del modelo, puedan presentar covariación entre ellas.

$$\Phi_{ij} \neq 0 \text{ para todo } i, j \quad (4)$$

Estas cuatro presunciones vienen a plantear las condiciones de funcionamiento del modelo, orientando a su vez sobre los posibles problemas que éste muestra en su ajuste a los datos. Sin embargo, no es habitual que el sistema se formule para las variables expresadas en términos “brutos”, sino que éstas sufren una serie de transformaciones. Como veremos, la finalidad de estas transforma-

ciones es conseguir una mayor facilidad de estimación de parámetros, así como mejorar la comparabilidad entre los coeficientes. A su vez, dichas transformaciones dejarán su huella sobre las presunciones.

#### 5.4. TRANSFORMACIONES

La primera de las transformaciones produce efectos interesantes en el sistema de ecuaciones. En primer lugar suprime el coeficiente constante ( $\alpha$ ) de la ecuación. Debemos considerar que el coeficiente constante es un parámetro a estimar y sin embargo, con frecuencia, es un mero apoyo matemático para ajustar la solución. De hecho, al expresar el valor de la dependiente para determinadas combinaciones de valores de las que la explican, puede estar asociada a una situación sin significado. Eliminarla no supone ningún problema porque puede recuperarse en caso de que se necesite.

Otro efecto interesante es que las medias de las variables se hacen igual a 0. ( $\mu y_i^d = 0$  y  $\mu x_i^d = 0$ ). No obstante, no produce ningún efecto sobre los coeficientes, que permanecen expresados al igual que en la ecuación original. Para transformar las variables calculamos su desviación a la media.

$$y_i^d = y_i - \mu y_i \text{ para todos los } i$$

$$x_i^d = x_i - \mu x_i \text{ para todos los } i$$

El impacto sobre la notación es una  $d$  como superíndice sobre las variables.

$$y_1^d = \gamma_{11}x_1^d + \gamma_{12}x_2^d + \zeta_1$$

$$y_2^d = \beta_{21}y_1^d + \gamma_{21}x_1^d + \zeta_2$$

$$y_3^d = \beta_{31}y_1^d + \beta_{32}y_2^d + \gamma_{32}x_2^d + \zeta_3$$

$$y_4^d = \beta_{41}y_1^d + \zeta_4$$

$$y_5^d = \beta_{52}y_2^d + \beta_{53}y_3^d + \beta_{54}y_4^d + \zeta_5$$

Y una modificación en la primera presunción, para indicar que la media de todas las variables en la ecuación es igual a 0

$$\mu y_i^d = \mu x_i^d = \mu \zeta_i = 0 \text{ para todo } i \quad (1)$$

$$\text{Cov}(\zeta_i, x_j^d) = 0 \text{ para todo } i \quad (2)$$

$$\psi_{ij} = 0 \text{ para todo } i \neq j \quad (3)$$

$$\phi_{ij} \neq 0 \text{ para todo } i, j \quad (4)$$

Otra transformación muy frecuente consiste en normalizar las variables mediante la división de éstas, expresadas en desviación a la media, por la desviación típica de la variable.

$$y^s_i = y^d_i / \sigma_{y_i}$$

$$x^s_i = x^d_i / \sigma_{x_i}$$

Es importante notar que la transformación mediante la división de las variables por la desviación típica afecta a los parámetros y a su interpretación. Así, para un coeficiente normalizado la interpretación de  $\beta^s_{ij}$  es que  $y^s_i$  cambiará  $\beta^s_{ij}$  desviaciones típicas cuando  $y^s_j$  cambie una desviación típica, para todas las demás variables permaneciendo sin cambios. Tenemos una interpretación equivalente para el caso de los coeficientes normalizados que expresan los efectos directos de las variables exógenas  $y^s_{ij}$ . Así, una variable  $y^s_i$  cambiará  $y^s_{ij}$  desviaciones típicas cuando  $x^s_j$  cambie una desviación típica, para todas las demás variables permaneciendo constantes.

El sistema de ecuaciones se expresa, con notación simplificada para las variables normalizadas, introduciendo un superíndice “s” en todas las variables y parámetros así como un apóstrofe en el error.

$$y^s_1 = \gamma^s_{11} x^s_1 + \gamma^s_{12} x^s_2 + \zeta^s_1 y^s_2 = \beta^s_{21} y^s_1 + \gamma^s_{21} x^s_1 + \zeta^s_2$$

$$y^s_3 = \beta^s_{31} y^s_1 + \beta^s_{32} y^s_2 + \gamma^s_{32} x^s_2 + \zeta^s_3$$

$$y^s_4 = \beta^s_{41} y^s_1 + \zeta^s_4$$

$$y^s_5 = \beta^s_{52} y^s_2 + \beta^s_{53} y^s_3 + \beta^s_{54} y^s_4 + \zeta^s_5$$

En lo referido a las presunciones, es necesario añadir una quinta. Esta presunción indica que la variabilidad de las variables en el modelo (normalizadas) es igual a 1.

$$\mu y^s_i = \mu x^s_i = \mu \zeta^s_i = 0 \text{ para todo } i \quad (1)$$

$$\text{Cov}(\zeta^s_i, x^s_j) = 0 \text{ para todo } i, j \quad (2)$$

$$\psi^s_{ij} = 0 \text{ para todo } i \neq j \quad (3)$$

$$\Phi^s_{ij} \neq 0 \text{ para todo } i, j \quad (4)$$

$$\sigma^s_{x_i} = \sigma^s_{y_i} = 1 \text{ para todo } i \quad (5)$$

Consideremos las ventajas y desventajas de las diferentes formas de expresar los coeficientes, normalizados o no: una de las ventajas de emplear coeficientes no normalizados es que, en el caso de aplicar el modelo a diferentes poblaciones, éstos tenderán a ser los mismos, aún cuando la variabilidad interna de las variables no lo sea. Los coeficientes normalizados pueden cambiar más fácilmente (menos robustos al cambio) cuando se trata con poblaciones diferentes. Esto es debido a que los coeficientes normalizados son función de la desviación típica. Si varía la distribución típica de una variable al comparar

poblaciones distintas, provocará cambios en los coeficientes inducidos por dichas diferencias en la variabilidad. En ese sentido, cabe recomendar el uso de los coeficientes sin normalizar para ajustar modelos sobre diferentes poblaciones.

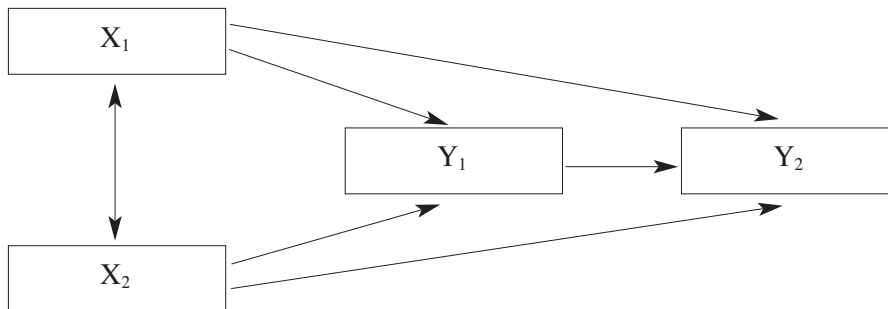
Por otra parte, si en el modelo se mezclan diferentes tipos de variables con rangos muy dispares y distintas escalas (tanto en el mismo modelo o para comparar entre modelos que provienen de diferentes investigaciones) será conveniente el empleo de coeficientes normalizados, dado que ello facilita la comparación entre modelos.

### 5.5. PARÁMETROS TEÓRICOS Y ESTIMADOS EMPÍRICOS

Se han definido los sistemas de notación de los modelos, así como los parámetros que los constituyen. Sin embargo, sobre la base de los datos sólo nos es posible obtener coeficientes de covarianza o de correlación, varianzas, etc. y desde ellos debemos definir los diferentes parámetros y efectos. Para ello se definen dos reglas de descomposición que vinculan teóricamente a los coeficientes y los parámetros del modelo.

Observemos el siguiente ejemplo. El modelo estructural tiene asociada un diagrama estructural, la matriz de correlaciones, un sistema de ecuaciones y sus presunciones estandarizadas.

*Diagrama estructural*



*Matriz de correlaciones*

$x_1$	$\rho_{x_1x_1}$			
$x_2$	$\rho_{x_2x_1}$	$\rho_{x_2x_2}$		
$y_1$	$\rho_{y_1x_1}$	$\rho_{y_1x_2}$	$\rho_{y_1y_1}$	
$y_2$	$\rho_{y_2x_1}$	$\rho_{y_2x_2}$	$\rho_{y_2y_1}$	$\rho_{y_2y_2}$
	$x_1$	$x_2$	$y_1$	$y_2$

*Sistema de ecuaciones*

$$y^s_1 = \gamma^s_{11} x^s_1 + \gamma^s_{12} x^s_2 + \zeta_1$$

$$y^s_2 = \beta^s_{21} y^s_1 + \gamma^s_{21} x^s_2 + \gamma^s_{22} x^s_2 + \zeta_2$$

*Presunciones estandarizadas*

$$\mu y^s_i = \mu x^s_i = \mu \zeta_i = 0 \text{ para todo } i \quad (1)$$

$$\text{Cov}(\zeta_i, x^s_j) = 0 \text{ para todo } i, j \quad (2)$$

$$\psi_{ij} = 0 \text{ para todo } i \neq j \quad (3)$$

$$\Phi^s_{ij} \neq 0 \text{ para todo } i, j \quad (4)$$

$$\sigma_{x^s_i} = \sigma_{y^s_i} = 1 \text{ para todo } i \quad (5)$$

## 5.6. PRIMERA REGLA DE DESCOMPOSICIÓN

Es denominada así dado que descompone la correlación observada entre variables en cuatro componentes de variación. Definición: el coeficiente de correlación entre dos variables es igual a la suma de los efectos directos, los efectos indirectos, las relaciones espurias y los efectos conjuntos. La diagonal de la matriz de correlaciones no se ve afectada por esta primera regla.

Fijémonos en el modelo del ejemplo anterior: la correlación observada entre las variables predeterminadas es igual al parámetro que expresa su covariación.

$$\rho_{x^s_1 x^s_2} = \phi^s_{12}$$

Del mismo modo observamos que la correlación entre  $x^s_1$  e  $y^s_1$  ( $\rho_{y^s_1 x^s_1}$ ) es igual a un efecto directo ( $\gamma^s_{11}$ ) más un efecto conjunto entre  $x^s_1$  e  $x^s_2$  ( $\gamma^s_{12} \phi^s_{21}$ ), por tanto su expresión sería:

$$\rho_{y^s_1 x^s_1} = \gamma^s_{11} + \gamma^s_{12} \phi^s_{21}$$

El último elemento de la matriz de correlaciones sería un efecto conjunto dado que no sabemos si es un efecto indirecto a través de  $x^s_2$  o espurio debido a  $x^s_2$ . De este modo terminaríamos de expresar cada una de los elementos de la matriz de correlaciones.

$$\rho_{y^s_1 x^s_2} = \gamma^s_{12} + \gamma^s_{11} \phi^s_{21}$$

$$\rho_{y^s_2 x^s_1} = \gamma^s_{21} + \beta^s_{21} \gamma^s_{11} + \gamma^s_{22} \phi^s_{21} + \beta^s_{21} \gamma^s_{12} \phi^s_{21}$$

$$\rho_{y^s_2 x^s_2} = \gamma^s_{22} + \beta^s_{21} \gamma^s_{12} + \gamma^s_{21} \phi^s_{21} + \beta^s_{21} \gamma^s_{12} \phi^s_{21}$$

$$\rho_{y^s_2 y^s_1} = \beta^s_{21} + \gamma^s_{21} \gamma^s_{11} + \gamma^s_{22} \gamma^s_{12} + \gamma^s_{22} \phi^s_{21} \gamma^s_{11} + \gamma^s_{21} \phi^s_{21} \gamma^s_{12}$$

## 5.7. SEGUNDA REGLA DE DESCOMPOSICIÓN

La segunda regla de descomposición responde de la variabilidad apreciada en la diagonal de la matriz de correlación. Definición: la varianza total de una variable endógena es igual a la cantidad de varianza explicada por las variables causantes de dicha variable endógena, más una cantidad de varianza no explicada. En definitiva lo que se viene a afirmar es que la varianza de la variable endógena estandarizada es igual a la varianza explicada por las variables estructurales y a la varianza no explicada por éstas. Dado que las variables están estandarizadas, su varianza es igual a 1. Debemos recordar que la varianza de las variables exógenas no se explica desde otras variables contenidas en el modelo, Por lo tanto, la varianza observada en las variables exógenas es igual a la varianza del modelo. Así, si utilizamos el ejemplo anterior, la descomposición de los elementos de la diagonal correspondientes a las variables exógenas sería la siguiente:

$$\rho_{x^s_1 x^s_1} = \phi^s_{11}$$

$$\rho_{x^s_2 x^s_2} = \phi^s_{22}$$

Por otra parte, en el ejemplo nos quedaría descomponer la varianza de las variables endógenas. La proporción de varianza explicada mediante un grupo de variables se denota:

$$R^2_{y_1 \cdot x_1, x_2}$$

Siendo  $R^2$  el coeficiente de determinación, la primera variable aquella endógena que se desea explicar y separada de las demás por un punto.

$$\rho_{y_1 y_1} = 1 = R^2_{y_1 \cdot x_1, x_2} + \psi'_{11}$$

$$\rho_{y_2 y_2} = 1 = R^2_{y_2 \cdot y_1, x_1, x_2} + \psi'_{22}$$

Es importante señalar que el coeficiente de determinación es una función de los parámetros del modelo estructural y no un nuevo parámetro del modelo. Se puede demostrar que "para cualquier variable endógena, la proporción de varianza explicada puede obtenerse sumando los productos de los efectos directos y los coeficientes de correlación entre la variable endógena y cada una de las variables causales que les afecta directamente".

$$R^2_{y_1 \cdot x_1, x_2} = \gamma^s_{11} \rho_{y_1 x_1} + \gamma^s_{12} \rho_{y_1 x_2}$$

$$R^2_{y_2 \cdot y_1, x_1, x_2} = \beta^s_{21} \rho_{y_2 y_1} + \gamma^s_{21} \rho_{y_2 x_1} + \gamma^s_{22} \rho_{y_2 x_2}$$

La proporción de varianza no explicada es igual a la varianza de las perturbaciones estandarizada  $\psi'_{ii}$

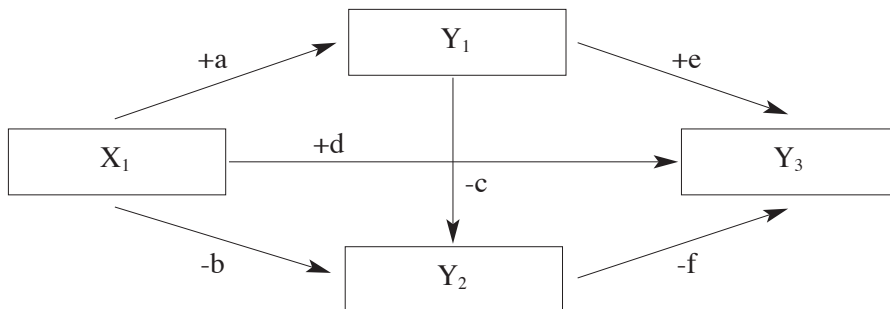


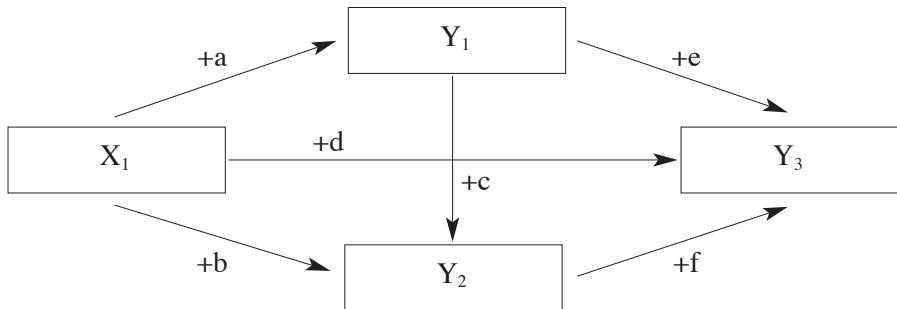
## 6. TIPOLOGÍA DE SISTEMAS

### 6.1. SISTEMAS SUPRESORES O DE REFUERZO

Como hemos visto antes, podemos establecer que las relaciones entre las variables tienen signo positivo o negativo, en función de que su covarianza sea positiva o negativa. En el contexto de un sistema que incluye múltiples efectos entre múltiples variables, y en función del signo que se establece en los diferentes senderos, podemos diferenciar entre sistemas consistentes o de refuerzo y sistemas inconsistentes o supresores. Esta clasificación expresa en qué medida las variables se potencian o no entre sí. En otras palabras, los efectos o relaciones entre variables tienen signos positivos o negativos, expresándose en función a la polaridad de las variables. Éste es un concepto importante. Supongamos dos variables con rango entre 1 y 10. El signo de su covariación puede ser positivo o negativo. Supongamos que es negativo, es decir, cuando una de ellas crece la otra decrece. Bastaría con “girar” la dirección de una de ellas para conseguir un signo positivo en la relación. Es importante mantener la significación en las relaciones y en algunas circunstancias la dirección de la escala es ciertamente arbitraria, como es en el caso de la ubicación ideológica (1 izquierda y 10 derecha o 10 izquierda y 1 derecha). Esta relación bivariable es extensible a los sistemas en su totalidad.

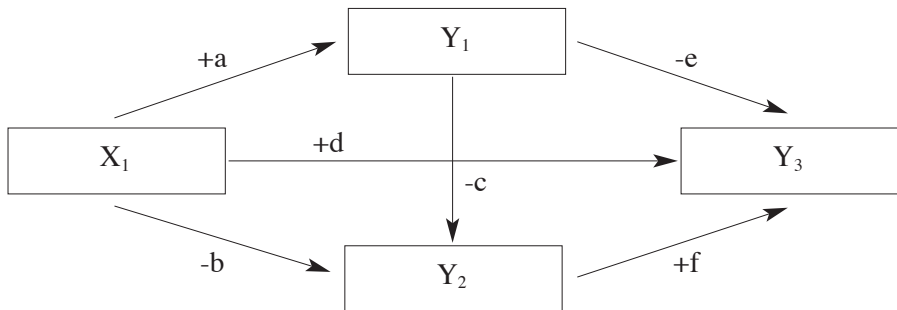
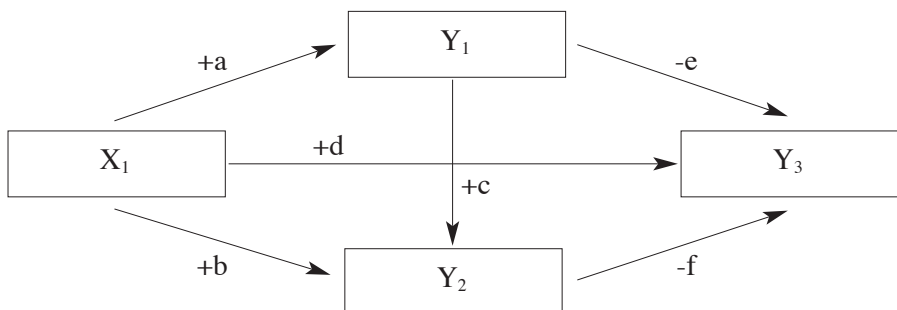
#### 1ª. Sistema consistente original

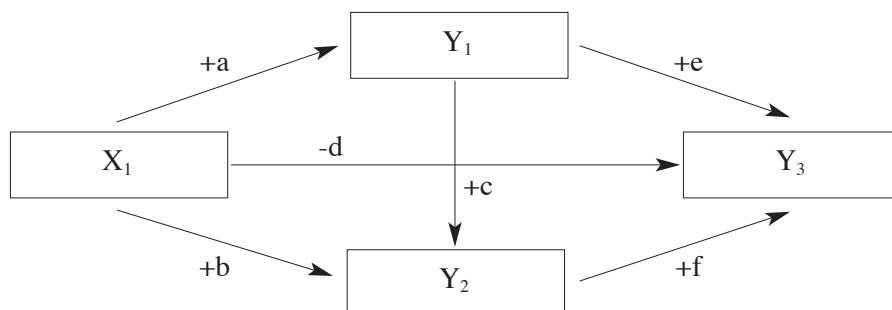


1b. Sistema consistente girando  $Y_2$ 

Se denominan sistemas inconsistentes a aquellos donde algunos de los componentes en una relación tienen signos contrarios o también supresores, en la medida que los efectos que influyen en sentido contrario reduce el efecto total presente en esa relación.

## 2ª. Sistema inconsistente original

2b. Sistema inconsistente girando  $Y_2$ 

2c. Sistema inconsistente girando  $Y_3$ 

El procedimiento para determinar el carácter supresor o de refuerzo de un sistema es el siguiente: un sistema es inconsistente si al menos un par de variables presenta simultáneamente signos positivos y negativos considerando tanto los efectos directos como los indirectos. Si no existe tal par de variables, el sistema es consistente. En un sistema consistente todos los coeficientes negativos pueden ser positivizados “girando” las variables (es decir, haciendo el mayor menor y el menor mayor). El procedimiento operativo se basa en esta última apreciación. En primer lugar se determina cuál es la variable que recibe más efectos negativos y se gira su polaridad. Se determina cuál es la siguiente que recibe más signos negativos y se procede igual. Si al proceder así se eliminan todos los signos negativos, el sistema es consistente o de refuerzo; si no es posible, es un sistema supresor.

Esta característica de efecto supresor o reforzador es importante tanto en sentido técnico como en términos de argumentación de una explicación. En lo que se refiere a la capacidad explicativa, los sistemas reforzadores tienden a expresar situaciones de “statu quo” al reforzarse el efecto de las variables entre sí dentro del sistema, como es el ejemplo de las clases sociales. Así, la clase social de los padres tiene un efecto directo positivo sobre la clase social de los hijos y los diferentes senderos que establecen las variables intervinientes refuerzan ese efecto. Inversamente, los sistemas supresores tienden a corresponder con la noción de “consecuencias no esperadas” de forma que X tiene un efecto directo positivo sobre Y, pero al mismo tiempo genera una cadena estructural que tiende a disminuir o reducir el efecto final. Por ejemplo, el nivel educativo tiende a producir una relación positiva con respecto a temas sociales (a mayor nivel, mayor comprensión). No obstante, el mayor nivel educativo también correlaciona bien con ingresos (a más nivel educativo, más ingresos), pero ingresos se relaciona negativamente con la aceptación de temas sociales (a más ingresos, menor aceptación). En ese sentido,

educación tiene un efecto directo positivo con respecto a la aceptación de políticas sociales (+) y uno indirecto negativo (+ por - = -) mediante la variable ingresos. De este modo, nivel educativo y aceptación de políticas sociales forman un sistema inconsistente.

La importancia de la congruencia del sistema también afecta cuestiones de carácter técnico, como son la determinación de los efectos totales. En un sistema de refuerzo el efecto directo de  $X_i$  sobre  $Y_j$  siempre será de una magnitud igual o inferior al efecto total. Por el contrario, en un sistema supresor el efecto directo entre dos variables puede ser superior al efecto total de dicha variable. El efecto total se refiere a la suma de todos los efectos (directos e indirectos) de una variable sobre otra.

No obstante debe destacarse que aunque la polaridad (qué cifra se atribuye a lo que es mayor y a lo que es menor) en que se expresa una variable es arbitraria, es muy interesante intentar mantener una coherencia lógica argumental que no violente la explicación en dependencia de la polaridad de la variable. Ambos factores deben de ser tenidos en consideración.

## 6.2. MODELOS RECURSIVOS Y NO RECURSIVOS

Las diferentes tipologías de sistemas estructurales se establecen sobre la base de diferentes criterios que dan pie a conjuntos específicos de terminologías. No obstante, tal y como advirtiera Bentler (1994), todas las tipologías se apoyan sobre la noción básica de un conjunto de ecuaciones estructurales lineales. Las variantes simplemente expresan las diferentes formas que este conjunto de ecuaciones adquiere en función a la finalidad de su utilización. Así, se diferenciará entre sistemas recursivos o no recursivos en función a la direccionalidad del sistema, según esté totalmente ordenado o no. Uno de los aspectos principales de esta diferencia es el problema de la identificación. Es decir, de la complejidad que puede suponer la resolución matemática del sistema.

El análisis estructural puede emplearse combinando variables latentes, (del mismo modo que el análisis factorial), junto con otras variables dentro del modelo explicativo; así mismo, puede referirse a datos en un solo momento del tiempo o en varios (como en el análisis de panel) o en simulaciones mediante ecuaciones simultáneas, etc. En cualquiera de estas formas de utilización, el elemento básico es la idea de estructura.

