

CAPÍTULO 9

INDICADORES, VALIDEZ, CONSTRUCCIÓN DE ÍNDICES

9.1. ¿Por qué indicadores?

En el capítulo 7 se dio por sentado que dar una definición operativa¹ sólo presenta problemas prácticos: encontrar la tabla del censo con los datos requeridos, calcular porcentajes, formular el texto de una pregunta y los textos de las respuestas predefinidas que se ofrecen a la elección del entrevistado, etcétera.

Empero no es siempre así. A veces la naturaleza de la propiedad bajo examen no permite imaginar una definición operativa directa;² otras veces ésta se puede imaginar, pero nuestro sentido común, nuestros conocimientos de la situación, nuestra experiencia de investigaciones previas, nos dicen que esa definición operativa sería inadecuada para registrar la propiedad en cuestión.

Imaginemos por ejemplo que nuestra unidad de análisis sea el Estado, nuestro ámbito sea América y la propiedad que interesa sea el nivel de libertad política de que gozan los ciudadanos de cada Estado. Una definición operativa directa sería algo como:

- 1) busque al señor Estados Unidos, a la señora Argentina, al señor Brasil, etcétera, y pregúnteles cuál es el nivel de libertad política de sus ciudadanos;
- 2) pídale contestar eligiendo uno de los siguientes niveles: ninguno / caso / insuficiente / suficiente / bueno / óptimo.

Sin embargo una cosa de este tipo es manifiestamente imposible, porque el señor Estados Unidos, la señora Argentina, el señor Brasil, etcétera, no existen. Se podría imaginar hacer la pregunta al presidente de cada Estado; pero su respuesta no sería confiable, porque incluso el peor dictador declararía que los ciu-

¹ Es decir, establecer el conjunto de acciones, reglas y convenciones que permiten convertir una propiedad de algunos objetos de un cierto tipo en una variable en la matriz de datos.

² Es decir, estrechamente vinculada con la propiedad en cuestión y directamente sugerida por ella, sin posibilidad de dudas.

dadanos de su Estado gozan de una plena libertad política. Se registraría la imagen que los presidentes quieren dar del nivel de libertad política de sus países, que es una cosa distinta del nivel en sí mismo.

Una idea aparentemente mejor sería hacer la pregunta a muestras de ciudadanos de aquellos Estados pidiendo sus opiniones. Por el contrario, esta solución presenta problemas aún más grandes, que se pueden sintetizar con la pregunta: "¿la opinión de quiénes?" A la que seguirían otras preguntas, como: ¿Se saca una muestra de todos los ciudadanos adultos o sólo de los que tienen una idea determinada de lo que se entiende por 'libertad política'? ¿Y cómo juzgar qué ciudadanos son efectivamente competentes? ¿Se saca una muestra de todos los ciudadanos adultos o sólo de los que no tienen ningún miedo a declarar su opinión sobre un problema tan delicado a un ajeno desconocido? ¿Y cómo juzgar qué ciudadanos declaran efectivamente lo que piensan?

Además, considerando que las opiniones varían de un sujeto a otro, ¿cuál se tomará como estado en la propiedad "nivel de libertad política" de cada Estado investigado? Se podría tomar la opinión modal (es decir, compartida por el mayor número de sujetos) o la opinión mediana, si se propuso a los entrevistados una especie de escala ordinal como en la definición anterior.³

Al final, incluso admitiendo poder tomar en cada Estado muestras de ciudadanos efectivamente competentes y sinceros en manifestar sus juicios, lo que se registre será sólo la opinión modal (o mediana) de estos ciudadanos acerca del grado de libertad política del respectivo Estado. No se puede negar que tal opinión esté relacionada con el nivel de libertad política en sí; pero tampoco se puede afirmar que sean la misma cosa. Se busca el nivel de libertad política pero se obtiene —en el mejor caso— la opinión modal de ciudadanos competentes y sinceros. Como todas las opiniones, ésta puede ser equivocada, ilusoria.

Imaginemos ahora que nuestro ámbito sea la Argentina, nuestra unidad de análisis sea el varón adulto, y la propiedad que interesa sea su grado de autoritarismo. En este caso, hay un camino sencillo para una definición operativa directa. Sin esfuerzo, se pueden imaginar preguntas como: "¿Cuál es a su juicio su grado de autoritarismo?" y respuestas como: inexistente / bajo / medio / alto. Pero a veces los caminos sencillos no conducen al lugar debido, y éste sería un caso. En efecto, no se registraría el grado de autoritarismo, sino la propensión de cada varón a declararse autoritario —una propensión rara entre los individuos en general, y quizá rarísima entre los varones efectivamente autoritarios.

Se podrían ofrecer otros ejemplos; pero parece mejor tratar de establecer algunos criterios generales:

- i) No se puede imaginar una definición operativa directa cuando:
 - a) la unidad de análisis es algo que no se puede interrogar (una institución, un agregado territorial, etcétera) y

³ Nótese que la opinión modal y la opinión mediana coinciden necesariamente sólo si la distribución de las respuestas tiene una forma campanular y simétrica.

- b) los estados en la propiedad investigada no se pueden registrar directamente (como por ejemplo la superficie, etcétera) o sacar de documentos oficiales (como se puede hacer para el número de habitantes, el tipo de sistema electoral, el número de votos de un partido, etcétera).
- 2) Una definición operativa directa se puede imaginar, pero no es confiable (en el sentido de que va producir muchos datos no fidedignos) cuando la unidad de análisis es un ser humano y la propiedad investigada:
 - a) tiene respuestas socialmente deseables (porque casi todos van a afirmar que son honestos y generosos y negar que son autoritarios, obsesivos, tediosos, etcétera);
 - b) es algo familiar al científico pero no al lego (no cabe preguntar al hombre de la calle cuál es su grado de anomia, de particularismo, de introversión);
 - c) es algo que la moral dominante en la comunidad del entrevistado considera reprobable (como costumbres sexuales atípicas, consumo de drogas, evasión impositiva, etcétera).

El hecho de que no se pueda dar una definición operativa directa no conlleva una renuncia a recoger informaciones sobre esa propiedad: para hacerlo, hay que buscar una o más propiedades que:

- a) admitan definiciones operativas directas que sean aceptables;
- b) tengan una fuerte relación semántica con la propiedad que interesa.

Estas propiedades se llaman INDICADORES de la propiedad X, y la relación que se establece entre ellos y la propiedad X se llama "relación de indicación". Sobre ella volveremos pronto.⁴ Queremos con anterioridad destacar el hecho de que el recurso a indicadores no es específico de las ciencias sociales. Se usan indicadores también en la vida cotidiana, cada vez que algo manifiesto se considera un indicio, una señal de algo que no es manifiesto pero es interesante. La oración "el chico se puso colorado. Eso es un claro indicador de su culpa" no sólo muestra que el concepto es familiar, sino

⁴ Muchos autores (Merton 1948; Lazarsfeld y Barton 1951; Galtung 1967; Przeworski y Teune 1970; McKennell 1973; Nowak 1976; Carmines y Zeller 1979) escriben que los indicadores son necesarios cuando el concepto es general y/o tiene gran importancia teórica. Pero, como observó Blalock, "algunas variables teóricamente importantes pueden ser fácilmente operativizadas: la edad, el sexo, la raza, la confesión religiosa" (1961: 163). A esta corta lista se pueden agregar la nacionalidad, el lugar de residencia, el título de estudio, el estado civil, la condición laboral, el tipo de ocupación y —con unidades territoriales— la superficie, la población, la forma de régimen político, etcétera.

Los autores mencionados al inicio de esta nota cometieron un clásico error lógico: de la observación indiscutible que los indicadores son conceptos más específicos de los conceptos que indican (el motivo de esto se verá en el apartado 9.3), sacaron la conclusión de que todos los conceptos generales necesitan indicadores. Pero esto es claramente un *non sequitur*, es decir una deducción equivocada, porque la premisa no conlleva la conclusión.

también que el término existe en el lenguaje ordinario. Y en la tecnología el término se usa para designar instrumentos que revelan el estado de un proceso (indicadores de nivel, de presión, etcétera).

En las ciencias sociales el concepto —no el término— de indicador aparece a mediados del siglo XIX en los escritos de algunos positivistas de habla francesa (Villermé 1840; Quetelet 1869). A finales del siglo, Durkheim habla de *indices extérieurs* que simbolizan hechos interiores y fenómenos morales (1893; 1896). El positivista italiano Niceforo, emigrado a París, dedica una obra a los “índices numéricos de la civilización y del progreso” (1921), usando el término en el mismo sentido que se da ahora a la oración ‘indicadores sociales’.