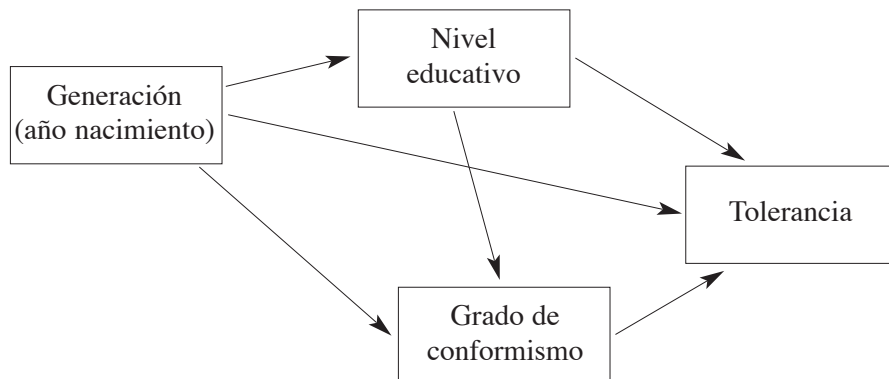
En general, podemos considerar una distinción importante entre dos tipos de sistemas, los sistemas recursivos y los no recursivos. Los modelos recursivos son aquellos modelos estructurales en los que todos los efectos estructurales se establecen en una sola dirección; es decir, se determinan relaciones asimétricas unidireccionales (donde los errores o perturbaciones están

incorrelacionados entre las diferentes ecuaciones). Es decir, un modelo recursivo será:

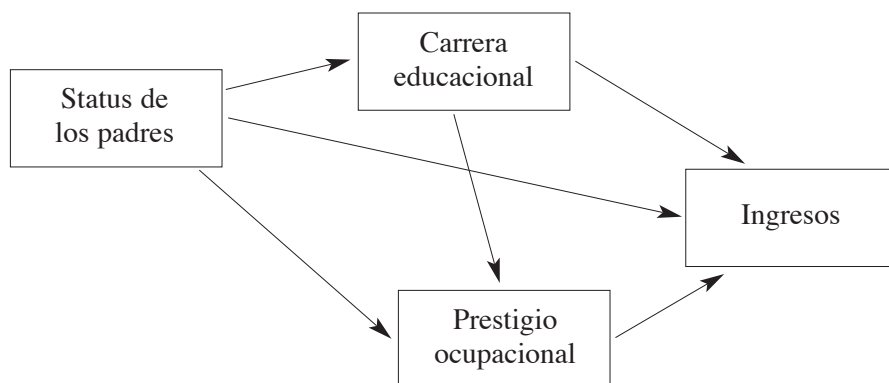
1) jerárquico, donde todas las variables en el modelo pueden ser ordenadas y etiquetadas en una secuencia  $y_1, y_2, y_3, y_4, \dots, y_n$  de tal modo que para todo  $y_i$  e  $y_j$  donde  $i < j$ ,  $y_j$  no se presenta como causa de  $y_i$ . Por lo tanto  $\beta_{ij}$  será igual a cero. Según esto, la primera variable endógena sólo podrá ser influida por una variable exógena. La segunda endógena sólo podrá ser influida por una exógena o la endógena anterior y así sucesivamente. Según este criterio de jerarquía, en un modelo recursivo no pueden aparecer relaciones recíprocas entre dos variables ni puede pasar que una variable endógena pueda influir mediante un efecto indirecto sobre otra anterior.

### Modelos recursivos

#### a) Estatus socioeconómico



#### b) Tolerancia a lo distinto



2) los errores deben de estar incorrelacionados entre sí y respecto a las variables exógenas. Esta característica permite estimar los coeficientes mediante Mínimos Cuadrados Ordinarios de forma insesgada<sup>20</sup> y consistente. En ese sentido, los modelos estructurales recursivos son fáciles de estimar.

No obstante, en muchas ocasiones ambas presunciones son poco realistas. Con frecuencia, en muchos análisis es dudoso que las presunciones sean apropiadas. Por ello, no debe optarse por un modelo recursivo a la ligera, por comodidad o por conveniencia. A menos que se esté perfectamente convencido de que las relaciones son estrictamente unidireccionales (jerárquicas) y de que los factores (o variables no incluidas en el análisis) que están contribuyendo al error de cada ecuación son distintos para cada ecuación (es decir, que no hay factores que influyan en común sobre ambas ecuaciones) no debe optarse por un modelo recursivo. El problema no debe ser de comodidad sino de acierto en la descripción completa y realista de un fenómeno social. Consideremos que si las presunciones no son ciertas (jerarquía e independencia de los errores) los estimados de los coeficientes (mediante Mínimos Cuadrados Ordinarios, OLS) serán inconsistentes y sesgados, con lo cual no sólo no habremos esclarecido nada, sino que lo habremos oscurecido.

Los modelos no recursivos, por el contrario, postulan la posibilidad de efectos recíprocos, o con carácter más general, que se produzcan efectos en ambas direcciones dentro del sistema. Un caso límite de no-recursividad lo plantean los modelos completamente no recursivos. En un modelo completamente no recursivo, todas las variables endógenas se ven afectadas por todas las demás variables endógenas y exógenas presentes en el modelo. No obstante, independientemente de su utilidad para la investigación, no es conveniente definir modelos estructurales completamente no recursivos dado que dichos modelos son siempre subidentificados. Por el contrario, alguno de los parámetros del modelo no recursivo se supone que es igual a cero. Recordemos que un parámetro fijado a cero implica que hemos postulado que no existe un efecto entre dos variables. Como tendremos ocasión de comprobar cuando se considere el problema de la identificación, las presunciones que se adopten en el modelo recursivo serán de gran importancia para sus posibilidades de identificación. En general, las presunciones que empleemos serán que la media de las variables y la de los errores serán igual a cero (transformación mediante desviación a la media) y que los errores están incorrelacionados con las varia-

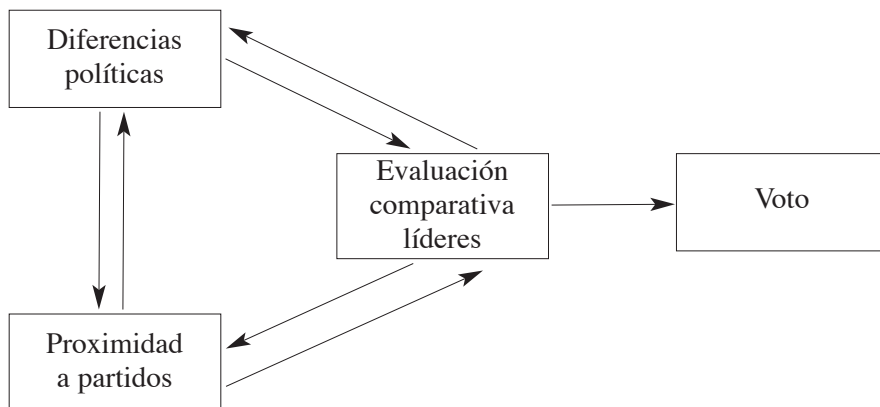
---

20. El término insesgado se refiere a aquel estimado que, como media, es igual al valor real del parámetro. Por otra parte, el término consistente se refiere a aquel estimado cuya distribución, cuando la muestra se aproxima a infinito, se aproxima a una distribución con la mayor probabilidad de estar centrada sobre el parámetro.

bles independientes. En un modelo no recursivo no tiene mucho sentido plantear que todos los errores están incorrelacionados con todas las variables endógenas: siempre hay algún error que estará relacionado con alguna variable endógena, por el mismo planteamiento del modelo. Por el contrario, la presunción útil y que puede tener sentido teórico en un modelo no recursivo es que los errores están incorrelacionados entre sí.

### *Modelos no recursivos*

#### a) Comportamiento electoral (Page y Jones)



Podemos apreciar que mientras en los modelos recursivos la explicación está ordenada de forma asimétrica en una sola dirección, en los modelos no recursivos, aparecen relaciones que invierten el orden de la causalidad, estableciendo relaciones recíprocas. Esta distinción es especialmente eficaz en términos de identificación del sistema, es decir, implica aspectos esencialmente técnicos en tanto permite o no tener soluciones. Desde el punto de vista de la explicación es evidente que los modelos estructurales no recursivos son bastante más realistas que los modelos recursivos. No obstante y como veremos, los problemas que plantean en términos de identificación los hace bastante poco frecuentes.



## 7. EL PROBLEMA DE LA IDENTIFICACIÓN

El concepto de identificación está ligado a las operaciones matemáticas que se realizan para efectuar el ajuste del modelo sobre los datos de que se disponen. En función al estado de identificación del modelo podrá o no tener un conjunto de soluciones que sean operativas para el investigador. De este modo, podremos afirmar que una ecuación está identificada (y un modelo estructural en general) cuando sus parámetros se pueden determinar de modo único a partir del conocimiento que se puede extraer de un conjunto de observaciones completas y adecuadas. Lo primero que debe destacarse es que el problema de la identificación del sistema no es un problema de inferencia estadística. Un modelo no tendrá problemas de identificación por más inestable que sea la muestra que facilita la información para ajustar el modelo. El problema de la identificación se refiere a la relación entre información y parámetros a estimar. Se trata en definitiva de poseer más hipótesis que información para testarlas. En resumen, la identificación del sistema no es un concepto que esté relacionado con la calidad de los datos o de la medición. Incluso con los mejores datos, es decir, con indicadores válidos y fiables procedentes de una gran muestra, puede surgir el problema de la identificación. La identificación está directamente relacionada con la especificación del sistema, es decir, con las relaciones que planteamos que existen entre variables a efectos de explicar un fenómeno social.

Podemos efectuar un planteamiento intuitivo desde el álgebra mediante el examen de un sistema de ecuaciones. Básicamente, la cuestión de la identificación se refiere a tener la suficiente información para obtener una solución única a un conjunto de incógnitas. Así, por ejemplo:

a) **Identificación exacta.** El siguiente sistema de ecuaciones constituye un sistema exactamente identificado:

$$\begin{aligned}2x + 3y &= 7 \\ x - 4y &= -2\end{aligned}$$

Ello es así porque hay tantas ecuaciones linealmente independientes entre sí como incógnitas. Así, obtenemos una solución única donde:

$$x = 2$$

$$y = 1$$

**b) Subidentificación.** La subidentificación aparece cuando poseemos más incógnitas que ecuaciones linealmente independientes entre sí. Por ejemplo, en el sistema de ecuaciones:

$$2x + 3y = 7$$

$$4x + 6y = 14$$

Este sistema está subidentificado, dado que aún cuando hay dos ecuaciones con dos incógnitas, la segunda es simplemente la primera multiplicada por dos (son linealmente dependientes). Es decir, al ser la segunda ecuación simplemente la primera multiplicada por dos, no aporta ninguna información nueva que ayude a resolver de un modo único las incógnitas X y Y. De hecho, sólo tendremos una ecuación con dos incógnitas, lo que lleva a un conjunto infinito de soluciones. Por ejemplo, las siguientes pueden ser soluciones al sistema anterior:

$$x = 2 \qquad y = 1$$

$$x = 3,5 \qquad y = 0$$

$$x = 5 \qquad y = -1$$

Como nuestra intención es obtener unos estimados con significado teórico para ese conjunto de incógnitas, la existencia de infinitas soluciones es una situación indeseable.

**c) Sobreidentificación.** Una situación semejante puede aparecer cuando poseemos un número mayor de ecuaciones que de incógnitas. Por ejemplo, el sistema de ecuaciones linealmente independiente que mostramos a continuación posee dos incógnitas y tres ecuaciones.

$$2x - y = 7 \qquad (1)$$

$$x + 3y = 0 \qquad (2)$$

$$3x - 2y = 2 \qquad (3)$$

Si se emplearan las ecuaciones (1) y (2) para resolver el sistema, obtendremos una solución única para ese sistema de dos ecuaciones:

$$x = 3$$

$$y = -1$$

Las ecuaciones (1) y (3) dan como resultado:

$$x = 12$$

$$y = 17$$

Las ecuaciones (2) y (3) ofrecen como resultado las soluciones:

$$x = 6/11$$

$$y = -2/11$$

El término identificación y cada uno de los estados posibles (sub, exacta y sobre) se refieren tanto a las ecuaciones estructurales por separado como al conjunto del sistema de ecuaciones. Así, diremos que un modelo estructural está identificado si todas y cada una de las ecuaciones que lo componen están identificadas. Por el contrario diremos que un sistema no está identificado cuando alguna de sus ecuaciones esté subidentificada. Como podemos apreciar, la solución del sistema depende de la relación entre información e incógnitas. En el caso de la subidentificación, no tendremos solución posible (cualquier solución será indeterminada); en el caso de la identificación exacta tendremos la posibilidad de estimar los parámetros mediante una solución única. Sin embargo, la situación más interesante se produce en caso de la sobreidentificación. En este caso, como podremos apreciar, se presenta la posibilidad de testar la bondad del ajuste del modelo.

Seguidamente vamos a considerar en primer lugar algunos criterios para identificar el estado del sistema de ecuaciones, para después plantear las posibles alternativas de los modelos no identificados. Evidentemente, en la medida que el problema de la identificación es un problema de especificación, sólo una reelaboración de la explicación (es decir del modelo y de la teoría) puede ofrecer soluciones.

### 7.1. LA DETERMINACIÓN DEL ESTADO

El problema de la identificación tiene consecuencias diferentes según se trate de sistemas recursivos o no recursivos. Como veremos, en el caso de los sistemas recursivos siempre existe la posibilidad de establecer restricciones que permitirán identificar (y solucionar) el sistema en todos los casos. No sucede lo mismo con los sistemas no recursivos, donde en determinadas situaciones su identificación requerirá necesariamente la modificación de éste (introduciendo nuevas variables o restricciones de coeficientes o covarianzas).

Podemos preguntarnos qué es lo que hace a los modelos no recursivos especiales en términos de identificación. En principio, de forma intuitiva podríamos pensar que contando con suficiente datos el sistema debería tener solución. Sin embargo consideremos el ejemplo siguiente:



$$Y_1 = \beta_{12}Y_2 + \zeta_1$$

$$Y_2 = \beta_{21}Y_1 + \zeta_2$$

Este modelo no está identificado; pero esto es evidente, en la medida que es imposible determinar en qué sentido se desplaza la causalidad (cuando sólo tenemos datos referidos a un solo punto en el tiempo). Así, solamente con el dato de la covariación entre ambas variables no existe ninguna forma matemática de distribuir cuánta covarianza corresponde al efecto de  $y_1$  sobre  $y_2$  y cuánta corresponde al efecto inverso, de  $y_2$  sobre  $y_1$ . Esto puede pasar perfectamente en un modelo no recursivo más complejo. Además, en relación con los modelos recursivos, en un modelo no recursivo existen en general más parámetros a estimar, incluso poseyendo el mismo número de variables. Otro aspecto que influye en el problema de la identificación de modelos no recursivos es el hecho de no postular que los errores son independientes entre sí. Por lo tanto, los criterios de identificación que introduciremos seguidamente son especialmente pertinentes en el caso de los modelos no recursivos.

Vamos a considerar tres criterios de evaluación del estado del modelo. El primero de ellos va a considerar el sistema en conjunto y por lo tanto aportará un diagnóstico global. En estas condiciones se tiene poca información para intervenir sobre el modelo, si bien es un procedimiento rápido de diagnóstico. Una mayor utilidad a efectos de intervenir en el caso de no identificación del sistema son los procedimientos que evalúan el estado de cada una de las ecuaciones del sistema. De este modo, identificando las ecuaciones problemáticas es posible intervenir sobre las relaciones de forma que se posibilite la identificación del sistema. En los modelos no recursivos se emplearán otros dos medios para evaluar las posibles restricciones de coeficientes, las condiciones de rango y las condiciones de orden. Como se ha dicho, estos procedimientos operan evaluando cuál es la situación de cada ecuación; esto viene dado porque en un modelo podrían existir ecuaciones subidentificadas junto a otras identificadas exactamente y otras sobreidentificadas. El poder detectar cual es la situación de cada ecuación dentro del sistema ayuda claramente en el procedimiento de identificación global del sistema de ecuaciones. El procedimiento que evalúa directamente el sistema de ecuaciones en conjunto no ofrece una orientación con respecto a cómo corregir el sistema de modo

que, como mínimo, se determine un conjunto finito de soluciones para las incógnitas (parámetros) a estimar. Por último, recordar que la identificación, en la medida que depende de la especificación, se verá afectada por las presunciones sobre el error. Aquí consideraremos, tal como se advirtió inicialmente, que las medias de los errores y las variables son cero (desviaciones o normalización) y que los errores son independientes de las variables exógenas (es decir, no covarían). En los sistemas donde se planteen otras presunciones la identificación por los siguientes procedimientos puede verse afectada. Los dos primeros procedimientos son condiciones<sup>21</sup> necesarias pero no suficientes. El último procedimiento es condición suficiente.

## 7.2. IDENTIFICACIÓN DEL SISTEMA

Como sabemos, el problema que queremos solucionar es si los parámetros estructurales de un modo pueden ser determinados de forma única sobre la base de la información que se disponga de varianzas y covarianzas entre las variables observadas. Una regla general en álgebra es que una condición necesaria para resolver las incógnitas en un sistema de ecuaciones es que el número de incógnitas debe ser igual o inferior que el número de ecuaciones (linealmente independientes entre sí, claro está). Las incógnitas, en este caso, son los parámetros estructurales. Es posible determinar el número de ecuaciones (en términos de descomposición de efectos, es decir las varianzas y las covarianzas por un lado y por el otro los parámetros). Las ecuaciones, insistimos, se refieren a las correspondientes a la relación entre parámetros y varianzas y covarianzas. En ese sentido, es fácil apreciar que tendremos tantas ecuaciones como varianzas y covarianzas. Así, si en un modelo tenemos 4 variables (tanto exógenas como endógenas) el número de ecuaciones será igual a  $\frac{1}{2}n(n+1)$ , siendo  $n$  el número de variables. Así, tendríamos  $\frac{1}{2}4(4+1) = 10$  ecuaciones. La diferencia entre el número de ecuaciones y el número de parámetros estructurales a estimar se denomina *grados de libertad* y se notan como  $df$  (del inglés *degrees of freedom*). Una vez definidos estos términos, el criterio para identificar el sistema puede formularse como sigue:

Una condición necesaria para la identificación de un modelo de ecuaciones estructurales es que los grados de libertad deben ser iguales o mayores que cero, es decir  $df \geq 0$ . Los grados de libertad resultan de comparar la información de que se dispone (varianzas y covarianzas) con los parámetros del modelo que deben estimarse.

---

21. Necesaria pero no suficiente. Quiere decir que si no se cumple esa condición la ecuación no puede ser identificada. Si la condición se cumple, la ecuación puede o no puede ser identificada, pero existe la posibilidad.

Como ya hemos dicho, la forma de contabilizar el número de ecuaciones (varianzas y covarianzas) es directo, contabilizando las variables exógenas y endógenas del modelo, dividiendo por dos y multiplicando por el número de variables más uno:

$$\frac{1}{2} n(n+1)$$

Sin embargo, el número de incógnitas puede ocasionar dudas, dado que depende de la especificación del modelo. En principio, dado que consideramos las variables expresadas en desviación a la media o normalizadas, la constante  $\alpha$  desaparece de la ecuación, eliminando una incógnita a estimar. Sin embargo, permanecen como incógnitas los parámetros  $\beta$ ,  $\gamma$ ,  $\psi$ ,  $\phi$ . El número de parámetros  $\beta$ ,  $\gamma$  puede determinarse directamente de las ecuaciones o de los diagramas. El número de parámetros  $\phi$  (correspondientes a las varianzas y covarianzas de las variables exógenas) es igual a  $\frac{1}{2} q(q-1)$ , siendo  $q$  el número de variables exógenas ( $x$ ). El número de parámetros  $\psi$  es como mínimo  $p$ , siendo  $p$  el número de varianzas de los errores. El total que resulta de sumar los parámetros anteriores expresa el número de incógnitas a resolver. Es decir, que trabajando con variables expresadas en desviación sobre la media sólo se trata de contar los parámetros  $\beta$ ,  $\gamma$ ,  $\psi$ ,  $\phi$ .

### 7.3. CONDICIONES DE ORDEN

Técnicamente, la condición de orden es una condición necesaria, pero no suficiente, para la identificación de una ecuación. Sin embargo, en muchas de las situaciones que se producen en la práctica al analizar datos, esta condición funciona como necesaria y suficiente. La condición de orden afirma que si tenemos un modelo consistente en  $K$  ecuaciones lineales, para que cualquier ecuación en el modelo esté identificada debe de excluir como mínimo un número de variables igual (o mayor) a  $K-1$ , de entre todas las variables que aparecen en el modelo. Evidentemente el número de ecuaciones será igual al número de variables que tratamos de explicar, es decir, al número de variables endógenas o endógenas intervinientes.

Por ejemplo, en el caso que un sistema posee 12 ecuaciones y 15 variables, para que una ecuación cualquiera esté identificada debe de excluir 11 variables ( $12-1$ ) de entre todas las que aparecen en el modelo. Es decir, las ecuaciones identificadas deben de excluir 11 variables (sus coeficientes = 0) y retener 4 variables (con coeficientes  $\neq 0$ ).

### 7.4. CONDICIONES DE RANGO

La condición de rango afirma (Christ, 1966) que una ecuación en un modelo de  $K$  ecuaciones lineales está identificada si existe un determinante de cual-

quier submatriz de coeficientes  $K-1$  dentro de la matriz que resta después de omitir todas las columnas donde la ecuación a identificar posea coeficientes distintos de 0 y omitiendo la ecuación a identificar. El proceso se repetirá hasta identificar cada ecuación. Hecho esto, evaluaremos de la siguiente manera:

- Si no existiese ninguna submatriz de rango  $K-1$  con determinante  $\neq 0$  la ecuación esta subidentificada.
- Si existe sólo una submatriz de rango  $k-1$  con determinante  $\neq 0$  la ecuación está identificada exactamente.
- Si existe más de una submatriz de rango  $k-1$  con determinante  $\neq 0$  la ecuación está sobreidentificada.

La condición de rango es una condición necesaria y suficiente para identificar una ecuación.

Una vez considerados los diferentes procedimientos para comprobar el estado del sistema de ecuaciones, podemos ofrecer algunas conclusiones generales que servirán para diagnosticar en que status se encuentran en la práctica los modelos:

(1) Los modelos de una sola ecuación estructural donde el error no covaría con las variables exógenas (son independientes) están siempre identificados. Son conocidos como modelos de regresión. ( $\sigma_{\zeta_{ixj}} = 0$ )

(2) Los modelos de ecuaciones sin efectos estructurales recíprocos y con las presunciones  $\sigma_{\zeta_{ixj}} = 0$  para todo  $i, j$  y que  $\sigma_{\zeta_{i\zeta_j}} = 0$  para todos los  $i \neq j$ , siempre están identificados. Son denominados modelos recursivos.

(3) Los modelos donde existen efectos de  $x_i$  sobre  $y_j$ , existiendo covariación entre las variables exógenas y el error, no están identificados.  $\sigma_{\zeta_{ixj}} \neq 0$ .

(4) Los modelos estructurales con efectos estructurales recíprocos (modelos no recursivos) no estarán identificados en el caso particular en que el conjunto de variables endógenas se vean afectadas todas ellas entre sí.

Esencialmente, las conclusiones que pueden extraerse de las observaciones anteriores es que todos los modelos recursivos de ecuaciones estructurales (1) (2) están identificados, siempre que las variables importantes se encuentren presentes en el modelo. Es decir, que la especificación sea la correcta. Es fundamental que todas las variables importantes estén en el modelo como garantía de que las presunciones se podrán cumplir; es decir, que no covariarán las variables exógenas con el error y los errores estarán incorrelacionados entre sí.

La conclusión tercera expresa la importancia de la presunción acerca de que las variables exógenas no deben covariar con el error. Como sabemos,

la covariación entre las variables exógenas y el error indicará la existencia de causas comunes omitidas. Cuando exógena y error covarían puede ignorarse tal covariación, pero la consecuencia será que los parámetros  $\gamma$  estimados serán erróneos. La otra opción es introducir la covariación entre exógena y error en el modelo, entre las presunciones, pero entonces es probable que el modelo se convierta en subidentificado. En definitiva, todo apunta al hecho de que la incorporación de causas comunes es realmente vital para la consistencia explicativa y matemática del modelo.

Por último, la cuarta conclusión indica en qué condiciones un modelo no recursivo no puede ser identificado.

Como hemos podido apreciar, la identificación aparece como un problema especialmente en los sistemas no recursivos. En todo caso, podemos plantear algunas orientaciones para atenuar los problemas de identificación y sabiendo de antemano que restarán modelos matemáticamente no identificables. Dado que las condiciones de orden y de rango nos indican si una ecuación está subidentificada, identificada exactamente o sobreidentificada, el problema consiste en cómo actuar sobre las ecuaciones que plantean problemas.

#### 7.5. LOS PROCEDIMIENTOS DE RESTRICCIÓN

Existen dos procedimientos básicos de tipo restrictivo para intentar que un sistema de ecuaciones esté identificado, las restricciones de coeficientes y las restricciones de covarianzas. Un tercer procedimiento consiste en la introducción de nuevas variables explicativas en el modelo. Las restricciones de coeficiente actúan imponiendo limitaciones sobre los coeficientes que unen las variables medidas, por ejemplo fijándolos a cero. Por su parte, las restricciones de covarianza efectúan presunciones sobre la correlación entre las variables residuales.

En los modelos recursivos la identificación es más simple dado que, por ejemplo, en este caso la mitad de los coeficientes son igual a cero (puesto que no hay efectos recíprocos). Así, en un modelo recursivo afirmar que existe una relación entre  $Y_1$  e  $Y_2$  implica que el efecto inverso no se va a dar. Además, sabemos que en los modelos recursivos se efectúan presunciones sobre el error que si bien no son realistas, sí se corresponden con la estructura teórica del modelo que se propone (asimétrico). En ese sentido, efectuando las presunciones habituales en un modelo recursivo tendremos garantía de que estará identificado (Boudon, 1971).