El primero en usar el término ‘indicador’ en el sentido que se le da en la aproximación estándar a las ciencias sociales fue el sociólogo norteamericano Stuart Dodd (1942). Pero ha sido el metodólogo Paul Lazarsfeld quien codificó este uso, distinguiendo entre ‘indicador’ e ‘índice’ (en cuanto combinación de indicadores).⁵

Cabe destacar tres semejanzas entre el concepto ordinario de indicador y el concepto que se usa en la aproximación estándar en las ciencias sociales:

- en ambos casos el indicador es algo manifiesto (o registrable) que da informaciones sobre algo que no es manifiesto (o directamente registrable);
- en ambos casos el fenómeno manifiesto / la propiedad registrable puede interesar en sí o como indicador de algo que no lo es. La asertividad de un joven puede interesar en sí o como indicador de un complejo de inferioridad; la tasa de divorcios puede interesar en sí o como indicador de secularización; el porcentaje de votos a un partido de oposición puede interesar en sí o como indicador de insatisfacción con la política del gobierno;
- en ambos casos el fenómeno manifiesto / la propiedad registrable puede ser considerada —por diferentes observadores e incluso por el mismo— indicador de dos o más fenómenos/propiedades. El hecho de ponerse colorado puede ser considerado un indicador de sensación de culpa o de naturaleza retraída. La tasa de divorcios puede ser considerada un indicador de secularización o de modificación de las estructuras familiares.

9.2 La naturaleza de los indicadores en las ciencias sociales

Las diferencias entre los indicadores que usamos en la vida cotidiana y los que usamos en la aproximación estándar a la investigación social se derivan del hecho de que en esta última el indicador debe su existencia a la necesidad de re-

⁵ A los índices se dedicarán los dos párrafos finales de este capítulo.

gistrar propiedades que no se pueden definir operativamente de forma directa. Recuérdese que la definición operativa es el instrumento que nos permite transformar una propiedad del mundo real en una variable que ocupa un vector-columna de la matriz de los datos (véase apartado 7.1). Por consiguiente, la naturaleza de un indicador en las ciencias sociales está estrechamente vinculada con la de los vectores-columna de la matriz. De eso se sigue que en la aproximación estándar a las ciencias sociales:

- sólo se pueden concebir como indicadores conceptos que se refieren a propiedades;
- estas propiedades tienen que presentar estados en todos⁶ los ejemplares de la unidad de análisis de la investigación, es decir en todos los casos que ocupan las filas de la matriz.

Un ejemplo mostrará mejor lo que se quiere decir. En una clase italiana de metodología, los alumnos tenían que proponer —a título de ejercicio— indicadores de estatus social. La unidad era el individuo adulto, el ámbito era la Italia contemporánea. Fue propuesto el tipo de estación de esquí preferido por el individuo. Pero eso no se puede aceptar como indicador de estatus social porque la mayoría de la población italiana no sabe esquiar, o no tiene medios para frecuentar estaciones de esquí.

Sólo pueden ser elegidos como indicadores (de otro concepto) conceptos que se refieran a propiedades, que sugieran una definición operativa directa y cumplan con el requisito (b) señalado arriba. Naturalmente, éstas son condiciones necesarias pero no suficientes: la condición esencial es que el investigador⁷ perciba una estricta relación semántica entre el indicador y el concepto indicado.

Acerca de la naturaleza de esta relación semántica (relación de indicación) se pueden leer las opiniones más variadas.

Muchos autores⁸ hablan de una relación causal, generalmente sin establecer si el concepto indicado es la causa o el efecto del concepto indicador. Como veremos enseguida, hay casos en que los conceptos que fueron y son elegidos como indicadores se pueden considerar razonablemente causas del concepto que estarían indicando, y casos en que se pueden considerar efectos; en algunos casos los indicadores que fueron y son elegidos no se pueden considerar ni causa ni efecto del concepto indicado. Aun en los casos en los que se puede imaginar qué relación exista, establecer su dirección no es siempre tan obvio como parece.

⁶ Al menos en principio, y con la posible salvedad de unas excepciones marginales.

⁷ Naturalmente, esta percepción es sometida al juicio de la comunidad científica. Si la elección no es obvia —como en efecto lo es en la mayoría de los casos— sería de esperar que el investigador la comente para justificarla.

⁸ Por ejemplo, Stevens (1951), Blalock (1961), Sullivan (1974), Smelser (1976).

Algunos ejemplos: la tasa de contaminación del aire o la tasa de criminalidad son algunas causas de la percepción de la calidad de vida en un municipio, y a menudo se eligen como indicadores de dicho concepto. Al contrario, la tasa de apoyo hacia un gobierno militar puede considerarse un efecto de una actitud autoritaria más general del sujeto, y a menudo se elige como indicador de autoritarismo.

Veamos un caso menos obvio. En una investigación sobre municipios de Italia central afloró una altísima correlación negativa entre el ingreso per cápita del municipio y la proporción de casas poseídas por la familia que las habitaba. El signo de esta correlación parecía ir en contra del sentido común (cuanto más rica es una familia, tanto más debería estar en condiciones de comprar la casa donde habita). Sin embargo, la relación negativa era tan alta y generalizada que indujo a reflexionar. Y así se comprendió que esa relación no era aleatoria, sino el efecto de una verdadera cadena causal: 1) en los municipios de la montaña, o de alguna otra forma aislados, las oportunidades de trabajo son limitadas; 2) por ese motivo, los ingresos medios son bajos y 3) los jóvenes tienden a desplazarse a municipios urbanos y céntricos. 4) Esta migración interna reduce la demanda de alojamiento en los municipios aislados y pobres y la acrecienta en los municipios ricos y céntricos; 5) por eso, los precios de los alojamientos tienden a bajar en los primeros y a subir en los segundos. 6) Por consiguiente, incluso muchos habitantes pobres de municipios aislados tienen bastante plata como para comprar la casa donde viven, mientras los habitantes más ricos de los municipios céntricos no consiguen hacer lo mismo. Al haber detectado esta cadena causal, el investigador puede estar más confiado en elegir la tasa de casas que son de propiedad de los que las habitan sobre el total de las casas como indicador de subdesarrollo económico.

Pero a veces no se logra detectar la relación que liga una propiedad fácil de operativizar con otra no directamente operativizable, y se elige la primera como indicador de la segunda basándose en una simple asociación estadística. El concepto de asociación será ilustrado en el capítulo 15; ahora baste con un ejemplo, extraído de un renombrado test psicológico, el *Minnesota Multiphasic Inventory*. En ese test, muchas de las 480 preguntas a las que los sujetos tienen que contestar parecen bizarras: un ejemplo es: "¿Siente usted un prurito en su dedos cuando abre un picaporte?". Es posible que los psicólogos hayan detectado relaciones estadísticas entre sensaciones tan específicas y extrañas e importantes propiedades psíquicas. Basándose en ellas, han decidido elegir tal u otra sensación como uno de los indicadores de autoritarismo (o de anomia, o de inclinaciones paranoicas, u otro). Naturalmente, éstos son casos límite, que se pueden justificar sólo si necesitan indicadores de un concepto X y no se encuentran conceptos cuya asociación con el concepto X se pueda justificar semánticamente. Como se decía, los planteamientos acerca de la naturaleza de la relación de indicación son variados. Lazarsfeld (1958; 1966) y otros que lo siguen⁹ hablan de una relación probabilística. Pero esto es un burdo error gnoseológico-

⁹ Por ejemplo, Verba (1969); Nowak (1976).

co, ya que una relación probabilística puede darse sólo entre un evento (o clase de eventos) y otro evento (o clase de eventos).

Otros¹⁰ hablan de inferencia; esto también es un error gnoseológico, ya que la inferencia es un posible vínculo entre aseeraciones, nunca entre conceptos.

Guttman (1950) y otros¹¹ hablan de "muestreo de un universo de contenido". Pero esto es otro error gnoseológico, ya que el muestreo supone que la población (y por lo tanto la muestra) esté formada por individuos (u otros objetos físicos) delimitados, no por entidades vagas y no delimitadas como los conceptos.

Menos grave parece el error de los que hablan de una relación lógica o analítica.¹² En efecto, una relación lógica puede establecerse entre conceptos (por ejemplo la relación entre un género y sus especies es de este tipo); pero la relación entre un concepto y sus indicadores no tiene la misma naturaleza que la relación género-especie, que es necesaria (todos los gatos son felinos, pero hay felinos que no son gatos). Cada investigador, o grupo de investigadores, la establece según su conocimiento de la población estudiada y de la literatura sobre el tema, sus experiencias de investigaciones previas, y en definitiva según su preferencia. Otros investigadores —tanto como el mismo investigador en una ocasión sucesiva— pueden elegir otros indicadores para el mismo concepto, y a menudo lo hacen.

Esto no significa que la elección sea arbitraria. Obviamente el investigador, o grupo de investigadores, tiene el máximo interés en elegir indicadores que sean válidos¹³ —es decir que les permitan recoger fielmente al menos algunos aspectos de la intensidad del concepto indicado. Por esto no se atreverá a realizar su elección sin conocer la población que va ser objeto de su investigación y la literatura científica sobre el problema que le interesa.¹⁴ No está nada obligado a repetir las elecciones de los que han estudiado el mismo problema de antemano, y todavía menos si estudia el mismo fenómeno en otro ámbito: se asume que el mismo fenómeno pueda tener significados radicalmente diferentes en ámbitos diferentes.¹⁵ Un ejemplo clásico es la tasa de participación electoral,

¹⁰ Por ejemplo, Galtung (1967); McKennell (1973); Singer (1982).

¹¹ Por ejemplo, Cronbach y Meehl (1955).

¹² Por ejemplo, Nowak (1976); Sullivan y Feldman (1979).

¹³ Al concepto de validez se dedicará el apartado 9.4.

¹⁴ Esto no significa que deba elegir sus indicadores en función de su teoría, como sostienen muchos. Sostener que la teoría que se está controlando empíricamente deba dictar, o inspirar, la elección de los indicadores que sirven para controlarla presenta —además de un difuso aire de circularidad— el obvio riesgo de que se elijan los indicadores que ofrecen la mayor probabilidad de corroborarla.

¹⁵ Esta posibilidad no escapó a muchos autores de comparaciones internacionales (Teune 1968; Verba 1969; Frey 1970; Mokrzycki 1983), que subrayaban cómo indicadores del mismo concepto tienen que ser "equivalentes", no iguales. Sin embargo, esta recomendación no es siempre escuchada, y "muchas instituciones que colaboran con los norteamericanos en investigaciones comparadas se quejan de que los cuestionarios que se les envían desde los Estados Unidos son a menudo acompañados por el pedido de que la traducción sea absolutamente literal" (Brislin, Lonner y Thorndike 1973: 39).

que puede ser considerada un indicador de fortaleza de la vida democrática en los países donde hay elecciones libres, mientras que era un indicador de la capacidad de movilización del partido único en los países del Este europeo; e incluso podría ser considerado un indicador de cumplimiento de las normas en los países donde votar es un deber jurídico, como la Argentina.

Otra característica de una investigación que puede cambiar radicalmente el significado de un fenómeno es el nivel de la unidad de análisis cuando ésta sea territorial. Un ejemplo de este cambio lo proporciona el diferente significado del porcentaje de ancianos en la población. Si la unidad de análisis es el municipio, esto es un indicador de marginalidad y de estancamiento económico, porque —como se decía antes— los jóvenes tienden a trasladarse a lugares donde encuentran más oportunidades de trabajo. Si la unidad de análisis es más grande e incluye muchos municipios, los municipios periféricos y los centrales se funden en un mismo caso, y por lo tanto el porcentaje de ancianos deja de estar vinculado a la diferencia entre municipios centrales y periféricos. A nivel de esta unidad más grande el porcentaje de ancianos puede ser —en naciones ricas— un indicador de calidad del clima y más en general de la vida, ya que en muchos países los jubilados tienden a trasladarse —si pueden— a las mejores áreas desde estos puntos de vista.

Más allá de estas consideraciones, si un investigador se aleja de las elecciones de indicadores hechas por los que han estudiado el mismo fenómeno en el mismo ámbito y con (una) unidad de análisis del mismo nivel, él debería justificar su rechazo de los indicadores tradicionalmente elegidos para el concepto que está estudiando y su elección de indicadores diferentes.

9.3. Aspectos indicativos y aspectos extraños: la doble pluralidad de las relaciones entre conceptos e indicadores

Al final del apartado 9.1 se observó que una propiedad directamente registrable con una definición operativa puede ser considerada —por diferentes observadores e incluso por el mismo— un indicador de dos o más propiedades. Algunos¹⁶ ven este fenómeno con desaprobación, como una señal de la insuficiente madurez de las ciencias sociales; otros, basándose en una reseña de investigaciones publicadas, se limitan a documentarlo.¹⁷ Trataremos de profundizar este punto destacando sus causas.

Cuando dos o más personas conversan entre sí, ellas dan por sentado que cada término tiene el mismo significado para todos. Como lo mostraron los etnometodólogos (Garfinkel 1964; 1967), esta convención es necesaria porque si cada interlocutor pidiera el significado de cada término, se abriría un retorno *ad infinitum* y la vida social se paralizaría. Pero si no se tienen

¹⁶ Por ejemplo, Zetterberg (1954), que lamentaba la extrema variedad de los indicadores elegidos en la investigación sobre grupos.

¹⁷ Por ejemplo, el psicólogo Mc Nemar (1946).

finés prácticos sino cognoscitivos, y se pide a los interlocutores que definan un término cualquiera, casi siempre se encuentra que las definiciones son sensiblemente diferentes —lo que significa que el mismo término designa conceptos distintos para cada sujeto. Por ejemplo, al escuchar el término 'autoritarismo' algunos pueden pensar exclusivamente (o sobre todo) en fenómenos políticos, otros preferentemente en la actitud de los varones con respecto a sus esposas, otros preferentemente en la actitud de los gerentes de una empresa u oficina con respecto a los empleados. Y profundizando más, se encuentra que para algunos el autoritarismo político significa sobre todo el apoyo a un gobierno militar, para otros el deseo de que haya un solo partido, o sólo dos, para otros la opinión de que la toma de todas las decisiones políticas pertenezca al ejecutivo central en desmedro de las autonomías locales, etcétera.

No se pretende decir que cada sujeto da un único y diferente significado a cada término; al contrario, que cada sujeto puede darle muchos significados (algunos centrales, otros marginales, otros ausentes) y la distribución de aspectos centrales, marginales y ausentes varía de sujeto a sujeto. En filosofía del conocimiento, el conjunto de los aspectos centrales y marginales se llama INTENSIÓN de un concepto; toda la argumentación previa se puede sintetizar diciendo que cada término designa —o, al menos, puede designar— conceptos con intensiones diferentes para cada sujeto que lo pronuncia o lo escucha.¹⁸

Al esclarecer esto, no debería extrañar el hecho de que el mismo concepto A, directamente operativizable, sea elegido por diferentes autores (e incluso por el mismo autor en situaciones diferentes) como indicador de dos o más conceptos no operativizables distintos. Esto deriva naturalmente del hecho de que la intensión de este concepto A varía de sujeto a sujeto y también varía en el tiempo para el mismo sujeto.

Tomemos como ejemplo la tasa de abortos registrados en cada provincia de una nación de tradición católica donde el aborto es legal. Habitualmente esto se considera un indicador de secularización, es decir, de autonomía con respecto a los preceptos de la Iglesia. En países como la Argentina, donde el aborto no está legalizado, los abortos registrados sólo son un subregistro del total de abortos practicados. La tasa de abortos registrados podría ser un indicador de las malas condiciones sanitarias en que se practican los abortos clandestinos en sectores populares, ya que sólo se registran aquellos casos que derivan en la hospitalización de la mujer en el sistema público. Y en una nación —como por ejemplo Italia, que ha legalizado el aborto pero reconoce a los médicos de los hospitales públicos el derecho a la objeción de conciencia antes de practicarlos—, un investigador podría

¹⁸ En rigor, se podría decir que cada intensión diferente corresponde a un concepto diferente, lo que conllevaría que, mientras que se puede fácilmente averiguar que dos hablantes usan el mismo término, no se puede averiguar —por falta de acceso directo a sus mentes— si están pensando en el mismo concepto.

elegir la tasa de abortos registrados como un indicador de la propensión a la objeción de conciencia de la clase médica de cada provincia, ya que las mujeres, frente a un rechazo (o probable rechazo) de practicarles un aborto, se desplazarían a otras provincias donde la clase médica no tiene problemas de conciencia frente al aborto.

Cada investigador puede considerar central un aspecto diferente de la intensidad del concepto de tasa de aborto, y gracias a este aspecto —que podríamos llamar ASPECTO INDICATIVO— elige la tasa de abortos registrados como indicador de diferentes conceptos. En cuanto a los otros aspectos, que no le sirven para establecer la relación de indicación —y que por lo tanto podríamos llamar ASPECTOS EXTRAÑOS—, o los ignora, o los considera no importantes, o se da cuenta de su presencia e importancia pero necesita aquel indicador y espera que los aspectos extraños no tergiversen demasiado la relación de indicación que quiere establecer. Sea cual fuere el reconocimiento que un investigador tiene de los aspectos extraños del indicador que elige, esos aspectos permanecen y contribuyen a influir, por ejemplo, la manera en que una pregunta dada será entendida por los entrevistados. Por ese motivo, la minimización de los aspectos extraños tiene que ser sin duda el primer criterio a considerar en la elección de indicadores. Y dado que no parece posible encontrar un indicador totalmente exento de aspectos extraños, un investigador debería tenerlos muy en cuenta al interpretar sus datos.