Por ejemplo, consideremos un sistema no recursivo con tres variables con efectos recíprocos entre ellas. De acuerdo al criterio de sistema  $\frac{1}{2}3(3+1)$ , obtenemos 6 ecuaciones. Por otro lado tenemos 9 incógnitas, compuestas por 6

coeficientes y 3 errores. Tendríamos más incógnitas que ecuaciones. Este es, como sabemos, un caso evidente de subidentificación o falta de información. Un procedimiento para solucionar el sistema es mediante restricciones de coeficientes y de covarianzas. Si lo convertimos en un modelo recursivo fijaremos tres coeficientes a 0. Contando con la presunción recursiva donde las covarianzas de las tres variables residuales están incorrelacionadas, obtendremos finalmente seis ecuaciones con seis incógnitas. Tendríamos con ello una identificación exacta. Como ya sabemos, el planteamiento de un modelo recursivo es correcto siempre que tengamos seguridad de que las causas comunes han sido incluidas en él. En ese sentido, la especificación del modelo es una fase especialmente ligada a la verosimilitud y fiabilidad de los coeficientes estimados finalmente. Es decir, de la fiabilidad del modelo.

En los modelos no recursivos, por el contrario, la situación se complica, en la medida que la inclusión de todas las causas comunes no garantiza la identificación del modelo, y por lo tanto su resolución. Cuando estamos considerando un modelo no recursivo no es posible efectuar las restricciones de los modelos recursivos. Por ejemplo, en este tipo de modelos no son practicable las presunciones que establecíamos en los modelos recursivos acerca de las covarianzas entre las variables residuales. Ello convierte los modelos no recursivos en modelos que se aproximan más a la realidad, dado que no presumen que las variables residuales están incorrelacionadas. Sin embargo, al eliminar esa restricción sobre las covarianzas de los errores se complica la tarea de la identificación. En los modelos no recursivos imponemos menos restricciones sobre los coeficientes y covarianzas, lo que conlleva un número mayor de incógnitas y una mayor dificultad para obtener soluciones únicas. Además, el problema más frecuente se refiere a la situación donde las variables que explican en cada ecuación a las distintas endógenas tienden a repetirse en las diferentes ecuaciones.

La manera de intentar identificar los modelos no recursivos (y que por lo tanto tengan solución) pasa por aplicar las condiciones de orden y de rango, de forma que se pueda identificar las ecuaciones infraidentificadas. Cuando una ecuación está subidentificada no existe ninguna técnica de estimación que ofrezca estimados válidos. Por ello, hay que intentar transformar una ecuación subidentificada en otra identificada, generalmente introduciendo nuevas variables en el modelo. Estas variables nuevas a introducir en el modelo deberán afectar (explicar) sólo a determinadas variables (aquellas con ecuación infraidentificada). En ese sentido, la identificación mediante la introducción obligatoria de nuevas variables y condicionadas a una relación concreta supone en la mayoría de los casos una cierta violencia y forzamiento teórico del modelo. Por ello, aún cuando las modificaciones del modelo vengan

impuestas desde la necesidad de identificación, la introducción de nuevas variables debe estar, en primer lugar, teóricamente orientada. Es la teoría la que debería tener la última palabra en el sentido de indicar si es posible introducir nuevas variables, cuáles deban de ser éstas, así como su relación con las variables endógenas del modelo.

Esto último es una cuestión importante, dado que no por el hecho de introducir nuevas variables se va a facilitar la identificación del sistema de ecuaciones, sino que esto dependerá de las pautas de asociación propuestas para las nuevas variables. Una asociación u otra facilitará la identificación o no<sup>22</sup>.

En una segunda instancia, es conveniente que esas nuevas variables posean determinadas propiedades estadísticas, algunas de las cuales son consecuencia directa de la sensatez teórica. En primer lugar, es conveniente que las nuevas variables sean variables exógenas, y no correlacionadas con el error de las variables endógenas. Además, deben de estar fuertemente asociadas con aquellas variables a las que están afectando teóricamente (Fisher, 1971). La búsqueda de variables exógenas (predeterminadas) con dichas características no siempre es fácil. Las alternativas son desfigurar el modelo explicativo o abandonar cualquier esperanza de solución. En cualquier caso, la introducción de nuevas variables aparece como alternativa a la supresión de efectos (coeficientes = 0). En principio no deberían suprimirse relaciones entre variables que supongan una especificación importante del modelo, especialmente porque la supresión de efectos importantes, si realmente los son, puede sesgar la fiabilidad de los demás parámetros dentro del modelo.

Como podemos apreciar, las condiciones que la identificación impone sobre el modelo explicativo son bastante importantes. En el caso de los modelos recursivos, porque presume condiciones drásticas de jerarquía y de completitud de la especificación (incluyendo todas las causas comunes importantes). En el caso de los no recursivos, al imponer la introducción de nuevas variables y además con una función relacional obligada, es decir afectando a determinadas variables y no a otras. Por otro lado, la supresión de coeficientes (mediante la fijación de los efectos a cero) que fueron introducidos previamente en la fase de especificación del modelo implica la amenaza de sesgar los resultados estimados. Como puede verse, no son despreciables las consecuencias de la identificación (relación información e incógnitas) en la explicación que se pretende ofrecer.

---

22. Muy probablemente, el principio de parsimonia ha establecido teóricamente la conveniencia de simplificar el modelo. Puede ser conveniente a efectos de la identificación del sistema de ecuaciones recuperar variables interesantes, pero que han sido previamente descartadas por ese criterio de simplificación.

El dilema es evidente, mantener una explicación que no podrá ser testada o degenerar, por imposición matemática, el modelo explicativo en función a sus posibilidades de solución. No se trata de modificaciones introducidas por el ajuste del modelo, donde las covarianzas encontradas en la estructura de los datos (entre errores y entre variables) imponen una revisión de lo que se pensaba, sino modificaciones conducidas por la mecánica interna del modelo propuesto. No es demasiado atractivo que la técnica de modelado de la realidad (explicación de ésta) determine las características finales de esta explicación. Evidentemente, las transformaciones del modelo explicativo son un aspecto crucial de la tarea de investigar. Éstas deberán desarrollarse siempre que sean teóricamente aceptables en el caso de sistemas subidentificados. No sería aceptable que algo tan importante como es una explicación de los fenómenos sociales se vea sesgada por la necesidad de modificarla a efectos de ser matemáticamente solucionable. La prioridad debe ser siempre la mejor explicación, no la explicación que mi método de análisis de la realidad me ha permitido o me ha obligado a producir.

Éste es un fenómeno que supone un riesgo evidente, en la medida que la dinámica de modelado puede conducirnos fuera de la explicación, a un terreno donde las reglas de juego las imponen las matemáticas. No debería actuarse con timidez o complacencia, y una opción a plantearse seriamente, dependiendo de las condiciones teóricas que imponga la identificación del sistema, sería optar por otra estrategia de modelado que permita vías alternativas de testar la explicación. La subidentificación supone riesgos teóricos importantes, donde una de las principales ventajas es una nueva oportunidad para repensar el modelo (la explicación que se ofrece).

Cuando las ecuaciones estructurales presentan una identificación exacta o sobreidentificación no existe problema en términos de la especificación del modelo, de modo que éste (la explicación que se propone) no se verá modificado por las condiciones matemáticas del sistema (relación incógnitas/información). En ambas condiciones del sistema, la sobreidentificación ofrece posibilidades importantes. Un riesgo evidente de un sistema exactamente identificado es que alguno de los efectos propuestos sea cero, con lo que la identificación se verá amenazada. Por el contrario, en los modelos sobreidentificados este riesgo no se da. Por otra parte, para la solución del sistema se requiere igual número de ecuaciones (información) que de incógnitas (coeficientes), lo que hace que la información extra (ecuaciones no utilizadas) puedan emplearse para testar el modelo.

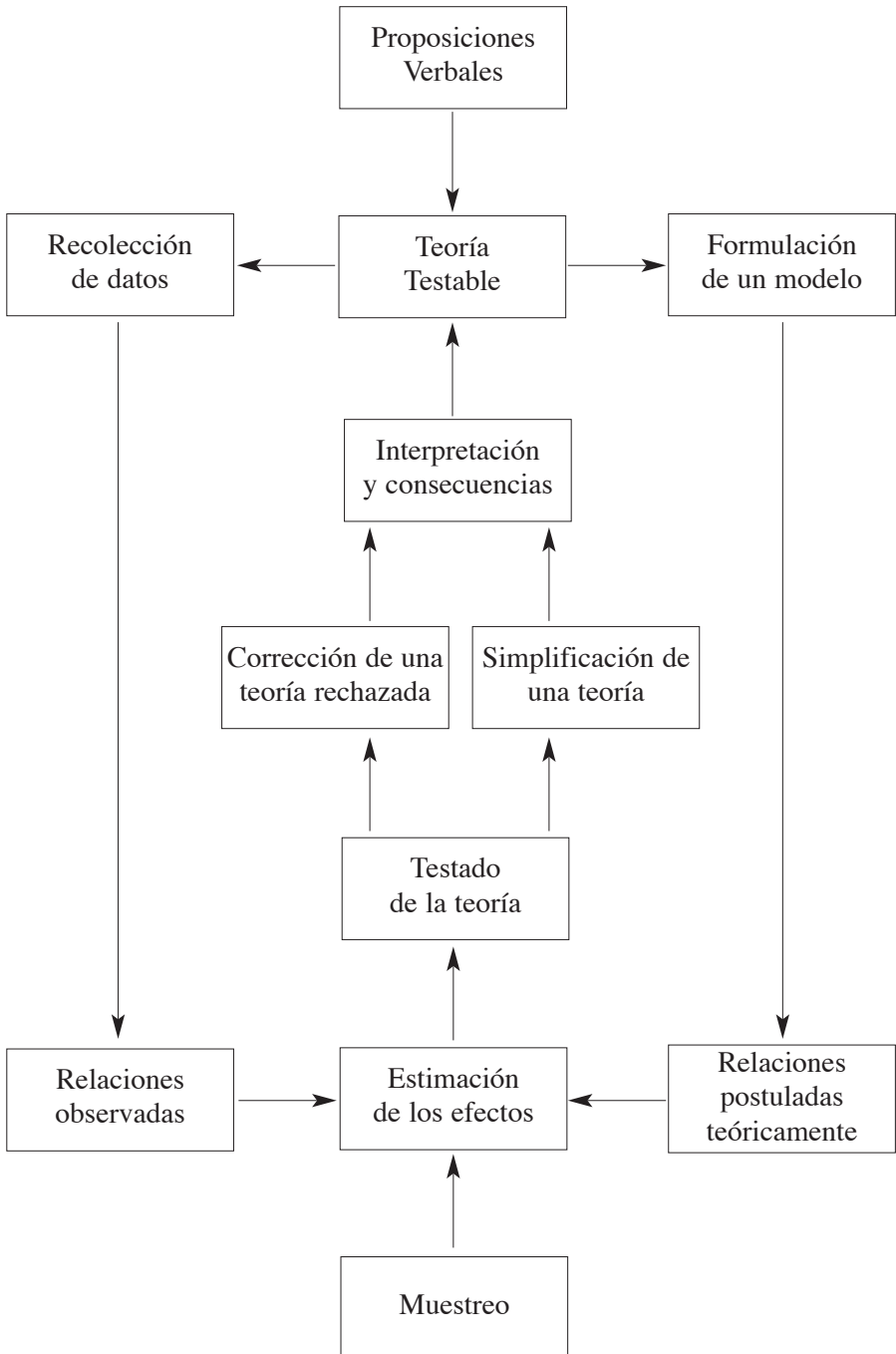


## **8. TESTADO DE MODELOS. LA BONDAD DE AJUSTE**

Como indicábamos, el exceso de información supone una oportunidad especial para testar el modelo estructural. Para ello debemos considerar que el modelo se ajusta sobre un conjunto de datos y la matriz relacional que estos datos ofrecen. El modelo, realmente, aspira a reproducir dicha estructura relacional empírica, pero en el contexto de una estructura teórico explicativa. Es decir, el modelo opera aspirando a traducir una estructura relacional empírica (generada por una definición previa y una selección de lo que es importante para definir el fenómeno en estudio) en una estructura relacional de conceptos vinculados por una argumentación explicativa. En este contexto, un aspecto muy importante ligado al concepto de identificación es la posibilidad de testar el modelo; testar el modelo consiste en comparar la estructura empírica que ha sido integrada en el modelo con la estructura empírica original y sobre la que éste se apoyó.

Una primera aproximación consiste en comparar la matriz de correlaciones observada con la matriz de correlaciones reproducida según los diferentes modelos. En efecto, una vez que han sido estimados los diferentes parámetros para un modelo, es factible estimar la matriz de covarianza o de correlaciones.

En ocasiones, la comparación entre las dos matrices de correlación concluye en que son idénticas. Esta igualdad de las dos matrices, la obtenida directamente desde los datos y la reproducida a partir del modelo, no es incidental. Esto sucede siempre que el modelo es un "modelo recursivo saturado". En este tipo de modelo se incluyen todos los efectos posibles (no recíprocos). En los modelos saturados el número de grados de libertad es igual a 0, lo que significa que el número de correlaciones es igual al número de parámetros que deben ser estimados. En este tipo de modelos exactamente identificados solamente es posible hallar una solución, que además ajustará perfectamente con los datos y como consecuencia de ello la matriz de correlacio-



nes estimada coincidirá exactamente con la observada. Este resultado puede encontrarse generalmente en cualquier tipo de modelo con 0 grados de libertad<sup>23</sup>.

Una situación adecuada para este test de reproducibilidad de la matriz de relaciones empíricas aparece con los sistemas sobreidentificados. Como se trató previamente, para la solución de los parámetros se hacen necesarias tantas ecuaciones como número de parámetros. Por ello, cuando tenemos más ecuaciones que parámetros es posible emplear el exceso de información para testar el modelo. Evidentemente, este test no es posible cuando el modelo está exactamente identificado, dado que no quedan ecuaciones libres para testar. Veamos seguidamente el procedimiento de testado de los modelos explicativos sobreidentificados.

En primer lugar debemos distinguir, como es habitual, entre los coeficientes teóricos propuestos y los coeficientes estimados. Partiendo de los valores de los coeficientes estimados y mediante la relación que el modelo propone entre coeficientes y parámetros teóricos es posible reproducir las covarianzas y varianzas. Las varianzas y covarianzas pueden derivarse de forma única desde los parámetros estructurales del modelo (como sabemos, lo contrario no es cierto y de ahí el problema de la identificación).

El grado en que las varianzas y covarianzas reproducidas desde el modelo se parezcan a las obtenidas directamente y sobre las que se apoya el modelo actuará como test del modelo propuesto. Un mismo conjunto de varianzas y covarianzas puede ser el punto de partida para ajustar diferentes modelos, en tanto que cada modelo postulará un tipo distinto de relaciones. En la medida que la reconstrucción de varianzas y covarianzas depende de las relaciones propuestas, cada modelo generará una reproducción distinta de la matriz relacional empírica.

En ese sentido cabe la noción de testar los diferentes modelos, en la medida que diferentes modelos generarán diferentes varianzas y covarianzas reproducidas. De hecho, si el modelo propuesto diese cuenta completa de las covarianzas y varianzas originalmente obtenidas, empíricamente no habría diferencias entre éstas y las varianzas y covarianzas reproducidas. Podemos entender que un modelo es correcto en la medida que las diferencias entre varianzas y covarianzas de partida se correspondan con las generadas desde el modelo. Por el contrario, si el modelo propuesto fuese erróneo, los dos conjuntos de varianzas y covarianzas serían muy diferentes, indicando de este modo la poca bondad del ajuste del modelo sobre los datos. No tratamos aquí de la realidad, sea la que sea, sino de cómo ésta ha sido expresada mediante datos (hablada) y cómo un modelo puede ajustar mejor o peor sobre esos datos.

---

23. Con la excepción de aquellos en los que algún parámetro no puede ser identificado.

En todo caso, cuando se compara la matriz reproducida de varianzas y covarianzas con la matriz de datos original, es importante evaluar si las diferencias se reparten de forma igual entre todos los coeficientes. La concentración de diferencias importantes en unos coeficientes en concreto nos indicarán posibles problemas de especificación del modelo en esas variables en concreto.

Cuando los grados de libertad son mayores de cero ( $df > 0$ ), es decir cuando tenemos un modelo sobreidentificado, la similaridad entre las dos matrices dependerá de las restricciones que definamos en el modelo. En principio, el tamaño de las desviaciones entre las dos matrices dependerá de la magnitud del error que exista en el modelo especificado. Así, un indicador importante se refiere a la matriz de residuales. Esta matriz de residuales se construye mediante la comparación entre la matriz de covarianzas originales y la matriz de covarianzas estimadas a partir del modelo. El valor de estos residuales debería de ser relativamente pequeño, y distribuido de modo parecido entre las diferentes variables, si el modelo se ajusta de modo razonable a los datos. Unos valores residuales excesivamente grandes asociados a variables concretas son un indicador de un mal ajuste.

La matriz de residuales estandarizada (basada en las correlaciones) es más fácil de interpretar que la matriz no estandarizada (basada en covarianzas), dado que el tamaño de los residuales no depende de la escala de medida de las variables observadas.

Aún cuando la comparación entre la matriz de covarianza observada y reproducida ofrece un buen indicador del ajuste del modelo a los datos, es evidente la necesidad de un test más preciso, especialmente en la medida que es previsible que se produzcan fluctuaciones aleatorias procedentes de la naturaleza muestral de los datos. En ese sentido, es necesario algún tipo de test que ayude a decidir si las desviaciones entre las covarianzas observadas y las reproducidas son fluctuaciones muestrales o si son consecuencia de que el modelo está mal especificado. Este tipo de medidas se denominan test estadísticos. Cuando la distribución de un test estadístico es conocida (indicando para un modelo correcto que parte de la fluctuación de la medición corresponde al muestreo) es posible efectuar la comparación entre los valores observados y el conjunto de valores esperados si el modelo es correcto.

Es posible considerar tres test estadísticos diferentes, asociados con los tres procedimientos de estimación considerados. El mejor candidato para los mínimos cuadrados no ponderados (ULS) es la suma de los residuales al cuadrado, definidos los residuales como la diferencia entre las covarianzas observadas y las covarianzas reproducidas. Es evidente que este estadístico  $T_{\text{ULS}}$  crece cuando las desviaciones entre las covarianzas observadas y las repro-

ducidas se incrementan. Es decir, el valor del estadístico es mayor cuanto mayor es la diferencia entre la matriz observada y la matriz de covarianza reproducida. Desafortunadamente no es conocida la distribución de este estadístico.

Para los otros dos procedimientos de estimación sí es posible formular estadísticos cuya distribución es conocida.  $T_{GLS}$  también depende del tamaño de los residuales: tendrá un valor mayor si los residuales son grandes, y menor si las diferencias entre las covarianzas observadas y las reproducidas son pequeñas. Los residuales son ponderados en el mismo modo que en los procedimientos de estimación, introduciendo un factor de ajuste:  $\frac{1}{2} (N - 1)$ . Por último tenemos el estadístico  $T_{ML}$ , que al igual que los dos estadísticos anteriores está más lejos de un buen ajuste cuanto más grande es su valor. Cuando el coeficiente se aproxima a cero el modelo ofrece un ajuste perfecto.

El coeficiente ***CHI-cuadrado para el modelo de independencia*** nos ofrece un test de significación para la hipótesis de que las diferentes variables observadas consideradas en el modelo son mutuamente independientes. Si el coeficiente indica que las variables observadas son realmente independientes, no tiene ningún sentido continuar explorando modelos de dependencias entre ellas, ya sea porque no existe realmente una estructura de covarianzas que analizar o porque el tamaño muestral es demasiado pequeño para poder demostrar algo.

Otro de los test se refiere a CHI-cuadrado basado en un número determinado de grados de libertad. La distribución de Chi cuadrado ha resultado ser de gran utilidad en múltiples situaciones, entre ellas el testado de la bondad de ajuste en modelos estructurales. Bajo ciertas condiciones, el test estadístico para comprobar la adecuación de un modelo se distribuirá aproximadamente como la distribución de Chi cuadrado, con el valor  $K$  igual a los grados de libertad de un modelo. Por ejemplo, con 10 grados de libertad la probabilidad de encontrar un estadístico próximo a cero es muy baja. Es mucho más probable que el valor obtenido se encuentre entre cinco y quince, mientras que valores mayores que veinte son bastante improbables. Así, en general, es posible esperar, como media, que el valor del estadístico en un buen modelo se aproximara al valor de los grados de libertad. Dado que la distribución de Chi cuadrado es conocida exactamente para cualquier grado de libertad, es posible efectuar afirmaciones más precisas aún. Si comprobamos la tabla de valores de Chi cuadrado, podemos apreciar cómo para un modelo con diez grados de libertad, en un 95% de las muestras de posibles estadísticos se espera que el valor sea menor que 18.3, o para 1 grado de libertad se espera que el valor sea menor que 3.84. Gracias a esta relación entre el test estadístico y la distribución Chi cuadrado, podemos evaluar qué valores

distintos de cero son esperables debido a las fluctuaciones muestrales. Por ello, es factible considerar un test estadístico que indique la bondad del ajuste de un modelo teórico sobre los datos empíricos así como tener en cuenta las variaciones debidas a las oscilaciones muestrales que caracterizan a los datos.

En resumen, podemos considerar un test estadístico que nos ofrezca información sobre el ajuste del modelo a los datos. La tabla de Chi cuadrado nos facilita la información sobre lo grande que debería de ser el valor de un test estadístico en el 95% de las muestras en caso que el modelo sea correcto. Es decir, si al ajustar un modelo sobre una muestra particular el valor es menor de este valor crítico, el modelo no se rechaza dado que puede haberse producido debido a fluctuaciones muestrales. Si el valor del estadístico es mayor que este valor crítico, sabemos que existe un 5% de probabilidad de que sea debido a fluctuaciones muestrales, asumiendo que el modelo es el correcto. Por ello, se rechaza la hipótesis de que el modelo es correcto cuando el valor excede el valor crítico. Lo más probable es que sea un problema de especificación del modelo y no de fluctuaciones muestrales.

En cualquier caso, debemos considerar que cualquier conclusión referida al modelo puede ser errónea dado que cualquier valor puede ser debido a fluctuaciones muestrales. De acuerdo al procedimiento descrito, también es factible el caso opuesto. Es decir, la posibilidad de rechazar un modelo cuando la teoría es correcta es del 5%. Una forma de rehuir este posible error es testando al 0.01 en vez del 0.05. En ese caso, la probabilidad de rechazar un modelo correcto se reduce al 1%. En general, en ciencias sociales se considera aceptable un riesgo del 5%. Este nivel de riesgo se denomina nivel de significación del test.

Vamos a considerar varios indicadores de bondad de ajuste. **RMR** es un coeficiente basado directamente sobre los residuales. Esta medida es próxima a 0 cuando todos los residuales están próximos a cero. Esto implica que tiene una utilidad especial para comparar el ajuste de un modelo con el de modelos alternativos para un mismo conjunto de datos. Por otra parte, no está claro cómo debe de ser de grande esta medida para que exprese un mal ajuste del modelo sobre la datos. Esto significa que este coeficiente es especialmente útil para comparar diferentes modelos alternativos ajustados sobre los mismos datos, dado que sus valores solamente puede ser comparados en esas condiciones. Para intentar evitar el problema de la interpretación de dichos coeficientes, es posible normalizarlos, de forma que su valor posible oscile entre 0 y 1. Es lo que sucede con los coeficientes denominados Índices de Bondad de Ajuste (**GFI**). Es evidente que una de las ventajas principales de este coeficiente de ajuste es que su valor es más fácil de interpretar que el anterior. Cuando GFI es cero el ajuste es malo, mientras que cuando GFI está próxi-

mo a 1 el ajuste es bueno. Este coeficiente puede ser utilizado para comparar modelos alternativos ajustados sobre el mismo conjunto de datos y también para comparar modelos ajustados sobre diferentes datos. Actualmente el problema con estas mediciones es que están poco estudiadas y no está muy claro cuándo es factible hablar de un ajuste aceptable o no.

Una cuestión importante en estos coeficientes es que la medición de la bondad del ajuste debería considerar el número de parámetros empleado en el modelo. En general, es posible obtener un mejor ajuste con un número mayor de parámetros. Por ello, es interesante considerar una medición que también tenga en cuenta el número de parámetros empleados o, de modo equivalente, el número de grados de libertad. Un coeficiente que tiene en cuenta los grados de libertad es el denominado *AGFI*. Este coeficiente también oscila entre 0 y 1, donde cero expresa un ajuste deficiente y 1 un ajuste excelente. No obstante, también en este caso es difícil considerar los valores intermedios de la medición, en el sentido de los valores que expresan un buen ajuste o un mal ajuste. Entre las ventajas de estos estadísticos se encuentra la no dependencia del tamaño muestral. Los test estadísticos basados en Chi-cuadrado son muy sensibles a los errores pequeños en el caso de muestras grandes, y poco sensitivos en caso de muestras pequeñas. Estos coeficientes nos ofrecen un nivel absoluto de ajuste del modelo a los datos, mientras que los test basados en Chi cuadrado ofrecen básicamente una apreciación relativa. Esto se hace evidente, por ejemplo, si consideramos que en el caso de Chi cuadrado la comparación se efectúa con el modelo nulo (independencia), lo que en definitiva es una elección arbitraria. No obstante, poseen por ello la ventaja de una fácil interpretación: la mejora proporcional supone el ajuste de un nuevo modelo, comparado con el modelo nulo.