A esta pluralidad “desde abajo” en las relaciones de indicación (en el sentido de que un concepto puede ser razonablemente elegido como indicador de otros varios conceptos), le corresponde una pluralidad “desde arriba”, en el sentido de que para cada concepto que interesa, pero que no sugiere directamente una definición operativa, cabe elegir más de un indicador.

Veamos el porqué de esto partiendo de un ejemplo. Al inicio de este capítulo se dijo que la libertad política no se puede definir operativamente de manera directa (también en el sentido más específico de “nivel de libertad política que gozan los ciudadanos de varios Estados”). Para permitir un registro de situaciones empíricas cabe descender hacia conceptos más específicos, que son aspectos de la intensidad del concepto de libertad política. Los filósofos políticos convienen sobre al menos tres aspectos de esta intensidad: libertad de opinión, de asociación, de prensa. Cada uno de estos aspectos —como otros,¹⁹ cuya pertenencia a la intensidad no goza sin embargo de un acuerdo unánime— indica varias direcciones en las que podemos buscar un indicador, es decir, un concepto que se pueda definir operativamente de manera directa y que esté semánticamente vinculado a la libertad política.

El concepto de libertad de opinión sugiere, por ejemplo, el total de días de cárcel sufridos por la expresión de las propias opiniones durante un año

dado, dividido por el número de adultos en la población de cada Estado.²⁰ La definición operativa indicaría dónde proporcionarse la documentación necesaria, cómo efectuar controles consultando agencias internacionales, etcétera.

El concepto de libertad de asociación sugiere, por ejemplo, el número de asociaciones que han sido disueltas —o cuya constitución ha sido prohibida— por motivos políticos (dividido por el número de asociaciones admitidas y no molestadas) durante un período dado.

El concepto de libertad de prensa sugiere, por ejemplo, el total de ejemplares de diarios y periódicos secuestrados por motivos políticos en un año dado dividido por el total de ejemplares circulantes en ese año.

El concepto de libertad de ir al extranjero sugiere, por ejemplo, el número de pasaportes solicitados y no otorgados dividido por el total de pasaportes solicitados durante un año dado.

La definición operativa de los últimos tres indicadores sería análoga a la de arriba (días de cárcel), cambiando obviamente las fuentes de información y la manera de controlarlas.

Como se vio, para posibilitar cualquier registro de informaciones empíricas semánticamente relevantes para el concepto de libertad política cabe recurrir a conceptos mucho más específicos y acotados, que conduzcan directamente a estas informaciones. Por esta razón se habló de un pasaje “desde arriba” del concepto a sus indicadores, y de un pasaje “desde abajo” de cada indicador elegido al concepto relativo. Y ya que los aspectos de la intensidad de un concepto general pueden ser muchos, para no reducir demasiado la generalidad de este concepto, y a final de cuentas cambiarlo, es necesario buscar indicadores para muchos aspectos; lo ideal sería encontrar indicadores para cada aspecto relevante de su intensidad.²¹

Esto es lo que se entiende cuando se habla de “pluralidad desde arriba” de la relación de indicación. De una forma rápida, la doble pluralidad puede sintetizarse así: cada concepto que no sugiere directamente una definición operati-

²⁰ Si no se hiciera esta división, los Estados más poblados tendrían una obvia desventaja. Esta operación, que apunta a colocar un fenómeno en sus dimensiones exactas, es denominada ‘normalización’ y está ilustrada en detalle más adelante.

²¹ Esto es solamente un ideal, porque los aspectos relevantes de la intensidad de un concepto no son fácilmente enumerables, y además cualquier intento por establecerlos va a ser inevitablemente controvertido. En el apartado 9.5 se aclararán otras dificultades que se encuentran al buscar indicadores para cada aspecto relevante de la intensidad de un concepto. Por otro lado, si se dejan sin indicar muchos aspectos importantes de la intensidad de un concepto, sería necesario cambiar su denominación. Lo que raramente se hace, como lamentan McNemar (1946), Coombs (1953), Cook y Selltiz (1964), Lutynski (1978). Una consecuencia del no hacerlo la señalan Cronbach y Meehl: “Si el investigador A, para un concepto que denomina ‘agresividad’, elige indicadores asociados con ataques abiertos a los demás y el investigador B, por otro concepto que también denomina ‘agresividad’, elige indicadores asociados con la hostilidad reprimida, los mismos resultados que corroboran la teoría del investigador A falsean la teoría del investigador B, y viceversa” (1955, 291).

¹⁹ Como por ejemplo la libertad religiosa, o la libertad de ir al extranjero sin restricciones políticas.

va necesita una pluralidad de indicadores,²² y cada concepto que puede ser directamente operativizado puede ser elegido como indicador de una pluralidad de otros conceptos.

Una combinación particular de pluralidad desde arriba y pluralidad desde abajo se encuentra a menudo en el ANÁLISIS SECUNDARIO. Con esta expresión se designa el hecho de que muchos investigadores, para profundizar un problema cognoscitivo que les interesa, en lugar de recolectar directamente datos analizan aquellos ya registrados por otros investigadores y depositados en archivos apropiados.²³

Naturalmente, cada vez que se analizan los datos de una investigación ajena, estamos limitados a las propiedades que interesaron al autor originario, o autores, de aquella investigación. Por lo tanto, puede pasar que algunas propiedades que interesan no se encuentran en la matriz de datos.²⁴ Antes de renunciar a considerarlas en sus modelos,²⁵ muchos investigadores examinan atentamente toda la matriz de la investigación originaria para ver si encuentran una o más variables cuyo concepto correspondiente pueda ser considerado un indicador de una u otra de las variables que les interesan. De esta forma, datos que fueron recogidos para tener información acerca de una propiedad C, interesante en sí o como indicador de una propiedad A, son explotados para tener informaciones acerca de una propiedad B, ésa tampoco directamente operativizable como la propiedad A.

Esta operación — como pasa con todas las relaciones de indicación que un investigador establece — está, al menos en teoría, sometida al juicio de la comunidad científica en cuanto a su validez — un concepto que examinaremos pronto.

9.4. La validez como juicio acerca de la proximidad semántica entre un concepto y su indicador

En rigor, la VALIDEZ es una propiedad del concepto I en cuanto posible indicador del concepto C en un ámbito espacio-temporal determinado con una unidad de análisis dada. El grado en que el concepto I posee esta propiedad no se puede medir,²⁶ ni averiguar de ningún modo "objetivo". Este grado sólo puede ser evaluado con criterios semánticos por el investigador, el equipo de investigadores, la comunidad de investigadores de una disciplina.

²² La necesidad de elegir más de un indicador para cada concepto ha sido destacada también por Lazarsfeld: "Cada indicador tiene un carácter específico y no puede ser considerado exhaustivo del significado de otro concepto" (1958, 107).

²³ Estos archivos de datos se difundieron en las últimas décadas en sociología y ciencia política, según el ejemplo de los archivos de relatos y otras informaciones etnográficas que se remontan a fines del siglo XIX, en la época del evolucionismo.

²⁴ Sobre esto y otros problemas del análisis secundario, véase Hyman (1972).

²⁵ El modelo es la herramienta básica del análisis de datos en la aproximación estándar. Véase capítulo 15.

²⁶ Contrariamente a lo que sostienen los metodólogos de inspiración conductista, que propusieron varios procedimientos para "medir" la validez (véase más adelante en este mismo apartado).

Este juicio se da — o se debería dar — teniendo en cuenta los conocimientos del problema y de la población estudiada, los resultados de investigaciones previas, la literatura científica sobre el tema. Una vez registrados los datos, el juicio debe tener en cuenta la fuerza y la forma de la relación (no sólo los coeficientes de asociación)²⁷ de aquella variable con otras.

En efecto, los metodólogos de inspiración conductista tienden a desacreditar un juicio basado en criterios meramente semánticos (llamado *content validation*, validación de contenido, o — sarcásticamente — *face validation*, validación a simple vista). Ellos juzgan también poco científica una forma de control de validez que usaban los psicólogos de la primera mitad del siglo XX: la validación "por grupos conocidos" (*known groups*). Para ver si un test dado era un indicador válido, por ejemplo, de anomia, se lo aplicaba a una población conocida (habitualmente, los alumnos de una clase de psicología). El test era juzgado válido si el grupo de alumnos que el profesor juzgaba "anómicos" obtenía en el test un puntaje medio significativamente más alto que el resto de la clase (Thurstone y Chave 1929; Frey 1970).