- Degrees of Freedom = 2
- Minimum Fit Function Chi-Square = 0.40 (P = 0.82)
- Normal Theory Weighted Least Squares Chi-Square = 0.39 (P = 0.82)
- Estimated Non-centrality Parameter (NCP) = 0.0
- 90 Percent Confidence Interval for NCP = (0.0 ; 2.81)
- Minimum Fit Function Value = 0.0025
- Population Discrepancy Function Value (F0) = 0.0
- 90 Percent Confidence Interval for F0 = (0.0 ; 0.018)
- Root Mean Square Error of Approximation (RMSEA) = 0.0
- 90 Percent Confidence Interval for RMSEA = (0.0 ; 0.094)
- P-Value for Test of Close Fit (RMSEA < 0.05) = 0.88
- Expected Cross-Validation Index (ECVI) = 0.11
- 90 Percent Confidence Interval for ECVI = (0.11 ; 0.13)

- ECVI for Saturated Model = 0.13
- ECVI for Independence Model = 5.45
- Chi-Square for Independence Model with 6 Degrees of Freedom = 858.40
- Independence AIC = 866.40
- Model AIC = 16.39
- Saturated AIC = 20.00
- Independence CAIC = 882.70
- Model CAIC = 49.00
- Saturated CAIC = 60.
- Normed Fit Index (NFI) = 1.00
- Non-Normed Fit Index (NNFI) = 1.00
- Parsimony Normed Fit Index (PNFI) = 0.33
- Comparative Fit Index (CFI) = 1.00
- Incremental Fit Index (IFI) = 1.00
- Relative Fit Index (RFI) = 1.00
- Critical N (CN) = 3681.52
- Root Mean Square Residual (RMR) = 0.0044
- Standardized RMR = 0.00079
- Goodness of Fit Index (GFI) = 1.00
- Adjusted Goodness of Fit Index (AGFI) = 0.99
- Parsimony Goodness of Fit Index (PGFI) = 0.20

El criterio de información de Akaike (**AIC**) y la versión de Bozdogan del estadístico (**CAIC**), toman en cuenta tanto los indicadores de ajuste como el número de parámetros a estimar empleados para lograr ese ajuste. El modelo que produzca *el valor más pequeño* en ambos estadísticos es el que puede considerarse potencialmente el más útil.

Podemos considerar, así mismo, otros tres índices de ajuste **NFI**, **NNFI**, y **CFI**. Todos ellos están basados en los valores de la función de ajuste y tienen como límite superior el valor 1. El NNFI presenta la ventaja de funcionar bastante bien independientemente de los tamaños muestrales. En general, la experiencia demuestra que todos estos índices necesitan *valores superiores a .9* para reflejar un modelo con ajuste adecuado.

<b>COEFICIENTES DE BONDAD DE AJUSTE</b>		
<b>GENERAL</b>		
GFI	0-Sin ajuste 1-Ajuste Perfecto	Los valores próximos a .90 expresan un buen ajuste del modelo
RMR Stand.RMR	El Investigador decide el nivel	Indica la proximidad entre la matriz de covarianzas original y la reproducida por el modelo propuesto
<b>INCREMENTO</b>		
NNFI (TLI) NFI	0 Sin ajuste 1 ajuste perfecto	Los valores superiores a .90 expresan un buen ajuste del modelo
<b>CRITERIO INFORMACIÓN</b>		
AIC CAIC ECVI	DEPENDE DEL MODELO	Compara los valores en modelos alternativos Emplear ECVI con muestras pequeñas
<b>ÍNDICE PARSIMONIA</b>		
AGFI	0 Sin límite inferior 1 ajuste perfecto	Valor ajustado para grados de Libertad, con .90 como buen ajuste del modelo
RMSEA	< 0.05 > 0.05 y < 0.08  > 0.08	Buen ajuste del modelo Ajuste moderado del modelo Mal ajuste del modelo
PNFI PGFI	0 Sin ajuste 1 ajuste perfecto	Compara los valores en modelos alternativos



## 9. MODELOS ESTRUCTURALES DE MEDICIÓN

Las variables latentes son aquellas que representan conceptos teóricos que no pueden ser observados directamente. Estas variables no observables directamente darían cuenta de la variabilidad apreciada. Son también conocidas como factores, y han sido tradicionalmente de gran interés para los investigadores en las ciencias sociales y del comportamiento. Como hemos visto en capítulos anteriores, ejemplos de variables latentes en sociología son la anomia, la clase social, el conservadurismo, la ideología política, etc. Dado que las variables latentes no son observables, su medición debe efectuarse indirectamente. Para ello el investigador intenta vincular la variable no observada con alguna que sí lo sea. Así, una variable latente es definida operacionalmente en función a los comportamientos u opiniones que muestran aquellas otras variables presentes que se supone la representan. En ese sentido, la medición de un comportamiento u opinión constituye, desde esta estrategia, tanto la medición directa de una variable observada como la medición indirecta de una variable que no es observada. Debemos aclarar que el concepto “comportamiento” tal y como se utiliza aquí incluye toda aquella respuesta dada por un individuo a un instrumento de medición. Por ello puede incluir opiniones, afirmaciones sobre comportamientos (presentes, pasados o futuros) o puntuación resultantes de la administración de un test.

El contexto de la metodología SEM (*Structural Equation Models* o Modelos de Ecuaciones Estructurales) las variables observadas actúan como indicadores del constructo teórico subyacente. En ese sentido se hace obvia la importancia de seleccionar las variables indicadoras adecuadas. El procedimiento estadístico mejor conocido y más antiguo para investigar la relación entre un conjunto de variables observadas y otras latentes es el análisis factorial. Mediante esta técnica, el investigador examina la covariación entre un conjunto de variables observadas para obtener información sobre la variable latente.

Existen dos tipos básicos de análisis factorial: el análisis factorial exploratorio y el análisis factorial confirmatorio. El primero se aplica cuando la

relación entre las variables observadas y la latente es incierta o desconocida. El análisis funciona de manera exploratoria para determinar cómo las variables observadas pueden relacionarse con el factor latente. Habitualmente, el investigador intenta localizar un número mínimo de factores que expliquen la correlación entre el conjunto de variables observadas. En el análisis factorial, la relación entre las variables observadas y la variable latente se denomina carga factorial. De este modo, se espera que una variable observada presente una carga factorial importante en la variable latente que representa, y mínima en aquellas variables latentes con las que no tiene relación teórica. Esta aproximación analítica se denomina exploratoria dado que el investigador no posee un conocimiento previo de aquellos indicadores que verdaderamente son útiles para medir la variable latente. En relación a la utilidad del análisis factorial exploratorio puede consultarse a Comrey, 1992; Gorsuch, 1983; Mulaik, 1972.

A diferencia del análisis factorial exploratorio, el análisis factorial confirmatorio responde a la situación en la que el investigador desea testar hipótesis sobre la relación entre las variables observadas y las variables latentes. En ese sentido, basándose en el conocimiento teórico o en la investigación empírica previa, se propone a priori una relación estadística entre las variables observadas y la variable latente. Posteriormente, el modelo puede evaluarse mediante procedimientos estadísticos para determinar la bondad del ajuste sobre los datos. En definitiva, el modelo factorial se limita a considerar cómo se vinculan las variables observadas con la variable latente o factor. Es decir, su preocupación principal es determinar en qué modo y grado las variables observadas han sido generadas desde las variables latentes subyacentes. En ese sentido, los coeficientes de regresión de los factores sobre las variables observadas son fundamentales. Por ello, aun cuando se considere la estructura de correlaciones entre los factores, habitualmente no se considera la relación estructural entre ellos. Debido a este interés primario en la medición, los modelos factoriales confirmatorios son considerados en el modelado estructural como modelos de medición.

Consideremos el ejemplo siguiente, donde varios estudiantes estiman la altura de sus compañeros. Es evidente que dicha estimación experimentará la variabilidad de las diferentes alturas de los individuos medidos, pero también una variabilidad en las estimaciones de los diferentes individuos que miden. Así, en la tabla siguiente podemos apreciar cómo a un individuo con una altura de 1,67 cm, le atribuye el estudiante Y1 una altura estimada de 1,68 cm, el estudiante Y2 le atribuye 1,65 cm, el estudiante Y3 estima 1,66 cm, el estudiante Y4, 1,64 cm y por último el estudiante Y5 estima que su altura es de 1,65 cm.

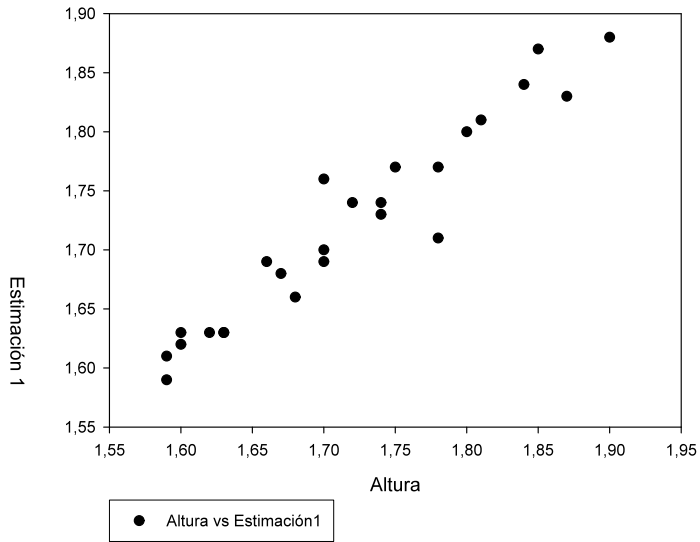
*Altura estimada*

<b>Altura real</b>	<b>Y1</b>	<b>Y2</b>	<b>Y3</b>	<b>Y4</b>	<b>Y5</b>
1,67	1,68	1,65	1,66	1,64	1,65
1,63	1,63	1,63	1,68	1,63	1,66
1,70	1,76	1,75	1,77	1,72	1,79
1,81	1,81	1,79	1,82	1,79	1,82
1,68	1,66	1,65	1,71	1,68	1,69
1,74	1,74	1,75	1,76	1,72	1,76
1,87	1,83	1,79	1,86	1,85	1,82
1,60	1,62	1,60	1,64	1,65	1,65
1,59	1,59	1,60	1,57	1,61	1,58
1,90	1,88	1,85	1,90	1,88	1,85
1,62	1,63	1,63	1,65	1,64	1,60
1,78	1,77	1,80	1,79	1,78	1,79
1,70	1,69	1,68	1,73	1,67	1,72
1,70	1,70	1,70	1,67	1,71	1,70
1,72	1,74	1,70	1,70	1,72	1,72
1,66	1,69	1,68	1,70	1,68	1,68

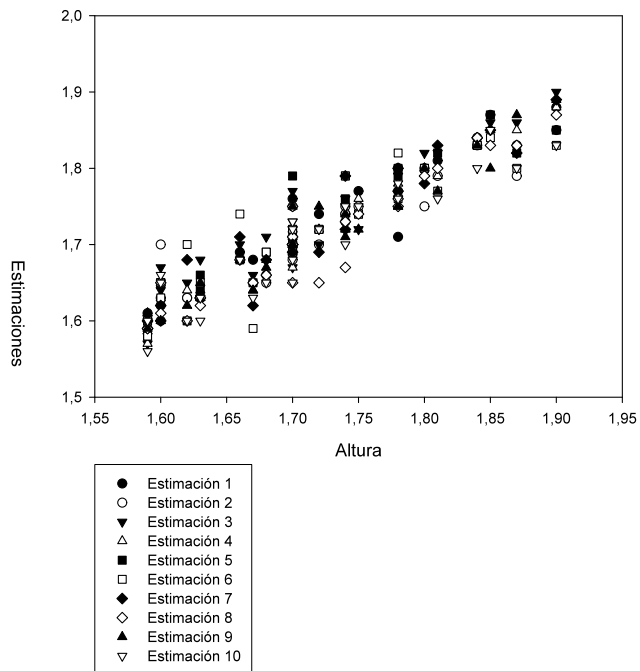
Resulta evidente que en este caso no conocemos la altura real del individuo, en primer lugar porque lo que tenemos es una información sobre la altura que tenía la última vez que se midió. No obstante, las circunstancias de la medición influyen notablemente en la altura medida. Es un caso equivalente, pero menos extremo que cuando se considera el peso de un individuo: la variabilidad del peso es elevada, no sólo a largo plazo, incluso en el transcurso del día. En ese sentido, la información que pueda dar un individuo sobre su peso es una estimación “basada en hechos reales”, pero posiblemente aproximada y no real. La estimación que efectúen sus compañeros será, así mismo, aproximada y subjetiva, según su percepción de la relación tamaño y peso.

La estimación y la altura que informa el individuo están relacionadas de forma clara, tal y como muestran los gráficos de dispersión siguientes.

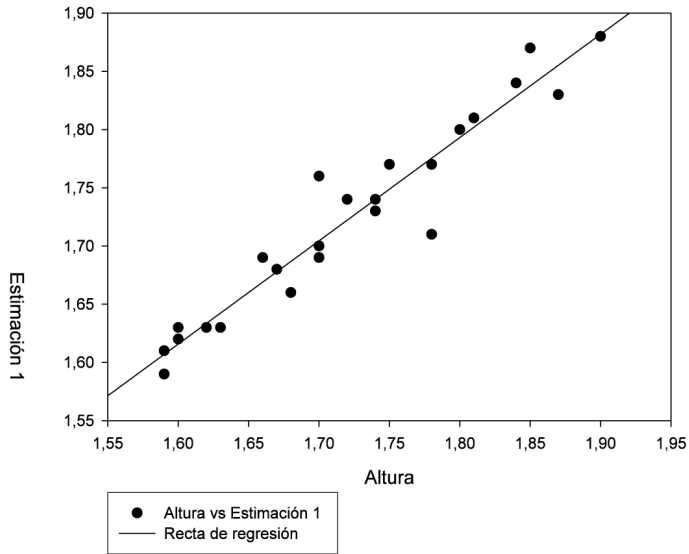
Altura y estimación 1



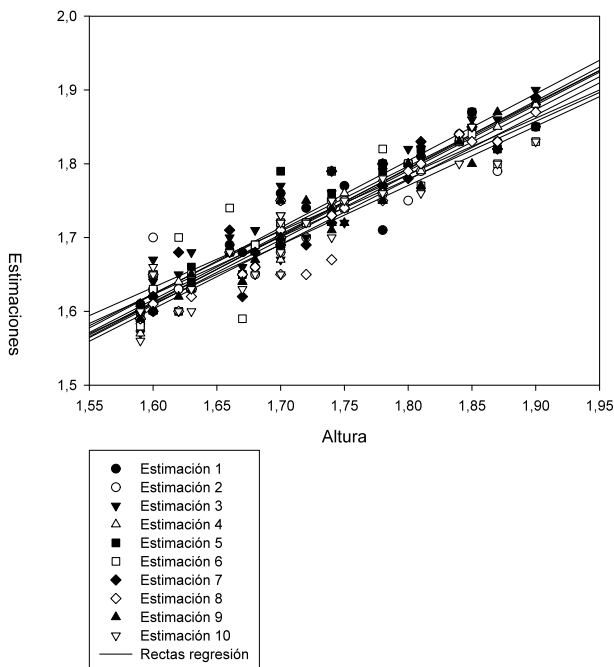
Altura y estimaciones



Recta de regresión de Altura vs Estimación 1



Recta regresión Altura vs Estimaciones



Tenemos, en principio, dos mediciones interesantes que efectuar. En primer lugar, cómo varían las diferentes mediciones que efectúa un individuo: ¿serán refinadas, buscando aproximarse, o a “grosso modo” y redondeando las mediciones? En ese sentido, podemos apreciar los esfuerzos por realizar una buena medición. Es decir, quien, en expresión popular “tiene ojo de buen cubero”. Una segunda información que nos interesa es cómo se relacionan las mediciones que efectúan los diferentes estudiantes con las alturas reales que indican los sujetos que son medidos. Para poder responder a estas preguntas necesitamos calcular la varianza (dispersión interna en las mediciones efectuadas por cada individuo) y la covarianza (medida de relación entre las dos mediciones, estimación y altura que indica el individuo).

El cálculo de la varianza es bastante simple. En primer lugar, debemos establecer una referencia común sobre la que establecer la dispersión, algo así como un “kilometro cero” de la distribución. El valor adoptado para cada variable es, en este caso, la media aritmética. Es conocida la afición en la estadística descriptiva por la media como coeficiente sintético de valor central. Tomamos pues, la media como valor de referencia para determinar la dispersión y calculamos la media de la variable. Después, determinamos la distancia de los valores a la media. Tomamos esas diferencias y las elevamos al cuadrado (es decir, las multiplicamos por sí mismas, de modo que eliminamos los posibles signos negativos). De ese modo, todas las diferencias a la media estarán ordenadas en una sola dirección, partiendo desde cero (cero es el caso en que el valor del caso en la variable coincida con la media). Después las sumamos y el sumatorio lo dividimos, usualmente por el número de casos menos uno (N-1).

Altura	Altura - Media	(Altura- Media) <sup>2</sup>
A	B	C
1,67	$(1,67 - 1,7106) = -0,040625$	$(-0,040625)^2 = 0,00165039$
1,63	-0,080625	0,00650039
1,7	-0,010625	0,00011289
1,81	0,099375	0,00987539
1,68	-0,030625	0,00093789
1,74	0,029375	0,00086289
1,87	0,159375	0,02540039
1,6	-0,110625	0,01223789
1,59	-0,120625	0,01455039
1,9	0,189375	0,03586289
1,62	-0,090625	0,00821289
1,78	0,069375	0,00481289

1,7	-0,010625	0,00011289
1,7	-0,010625	0,00011289
1,72	0,009375	8,7891E-05
1,66	-0,050625	0,00256289
M. aritmética de <i>Altura</i> (A) 1,710625		
Sumatorio de la columna (C) dividido por N-1 o Varianza de <i>Altura</i>		0,00825958

Las varianzas de las diferentes mediciones son las siguientes

Varianza	
Altura	,008
Y1	,007
Y2	,006
Y3	,008
Y4	,006
Y5	,01

Es importante recordar que según el procedimiento de cálculo de la varianza (elevando al cuadrado las diferencias) no es esperable obtener una varianza con valor negativo. La varianza viene expresada en el nivel de medición de la variable original. Esto implica que variables con unidades de medición de mayor magnitud (kilómetros por ejemplo), tendrán un coeficiente de varianza que en magnitud numérica será posiblemente superior al coeficiente de la varianza de la edad de un individuo en años. Por ello, no cabe comparar directamente las varianzas que proceden de mediciones de diferente escala. En el caso de nuestro ejemplo, todas las mediciones se referían a la altura de los individuos y se expresan en centímetros, lo que permite comparar las magnitudes de la varianza de las mediciones de “buen cubero” efectuadas por los individuos.

El cálculo de la covarianza, es decir, la coordinación en la variación entre dos variables, es simplemente una extensión de la situación anterior. Ahora se calcularán las diferencias de cada medición a su media, y el resultado para cada variable se multiplicará por el de la otra. La lógica es la misma. No obstante, mientras que era imposible que la varianza adquiriese valores negativos (al multiplicar cada valor por si mismo), la covarianza sí puede, en la medida que los valores de una variable se multiplicarán por los de la otra. Por ello, si la covarianza tiene valor negativo implica que cuando en una variable las

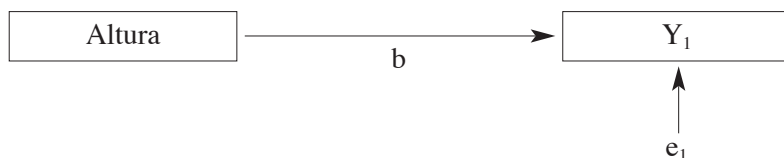
diferencias a su media son positivas, las diferencias a la media en la otra variable son negativas. Es decir, crecen en direcciones opuestas. Cuando el signo es positivo implica que en la media de los productos (de las diferencias a la media) predominan los valores positivos. Es decir, que cuando la media de las diferencias a la media en una variable es positiva, en la otra también lo es. O dicho de otro modo, que cuando los valores crecen en una de las variables los valores también tienden a crecer en la otra.

Altura	Y1	(Altura – media)	(Y1 – media)	
A B	C	D	(C*D)	
1,67	1,68	-0,040625	-0,03375	0,00137109
1,63	1,63	-0,080625	-0,08375	0,00675234
1,7	1,76	-0,010625	0,04625	-0,00049141
1,81	1,81	0,099375	0,09625	0,00956484
1,68	1,66	-0,030625	-0,05375	0,00164609
1,74	1,74	0,029375	0,02625	0,00077109
1,87	1,83	0,159375	0,11625	0,01852734
1,6	1,62	-0,110625	-0,09375	0,01037109
1,59	1,59	-0,120625	-0,12375	0,01492734
1,9	1,88	0,189375	0,16625	0,03148359
1,62	1,63	-0,090625	-0,08375	0,00758984
1,78	1,77	0,069375	0,05625	0,00390234
1,7	1,69	-0,010625	-0,02375	0,00025234
1,7	1,7	-0,010625	-0,01375	0,00014609
1,72	1,74	0,009375	0,02625	0,00024609
1,66	1,69	-0,050625	-0,02375	0,00120234
Medias	1,710625	1,71375		
Media de (C*D) o covarianza (sumatorio del producto de las diferencias a su media dividido por N-1)				0,0072175

La media del producto en la columna final viene dividida por el tamaño muestral menos uno  $(N-1)^{24}$ , mientras que las medias aritméticas de las mediciones lo son por N. Hasta el momento hemos considerado cómo varían las variables

24. La razón para dividir por N-1 en lugar de por N es de carácter técnico. Puede demostrarse que el dividir por N resulta en un coeficiente estadístico que, considerado como estimación del correspondiente parámetro poblacional, está ligeramente sesgado. La división por N-1 rectifica este sesgo en la estimación.

(varianza) y como covarían entre ellas. Pero no debemos confundir covarianza con explicación. El coeficiente de covarianza expresa la coordinación estadística entre dos variables, pero no indica nada sobre qué variable explica y en qué modo. Explicar una variable significa introducir el concepto de error.



El diagrama anterior nos muestra el conocido modelo de regresión simple. En él postulamos que el valor real (es decir la altura que cada individuo exhibe ante el sujeto que la mide subjetivamente) influye en la estimación que se efectúa ( $Y_1$ ). Esta estimación aparece influida por un error, que suponemos aleatorio.

$$Y_1 = a + b \text{ Altura} + e_1$$

En este modelo explicativo de la estimación subjetiva efectuada por el individuo  $Y_1$ , **a** y **b** son coeficientes constantes (parámetros). Corresponden con la intersección y la pendiente y miden el *error sistemático* que contiene la estimación de  $Y_1$ . Si la estimación de altura no contuviese sesgos, **a** sería cero y **b** tomaría el valor de uno.  $e_1$  es una nueva variable aleatoria que representa el *error no sistemático* en la predicción. En este modelo,  $Y$  se denomina la variable dependiente y *Altura* ( $X$ ) la variable independiente, e implica ciertas previsiones acerca de la varianza de  $Y$ , así como de la covarianza entre  $X$  y  $Y$ . El objetivo del análisis es comparar esta expectativa con los valores observados para poder estimar los valores de **a** y **b**. La varianza del error,  $\text{Var}(e)$ , es también un parámetro interesante para estimar. La varianza de  $e_1$  refleja la precisión de la estimación (menor varianza corresponde a una mayor precisión). En este modelo se asume que el valor real (*Altura*) y el error ( $e$ ) no están correlacionados y que la media de  $e$  es cero y su varianza homogénea,  $\text{Var}(e)$ .

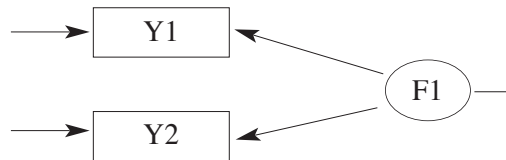
En este modelo de medición tenemos una información que puede estar anclada en la realidad o muy próxima a ella, como es la altura que indica el individuo. No obstante, cabe pensar una situación donde sólo pudiéramos apoyarnos en estimaciones subjetivas sobre lo que pueda ser el valor real. Ésta es una situación frecuente en algunas disciplinas como la psicología o la sociología. Continuando con el ejemplo de las estimaciones de la altura de los com-

pañeros de clase, supongamos que conocemos las estimaciones que efectúan dos alumnos pero no sabemos la altura real. El modelo, en este caso, sustituye la variable altura por una variable a estimar, latente, y que daría cuenta de la covariación entre las mediciones efectuadas por los estudiantes (explicaría el parecido entre ambas mediciones). Consideramos, así mismo, que ambas estimaciones de altura efectuadas por los dos estudiantes contienen error.

$$Y_1 = a + b F + e_1$$

$$Y_2 = c + d F + e_2$$

En este modelo postulamos la existencia de una variable latente que denominamos  $F$ . Al igual que sucedía en el ejemplo anterior, poseemos estimados de la media de  $Y_1$  y de  $Y_2$ , así como de sus varianzas y sus covarianzas. Habitualmente en las ciencias sociales y del comportamiento las mediciones son, en el mejor de los casos, intervalales, donde la posición del 0 es arbitraria. Esta es una de las razones por la que la intersección  $a$  y  $c$  no son de interés. Sí lo son, sin embargo los parámetros de pendiente  $b$  y  $d$ . El modelo de medición especificado por las dos ecuaciones aparece ilustrado en el diagrama siguiente. En este modelo, los parámetros que deben ser estimados son  $b$ ,  $d$ ,  $\text{Var}(F)$ ,  $\text{Var}(E_1)$  and  $\text{Var}(E_2)$ .

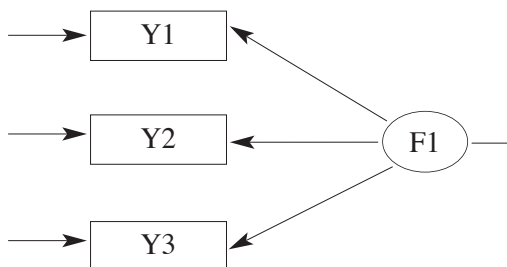


F\_A\_T\_A\_L\_E\_R\_R\_O\_R: Degrees of freedom is negative.

En este modelo partimos asumiendo que las  $\text{Cov}(E_1, E_2)$ ,  $\text{Cov}(F, E_1)$  y  $\text{Cov}(F, E_2)$  son todas 0. En otras palabras, asumimos que los errores de medición no están correlacionados y que su vez están incorrelacionados con la variable latente  $F$ . Estas pueden que no sean presunciones completamente realistas, pero de no plantearse, sería imposible estimar cualquier otro parámetro de interés. En este modelo tenemos cinco parámetros para estimar, pero sólo tres estadísticos nos aportan información relevante,  $\text{Var}(Y_1)$ ,  $\text{Var}(Y_2)$  y  $\text{Cov}(Y_1, Y_2)$ . Ciertamente, esto supone un problema importante. Cuando poseemos más incógnitas a resolver que información, no existe un conjunto de soluciones únicas para estimar los parámetros.

Por tanto lo que tenemos es que el modelo se encuentra infraidentificado. Es el problema clásico en regresión cuando ambas variables están sujetas a error. La única estrategia para encontrar soluciones es imponerse algunas restricciones en el valor de los parámetros. Por ejemplo, podemos plantear que las mediciones  $Y_1$  y  $Y_2$  sólo pueden ofrecer información sobre los sesgos que cometen los dos estudiantes uno en relación al otro. Esto supone que podría estimarse la ratio entre  $b$  y  $d$ , pero no  $b$  y  $d$ . El problema también se puede intentar solucionar fijando arbitrariamente  $b = 1$  (o de modo equivalente,  $d = 1$ ) o fijando la varianza de la variable latente  $F = 1$ . Cualquiera de las opciones anteriores no produciría diferencias significativas en la interpretación de resultados. No obstante, a pesar de ellas, el modelo continúa infraidentificado. Cualquier otra restricción implicaría prescindir de la estimación de valores y parámetros en los cuales estamos especialmente interesados. El problema de la identificación del sistema, cuando el modelo de medición no contiene suficiente información, debería de ser tomado en cuenta en la fase del diseño de la investigación. Esto obviamente no es factible cuando los datos empleados proceden de fuentes secundarias. Una forma de evitar desde el diseño de la investigación este problema consiste, por ejemplo, en utilizar tres indicadores independientes en lugar de sólo dos. Otra opción posible podría consistir en repetir el modelo temporalmente.

Con tres indicadores



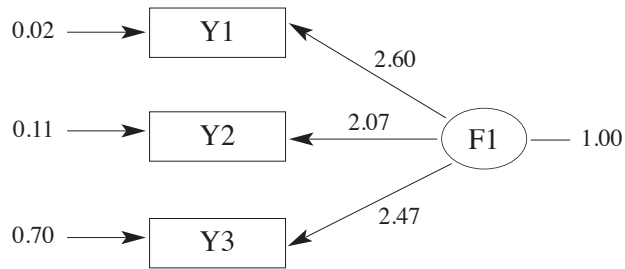
Goodness of Fit Statistics

Degrees of Freedom = 0

Minimum Fit Function Chi-Square = 0.0 (P = 1.00)

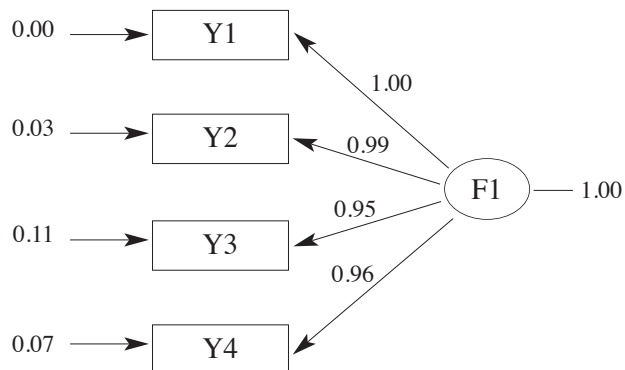
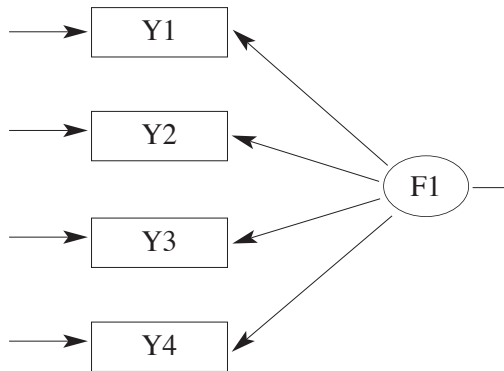
Normal Theory Weighted Least Squares Chi-Square = 0.00 (P = 1.00)

The Model is Saturated, the Fit is Perfect!



Chi-Square=0.00, df=0, P-value=1.00000, RMSEA=0.000

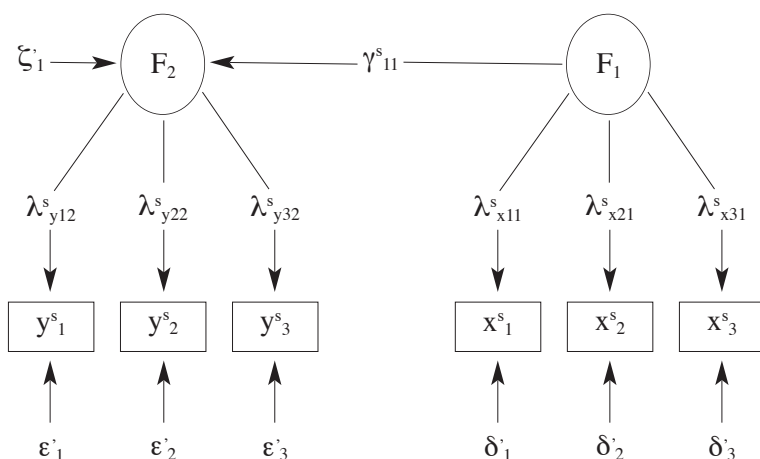
Con cuatro indicadores



Chi-Square=0.39, df=2, P-value=0.82107, RMSEA=0.000

## 9.1. LA EXPLICACIÓN ENTRE LATENTES

Es posible realizar análisis estructurales empleando variables latentes<sup>25</sup>. Veamos seguidamente un modelo simple de efecto entre variables latentes. El siguiente fue el primer modelo discutido en 1969 y donde se puede apreciar claramente su relación con los modelos estructurales construidos exclusivamente con variables manifiestas. En este modelo podemos apreciar dos variables latentes  $F_1$  y  $F_2$ , donde  $F_1$  es la variable latente causa y  $F_2$  la variable latente efecto.



Podemos apreciar como la variable latente  $F_1$  influye las variables indicadoras  $x_1, x_2, x_3$ , mientras que la variable latente  $F_2$  influye en las variables indicadoras  $y_1, y_2, y_3$ . Se han notado los errores ecuacionales como  $\delta_1, \delta_2, \delta_3$  el error para las variables  $x$ , y  $\epsilon_1, \epsilon_2, \epsilon_3$  para las variables  $y$ . En este tipo de modelos la notación de las variables y de los errores varía en la medida que se introduce el concepto de variable latente y variable indicadora.

En este modelo en particular, el interés está centrado en la relación entre variables latentes y no entre los indicadores. Sin embargo, la correlación entre indicadores es la única información que poseemos empíricamente desde los datos. La relación entre variables latentes es algo que debemos estimar desde la correlación entre indicadores. Precisamente, este fue el descubrimiento importante desde el análisis de senderos: que es posible estimar la relación entre las variables latentes mediante la correlación conocida entre las variables indicadoras. El modo como se deriva el efecto  $\gamma_{21}$  es exactamente igual a como hemos considerado en las reglas de descomposición.

25. Las variables latentes son denominadas factores en el ámbito de la psicología.

En primer lugar, las correlaciones apreciadas entre las variables indicadoras se expresan en términos de los parámetros del modelo y en segundo lugar, se estiman los parámetros empleando la información que les asocia a las correlaciones.

$$\rho_{y_2y_1} = \lambda_{y_{22}}^s \lambda_{y_{12}}^s \quad (3.1.1.)$$

$$\rho_{y_3y_1} = \lambda_{y_{32}}^s \lambda_{y_{12}}^s \quad (3.1.2.)$$

$$\rho_{x_1y_1} = \lambda_{x_{11}}^s \gamma_{21}^s \lambda_{y_{12}}^s \quad (3.1.3.)$$

$$\rho_{x_2y_1} = \lambda_{x_{21}}^s \gamma_{21}^s \lambda_{y_{12}}^s \quad (3.1.4.)$$

$$\rho_{x_3y_1} = \lambda_{x_{31}}^s \gamma_{21}^s \lambda_{y_{12}}^s \quad (3.1.5.)$$

$$\rho_{y_3y_2} = \lambda_{y_{32}}^s \lambda_{y_{22}}^s \quad (3.1.6.)$$

$$\rho_{x_1y_2} = \lambda_{x_{11}}^s \gamma_{21}^s \lambda_{y_{22}}^s \quad (3.1.7.)$$

$$\rho_{x_2y_2} = \lambda_{x_{21}}^s \gamma_{21}^s \lambda_{y_{22}}^s \quad (3.1.8.)$$

$$\rho_{x_3y_2} = \lambda_{x_{31}}^s \gamma_{21}^s \lambda_{y_{22}}^s \quad (3.1.9.)$$

$$\rho_{x_1y_3} = \lambda_{x_{11}}^s \gamma_{21}^s \lambda_{y_{32}}^s \quad (3.1.10.)$$

$$\rho_{x_2y_3} = \lambda_{x_{21}}^s \gamma_{21}^s \lambda_{y_{32}}^s \quad (3.1.11.)$$

$$\rho_{x_3y_3} = \lambda_{x_{31}}^s \gamma_{21}^s \lambda_{y_{32}}^s \quad (3.1.12.)$$

$$\rho_{x_2x_1} = \lambda_{x_{21}}^s \lambda_{x_{11}}^s \quad (3.1.13.)$$

$$\rho_{x_3x_1} = \lambda_{x_{31}}^s \lambda_{x_{11}}^s \quad (3.1.14.)$$

$$\rho_{x_3x_2} = \lambda_{x_{31}}^s \lambda_{x_{21}}^s \quad (3.1.15.)$$

Además de las ecuaciones anteriores, es posible especificar seis más correspondientes a las varianzas de las variables indicadoras, descomponiéndolas en varianza explicada y varianza no explicada. No nos ocuparemos de dichas ecuaciones en la medida en que el efecto que nos ocupa teóricamente es el que liga las dos variables latentes  $F_1$  y  $F_2$  ( $\gamma_{21}$ ). En resumen, tenemos 15 ecuaciones especificadas y sólo 7 parámetros desconocidos. En estos términos podemos concluir que la condición necesaria para la identificación del sistema se cumple.

Es evidente que con la información facilitada por los datos (en forma de correlaciones) y por las ecuaciones (3.1.1) (3.1.2.) (3.1.6) es posible estimar los coeficientes  $\lambda_{y_{12}}^s$ ,  $\lambda_{y_{22}}^s$ ,  $\lambda_{y_{32}}^s$ . Si consideramos la razón entre  $\rho_{y_2y_1}$  y  $\rho_{y_3y_1}$

$$\rho_{y_2y_1} / \rho_{y_3y_1} = \lambda_{y_{22}}^s / \lambda_{y_{32}}^s$$

luego

$$\lambda_{y_{22}}^s = \lambda_{y_{32}}^s (\rho_{y_2y_1} / \rho_{y_3y_1})$$

sustituyendo en la ecuación (3.1.6) obtenemos la solución para  $\lambda_{y32}^s$

$$\lambda_{y32}^s = \text{raíz} (r_{y3y1} r_{y3y1}) / r_{y2y1}$$

habiendo determinado  $\lambda_{y32}^s$  los otros coeficiente pueden hallarse por sustitución en las ecuaciones anteriores. Aplicando el mismo procedimiento pueden estimarse los coeficientes

$$\lambda_{x11}^s, \lambda_{x21}^s, \lambda_{y31}^s$$

Tras determinar los coeficientes que ligan las variables indicadoras a las variables latentes, aún nos quedan 9 ecuaciones para determinar el valor de  $\gamma_{21}$ .

De este modo, el parámetro puede resolverse desde 9 ecuaciones distintas. Podemos apreciar cómo una vez establecido el mecanismo estructural que da forma a las variables latentes es posible determinar la influencia de unas sobre otras. En este caso es incluso posible testar el modelo teórico propuesto dado que quedan 8 grados de libertad. Evidentemente, el modelado con variables latentes se ha desarrollado implicando modelos más complejos, donde se combinan variables latentes, indicadoras y manifiestas.



## 10. MODELOS ESTRUCTURALES CON LISREL

### 10.1. INTRODUCCIÓN: LISREL Y SIMPLIS

Las siguientes páginas pretenden ser un tutorial que ayude a usuarios nuevos o potenciales de LISREL a familiarizarse con su funcionamiento, y más concretamente en uno de los lenguajes de programación que admite: el lenguaje SIMPLIS.

LISREL es un paquete informático de estadística para el análisis multivariante que se utiliza habitualmente para estudiar las relaciones múltiples de dependencia que se establecen entre variables integradas en una estructura explicativa determinada. Desde este método de análisis, un modelo estructural se define como un conjunto de ecuaciones estructurales que pueden ser representadas mediante esquemas o diagramas. Para las ecuaciones especificadas que componen el modelo, LISREL estima los coeficientes desconocidos (efectos entre variables, residuales, etc.) y si se cumplen los requisitos es capaz de ofrecer estadísticos de bondad de ajuste, lo que permite valorar el ajuste empírico de la estructura explicativa formulada.

Como aproximación general al análisis, en los sistemas de ecuaciones podrán existir variables de dos tipos, observadas y latentes, de forma que es posible construir modelos estructurales utilizando únicamente variables observadas, variables latentes, o ambas a la vez. En su forma más desarrollada, el modelo asume que existe una estructura de dependencia entre las variables (lo que se denomina modelo estructural). En el caso de incorporar variables latentes, cada una de ellas dispondrá de una serie de variables observadas que actúan como indicadores (modelos de medición).

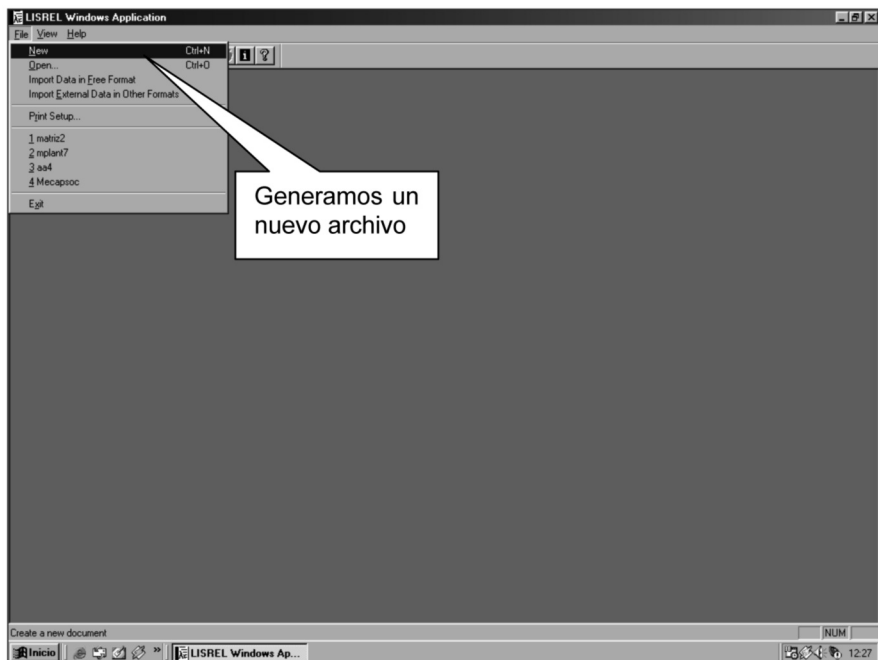
El programa LISREL permite construir y testar modelos en dos lenguajes distintos: el lenguaje SIMPLIS y el lenguaje LISREL. SIMPLIS es un lengua-

je de comandos basado en lenguaje natural, incorporado a las versiones de LISREL8 y diseñado para hacer más sencillo el modelado de ecuaciones estructurales que en las versiones anteriores, y su ventaja radica en la facilidad de uso y la flexibilidad que aporta en el trabajo con el programa. En este documento abordaremos solamente la utilización del lenguaje SIMPLIS para mostrar el funcionamiento del programa.

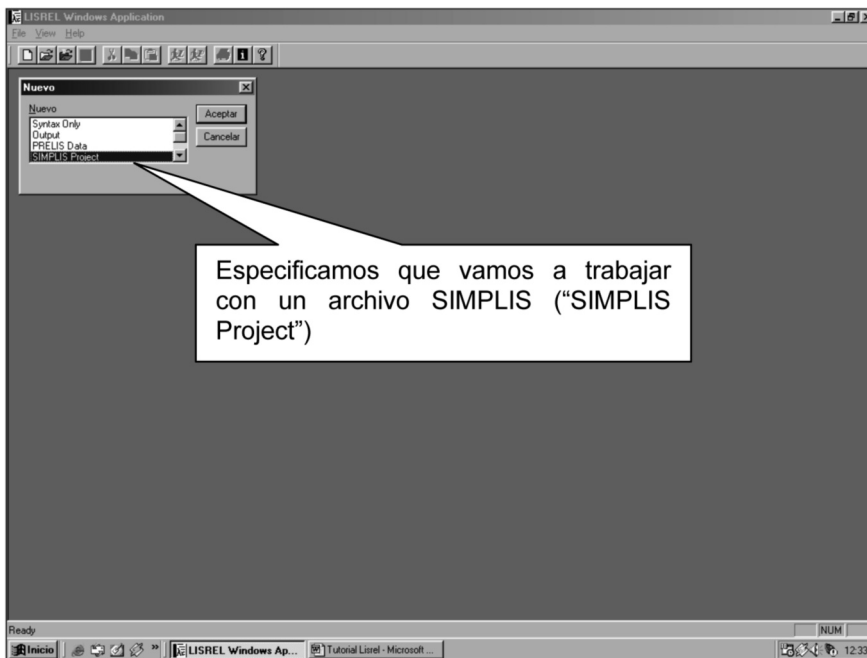
## 10.2. GENERACIÓN DE UN ARCHIVO SIMPLIS

Comencemos a familiarizarnos con el uso del programa. Al iniciar LISREL, lo primero que visualizaremos es una pantalla vacía que mostrará en su parte superior el menú principal del programa. Para generar un archivo SIMPLIS deberemos desplegar el menú “File” y seleccionar la opción “New”. En el caso de que ya dispongamos o hayamos creado previamente un archivo de trabajo, podemos abrirlo seleccionando la opción “Open” para trabajar con él.

**Figura nº 1. Generación de un archivo Simplis (1)**



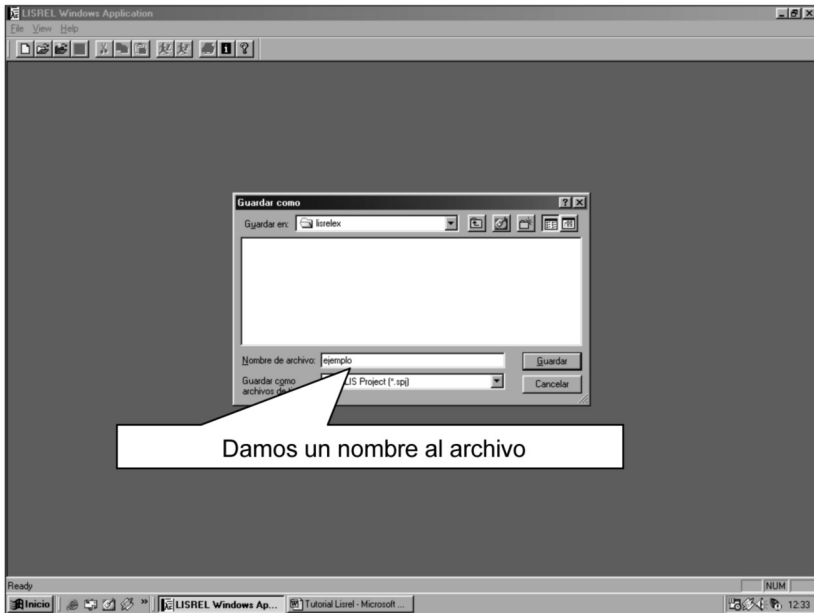
A continuación seleccionamos la opción “SIMPLIS Project”, con la que especificamos al programa que vamos a trabajar en lenguaje SIMPLIS.

**Figura nº 2. Generación de un archivo Simplis (2)**

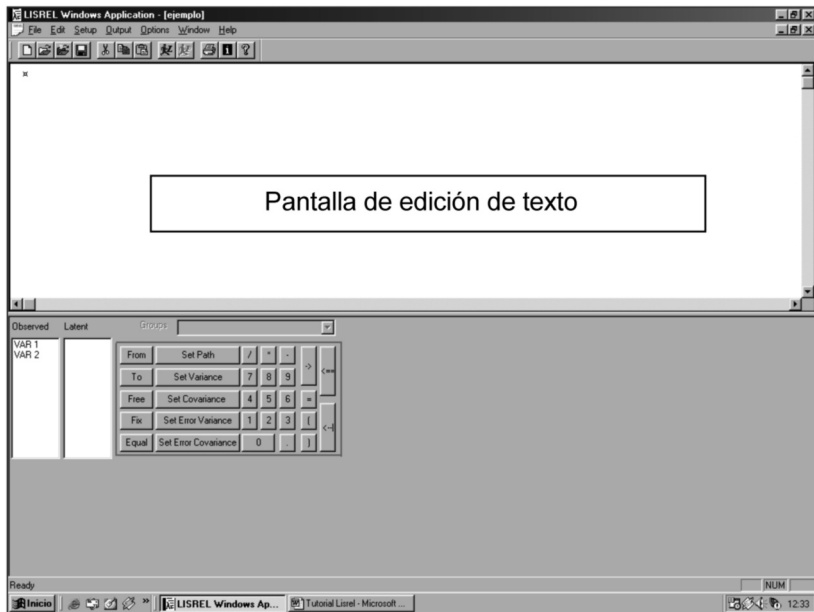
Seguidamente deberemos dar un nombre al proyecto que vamos a generar. Se creará un archivo con extensión .SPJ en el que se almacenarán todos los datos y comandos que especifiquemos. Cabe señalar que un proyecto SIMPLIS no se compone únicamente del archivo SPJ. El programa generará un conjunto de archivos complementarios cuando ejecutemos la programación, incluyendo archivos de output, ajuste, diagrama, etc., y todos estos archivos se guardarán en la misma carpeta que ocupa el archivo .SPJ. Por ello, es recomendable crear una carpeta específica para el proyecto, de forma que tengamos reunidos y organizados todos los archivos (ver Fig. 3).

El programa estará listo para comenzar a trabajar el archivo de input, cuya información iremos introduciendo en la pantalla de edición de texto. Podemos modificar el tamaño de esta pantalla todo lo que consideremos necesario, pinchando con el botón izquierdo del ratón en la línea divisoria inferior de la pantalla de edición de texto y arrastrando hacia abajo (ver Fig. 4).

**Figura nº 3. Generación de un archivo Simplis (3)**



**Figura nº 4. Generación de un archivo Simplis (4)**



### 10.3. CONVENCIONES DEL LENGUAJE SIMPLIS

El archivo de input representa el documento de trabajo en el que iremos introduciendo los datos y comandos necesario para la identificación, evaluación y ajuste del modelo.

SIMPLIS es un lenguaje de programación que utiliza lo que se denomina “lenguaje natural”, en este caso en inglés. Gracias a ello nos será más sencillo aprender su utilización, siempre que respetemos una serie de reglas y estructuras.

Un archivo básico de input en el lenguaje SIMPLIS consiste en un conjunto de líneas como las que se presentan en el ejemplo siguiente. Aunque este es un ejemplo relativamente completo, no incluye todas las opciones o toda la información que se puede incluir en las líneas. Algunas de ellas las detallaremos más adelante.