Todas las otras formas de control de validez se basan en coeficientes de asociación entre el vector que brinda los datos del indicador en cuestión y otros vectores de la matriz. Se vio en el apartado 7.4 que también la fiabilidad de los datos se controla — para la mayoría de los investigadores, de inspiración conductista o no — mediante correlaciones de dos a más vectores. Esta analogía ha conllevado una lamentable confusión entre el concepto de fiabilidad (la correspondencia entre situaciones reales y datos en la matriz) y el concepto de validez (el grado de correspondencia semántica entre dos conceptos): algunos usan sin cuestionamientos un término en lugar del otro y confunden estos dos problemas metodológicos distintos.²⁸ Otros llegan a calcular relaciones matemáticas entre los coeficientes correspondientes y elaboran técnicas para "enfrentar simultáneamente cuestiones de fiabilidad y de validez" (Siegel y Hodge 1968).²⁹

Se puede comentar que, en el fondo, ellos son los más coherentes: su idea de validez no es más que la consecuencia de un enfoque que reduce todos los problemas científicos a relaciones matemáticas entre vectores. Una reducción que tiene una justificación oficial — la ciencia tiene que ser "objetiva", y por cierto no hay nada más "objetivo" que la matemática — y una oculta: el deseo de quitarse de encima tareas tan minuciosas y tediosas como controlar el funcionamiento de las herramientas en el campo, ensuciándose las manos con una realidad siempre mucho más compleja que los modelos matemáticos,³⁰ y también la tarea intelectualmente exigente de reflexionar sobre las relaciones se-

²⁷ Sobre los coeficientes de asociación, véase el capítulo 15.

²⁸ Véanse, por ejemplo, Gulliksen (1936); Parry y Crossly (1950); Schuman (1966); Davies (1977); Carmines y Zeller (1979).

²⁹ Véanse también Cronbach (1949); Bohrnstedt (1970); Allen (1974). Para Lord y Novick (1968) la relación entre validez y fiabilidad es simplísima: la primera es la raíz cuadrada de la segunda.

³⁰ Sobre la infinita, y continuamente cambiante, complejidad de la realidad ha tenazmente insistido Max Weber (1904; 1906; 1913). Su lección debería ser meditada por los cultores de modelos matemáticos hipersimplificados.

mánticas que tienen dos conceptos en la mente de los individuos investigados en un ámbito espacio-temporal dado.

Veamos ahora las más difundidas de estas técnicas de validación. En la validación "concomitante" o "simultánea" (*concurrent validation*) se calculan coeficientes de asociación con otros supuestos indicadores³¹ del mismo concepto (Carmines y Zeller 1979), en la razonable expectativa de que dos o más indicadores del mismo concepto estén positivamente asociados entre sí. Pero incluso una fuerte asociación positiva no demuestra nada de forma conclusiva,³² porque nada impide que la asociación derive del hecho de que son todos indicadores de otro concepto. Además, si el indicador en cuestión está asociado con algunos y no con otros supuestos indicadores del mismo concepto, no podemos establecer cuál de los dos grupos está formado por indicadores válidos del concepto que nos interesa.

Relativamente más concluyente es la validación "predictiva", en la que se controla si, sobre la base de los puntajes en un indicador, somos capaces de prever correctamente los resultados de una prueba sucesiva (Scott 1968; Carmines y Zeller 1979). Por ejemplo, el vector que brinda los datos del test que trata de detectar la aptitud para conducir un helicóptero, dirigir una orquesta o bailar ritmos afrocubanos se asocia con el vector que brinda los resultados de la prueba final de un curso para conducir helicópteros, dirigir orquestas o bailar esos ritmos. Si la asociación es alta, el test se considera válido como indicador de las habilidades correspondientes, y puede ser adoptado por los directores de los cursos para seleccionar previamente los candidatos. Naturalmente, esta forma de validación sólo puede ser adoptada en situaciones como las descritas, es decir cuando el concepto a indicar puede tener una definición operativa directa bajo la forma de resultados de la prueba final de un curso.

La más interesante de estas validaciones mediante análisis estadístico de la matriz es la "validación de constructo" (*construct validation*), en la que el supuesto indicador no se asocia con indicadores del mismo concepto sino con variables que —según las expectativas teóricas o de sentido común— deberían tener una relación empírica fuerte (positiva o negativa) con el concepto a indicar (Cronbach y Meehl 1955). El indicador se juzga válido si las relaciones son parecidas a aquellas que se esperaban.

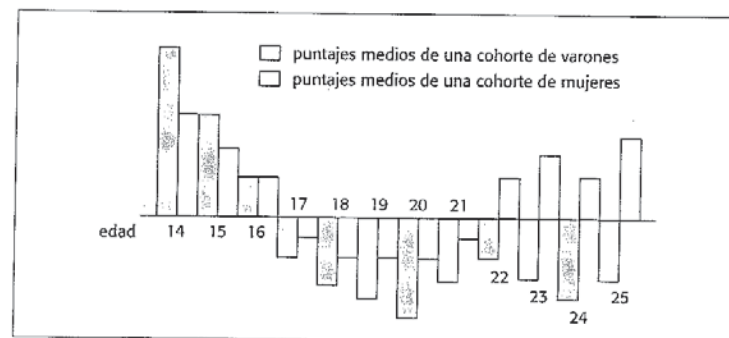
Un ejemplo aclarará lo que se entiende por validación de constructo. En un análisis secundario de una investigación de 7.000 jóvenes italianos en 1971, uno de los autores de este libro efectuó un análisis factorial de una larga batería de escalas de Likert. La primera dimensión extraída no era tan

³¹ Como lo observan Frey (1970) y Turner (1979), esto abre el problema de quién validó estos supuestos indicadores. Con otras palabras, la "validación concomitante", y en definitiva todas las formas de validación mediante asociación entre variables en la matriz, abren un regreso *ad infinitum*, al final del cual se encuentra —a pesar de la pretendida "objetividad"— la tan despreciada "validación de contenido": la variable X se considera un indicador válido del concepto C porque está estrechamente asociada con la variable Y; ¿pero quién validó la variable Y en cuanto indicador de C?

³² Con una feliz expresión debida a Winch y Campbell (1969), una fuerte asociación positiva es *evidence, not proof* (indicio empírico, no prueba).

fácil de interpretar, y fue bautizada "perbenismo" en cuanto distinguía a los jóvenes que aprobaban frases que parecían las recomendaciones de la mamá (juntarse con buenos compañeros, escuchar los consejos de los expertos, evitar posiciones extremas, confiar en los sabios, etcétera) de los jóvenes que las rechazaban.³³ La gran amplitud de la muestra permitió calcular el puntaje medio de cada cohorte de un año en el índice³⁴ que representaba este factor, dividiendo además entre mujeres y varones. Esos puntajes medios se representan en la figura 9.1. La línea horizontal representa el puntaje o (en el análisis factorial, esto es el puntaje medio de la muestra total). Si la columna de una cohorte está arriba de la línea horizontal, el puntaje medio en el índice de aquella cohorte es más alto que el puntaje medio de la muestra; si la columna está debajo de esa línea, el puntaje medio de la cohorte es inferior al promedio de la muestra. Las columnas tienen altura proporcional a la diferencia entre el promedio general y el promedio de aquella cohorte.

Figura 9.1: Un caso de validación de constructo



La validación de constructo que aflora claramente de la figura radica en el hecho de que el rumbo de las columnas es el que se esperaría si el índice registrase el verdadero nivel de aceptación que cada cohorte de jóvenes manifiesta para con las recomendaciones de la mamá. Es sabido en efecto que en los países latinos las mujeres se emancipan temprano de la influencia materna, mientras que los varones tardan más en hacerse independientes. El gráfico muestra que este desfase se reduce rápidamente, y a los 16 años los varones han alcanzado el mismo nivel de (no)aceptación de las mujeres. A continuación, los niveles de aceptación siguen bajando, pero el de los varones baja más rápidamente: un análisis de regresión múltiple del in-

³³ El término 'perbenismo' proviene de la expresión *per bene* (como se debe, adecuado a las normas culturales de la "buena burguesía").

³⁴ En las ciencias sociales se denomina 'cohorte' a un grupo de individuos que tienen la misma edad. No sólo los indicadores, sino también los índices —que son combinaciones de indicadores, y de los que se hablará pronto— son sometidos a un juicio de validez.

dice sobre otras variables mostró que esto depende de la mayor exposición que los jóvenes italianos tenían en aquellos años a influencias exteriores a la familia, como la escuela (que significa profesores, lecturas y —más que otra cosa— compañeros a imitar).³⁵

El nivel de aceptación de los varones alcanza su mínimo a los 20 años, y luego sigue siendo muy inferior al promedio general. Al contrario, el puntaje de las mujeres, tras detenerse por debajo del promedio a los 18-20 años, sube nueva y rápidamente por encima del promedio de la muestra, hasta alcanzar a los 25 años un nivel parecido al que tenía a los 14 años. Este rumbo se explica fácilmente si se piensa en el hecho de que las jóvenes italianas de 1971 tendían a casarse mucho más temprano que las de hoy en día, y por lo tanto a los 20 años empezaban a posicionarse ante las frases que hemos mencionado (juntarse con buenos compañeros, escuchar los consejos de los expertos, evitar posiciones extremas, confiar en los sabios, etcétera) no tanto como hijas sino como madres.