**Tabla n° 1. Ejemplo de archivo de input en LISREL**

LÍNEAS DEL ARCHIVO DE INPUT	Ejemplo de programación
<b>Etiqueta del modelo</b>	MODELO SOBRE DECISIONES PÚBLICAS ETIQUETAS CONFSOC = Confianza social CONFPPOL = Confianza en partidos políticos SATISDEM = Satisfacción con la democracia DECCIUD = Posicionamiento a favor de aumentar la capacidad de influencia de la ciudadanía en las decisiones públicas
<b>Observed Variables</b>	Observed variables CONFSOC CONFPPOL SATISDEM DECCIUD
<b>Covariance Matrix</b>	Covariance matrix 5.510 1.111 5.707 .906 2.478 6.369 -.172 -.741 -.596 5.902
<b>Sample Size</b>	Sample size 1432
<b>Relationships</b>	Relationships CONFPPOL = CONFSOC DECCIUD SATISDEM = CONFPPOL CONFSOC DECCIUD = SATISDEM
<b>Path Diagram</b>	Path diagram
<b>End of Problem</b>	End of Problem

### *Etiqueta del modelo*

Las primeras líneas son líneas libres de título. En LISREL pueden usarse tantas líneas como se consideren necesarias para el título, no existe limitación en este sentido. La primera línea para cada archivo de input por lo tanto puede ser una línea de título que contenga información como encabezamiento del problema. Se puede escoger no insertar una línea de título o utilizar una sola línea, pero con frecuencia se utilizan varias líneas para describir las características del modelo y los datos. Una recomendación útil puede ser consignar en este espacio también las etiquetas de las distintas variables que forman parte de la matriz de covarianzas para posteriores referencias.

### *Observed Variables*

Tras las líneas de títulos, si decidimos escribirlas, debemos incluir una línea con el comando *Observed Variables*. La razón de especificar el término *Observed* se debe a que en LISREL puede elaborarse modelos con variables no observables, es decir, variables latentes, cuyos nombres también tendríamos que definir con la etiqueta *Latent Variables* tras la línea de variables observables en el caso de que el modelo las incluyera. Tras el comando *Observed Variables* debemos dejar un espacio en blanco seguido de la lista de etiquetas de las variables observables correspondientes a los datos.

Es importante destacar que, mientras que en el caso de los comandos el programa no distingue entre mayúsculas y minúsculas (podremos escribir indistintamente “Observed Variables”, “observed variables” o “OBSERVED VARIABLES”), en los nombres de las variables sí se distingue entre mayúsculas y minúsculas, por lo que “edad”, “Edad” o “EDAD” serían consideradas por el programa como tres variables diferentes. También cabe decir aquí que, al menos en las versiones actuales del programa en el momento de escribir este manual, el programa truncará los nombres de las variables a los 8 primeros caracteres. Eso significa que si usamos dos variables como “NivelEstudiosPadres” y “NivelEstudiosHijos” el programa las interpretará como la misma variable, ya que únicamente leerá los 8 primeros caracteres, es decir, en ambos casos leerá “NivelEst” y daría un error por nombre duplicado. Por último, las variables no pueden incluir espacios ni caracteres especiales; si quisiéramos utilizar espacios tendríamos que poner la variable entre comillas, pero no es recomendable.

La selección de variables es automática en el sentido de que solamente las variables que forman parte del modelo se utilizan en el análisis, aunque existan más variables incluidas en la lista y en la matriz de datos. Es posible por lo tanto la introducción de un número determinado de variables y que se analice solamente un conjunto de ellas.

### *Covariance Matrix*

Las siguientes líneas del archivo de input se dedican a consignar los datos, que en el caso del ejemplo pertenecen a una matriz de covarianzas, y deben ir precedidas por el comando *Covariance Matrix*. Los elementos de la matriz deben aparecer separados, por ejemplo por espacios en blanco o tabuladores. Aunque LISREL lee los datos en diferentes opciones de formatos, los más usuales son matrices de covarianzas o correlaciones. La matriz de covarianzas o de correlaciones es una matriz de variables simétrica, de la cual se necesitan solo los elementos de la parte inferior y la diagonal. Naturalmente, el número de elementos en la matriz debe corresponder con el número de variables declaradas en la instrucción “Observed Variables” o el programa producirá un error de ejecución.

### *Sample Size*

La siguiente línea del archivo de input especifica el tamaño muestral, es decir, el número de casos en el que se basa la matriz de covarianzas. El tamaño muestral se debe introducir en forma de número tras las palabras *Sample Size* o en la siguiente línea.

### *Relationships*

En las siguientes líneas se ubican las ecuaciones de regresión a estimar. Aunque existen distintas formas de anotación en LISREL de estas ecuaciones, el ejemplo utiliza un formato en el que previo al signo igual se sitúa la variable dependiente, y tras el signo igual las variables independientes. Cada línea en este apartado generará una ecuación de regresión para el modelo, por tanto cada línea corresponde con una variable endógena. En lugar del comando “Relationships” podríamos utilizar también “Equations”.

### *Path Diagram*

El comando *Path Diagram* especifica que entre los outputs a generar se presentará el diagrama causal del modelo en sus distintas soluciones, junto con la estimación de los coeficientes de las relaciones entre variables.

### *End of Problem*

El comando *End of Problem*, que es opcional, especifica el final del archivo de input. En cualquier caso se recomienda su uso, especialmente cuando se incluyen varios problemas en el mismo archivo de input. A partir de este comando LISREL no interpretará ninguna otra instrucción.

#### 10.4. TESTADO Y AJUSTE DE MODELOS CON LISREL

Una vez descrito el contenido del archivo de input, estamos ya en condiciones de ejecutar el modelo. Para ello debemos pinchar con el cursor el icono de la barra de menús “Run Lisrel”:



Si todo va bien, y puesto que en el archivo de input hemos solicitado que el programa elabore el Path diagram, la primera pantalla que veremos al ejecutar el programa es el diagrama causal, aunque paralelamente el sistema creará tres archivos que son necesarios para el trabajo con modelos: un archivo de input (input file), un archivo de output (output file), y un archivo con el diagrama (path diagram). Podemos visualizar uno u otro abriendo el menú “Window” de la barra de menús desplegables.

El proceso de testado del modelo se realiza a partir de diferentes criterios, siendo las más usuales tres de ellas:

- a) En primer lugar, la valoración de los coeficientes globales de bondad de ajuste del modelo, con el fin de mostrar diferentes estadísticos que nos cercioran la validez empírica de la propuesta teórica y explicativa que propone el modelo.
- b) En segundo lugar, el examen de la significación de los coeficientes de las relaciones que componen el diagrama del modelo, que nos ayudan a evaluar de una forma gráfica las relaciones de dependencia entre las variables que forman parte del modelo.
- c) Finalmente, el análisis de las diferentes ecuaciones estructurales resultantes del modelo, que nos proporcionarán información entre otras cuestiones de la significación de cada una de las relaciones propuestas.

#### *Coefficientes globales de bondad de ajuste*

Como vimos en un capítulo anterior, los índices de ajuste más habituales que se utilizan para la validación de modelos son los siguientes: Chi-cuadrado, grados de libertad, RMSEA (Root Mean Square Error of Approximation) y p-Value.

El Chi-cuadrado es una medida de ajuste global del modelo sobre los datos. En cualquier caso, dada la sensibilidad que esta medida tiene con respecto al

tamaño muestral, muchos investigadores han propuesto toda una variedad de índices para evaluar el ajuste de los modelos. Buena parte de las medidas de bondad de ajuste son funciones de Chi-cuadrado y de los grados de libertad, y muchos de estos índices no solo consideran el ajuste del modelo, sino también su simplicidad: GFI (Goodness of Fit Index), AGFI (Adjusted Goodness of Fit Index), NFI (Normed Fit Index), NNFI (Non Normed Fit Index), CFI (Comparative Fit Index) o SRMR (Standardized Root Mean Square Residual). Si alguna de estas medidas indica un pobre ajuste sobre los datos, tendremos que proceder con cautela en la evaluación del ajuste del modelo.

**Tabla nº 2. Coeficientes globales de bondad de ajuste**

Estadísticos	Variación	Valores recomendables
RMSEA		< 0.05
<i>p</i> - Value	0 – 1	Valores cercanos a 1
GFI	0 – 1	> 0.90
AGFI	0 – 1	> 0.90
NFI	0 – 1	> 0.80
NNFI	0 – 1	> 0.95
CFI	0 – 1	> 0.95
SRMR		Valores cercanos a 0

Vemos los resultados del output en relación a los coeficientes globales de ajuste en el ejemplo propuesto (ver Tabla nº 3, pág. 124).

### *Estimados del Path Diagram*

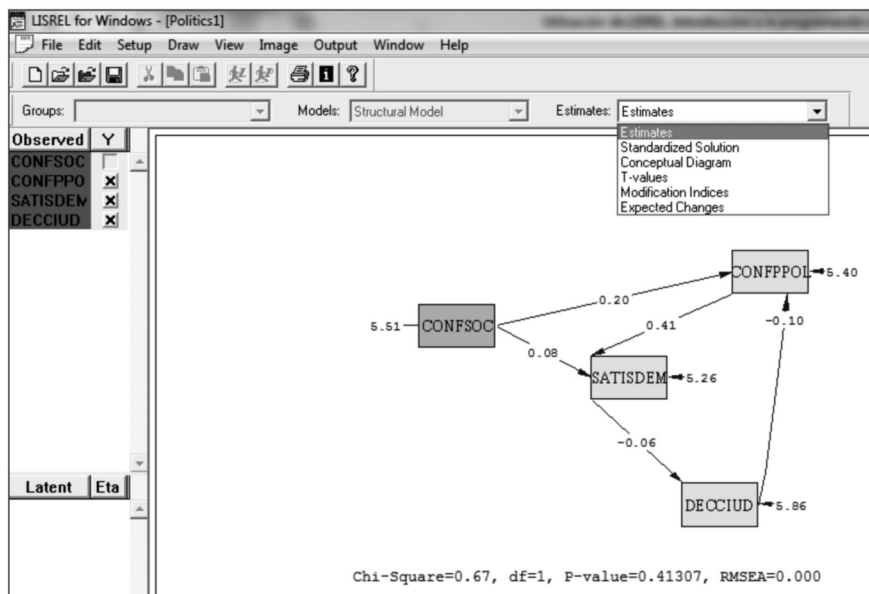
A través de la visualización del diagrama causal podremos valorar la existencia de algún parámetro anómalo o con un valor poco razonable. Los parámetros estimados deberían tener el signo y tamaño correcto de acuerdo con la teoría o las especificaciones que se realizan a priori. LISREL proporciona distintas soluciones que serán de gran ayuda en el ajuste y evaluación del modelo que generamos.

En la pantalla del diagrama, disponemos de un menú desplegable en la parte superior derecha denominado “estimates”. Este menú nos permitirá visualizar los distintos estimados del diagrama (ver Figura nº 5, pág. 125).

**Tabla nº 3. Archivo de output en LISREL**

Goodness of Fit Statistics	
Degrees of Freedom = 1	
Minimum Fit Function Chi-Square = 0.67 (P = 0.41)	
Normal Theory Weighted Least Squares Chi-Square = 0.67 (P = 0.41)	
Estimated Non-centrality Parameter (NCP) = 0.0	
90 Percent Confidence Interval for NCP = (0.0 ; 6.04)	
Minimum Fit Function Value = 0.00047	
Population Discrepancy Function Value (F0) = 0.0	
90 Percent Confidence Interval for F0 = (0.0 ; 0.0042)	
Root Mean Square Error of Approximation (RMSEA) = 0.0	
90 Percent Confidence Interval for RMSEA = (0.0 ; 0.065)	
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.86	
Expected Cross-Validation Index (ECVI) = 0.013	
90 Percent Confidence Interval for ECVI = (0.013 ; 0.018)	
ECVI for Saturated Model = 0.014	
ECVI for Independence Model = 0.26	
Chi-Square for Independence Model with 6 Degrees of Freedom = 369.54	
Independence AIC = 377.54	
Model AIC = 18.67	
Saturated AIC = 20.00	
Independence CAIC = 402.61	
Model CAIC = 75.07	
Saturated CAIC = 82.67	
Normed Fit Index (NFI) = 1.00	
Non-Normed Fit Index (NNFI) = 1.00	
Parsimony Normed Fit Index (PNFI) = 0.17	
Comparative Fit Index (CFI) = 1.00	
Incremental Fit Index (IFI) = 1.00	
Relative Fit Index (RFI) = 0.99	
Critical N (CN) = 14170.31	
Root Mean Square Residual (RMR) = 0.039	
Standardized RMR = 0.0069	
Goodness of Fit Index (GFI) = 1.00	
Adjusted Goodness of Fit Index (AGFI) = 1.00	
Parsimony Goodness of Fit Index (PGFI) = 0.100	

Figura nº 5. Estimados del Path Diagram



Los coeficientes estimados en el Path Diagram puede asumir distintas opciones:

- Solución estimada*
- Solución estandarizada*. Es la más habitual en la presentación de los modelos.
- T-values*. Esta opción es especialmente útil para el ajuste del modelo. Los t-values aparecen en el Path Diagram para cada parámetro estimado. En el Path Diagram, los t-values significativos aparecerán en color negro, y los t-values no significativos se mostrarán en un color diferente (por defecto en color rojo). En efecto, existe un umbral de valores que determina qué t-values se considerarán significativos y cuáles no serán significativos. Salvo que indiquemos lo contrario, valores inferiores a 1.96 de magnitud se mostrarán en color rojo. Esto quiere decir que los valores que se sitúen en la horquilla que comprende desde el  $-1.96$  hasta  $+1.96$  se considerarán no significativos empíricamente.
- Índices de modificación*. LISREL puede proponer modificaciones para cada parámetro del modelo. Cada índice de modificación mide cuánta disminución se espera en el Chi-square si se modifica este parámetro particular y el modelo es reestimado. Los índices de modificación se utilizan en el proceso de evaluación y ajuste del modelo a par-

tir de la valoración del investigador en torno a la pertinencia explicativa de cada índice propuesto por LISREL, comenzando normalmente por el que mayor magnitud presenta, lo que conduce a una reestimación dinámica del modelo.

Para la medición de las relaciones, existe además un error en cada variable observada, como se puede apreciar en las flechas anexas a los rectángulos que las representan. El término de error se interpreta en este caso como errores de medición (o errores observacionales) en las variables observables, aunque pueden contener otros componentes sistemáticos específicos.

#### *Análisis de las ecuaciones estructurales resultantes del modelo*

Las ecuaciones estructurales ofrecen una representación alternativa de las hipótesis causales presentes en el modelo que puede ayudarnos a sistematizar las relaciones propuestas, y se encuentran en el archivo de output de LISREL. El modelo utilizado en el ejemplo habrá producido tres ecuaciones estructurales, correspondientes a las variables endógenas del modelo. Escojamos una de ellas para ejemplificar la información contenida en ellas.

$$\text{SATISDEM} = 0.41 * \text{CONFPPOL} + 0.081 * \text{CONFSOC}, \text{ Errorvar.} = 5.26, R^2 = 0.17$$

(0.026)	(0.026)	(0.20)
15.89	3.08	26.74

El coeficiente de regresión aparece antes del signo \* de cada variable. Los números entre paréntesis debajo de los coeficientes de regresión indican los errores standard de los parámetros estimados. Debajo de los errores standard están los t-valores de cada parámetro estimado. La correlación múltiple al cuadrado ( $R^2$ ) se da también para cada relación, siendo esta última una medida de la fuerza de la relación lineal.

#### 10.5. MODELOS DE MEDICIÓN

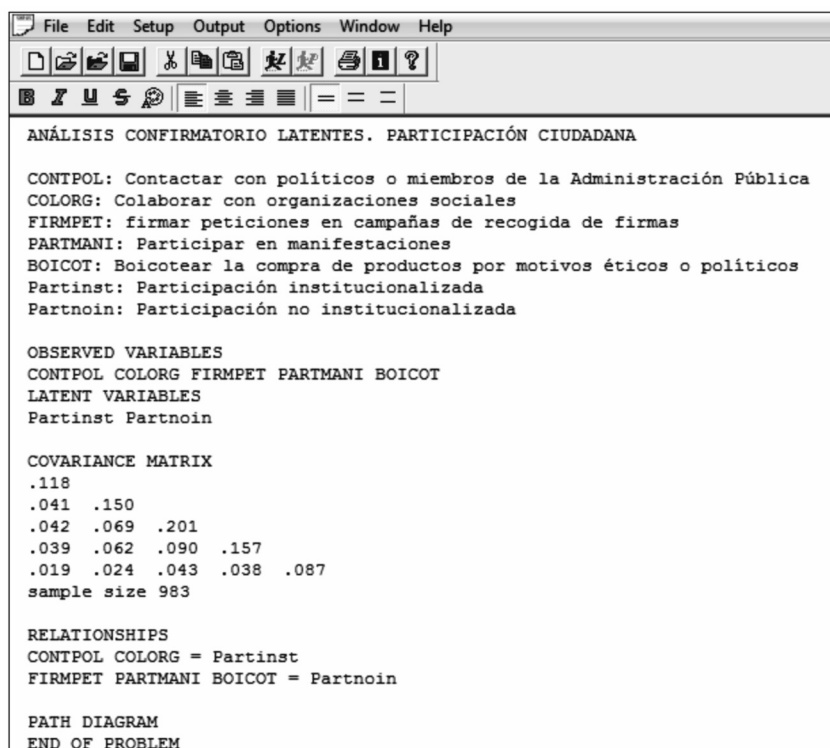
LISREL en ocasiones es útil para testar y ajustar la validez de los denominados modelos de medición. Como se comentó, los modelos de medición especifican y estiman las relaciones entre uno o varios factores (variables latentes o constructos teóricos) y las variables observables en que se anclan, con la condición previa del manejo por parte del investigador de la naturaleza factorial de la estructura a testar. La intención en un modelo de medición es, por lo tanto, describir hasta qué punto los indicadores observables sirven como un buen instrumento de medición de las variables definidas como latentes, lo que liga esta acción al análisis factorial confirmatorio. El análisis factorial confirmatorio en LISREL, al igual que en otros paquetes estadísticos, intentará identificar una serie de factores latentes predefinidos menor que el número de varia-

bles observables en los datos, a partir de la propuesta teórica del investigador. Estos factores latentes serán los que deberán dar cuenta de las intercorrelaciones entre las variables observables manejadas.

Aunque LISREL es indicado en estudios confirmatorios, puede ser también utilizado para la identificación de análisis factoriales exploratorios, siempre que al menos se manejen unas hipótesis de partida con la distribución de cargas factoriales, entendiéndolo así como paso previo o tentativo en el diseño de análisis factoriales confirmatorios.

Veamos un ejemplo a partir del siguiente archivo de input.

**Figura nº 6. Ejemplo de input para modelo de medición**



```

File Edit Setup Output Options Window Help
[Icons]
[Icons]
ANÁLISIS CONFIRMATORIO LATENTES. PARTICIPACIÓN CIUDADANA

CONTPOL: Contactar con políticos o miembros de la Administración Pública
COLORG: Colaborar con organizaciones sociales
FIRMPET: firmar peticiones en campañas de recogida de firmas
PARTMANI: Participar en manifestaciones
BOICOT: Boicotear la compra de productos por motivos éticos o políticos
Partinst: Participación institucionalizada
Partnoin: Participación no institucionalizada

OBSERVED VARIABLES
CONTPOL COLORG FIRMPET PARTMANI BOICOT
LATENT VARIABLES
Partinst Partnoin

COVARIANCE MATRIX
.118
.041 .150
.042 .069 .201
.039 .062 .090 .157
.019 .024 .043 .038 .087
sample size 983

RELATIONSHIPS
CONTPOL COLORG = Partinst
FIRMPET PARTMANI BOICOT = Partnoin

PATH DIAGRAM
END OF PROBLEM

```

A diferencia de ejemplos anteriores, este archivo de input añade una línea tras el comando de “Observed variables” en el que se incluyen las etiquetas de las variables latentes o factores a testar: “Latent variables”.

```

LATENT VARIABLES
Partinst Partnoin

```

Además, en las líneas relativas a la descripción de las ecuaciones se incluye la propuesta de distribución de las variables indicadoras entre los factores propuestos, de forma que antes del signo igual se ubica el conjunto de variables indicadoras y después del signo igual el factor establecido.

#### RELATIONSHIPS

CONTPOL COLORG = Partinst

FIRMPET PARTMANI BOICOT = Partnoin

Si bien en este ejemplo se han indicado en primer lugar las variables observadas seguidas del signo igual (=) y la variable latente, tenemos la opción de utilizar una notación alternativa, poniendo en primer lugar la variable latente, seguida de un guión y el signo “mayor que”, a modo de flecha (->) y después las variables indicadoras. Es decir:

#### RELATIONSHIPS

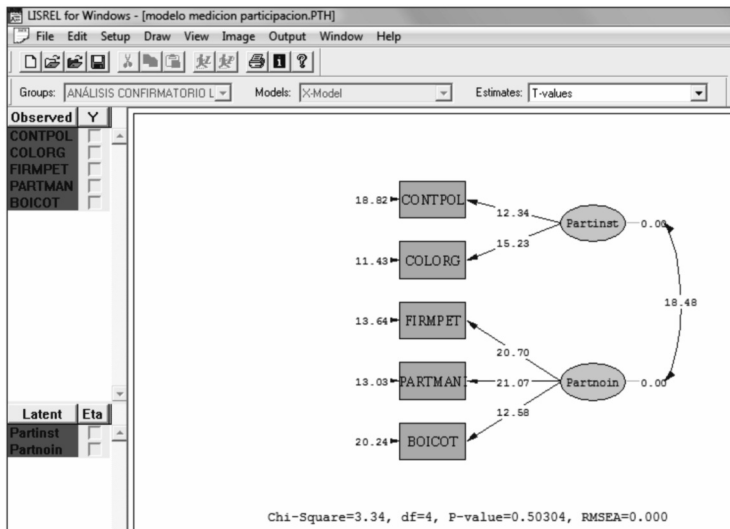
Partinst -> CONTPOL COLORG

Partnoin -> FIRMPET PARTMANI BOICOT

Del mismo modo, es recomendable establecer una convención sobre la utilización de mayúsculas y minúsculas. Una de estas convenciones es utilizar mayúsculas para los nombres de las variables latentes y minúsculas para variables observadas, o bien como en este ejemplo, minúsculas para las variables latentes y mayúsculas para sus variables indicadoras.

El resultado del Path Diagram es el que se detalla a continuación.

**Figura nº 7. Path Diagram de ejemplo de modelo de medición**



Como se puede observar, tanto los coeficientes de ajuste global (proporción entre Chi-square y grados de libertad, P-value, RMSEA) son adecuados, y además las cargas de los valores T para cada variable indicadora en relación con su factor están dentro de los márgenes aceptados como significativos, por lo que podemos concluir con que la estructura factorial propuesta ajusta adecuadamente sobre los datos. Como es posible imaginar, esta utilidad de LISREL permite tanto asumir el modelo de medición en términos de análisis finalista como utilizar los modelos de medición para definir variables latentes que serán integradas en la construcción de un modelo estructural.

#### 10.6. MODELOS ESTRUCTURALES CON VARIABLES LATENTES

Hasta el momento los ejemplos han mostrado las estrategias de construcción de modelos con variables, y posteriormente hemos presentado indicaciones para la construcción de modelos de medición de variables latentes. En este apartado abordamos la elaboración de modelos estructurales que asumen variables latentes como integrantes del modelo. Aquí, por lo tanto, un modelo se considera un sistema estructural de ecuaciones para un conjunto de variables latentes clasificadas como dependientes o independientes en función del lugar que ocupan en la propuesta del investigador.

Veamos un ejemplo a partir del archivo de input (ver Figura n° 8, pág. 130).

Podemos observar que en la parte del input destinada a la ubicación de las ecuaciones, aparecen tanto las líneas que definen los modelos de medición de las variables latentes utilizadas como las que conforman la formulación del modelo estructura propuesto, que establecen el orden de dependencia entre las variables latentes.

##### RELATIONSHIPS

IMPFAMIL IMPAMIGS IMPTPLIB = Sociab

IMPBARRI IMPCIUDA = Imploc

MIEMBRO PARTICIP = Actasoc

IMPSOLID IMPENDEM = Comprcom

Imploc = Sociab

Actasoc = Sociab

Comprcom = Sociab Imploc Actasoc

} Modelos de medición

} Ecuaciones estructurales

Si solicitamos el Path Diagram de este input, obtenemos la siguiente representación gráfica del modelo estructural (ver Figura n° 9, pág. 131).

**Figura nº 8. Ejemplo de input para modelo estructural con variables latentes**

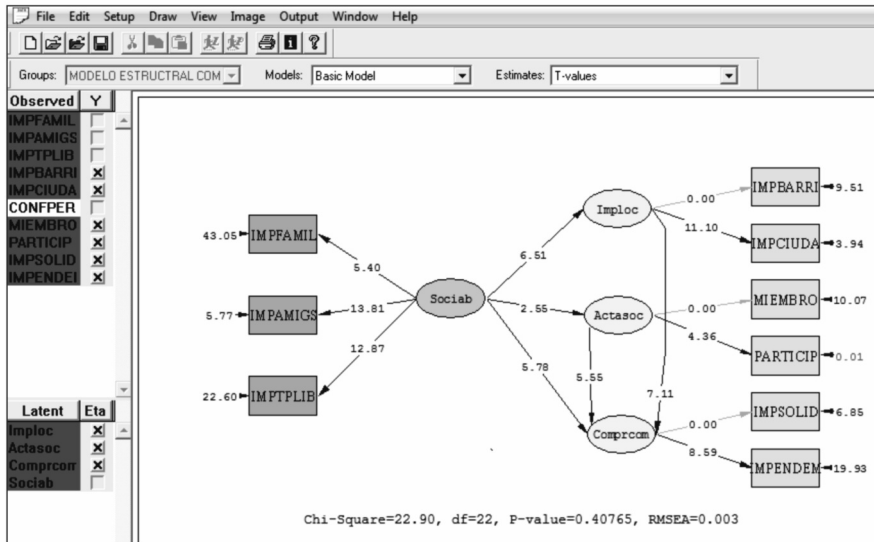
```

File Edit Setup Output Options Window Help
[Icons]
B I U S [Icons] = = =
MODELO ESTRUCTRAL COMPROMISO COMUNITARIO
IMPFAMIL: Importancia de la familia, IMPAMIGS: Importancia de Iso amigos
IMPTPLIB: Importancia del tiempo libre, IMPBARRI: Apego al barrio donde vive
IMPCIUDA: Apego al pueblo o ciudad donde vive, MIEMBRO: Es miembro de asociaciones
PARTICIP: Participa en actividades de asociaciones, IMPSOLID: Ser solidario con la gente que está peor
IMPENDEM: Pensar en los demás más que en uno mismo
Sociab: Sociabilidad, Imploc: Implicación en lo local
Actasoc. Actividad asociativa, Comprcom: Compromiso comunitario

OBSERVED VARIABLES IMPFAMIL IMPAMIGS IMPTPLIB IMPBARRI IMPCIUDA CONFERS
MIEMBRO PARTICIP IMPSOLID IMPENDEM
LATENT VARIABLES Sociab Imploc Actasoc Comprcom
Covariance Matrix
11.7314
.5462 3.2303
.2160 1.1468 3.4681
.2993 .4230 .1731 5.5185
.3613 .4432 .2349 3.1879 4.6368
-.0483 .4027 .1470 .0500 .1198 4.4563
-.0211 .0190 .0189 .0162 -.0002 -.0141 .1879
.0218 .0342 .0276 .0132 .0012 .0198 .0957 .1641
.1281 .2946 .2834 .4338 .4735 .0856 .0306 .0607 2.6588
.1966 .3383 .2290 .4098 .3941 .4285 .0585 .0694 1.4647 5.1864
SAMPLE SIZE = 3782
RELATIONSHIPS
IMPFAMIL IMPAMIGS IMPTPLIB = Sociab
IMPBARRI IMPCIUDA = Imploc
MIEMBRO PARTICIP = Actasoc
IMPSOLID IMPENDEM = Comprcom
Imploc = Sociab
Actasoc = Sociab
Comprcom = Sociab Imploc Actasoc
PATH DIAGRAM
END OF PROBLEM

```

**Figura nº 9. Path Diagram de ejemplo de modelo estructural con variables latentes**



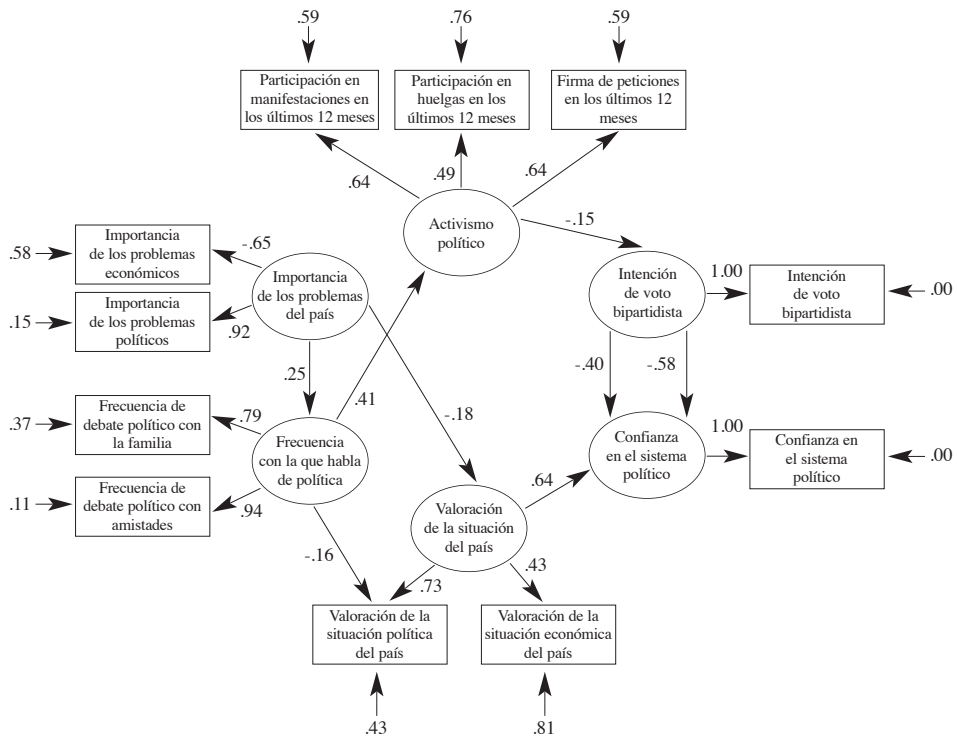
El modelo resultante está compuesto por la estructura múltiple de dependencia entre las variables latentes, ya sean exógenas o endógenas, junto con los coeficientes estimados para cada relación. Las variables observables en este caso actúan únicamente como componentes de los distintos modelos de medición que definen las variables latentes, no formando parte del modelo estructural en ningún caso.



## 11. EJEMPLOS DE MODELOS

A continuación se incluyen algunos ejemplos de modelos ajustados sobre datos reales.

### EJEMPLO 1. CONTEXTO SOCIOPOLÍTICO Y VOTO BIPARTIDISTA EN LA JUVENTUD ESPAÑOLA



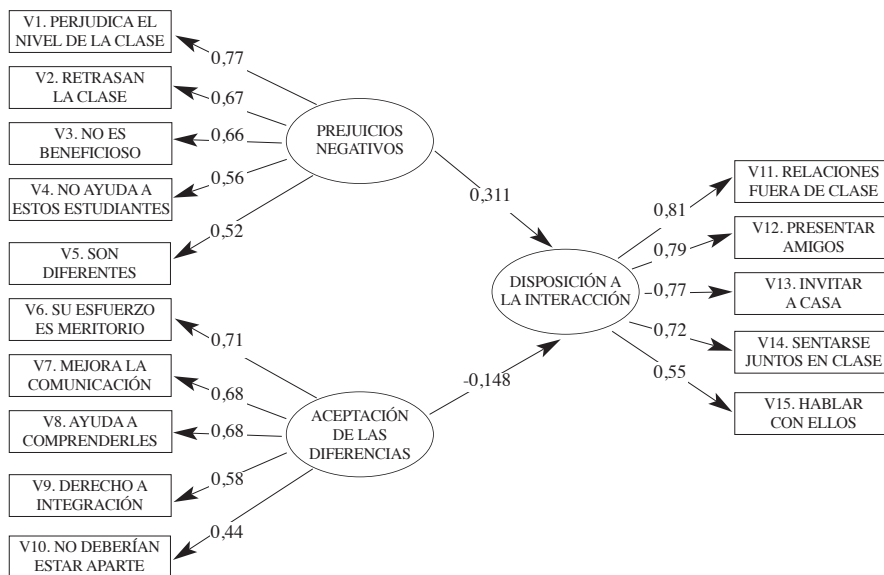
Chi-Square=35.93, df=37, P-value=0.51892, RMSEA=0.000

## RESUMEN

El modelo estructural muestra la tendencia creciente de los jóvenes en España hacia mecanismos no convencionales de expresión política, fundamentada en un proceso de repolitización activa en paralelo a una evaluación negativa de la situación política y económica del país y a un descenso de confianza en el sistema político. Así, es posible observar cómo la preocupación por los temas políticos ejerce un impacto positivo sobre el activismo de los jóvenes e inhibe la intención de voto a los dos grandes partidos políticos, todo ello en una estructura de interdependencia de variables propias del contexto sociopolítico.

Fuente: Francés, F. y Santacreu, O. (2014) “Crisis política y juventud en España: el declive del bipartidismo electoral”, *SocietàMutamentoPolitica*, 5 (10): 107-128.

## EJEMPLO 2. PREDISPOSICIÓN A LA INTERACCIÓN EN LA ESCUELA A PARTIR DE LAS ACTITUDES DE LOS ALUMNOS



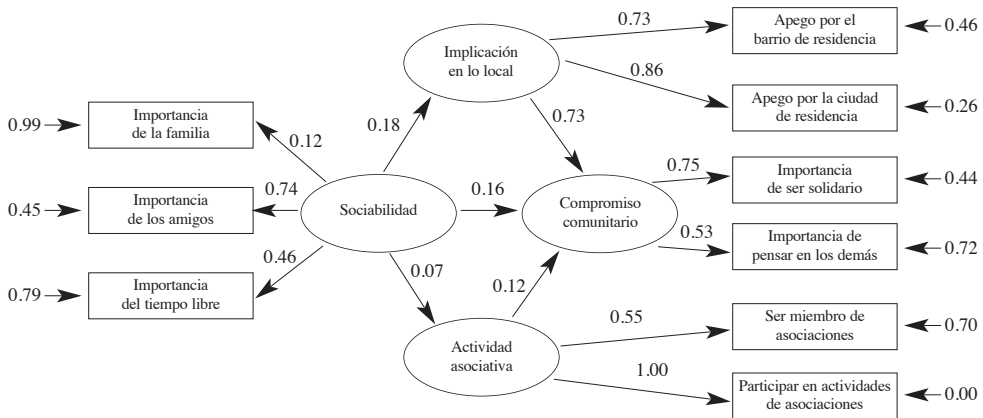
### RESUMEN

El modelo estructural explica la predisposición de los alumnos en la escuela a desarrollar comportamientos relacionados con la disposición a la interacción en función de las actitudes que presentan. En la estructura explicativa los prejuicios negativos inhiben o retraen la disposición a relacionarse de los alumnos. Por el contrario, la aceptación de las diferencias potencia la disposición a la interacción. En este sentido, puede observarse que los efectos de ambas variables latentes ejercen un impacto de signo distinto en la disposición a la interacción entendida como comportamiento.

Fuente: Alaminos, A.; Santacreu, O.; Mula, A.; Navas, L. et al. (2005) "Actitudes y expectativas sobre los procesos de integración a través de la escuela", en Alaminos, A.; Francés, F. y Santacreu, O. *Socialización, ideología y participación. Casos prácticos de análisis estructural de covarianzas*. Alicante: OBETS



### EJEMPLO 3. LA ESTRUCTURA DEL CAPITAL SOCIAL



Chi-Square=22.90, df=22, P-value=0.40738, RMSEA=0.003

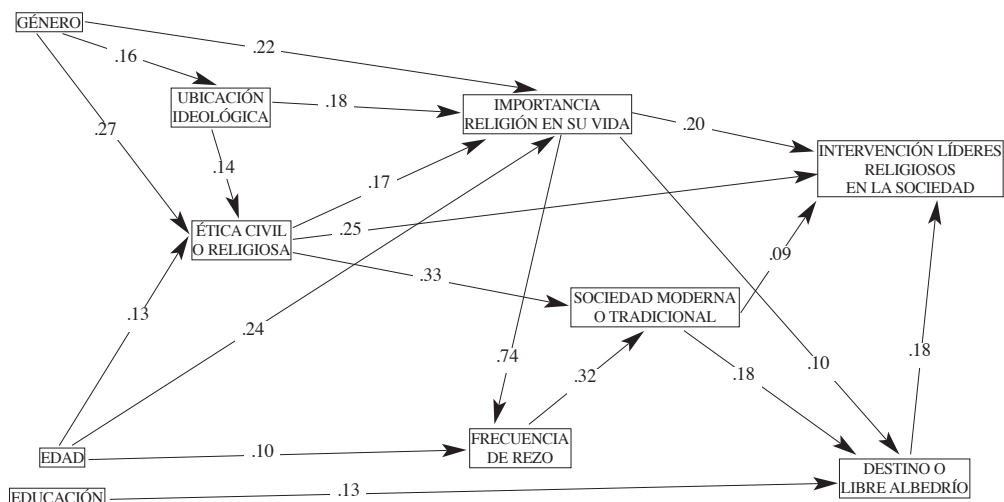
#### RESUMEN

El modelo estructural muestra la estructura múltiple de interdependencia entre cuatro variables clásicas de la noción de capital social: la sociabilidad informal, la pertenencia a asociaciones o colectivos, la implicación en el entorno cotidiano y el compromiso comunitario. En él se observa la capacidad explicativa que posee la variable latente de sociabilidad como elemento motriz del capital social, impactando positivamente sobre el resto de variables analizadas para explicar el desarrollo del compromiso comunitario de los sujetos. La participación asociativa, fundamento de gran parte de los enfoques del capital social, descansaría de acuerdo a ello en la esfera de las relaciones informales de los sujetos.

Fuente: Francés, F. (2004) “Una aproximación al concepto de capital social a través de la técnica de modelos estructurales”, en Alaminos, A.; Francés, F. y Santacreu, O. (eds.) *Reflexiones teóricas y modelos empíricos sobre identidad, diversidad y participación social*. Alicante: OBETS.



## EJEMPLO 4. ÉTICA CIVIL Y RELIGIOSA EN ESPAÑA



Chi-Square=17.42, df=23, P-value=0.798, RMSEA=0.000

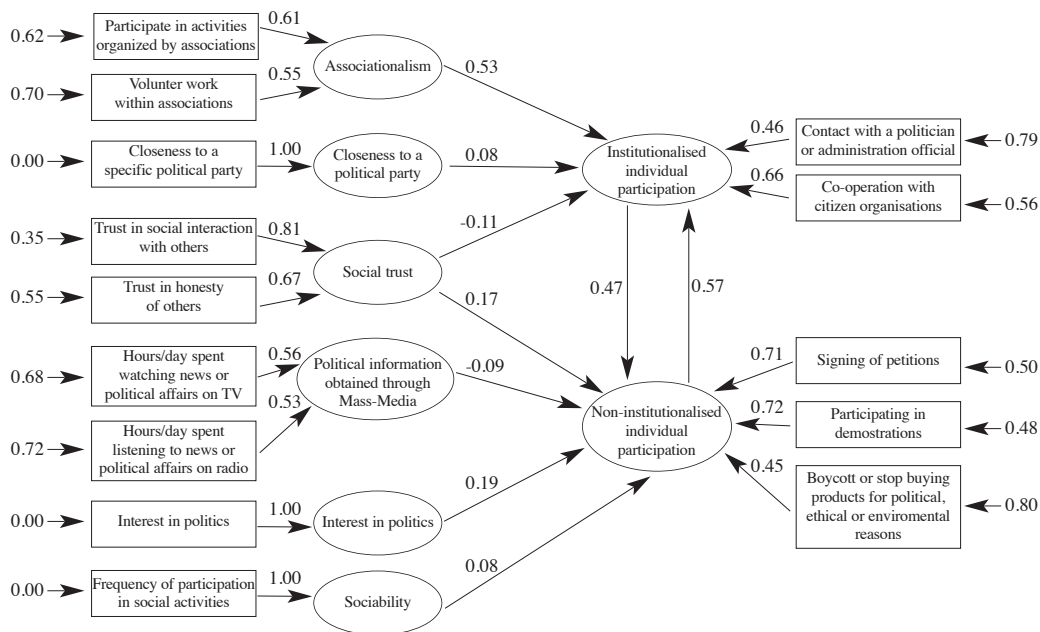
### RESUMEN

En el modelo estructural, realizado sobre la población española, la ética cívica y religiosa se encuentra en la base de las opiniones referidas a la presencia y función de los valores religiosos y la intervención de la iglesia católica en la vida cotidiana. Los resultados del modelo señalan cómo la edad muestra claramente el cambio social y deja entrever hacia dónde se dirige la religiosidad de la sociedad española. Las personas más jóvenes son las más partidarias de la separación entre Estado y religión, una menor práctica de la religión y un mayor peso de la no necesidad de creer para tener una conducta moral. Los datos indican que las opiniones favorables a la religión y mayor práctica íntima de la religión están más presentes en las mujeres. Por otro lado, la educación no es relevante a la hora de predecir la adopción de una ética civil o religiosa, pero sí para ser partidario de la separación entre Estado y religión y para asumir la autonomía frente al destino.

Fuente: Alaminos, A. y Penalva, C. (2012) "Ética civil y ética religiosa en España", *Revista Internacional de Sociología*, 70 (2): 347-373.



## EJEMPLO 5. PARTICIPACIÓN SOCIAL INDIVIDUAL EN LA ESFERA SOCIAL Y POLÍTICA



Chi-Square=58.52, df=56, P-value=0.3867, RMSEA=0.007

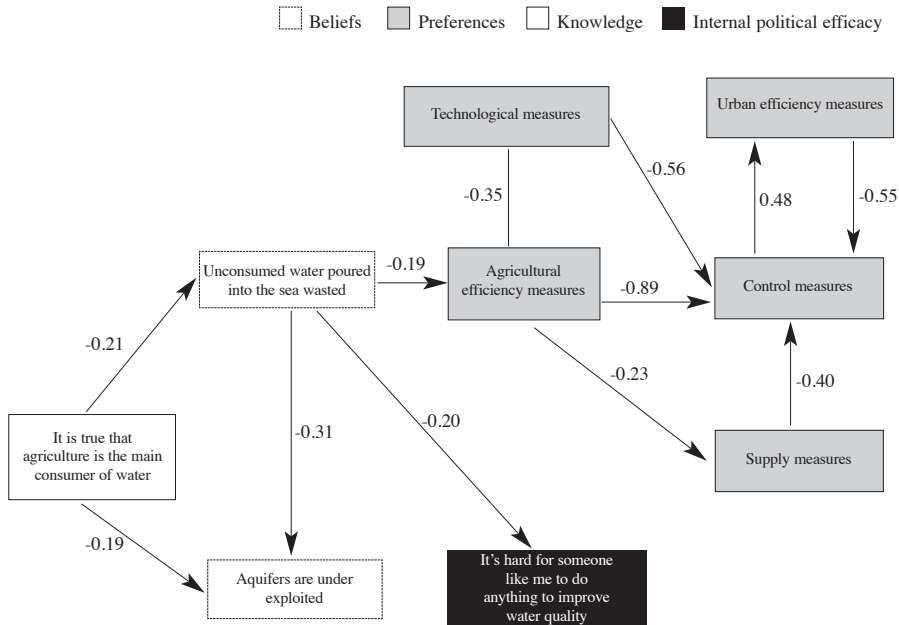
### RESUMEN

El modelo estructural muestra cómo distintas variables propias del contexto sociopolítico (asociacionismo, cercanía a partidos políticos, confianza social, sociabilidad, interés por la política y seguimiento de información política) ejercen relaciones diferenciadas de dependencia sobre dos formas básicas de participación social: la participación en los cauces previstos por las instituciones y la participación que se desarrolla al margen de ellas, que constituyen las variables endógenas del modelo. El resultado muestra, entre otras conclusiones, que ambas formas de participación no se desarrollan en paralelo sino que se retroalimentan mutuamente en el marco de la acción participativa, generando un modelo no recursivo.

Fuente: Ganuza, E. y Francés, F. (2008) "Rethinking the Sources of Participation: A Case Study of Spain", *European Sociological Review*, 24(4): 479-493.



## EJEMPLO 6. ESTRUCTURA ACTITUDINAL Y DE CREENCIAS SOBRE EL USO DEL AGUA



Chi-Square=21.65, df=23, P-value=0.54153, RMSEA=0.000

### RESUMEN

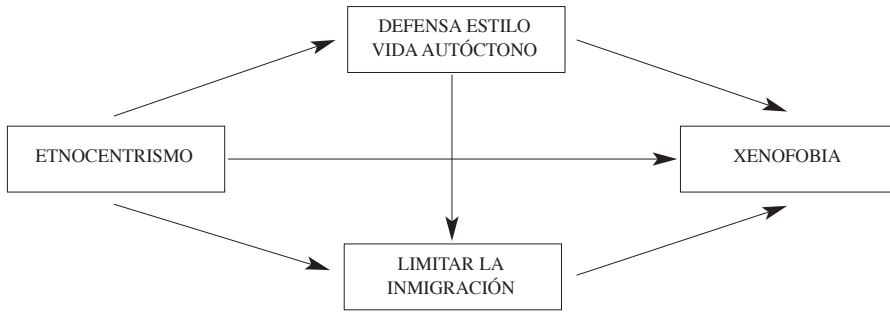
El modelo muestra la estructura múltiple de interdependencia en términos de variables relativas en torno a las creencias, preferencias y conocimientos sobre el uso del agua por parte de la población. En el resultado se observa el débil nexo existente entre la red de creencias y de preferencias por parte de los sujetos. Solo la medida relativa al ahorro agrario de agua se vincula al resto de dimensiones analizadas en la red actitudinal, aunque no se encuentra relacionada directamente con variables relativas al conocimiento. La estructura del modelo aborda por lo tanto una marcada disonancia entre los juicios objetivos y subjetivos de los usuarios del agua y las preferencias de éstos en torno a las medidas convenientes para controlar su uso.

Fuente: Ganuza, E.; Francés, F; Lafuente, R. y Garrido, F. (2012) "Do Participants Change Their Preferences in the Deliberative Process?", *Revista Española de Investigaciones Sociológicas*, 139: 225-246.

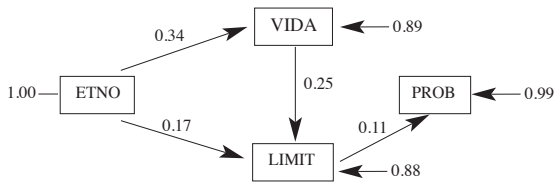


## EJEMPLO 7. ETNOCENTRISMO, XENOFOBIA Y MIGRACIONES

Figura 1: Modelo general

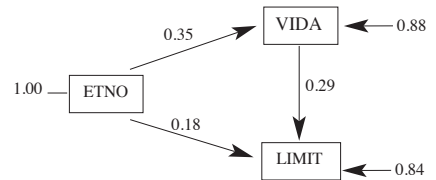


### MÉXICO



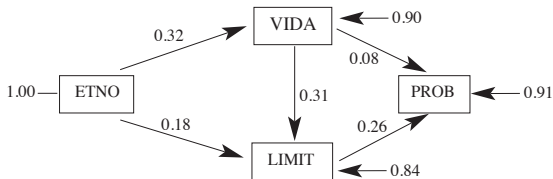
$\chi^2 = 0.44$ , GL = 2, P-valor = 0.80385, RMSEA = 0.000

### BOLIVIA



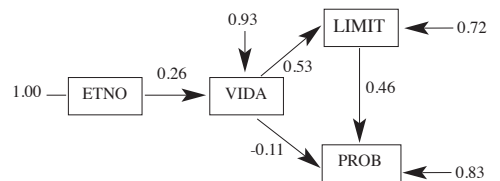
$\chi^2 = 0.00$ , GL = 0, P-valor = 1.00000, RMSEA = 0.000

### ARGENTINA



$\chi^2 = 0.14$ , GL = 1, P-valor = 0.70597, RMSEA = 0.000

### ESPAÑA



$\chi^2 = 0.39$ , GL = 2, P-valor = 0.82112, RMSEA = 0.000

## RESUMEN

El ejemplo plantea un análisis multigrupo de modelos estructurales por países en torno a las consecuencias sociales del etnocentrismo en relación a las migraciones. Los resultados muestran al etnocentrismo como un elemento común a las sociedades analizadas, de forma que los núcleos de opinión etnocéntrica se encuentran tanto en los países emisores de emigración como en las sociedades receptoras. Asimismo, se comprueba que la xenofobia, si bien se activa contextualmente tiene sus raíces emocionales en los prejuicios etnocéntricos, indistintamente de posibles experiencias negativas con los inmigrantes.

Fuente: Alaminos, A.; López, C. y Santacreu, O. (2010) “Etnocentrismo, xenofobia y migraciones internacionales en una perspectiva comparada”, *Convergencia*, 17(53): 91-124.