Cabe comentar que esta convincente valoración de constructo sólo se pudo lograr analizando en detalle y por separado el puntaje de cada cohorte, cotejando el rumbo de las cohortes femeninas y el de las masculinas. Ésta no se alcanzaría si se correlacionase el índice en cuestión, globalmente considerado, con cualquier otra variable globalmente considerada — como prescriben las técnicas estándar de valoración. El juicio sobre la validez de un indicador o de un índice, aun cuando se base en el análisis de sus relaciones con otras variables en una matriz de datos, a menudo tiene que ir más allá de lo que pueden brindar simples coeficientes globales de asociación.³⁶

9.5. Algunos ejemplos de elección de indicadores

Algunos ejemplos ayudarán a comprender mejor la naturaleza de los indicadores y podrán servir como modelos para la elección de nuevos.

Empezaremos con un concepto que se quiere operativizar en prácticamente todas las encuestas sociológicas sobre seres humanos adultos —el de estatus socioeconómico— y que sin embargo no sugiere una aceptable definición operativa directa (no se puede preguntar directamente a una persona de la calle cuál es su estatus socioeconómico porque muchos no entenderían la pregunta, y otros le darían significados muy diferentes de los que se le dan en sociología).

Como ya se dijo, en teoría se deberían buscar indicadores para cada aspecto relevante de la intensión del concepto de estatus socioeconómico. Por otro la-

³⁵ Hasta los años 1970 en Italia muchas familias retiraban a las niñas de la escuela a los 14-15 años, y a muchas no se les permitía salir de casa sin parientes o hermanos hasta los 18-20 años. El análisis de regresión múltiple es un desarrollo del análisis de regresión (del que se hablará en el apartado 15.6). Esa técnica permite valorar la influencia relativa de algunas variables en otra, llamada independiente. Todas las técnicas de regresión y correlación sólo se aplican a variables cardinales o cuasi cardinales.

³⁶ Sobre los coeficientes de asociación, véase el capítulo 15.

do, esta búsqueda encuentra obvios límites en el hecho de que no es siempre fácil encontrar conceptos que sugieran definiciones operativas directas —los únicos idóneos para ser elegidos como indicadores. Además, cuanto más grande es el número de indicadores, más difícil es la construcción de un índice que sintetice —como veremos en el apartado siguiente.

Por estos motivos no se pretende ofrecer una lista exhaustiva de indicadores, pero igualmente se sugiere más de un indicador para el mismo aspecto, o para aspectos cercanos, de forma tal que el lector que quiera inspirarse en la lista tenga un abanico dentro del cual elegir a su vez.

Aspectos relativos a la educación:

- título de estudio del sujeto;
- título de estudio del padre del sujeto;
- título de estudio de la madre del sujeto.³⁷

Aspectos relativos a la ocupación:

- ocupación principal del sujeto empleado o última ocupación del sujeto desempleado o jubilado;
- ocupación principal del jefe o jefa de la familia de procedencia.

Aspectos económicos:

- ingreso anual medio del sujeto en los últimos 5 años;
- ingreso anual medio de su familia en los últimos 5 años / número de individuos convivientes;
- metros cuadrados de espacio habitable a disposición de la familia / número de individuos convivientes;
- tipo al cual pertenece la vivienda principal de la familia.

Bienes duraderos poseídos por la familia:

- número de coches poseídos por la familia / número de adultos convivientes;
- número de equipos de TV poseídos por la familia / número de individuos convivientes;
- costo del coche más caro poseído por la familia.

Costumbres del tiempo libre:

- gasto mensual medio (en el último año) para comidas de la familia en restaurantes / número de individuos convivientes;
- gasto medio anual en vacaciones de miembros de la familia / número de individuos convivientes.

³⁷ El título del padre y el de la madre se pueden combinar en una sola variable, como veremos en el apartado siguiente.

El lector habrá notado que a menudo el indicador incluye la división por el número de individuos o de adultos convivientes en la familia del sujeto. Sólo así el indicador cobra su pleno sentido: se entiende en efecto que poseer tres coches o tres equipos de televisión cambia de significado si el poseedor es solo o si es parte de una familia de ocho miembros. El número de miembros tiende naturalmente a influir los niveles de muchos indicadores cuantitativos, pero no tiene ninguna relación con el concepto de estatus: es un aspecto ajeno a la relación de indicación, y dividiendo por el número de miembros se neutraliza su influencia.

Esta operación de dividir una cifra por una base relevante para neutralizar aspectos de un fenómeno que no interesan se llama NORMALIZACIÓN y se aplica habitualmente no sólo en la creación de indicadores, sino también en el análisis estadístico de los datos.

Pasemos ahora a un concepto que se usa en muchas investigaciones cuya unidad de análisis es la provincia: el nivel de dinamismo económico-demográfico.

Para los *aspectos económicos* en general se podrían elegir indicadores como:

- la variación³⁸ en el ingreso per cápita;
- la variación en los depósitos bancarios per cápita;
- la variación en el gasto en obras públicas.

Para los *aspectos relativos a empresas* se podrían elegir:

- el saldo demográfico de las empresas (total de nuevas empresas menos total de empresas quebradas) en el intervalo considerado;
- la variación del monto anual de inversiones per cápita;
- la variación del consumo anual per cápita de kW/h para uso industrial;
- la variación en la proporción de suelo ocupado por establecimientos industriales y comerciales.

Para los *aspectos relativos a la ocupación* se podrían elegir:

- la variación en la tasa de empleados sobre la población en edad laborable;
- la variación en la cuota de empleados procedentes de otras provincias o Estados.

Para los *aspectos demográficos* se podrían elegir:

- el saldo demográfico natural (total de nacidos menos total de fallecidos) en el intervalo considerado;
- el saldo demográfico debido a transferencias de otras provincias o Estados menos transferencias a otras provincias o Estados.

Considerando que el desarrollo económico se ve favorecido por la difusión de la instrucción, se podría agregar la variación en la cuota de residentes graduados.

Finalmente consideremos un concepto que se puede encontrar en investigaciones con unidades ya sea individuales ya sea territoriales: el grado de emancipación femenina.

Cuando la unidad es territorial (una provincia, un Estado) un abanico de indicadores podría incluir:

- el cociente entre la tasa de ocupación femenina y la tasa de ocupación masculina;
- el total de mujeres con cargos políticos dividido por el total de hombres con cargos semejantes;
- el total de conductoras de autobús dividido por el total de conductores;
- el total de niños menores de cinco años dividido por el total de mujeres de 20-50 años;
- la edad media en que las mujeres dejan la familia de origen para ir a vivir solas;³⁹

[sólo si la unidad es un municipio urbano]

- el gasto total (en el último año) en platos preparados dividido por el gasto total (en el mismo período) en productos alimenticios no preparados en los supermercados del municipio.

Cuando la unidad es individual, el concepto apropiado es el grado de apoyo a la emancipación femenina. Un abanico de indicadores de éste podría incluir:

- la actitud hacia la asignación de cargos políticos a mujeres;
- la actitud hacia la asignación de cargos religiosos a mujeres;
- la actitud hacia el aborto;
- la actitud hacia los anticonceptivos;
- la actitud hacia las mujeres empresarias;
- el juicio acerca de la conveniencia de que las mujeres trabajen fuera de su casa;
- el juicio acerca de la conveniencia de que las mujeres trabajen de noche.

9.6. La construcción de índices tipológicos con variables categoriales y ordinales

Se dijo varias veces que la intensión de un concepto de propiedad tiene varios aspectos, y que —si no es posible definir operativamente el concepto de forma directa— se necesita encontrar indicadores para los aspectos más importantes, so pena de recoger informaciones sobre un concepto muy diferente del

³⁸ Todas las variaciones se entienden de un año X a un año Y, definidos por el investigador según sus intereses.

³⁹ Los dos últimos indicadores están invertidos, en el sentido de que cuanto más alto es el número de niños menores de cinco años, y más alta la edad en la que las mujeres dejan la familia de origen, más bajo es el nivel de emancipación femenina.

que pensamos. Naturalmente, las informaciones recogidas tienen que ser sintetizadas para reconstituir de alguna manera la unidad del concepto que interesa.

La operación con la que se realiza esta síntesis se llama CONSTRUCCIÓN DE UN ÍNDICE y, tal como se adelantó, es la tarea más delicada de toda la investigación social que sigue una aproximación estándar.

El criterio de construcción de los índices se puede establecer en abstracto, a la hora de diseñar la investigación; pero se realiza cuando los datos sobre los indicadores estén recolectados, y tengamos las variables correspondientes en la matriz. El procedimiento que se sigue es totalmente diferente según se traten las variables como cardinales o no, a punto tal que es preciso presentarlos en párrafos diferentes.⁴⁰

Si no las consideramos cardinales, se atribuye plena autonomía semántica a cada categoría de las variables que van a formar el índice. De esto se sigue que la herramienta para realizar la síntesis es la tabla bivariable,⁴¹ en la que cada categoría de las variables involucradas se representa en una fila o una columna de la tabla, y en cada celda hay un tipo —es decir la combinación de una categoría de la variable en las filas de la tabla con una categoría de la variable en las columnas. El índice así construido se denomina TIPOLÓGICO.

Empecemos con un ejemplo simple, en el que las dos variables involucradas son ordinales (lo que nos ayudará a realizar la síntesis) y tienen pocas categorías (lo que reducirá la complejidad del problema). Imaginemos que la propiedad que nos interesa sea el "capital cultural" (en el sentido de Bourdieu 1979; 1994) que recibió cada individuo durante su formación, y que dos de los indicadores elegidos sean —muy simplemente— el título de estudio del padre y el de la madre, cada uno registrado con cuatro categorías:⁴² elemental (*el* en algunas tablas que siguen) / medio (*me*) / medio-superior (*ms*) / superior (*sup*).

Tabla 9.1: Combinación de dos indicadores y reducción de la tipología

TÍTULO DE LA MADRE	TÍTULO DEL PADRE			
	ELEMENTAL	MEDIO	MEDIO SUPERIOR	SUPERIOR
elemental	1	2	3	4
medio	2	2	3	4
medio superior	3	3	3	4
superior	4	4	4	5

⁴⁰ Hay naturalmente casos mixtos, en los que algunas variables son cardinales o cuasi cardinales y otras no. Un ejemplo se presentará en el apartado siguiente.

⁴¹ Naturalmente hay índices formados con más de dos variables categoriales. Pero también en estos casos, ya que una tabla trivariable no es fácil de concebir ni de representar, en la práctica se opera con una sucesión de tablas bivariables, construyendo índices parciales (véase un ejemplo más adelante).

⁴² Se usan categorías generales, sin hacer referencia al ordenamiento escolástico de un país particular. El ajuste de estas categorías generales a la estratificación de títulos de cada país es un típico problema de definición operativa.

La tabla 9.1 presenta los dieciséis tipos resultantes de la combinación (producto lógico) de cada título del padre con cada título de la madre. Estos tipos podrían ser las dieciséis modalidades de la nueva variable. Esta variable deberá ser cruzada con otros indicadores de capital cultural, y por lo tanto se necesitará una secuencia de otras tablas. Dieciséis modalidades en las filas o en las columnas de una tabla son demasiadas; por eso es preciso reducir el número de tipos en la tipología.⁴³ Si la contribución del padre y la de la madre al capital cultural de un individuo se consideran equivalentes, una manera obvia de reducir esta tipología se muestra en la tabla 9.1:

- el código 1 en la nueva variable se atribuye a todos los que tienen ambos padres con el título elemental;
- el código 2 se atribuye a todos los que tienen al menos uno de los padres con el título medio;
- el código 3 se atribuye a todos los que tienen al menos uno de los padres con el título medio-superior;
- el código 4 se atribuye a todos los que tienen al menos uno de los padres con el título superior;
- el código 5 se da a todos los que tienen ambos padres con el título superior.

Se logra así una nueva variable ordinal, que sintetiza dos indicadores. Naturalmente, se pueden concebir otras formas de realizar esta síntesis, por ejemplo atribuyendo un peso mayor al título del padre (códigos en la tabla 9.2, parte izquierda) o al de la madre (códigos en la tabla 9.2, parte derecha).

Tabla 9.2: Otras formas de combinar dos indicadores y reducir la tipología resultante

TÍTULO DE LA MADRE	TÍTULO DEL PADRE				TÍTULO DEL PADRE			
	EL	ME	MS	SUP	EL	ME	MS	SUP
elemental	1	2	3	4	1	1	2	2
medio	1	2	3	4	2	2	3	3
medio superior	2	3	4	5	3	3	4	4
superior	2	3	4	5	4	4	5	5

Antes de mostrar cómo se combina una de las variables así construidas con otro indicador, siguiendo con la construcción del índice de capital cultural, ca-

⁴³ Dieciséis modalidades en las filas o en las columnas de una tabla, cruzándose con las categorías de otra variable, producen un número demasiado alto de celdas (el número de celdas de una tabla es el producto del número de filas por el número de columnas). La reducción del número de tipos en una tipología, llamado por brevedad "reducción de una tipología", fue teorizada por Hempel y Oppenheim (1936) y por Lazarsfeld (1937). Este último la llamó "reducción del espacio de atributos" por miedo a que el término 'tipología', de noble origen filosófico, fuera rechazado por el mercado norteamericano, en el que estaba desembarcando al huir del nazismo.

be destacar el hecho de que la relación entre el título del padre y el de la madre se puede elegir como uno de los indicadores del nivel de tensión y conflicto potenciales en la familia de procedencia del sujeto, a su vez una posible causa (no un indicador) de su nivel de (in)seguridad. El supuesto es que cuanto mayor sea el desnivel cultural entre los padres, tanto mayores probablemente serán —en el largo plazo— la tensión y el conflicto potenciales entre los dos. En este caso, los códigos a asignar a los tipos en las celdas (véase la tabla 9.3) podrían ser:

- 0 en la diagonal, donde el título es el mismo para ambos padres;
- 1 en las celdas cerca de la diagonal, donde sólo hay un escalón de diferencia entre el título de uno y del otro;
- 2 en las celdas donde hay dos títulos de diferencia;
- 3 en la celda a la extrema derecha alta y a la extrema izquierda baja de la tabla, donde la diferencia es la máxima posible.

Tabla 9.3: Otras formas de usar las mismas informaciones para crear un indicador de otro concepto

TÍTULO DE LA MADRE	TÍTULO DEL PADRE			
	ELEMENTAL	MEDIO	MEDIO SUPERIOR	SUPERIOR
elemental	0	1	2	3
medio	1	0	1	2
medio superior	2	1	0	1
superior	3	2	1	0

Veamos ahora cómo se puede proceder en la construcción del índice usando la variable formada por los títulos de los padres en combinación con otro indicador de capital cultural, por ejemplo el tipo de lecturas preferidas por el sujeto durante su adolescencia y su juventud. Imaginemos que las diversas respuestas a estas preguntas hayan sido reagrupadas en cuatro niveles:

- 1: historietas, crónica deportiva, novelas "rosa" y crónica "social" en diarios y periódicos (hist en la tabla 9.4);
- 2: cuentos de aventura y de viajes, novelas policíacas, secciones policíacas de diarios y periódicos (av en la tabla 9.4);
- 3: novelas, *best sellers*, ciencia ficción (nov en la tabla 9.4);
- 4: ensayos, clásicos de la literatura y del teatro (cl en la tabla 9.4).

En este caso la reducción de la tipología es una tarea delicada, sobre todo porque las categorías en ambas variables incluyen situaciones reales bastante diferentes, que cada investigador (y cada lector) puede interpretar a su manera. En general, los criterios que guían la elección de los tipos que se van a reagrupar (y por consiguiente las categorías de la variable resultante, que puede ser el índice final o una etapa intermedia en su construcción) son los mismos que guían la reducción del número de las clases en una clasificación (véase apartado 8.1):

el primero es la máxima proximidad semántica entre los tipos que se reagrupan (complementada por la máxima distancia semántica entre las categorías resultantes). Naturalmente, el punto de referencia para juzgar todo esto será el concepto general que se quiere operativizar mediante un índice —en este caso el capital cultural. El criterio que ya se ha definido como auxiliar es la oportunidad de evitar desequilibrios excesivos entre las frecuencias de las categorías resultantes, para no perjudicar la sucesiva fase de análisis.

Como siempre, hay que balancear el criterio semántico con el numérico.⁴⁴ Este balanceo se realiza más fácil y eficazmente si ya tenemos la tabla ante nuestra vista, es decir, cuando los datos ya han sido recolectados. Por este motivo se aconseja proyectar los índices en general y en abstracto antes de recolectar los datos, pero esperar que ellos sean recolectados antes de encarar los detalles de la construcción.

Desafortunadamente, no se puede cumplir con esta máxima al proponer el ejemplo de reducción que sigue (tabla 9.4). Sin embargo, la propuesta tiene en cuenta una consideración empírica, es decir, el hecho de que muchos entrevistados, para dar la mejor imagen de sí mismos al contestar una pregunta semejante, tienden a indicar —entre todas sus lecturas— aquellas que perciben como más apreciadas culturalmente.