## BIBLIOGRAFÍA

- Abell, P. "Measurement in Sociology I. Measurement and systems". *Sociology*, 2, (1968) 1-20.
- Abell, P. "Measurement in Sociology II. Measurement structure and sociological theory". *Sociology*, 3, (1969) 397-411.
- Abercrombie, Nicholas; Hill, Stephen; Turner, Bryan S. *The penguin dictionary of sociology*. New York, Penguin, 1996.
- Ackoff, R. L. *Redesigning the Future. A Systems Approach to Societal Problems*. New York, John Wiley & Sons, 1974.
- Ackoff, R.L. "On pairs and trios: The smallest social systems". *Systems* Adams, D., A. Ahmad, D. Haynes & J. Sheehan, "Embodiments of Beer's Educational Model". *Kybernetes*, 22(6), (1993) 44-59.
- Adorno, T. *Dialéctica Negativa*, Madrid, Taurus-Cuadernos para el Diálogo, 1975.
- Adorno, T.W. *La disputa del positivismo en la sociología alemana*, Barcelona, Grijalbo, 1973.
- Aigner, D.J. ; Goldberger, A.S., *Latent variables in socioeconomic models*, Amsterdam, North holland, 1977.
- Alaminos, A; Santacreu, O.; Mula, A.; Navas, L. et al. "Actitudes y expectativas sobre los procesos de integración a través de la escuela", en Alaminos, A.; Francés, F. y Santacreu, O. *Socialización, ideología y participación. Casos prácticos de análisis estructural de covarianzas*. Alicante: OBETS, 2005
- Alaminos, A. *El análisis de la realidad social. Modelos estructurales de covarianzas*. Alicante, OBETS, 2005.
- Alaminos, A. *Introducción a la Sociología Matemática*. Alicante, SPES, 2005.
- Alaminos, A. *Mode effects*, RC33, Mannheim, 1998.
- Alaminos, A. *Sociología Matemática*, Alicante, Gamma, 1996.
- Alaminos, A. *Testing political ideology configural invariance in Europe*. RC33 Amsterdam. 2004.
- Alaminos, A. *The structures of don't know*. LSDA. Colonia, 1999.
- Alaminos, A. y Penalva, C. "Ética civil y ética religiosa en España", *Revista Internacional de Sociología*, 70 (2), 2012, pp. 347-373.
- Alaminos, A., Frances, F. y Santacreu, O. *Reflexiones teóricas y modelos empíricos sobre identidad, diversidad y participación social*. Alicante, OBETS, 2004.
- Alaminos, A., Frances, F. y Santacreu, O. *Socialización, ideología y participación: casos prácticos de análisis estructural de covarianzas*. Alicante, OBETS, 2004

- Alaminos, A.; López, C. y Santacreu, O. "Etnocentrismo, xenofobia y migraciones internacionales en una perspectiva comparada", *Convergencia*, 17(53), 2010, pp. 91-124.
- Alexander, J.C. "Looking for theory: "Facts" and "values" as the intellectual legacy of the 1970s". *Theory and Society*, 10, (1981) 279-292.
- Anheier, Helmut K. "Structural analysis and strategic research design: Studying politicized interorganizational networks". *Sociological Forum* 2, (1987) 563-582.
- Arabie, P., S.A. Boorman, and P.R. Levitt "Constructing blockmodels: How and why". *Journal of Mathematical Psychology*, Vol. 17, (1978) 21-63.
- Archer, M., "Structuration versus morphogenesis". In: S.N. Eisenstadt & H.J. Helle (eds.), *Macro Sociological Theory: Perspectives on Sociological Theory*. London, Sage, 1985, 59-68.
- Asher, H.B., *Causal modeling*, Beverly Hills, Sage, 1983.
- Aulin, A., *Foundations of Mathematical System Dynamics; The Fundamental Theory of Causal Recursion and its Application to Social Science and Economics*. Oxford, Pergamon, 1990.
- Axten, N. And Fararo, T.J. "The information processing representation of institutionalized social action". In P. Krishann (ed.), *Mathematical Models of Sociology*. Keele, U.K.: Sociological Review Monograph No. 24, 1977.
- Bagozzi, R.P., *Causal models in marketing*, New York, Wiley, 1980.
- Bartholomew, David J. *Mathematical Methods in Social Science*. Chichester: John Wiley & Sons, 1981.
- Bartholomew, David J. *Stochastic Models for Social Processes*, New York, NY: John Wiley, 1967.
- Beaton, A.E. "An inter-battery factor analytic approach to clique analysis". *Sociometry*, 29, (1966) 135-45.
- Beltrami, E. *Mathematical Models in the Social and Biological Sciences*. Boston: Jones & Bartlett Publishers, 1993.
- Beltran, M., *Ciencia y sociología*, Madrid, CIS, 1979.
- Berkowitz, S.D. "Afterword: Toward a formal structural sociology." In Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press, 1988, pp. 477-497.
- Berkowitz, S.D. *An introduction to structural analysis: The network approach to social research*. Toronto: Butterworths, 1982.
- Berlinski, David *On System Analysis: An Essay Concerning the Limitations of Some Mathematical Methods in the Social, Political, and Biological Sciences*. Cambridge, MA: MIT Press, 1976.
- Berry, W.D., *Nonrecursive causal models*, Beverly Hills, Sage, 1984.
- Bielby, W.T., Hauser, R.M. "Structural equation models". *Ann. Rev. Sociol.* 3, (1977) 137-61.
- Blalock, H.M., Aganbegian, A., Borodkin, F.M., Boudon, R. y Capecchi, V. (Eds) *Quantitative sociology. International perspectives on mathematical and statistical modelling*. New York: Academic Press, 1975.

- Blalock, H.M.Jr. *Causal models in the social sciences*. Chicago: Aldine, 1971.
- Blossfeld, H. P., Prein, G. (eds.) *Rational Choice Theory and Large – Scale Data Analysis*. Boulder: Westview Press, 1998.
- Blossfeld, H. P., Rohwer, G. *Techniques of Event History Modeling: New Approaches to Causal Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates, 1995.
- Boudon R. *L'analyse mathématique des faits sociaux*. Paris, Plon, 1971.
- Boudon, R. *Les mathématiques en sociologie*. Paris: Presses universitaires de France, 1971.
- Boudon, R. *Mathematical Structures of Social Mobility*. New York: Elsevier, 1973.
- Boudon, Raymond; Grémy, Jean-Paul *Les Mathématiques en sociologie / par Raymond Boudon ; avec la collaboration de Jean-Paul Grémy*. Paris : Presses universitaires de France, 1971.
- Boyd, J.P. “Structural similarity, semigroups and idempotents”. *Social Networks*, 5, (1983) 157-172.
- Bradley, Ian.; Meek, Ronald L. *Matrices and Society: Matrix Algebra and its Applications in the Social Sciences*. Harmondsworth: Penguin, 1986.
- Brams, S. J., Schotter, A. and Schwodiauer, G. (eds.) *Applied Game Theory*. Vienna: Phyca-Verlag, 1979.
- Bråten, S., E. Jahren & A. Jansen, “Social Networks and Multilevel Structure: System Description and Simulations”. In: Cavallo, *Systems Methodology in Social Science Research; Recent Developments*. Boston, Kluwer Nijhoff, 1982, 113-141.
- Broekstra, G., “Probabilistic Constraint Analysis for structure identification: an overview and some social science applications”. In: B.P. Zeigler et al (eds.), *Methodology in Systems Modelling and Simulation*. Amsterdam, North Holland, 1979.
- Bugeda, J. *Curso de Sociología Matemática* Madrid, Instituto de Estudios Políticos, 1976.
- Bugeda, J., *La medida en Ciencias Sociales*, Madrid, CECA, 1974.
- Bunge, M. *Method, Model and Matter. Ch. 5, Concepts of model*. Boston: Reidel, 1973.
- Bunge, M., *Causalidad. El principio de la Ciencia moderna*, Buenos Aires, Ed. Universitaria, 1965.
- Bunge, M., *Las teorías de la causalidad*, Salamanca, Sígueme, 1977.
- Burt, Ronald S. “Models of network structure”. *Annual Review of Sociology*, 6, (1980a) 79-141.
- Carnap, R. ; Otros, *Matemáticas en las ciencias del comportamiento*, Madrid, Alianza, 1962.
- Castells, M. ; Ipola, E., *Metodología y epistemología de las Ciencias sociales*, Madrid, Ayuso, 1975.
- Castillo, J. *Introducción a la Sociología*. Madrid, Guadarrama, 1968.
- Cattell, R.B. *Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientists*. New York: Harper & Brothers, 1952.
- Cerroni, U., *Metodología y Ciencia Social*, Barcelona, Martinez-Roca, 1970.

- Christ, C. F. *Econometric Models and Methods*. New York, John Wiley, 1966.
- Cicourel, A.V. *Cognitive Sociology: Language and Meaning in Social Interaction*. New York: Free Press, 1974.
- Cicourel, Aaron V. *Method and measurement in sociology*. New York: The Free Press, 1964.
- Clamer, M., *Métodos matemáticos en la estadística*, Madrid, Aguilar, 1953.
- Coleman, J.S. "The mathematical study of change", en H.M. Blalock Jr. y A.B. Blalock (eds.) *Methodology in social research*. New York, McGraw-Hill, 1968.
- Coleman, J.S. *The Mathematics of Collective Action*. Chicago: Aldine, 1973.
- Coleman, James S. *Introduction to Mathematical Sociology*. New York, NY: Free Press, 1964a.
- Coleman, James S. *Models of Change and Response Uncertainty*. Englewood Cliffs, NJ: Prentice Hall, 1964b.
- Coombs, D.R., *A theory of data*, New York, Wiley, 1965.
- Cooper, R.A. ; Weekes, A., *Data, models and statistical analysis*, Totowa, Barnes and Noble, 1982.
- Coser, L.A. *Masters of Sociological Thought. Second edition*. New York: Harcourt, Brace, Jovanovich, 1977.
- Coxon, A.P.M. "Mathematical applications in sociology: measurement and relations". *International Journal of Math. Educ. In Sci. and Tech.*, 1, (1970) 159-174.
- Davis, J.A., *The logic of causal order*, Beverly Hills, Sage, 1985.
- Del Campo, S., *La sociología científica moderna*, Madrid, IEP, 1969.
- Dijkstra, W. & J. van der Zouwen, "Testing auxiliary hypotheses behind the interview". *Annals of Systems Research* 6, (1977) 49-63.
- Doreian, P., *Las matemáticas y el estudio de las relaciones sociales*, Barcelona, Vicens-Vives, 1973.
- Doreian, P., *Modeling social process*, New York, Elsevier, 1976.
- Doreian, Patrick "Estimating linear models with spatially distributed data". In Samuel Leinhardt (ed.), *Sociological Methodology*. San Francisco: Jossey-Bass, 1981, pp. 359-388.
- Duncan, O.D., *Introduction to structural equations models*, New York, Academic Press, 1975.
- Durkheim, E., *Las reglas del método sociológico*, Madrid, Akal, 1972.
- Eco, U. y Sebeok, T.A. (eds.) *El signo de los tres*. Dupin, Holmes, Peirce. Barcelona, Lumen. 1989.
- Einstein, Albert *Ideas and Opinions*. New York: Crown, 1954, [1933].
- Elster, J. *Tuercas y tornillos* Barcelona, Gedisa. 1996.
- Emshoff, J. R., *Analysis of Behavioral Systems*. New York, Macmillan, 1971.
- Erbring, Lutz and Young, Alice A. "Individuals and social structure: Contextual effects as endogenous feedback". *Sociological Methods and Research*, 7, (1979) 396-430..
- Everitt, B.S., *An introduction to latent variable models*, London, Chapman and hall, 1984.
- Fararo T.J. "Nature of mathematical sociology". *Soc. Res.*, 36, (1969) 75-92.

- Fararo, T. J. & J. Skvoretz, "Action and Institution, Network and Function: The Cybernetic Concept of Social Structure". *Sociological Forum* 1(2), (1986) 219-250.
- Fararo, T.J. "An introduction to catastrophes". *Behavioral Science*, 23, (1978) 291-317.
- Fararo, T.J. "Biased networks and social structure theorems". *Social Networks*, 3, (1981) 137-159.
- Fararo, T.J. "Biased networks and the strength of weak ties". *Social Networks*, 5, (1983) 1-12.
- Fararo, T.J. "Catastrophe analysis of the Simon-Homans model". *Behavioral Science*, 29, (1984) 212-216.
- Fararo, T.J. and Doreian, P. "Tripartite structural analysis". *Social Networks*, (in press), (1984).
- Fararo, T.J. and Skvoretz, J. "Institutions as production systems". *Journal of Mathematical Sociology*, 10, (1984) 117-182.
- Fararo, T.J. and Skvoretz, J. "The biased net theory of social structures and the problem of integration". In J. Berger and M. Zelditch, Jr. (eds.), *Sociological Theories in Progress. Volume Three*. Pittsburgh, PA: The University of Pittsburgh Press, (forthcoming).
- Fararo, T.J. and Sunshine, M. *A Study of a Biased Friendship Net*. Syracuse, NY: The Syracuse University Youth Development Center and Syracuse University Press, 1964.
- Fararo, T.J. *The Meaning of General Theoretical Sociology: Tradition and Formalization*. New York: Cambridge University Press, 1989.
- Fararo, Thomas J. *Mathematical Sociology: An Introduction to Fundamentals*.
- Fisher F.M. *The Identification Problem in Economics*. New York, McGraw-Hill, 1966.
- Fox, J., *Linear statistical models*, New York, Wiley, 1984.
- Francés, F. "Una aproximación al concepto de capital social a través de la técnica de modelos estructurales", en Alaminos, A.; Francés, F. y Santacreu, O. (eds.) *Reflexiones teóricas y modelos empíricos sobre identidad, diversidad y participación social*. Alicante: OBETS, 2004.
- Francés, F. y Santacreu, O. "Crisis política y juventud en España: el declive del bipartidismo electoral", *Società Mutamento Política*, 5 (10), 2014, pp. 107-128.
- Friedman, James W. *Oligopoly and the Theory of Games*. Amsterdam: North-Holland, 1977.
- Friedman, S. R. "Game theory and labor conflict: Limits of rational choice models". *Sociological Perspectives*, 26, (1983) 375-397.
- Furfey, P.H., *The scope and method of sociology*, New York, Harper, 1953.
- Ganuzo, E. y Francés, F. "Rethinking the Sources of Participation: A Case Study of Spain", *European Sociological Review*, 24(4), 2008, pp. 479-493.
- Ganuzo, E.; Francés, F.; Lafuente, R. y Garrido, F. "Do Participants Change Their Preferences in the Deliberative Process?", *Revista Española de Investigaciones Sociológicas*, 139, 2012, pp. 225-246.
- García Ferrando, M. "La sociología matemática hoy: usos y abusos" *Revista Española de la Opinión Pública*, 45, (1976), pp. 77-90. Voz "Sociología matemática".

- Golderberger, A.S. ; Duncan, O.D., *Structural equations models in the Social Sciences*, New York, Seminar Press, 1974.
- Graybill, F.A., *Introduction to matrices with applications in statistics*, Belmont, Nadsworth, 1979.
- Green, P.E. ; Carroll, J.D., *Mathematical tools for applied multivariate analysis*, New York, Academic Press, 1976.
- Hage, Per and Frank Harary *Structural models in anthropology*. Cambridge: Cambridge University Press, 1983.
- Harary, F., Norman, R.Z., Cartwright, D. *Structural models: an introduction to the theory of directed graphs*. New York: Wiley, 1965.
- Harary, Frank and Helene J. Kommel “Matrix measures for transitivity and balance”. *Journal of Mathematical Sociology*, Vol 6, (1979) 199-210.
- Heise, D.R. *Causal Anaysis*. New York: Wiley, 1975.
- Hellevik, O., *Introduction to causal analysis: exploring survey data by crosstabulation*, London, Allen and Unwin, 1984.
- Heyse, D.R., *Causal analysis*, New York, Wiley, 1975.
- Hintikka, J. *Models for Modalities*. Dordrecht: Reidel, 1969.
- Hirsch, M.W. and Smale, S. *Differential Equations, Dynamical Systems, and Linear Algebra*. New York: Academic Press, 1974.
- Holland, P.W. and Leinhardt, S. “The structural implications of measurement error in sociometry”. *Journal of Mathematical Sociology*, 3, (1973) 1-27.
- Holland, Paul W. and Leinhardt, Samuel “Transitivity in structural models of small groups”. *Comparative Group Studies*, 2, (1971) 107-124.
- Horst, P., *Matrix algebra for social scientists*, New York, Rinehart, 1965.
- Horts, P., *Factor analysis of data matrices*, New York, Rinehart, 1965.
- Ibañez, J., *Del algoritmo al sujeto*, Madrid, Siglo XXI, 1985.
- Jain, V., “Structural Analyses of General Systems Theory”. *Behavioral Science*, Vol. 26, (1981), pp. 51-62.
- James, L.R., *Causal analysis: assumptions, models and data*, Beverly Hills, Sage, 1982.
- Kim, Ki Hang and Roush, Fred W. *Mathematics for Social Scientists*. New York, NY: Elsevier Science, 1980.
- Korm, F., *Conceptos y variables en la investigación social*, Buenos Aires, Nueva Visión, 1969.
- Krishnan, P. Ed. *Mathematical Models of Sociology. Sociol. Rev. Monograph 24*. Staffordshire: England, Keele Univ. Press, 1977.
- Kruskal, Joseph B. and Wish, Myron *Multidimensional Scaling*. Beverly Hills: Sage, 1978.
- Kruskal, William (ed.) *Mathematical sciences and social sciences*. Englewood Cliffs: Prentice Hall, 1970.
- Lamo de Espinosa, E. *Juicios de valor y ciencia social; una crítica interna del avalorismo*, Valencia, Fernando Torres, 1975.
- Lamo de Espinosa, E. *La sociedad reflexiva*, Madrid, CIS, 1990.

- Lave, C.A., March, J.G. *An introduction to Models in the Social Sciences*. New York: Free Press, 1975.
- Lazarsfeld P.F. "A conceptual introduction to latent Structure Analysis", en Lazarsfeld (ed.) *Mathematical thinking in the social sciences*. New York. Free Press, 1954.
- Lazarsfeld P.F. y Henry N.W. *Readings in mathematical social science*. Cambridge, Mass: MIT Press, 1968.
- Lazarsfeld, P.F. ; Pasanella, A.K. ; Rosenberg, M., *Continuities in the language of social research*, New York, Free Press, 1972.
- Lazarsfeld, P.F. ; Reitz, J.G., *An introduction to applied sociology*, New York, Elsevier, 1975.
- Lazarsfeld, P.F. and Rosenberg, M. (eds.) *The Language of Social Research*. New York: The Free Press, 1955.
- Lazarsfeld, P.F., ed. *Mathematical Thinking in the Social Sciences*. New York: Free Press, 1954.
- Lazarsfeld, P.F., Henry, N.W. *Latent Structure Analysis*. Boston: Houghton Mifflin, 1968.
- Lazarsfeld, Paul F. (ed.); Sewell, William H. (ed.); Wilensky, Harold L. (ed.) *The uses of sociology*. New York: Basic Books, 1967.
- Lazarsfeld, Paul F. and Menzel, Herbert "On the relation between individual and collective properties". In Amitai Etzioni and Edward W. Lehman (eds.), *A Sociological reader on Complex Organizations*, Third Edition. New York: Holt, Rinehart and Winston, 1980, pp. 508-521.
- Lehmann, E.L., *Testing statistical hypotheses*, New York, Wiley, 1959.
- Leik, R.K., Meeker, B.F. *Mathematical Sociology*. Englewood Cliffs, NJ: Prentice Hall, 1975.
- Lord, F. M. "An analysis of the Verbal Scholastic Aptitude Test using Bimbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 28, 1968, pp. 989-1020.
- Mac Iver, R.M., *Causación social*, México, F.C.E., 1949.
- Marsden, Peter and Nan Lin (eds.) *Social structure and network analysis*. Beverly Hills: Sage, 1982.
- Marsden, Peter V. "Models and methods for characterizing the structural parametres of groups". *Social Networks*, 3, (1981) 1-27.
- Massarik R. "Magic, models, man and the cultures of mathematics". In F. Massarik and P. Ratoosh (eds.), *Mathematical Explorations in Behavioral Science*. Homewood, Ill.: R.D. Irwin and Dorsey Press, 1965.
- Mills, C.W., *La imaginación sociológica*, México, F.C.E., 1969.
- Mora, M., *Medición y construcción de índices*, Buenos Aires, Nueva Visión, 1971
- Mosteller, F. ; Tukey J.W., *Data analysis and regression*, Reading, Addison-Wesley, 1977.
- Mosterin, J., *Conceptos y teorías en la ciencia*, Madrid, Alianza, 1984.
- Navarro, M. "Apuntes para una teoría de la cultura económica", en VV.AA. *Escritos de Teoría Sociológica en homenaje a Luis Rodríguez Zúñiga*, Madrid, CIS, 1992, pp. 777-798.

- Navarro, M., "Tendencias de desigualdad en el consumo" en J.F. Tezanos y R. Sánchez Morales (ed.), *Tecnología y sociedad en el nuevo siglo*, Madrid, ed. Sistema, 1998, pp.375-399.
- New York, NY: John Wiley, 1973.
- Olinick, Michael *An Introduction to Mathematical Models in the Social and Life Sciences*. Reading, MA: Addison-Wesley, 1978.
- Ostrom, C.W., *Time series analysis. Regression techniques*, Beverly Hills, Sage, 1977.
- Pardiñas, F., *Metodología y técnicas de investigación en las Ciencias sociales*, México, Siglo XXI, 1969.
- Pareto, W. *Manual de economía política*, Genève, Librairie Droz, 1964.
- Peay, E.R. "Hierarchical clique structures". *Sociometry*, 37, (1974) 54-65.
- Peay, Edmund R. "Structural models with qualitative values". *Journal of Mathematical Sociology*, 8, (1982) 161-192.
- Peter M. Bentler, P.M. and Dudgeon, P. COVARIANCE STRUCTURE ANALYSIS: Statistical Practice, Theory, and Directions. *Annual Review of Psychology*. Vol. 47, 1996, pp. 563-592 .
- Powers, C.H. and Hanneman, R.A. "Pareto's theory of social and economic cycles: A formal model and simulation". In R. Collins (ed.), *Sociological Theory*. San Francisco: Jossey-Bass, 1983.
- Rapoport A. "Uses and limitations of mathematical models in Social Science" en Gross, L. (ed.) *Symposium on sociological Theory*. New York: Row-Peterson, 1959.
- Rapoport, A. "Mathematical models of social interaction", in R. Luce, R. Bush and E. Galanter (eds.), *Handbook of mathematical psychology*, Vol. II. New York: Wiley, 1963, pp. 493-579.
- Rapoport, A. *Mathematical Models in the Social and Behavioral Sciences*. New York, NY: John Wiley & Sons, 1983.
- Rodríguez Zúñiga, L. "El desarrollo de la teoría sociológica", en: Salustiano del Campo (ed). *Tratado de Sociología*. Madrid, Taurus, 1984.
- Rodríguez, J.A. *Análisis estructural y de redes*, Madrid, Centro de Investigaciones sociológicas, 1995.
- Sanders, W.B., *The sociologist as detective*, New York, Praeger, 1974.
- Schiffman, Susan S., Reynolds, M. Lance and Young, Forrest W. *Introduction to Multidimensional Scaling*. New York: Academic Press, 1981.
- Shubik, M. *Game Theory in the Social Sciences*. Cambridge: MIT Press, 1982.
- Sierra Bravo, R. *Ciencias sociales. Análisis estadístico y modelos matemáticos*, Madrid, Paraninfo, 1981.
- Sierra Bravo, R. *Ciencias sociales. Epistemología, lógica y metodología*, Madrid, Paraninfo, 1983.
- Simon, H. "The meaning of causal ordering". In R.K. Merton, J.S.Coleman, and P.H. Rossi (eds.), *Qualitative and Quantitative Social Research: Papers in Honor of Paul F. Lazarsfeld*. New York: Free Press, 1979.

- Skvoretz, J. and Fararo, T.J. "Languages and Grammars of action and interaction: A contribution to the formal theory of action". *Behavioral Science*, 25, (1980) 9-22.
- Sorensen, Aa. B. "Mathematical models in sociology". *Annual Review of Sociology*, 4, (1978) 345-371.
- Stinchcombe A.L. *Constructing social theories*. New York. Harcourt, Brace and world, 1968.
- Szenberg M. (ed) *Grandes economistas de Hoy*, Madrid, Debate, 1994.
- Tamayo, M., *El proceso de la investigación científica. Fundamentos de investigación*, México, Limusa, 1981.
- Tezanos, J.F., *La explicación sociológica: Una introducción a la sociología*, Madrid, UNED, 1987.
- Thurstone, L. ; Chave, E.J., *The measurement of attitudes*, Chicago, U. of Chicago Press, 1929.
- Torgerson, W.S., *Theory and methods of scaling*, New York, Wiley, 1958.
- Troitzsch, Klaus "Mathematical modelling and computer simulation of social processes: problems and a new solution". *BMS-Bulletin de Methodologie Sociologique*, nr. 40, (1993) s. 16-42.
- Turner, M. E., "The statistical analysis of mutual causation". In: Krippendorff, *Communication and Control in Society*. New York, Gordon and Breach, 1979, 43-57.
- Von-Wright, G.H., *Explicación y comprensión*, Madrid, Alianza, 1979.
- Wallace, W.L., *La lógica de la ciencia en sociología*, Madrid, Alianza, 1976.
- Wellman, Barry "Structural analysis: From method and metaphor to theory and substance", in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press, 1988, pp. 19-61.
- Wellman, Barry and S.D. Berkowitz "Introduction: Studying social structures", in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press, 1988, pp. 1-14.
- Wellman, Barry and S.D. Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press, 1988.
- Whitehead, A.N. and Russell, B. *Principia Mathematica*. Cambridge: Cambridge University Press, 1964 [1910].
- Wilson, T. "On the role of mathematics in the social sciences". *Journal of Mathematical Sociology*, 10, (1984) 221-239.
- Winton, C.A., *Theory and measurement in sociology*, Cambridge, Schenkman, 1974.
- Woodcock, A. y Davis, M. *Teoría de las catastrofes* Madrid, Catedra, 1986.
- Young, P.V., *Métodos científicos de investigación social*, México, IIS, 1960.
- Zapatero, J. C. "K. Popper y la metodología de las ciencias sociales", en Cuadernos Económicos de ICE, 3-4 (1977), p. 103.
- Zarowich, S.S., *Calidad de los datos estadísticos*, Roma, FAO, 1970.
- Zeisel, H., *Dígalo con números*, México, F.C.E., 1962
- Zetterberg, H., *Teoría y verificación en sociología*, Buenos Aires, Nueva Visión, 1968.