Los tipos que emergen de la reducción podrían ser:

- 1: capital cultural consistentemente bajo: sujetos cuyos padres no llegaron al nivel medio-superior y que prefieren leer historietas, deportes, cuentos de aventura y de viaje, crónica "social" y novelas "rosas" o policíacas;
- 2: capital cultural descendente: sujetos cuyos padres alcanzaron el nivel superior y que no llegaron a indicar ensayos o clásicos al contestar la pregunta; sujetos con uno de los padres con instrucción superior y que no llegaron a indicar clásicos o novelas de buen nivel al contestar la pregunta; sujetos con uno de los padres con instrucción media-superior y que prefieren leer historietas, deportes, o crónica "social" y novelas "rosa";
- 3: capital cultural consistentemente medio: sujetos con uno de los padres con instrucción superior y que no llegaron a indicar ensayos o clásicos al contestar la pregunta; sujetos con uno de los padres con instrucción medio-superior y que prefieren leer cuentos de aventura y de viajes, crónica y novelas de tipos varios (no "rosa"); sujetos con uno de los padres con instrucción media y que prefieren leer novelas de buen nivel, *best sellers*, ciencia ficción;
- 4: capital cultural ascendente: sujetos sin ningún padre con instrucción superior que indican ensayos o clásicos como su lectura preferida, y sujetos con ambos padres con título elemental que indican ensayos, clásicos, novelas de buen nivel, *best sellers*, ciencia ficción;

⁴⁴ Lo destacaba también Galtung (1967: 253).

- 5: capital cultural consistentemente alto: sujetos con uno o ambos padres con instrucción superior que indican ensayos o clásicos como su lectura preferida.

Tabla 9.4: *Cómo proceder en la construcción de un índice de capital cultural*

TÍTULO DE LOS PADRES	LECTURAS PREFERIDAS			
	HIST	AV	NOV	CL
2 elemental	1	1	4	4
1 medio	1	1	3	4
1 medio superior	2	3	3	4
1 superior	2	2	3	5
2 superior	2	2	2	5

Hasta ahora, sólo se combinaron tres indicadores, y ya el ejemplo puede dar una idea de la cantidad y complejidad de problemas semánticos y de decisiones consiguientes que están involucrados en la construcción de un índice con indicadores cuya definición operativa lleva a una variable categorial u ordinal. Cada vez que se quiera combinar un indicador más, la complejidad semántica se acrecienta de forma casi exponencial, las decisiones involucradas en reducir la tipología son cada vez más cuestionables, y el control intelectual sobre toda la operación y sus resultados disminuye.

Por este motivo, cada vez que se puedan imaginar definiciones operativas que lleven a variables cardinales o cuasicardinales, los investigadores expertos las prefieren. Por otro lado, ser conscientes de las ventajas vinculadas a las variables cardinales no justifica tratar como cardinales variables que ni siquiera son ordinales —como a menudo se encuentra.

9.7. La construcción de índices aditivos con variables cardinales y cuasi cardinales

Construir índices con variables cardinales o cuasi cardinales es más simple que construirlos con variables categoriales u ordinales, y eso por dos motivos. El primero es obvio: la posibilidad de realizar con plena legitimidad⁴⁵ operaciones matemáticas con los códigos numéricos que representan los estados en estas variables. El segundo motivo —la autonomía semántica reducida o nula de las categorías (véanse los apartados 8.4 y 8.5)— no es tan obvio, y por eso no se

⁴⁵ En rigor, la legitimidad es plena sólo en el caso de variables cardinales. Como se destacaba en el apartado 8.5, las variables cuasi cardinales presentan diferencias nada despreciables con las cardinales; por eso tratarlas como cardinales no es plenamente legítimo, y se lo hace para aprovechar las grandes ventajas involucradas en este tipo de tratamiento. Con variables ordinales, la distorsión puede ser tan grande que muchos no consideran legítimo tratarlas con técnicas cardinales. Con variables categoriales este tratamiento es un verdadero sinsentido, ya que los códigos numéricos de las categorías son enteramente arbitrarios.

encuentra frecuentemente mencionado en los textos de metodología. Empero, no tiene menor importancia, porque es lo que permite pasar de un tratamiento mediante tablas, donde el foco de la atención se fija en cada celda, a un tratamiento mediante diagramas, donde la atención puede fijarse en el rumbo global de la relación.

Como veremos en el capítulo 15, esta consideración vale no sólo para la construcción de índices, sino también para el análisis de las relaciones entre variables. Desde el punto de vista técnico, los procedimientos son los mismos: lo que cambia radicalmente es el objetivo del tratamiento. En un caso, estamos construyendo variables de mayor generalidad y porte teórico; en el otro, las variables ya están construidas y estamos explorando un segmento de la tupida red de relaciones entre ellas.

Sin embargo, las operaciones matemáticas necesarias para construir índices son mucho más sencillas de las que se usan en el propio análisis estadístico. A veces son simples sumas de puntajes: el puntaje de cada caso (individuo, agregado territorial u otro) en el índice es igual a la suma de sus puntajes en los indicadores. Por eso los índices de este tipo se denominan *ADITIVOS*.⁴⁶ Naturalmente, este nivel de simplicidad sólo se puede alcanzar si se satisfacen cuatro condiciones: una factual, una numérica y dos semánticas.

La condición factual es que no falten datos en uno o más indicadores: en estas circunstancias el puntaje final de cada caso en el índice debe ser el promedio y no la suma de sus puntajes válidos. Si se sumasen, los puntajes finales de todos los casos que tienen datos faltantes serían indebidamente menores de lo que deberían ser: un sujeto podría resultar menos inteligente, o menos autoritario, o menos conservador, o menos progresista de lo que es realmente. Naturalmente, le corresponde al investigador decidir cuántos datos faltantes se pueden aceptar: si un índice está formado por cinco indicadores y para el sujeto X faltan informaciones acerca de tres de ellos, sería sabio no asignar al sujeto X ningún puntaje en ese índice.

La condición numérica es que todas las variables que se suman tengan la misma extensión de escala, o al menos una extensión muy parecida. Esta condición vale igualmente para índices formados con variables métricas (resultados de una medición), con variables cardinales naturales (resultados de un conteo) y con variables cuasi cardinales (resultados de algunas técnicas de construcción de escalas).

Veamos un ejemplo para cada tipo. Imaginemos que la federación de atletismo de una nación pobre disponga de recursos para enviar a las Olimpiadas sólo un lanzador. Para elegir quién será se decide sumar para cada candidato los resultados que alcanza en los cuatro tipos de lanzamiento (disco, jabalina, martillo y bala) y enviar el que tiene la suma más alta. Pero está claro que de esta forma se perjudica a los lanzadores de bala, cuyo margen con respecto de los no especialistas puede girar en torno de (los) cinco me-

⁴⁶ Este rótulo se usa también cuando los puntajes no se suman directamente, sino que antes se someten a las elaboraciones descritas en el texto.

tros, mientras los otros lanzadores pueden tener márgenes de treinta metros y más con respecto de los no especialistas. La solución correcta sería estandarizar los puntajes (véase apartado 8.3), o al menos normalizar⁴⁷ cada resultado dividiéndolo por el récord nacional o por el récord mundial en la especialidad respectiva.

Imaginemos ahora que se quiera construir un índice de posesión de bienes muebles duraderos, y para cada sujeto se sumen el número de coches, el número de equipos de TV, de heladeras, de lavarropas, etcétera, hasta al número de equipos de radio y de celulares. Pero está claro que un celular no tiene el mismo valor que un coche: antes de sumar estos artículos se debería ponderar cada uno por su costo específico, o al menos por el costo medio del tipo de artículo.

Por último, imaginemos construir un índice de progresismo. Ya que el progresismo es un concepto multifacético y no se puede medir directamente, hay que encontrar indicadores y operativizarlos mediante las técnicas de construcción de escalas que se ilustran en el apartado 8.5. Por ejemplo, usando los puntajes asignados por cada sujeto a líderes o símbolos progresistas con el "termómetro", su autocolocación con la escala derecha-izquierda, y las respuestas elegidas reaccionando a frases de orientación progresista en una batería de escalas de Likert. Para crear el índice no se podrían sumar directamente estos puntajes,⁴⁸ porque los puntajes en los termómetros van de 0 a 100, en la escala derecha-izquierda van de 0 a 10, en las escalas de Likert van de 0 (o 1) a 4 (o 5). Antes de sumar se deberá estandarizar los puntajes, o al menos normalizarlos dividiendo cada uno por el máximo puntaje previsto para la escala relativa.

Algunos ejemplos ya presentados sirven también para ilustrar la primera condición semántica. Imaginemos que la federación de atletismo también quisiera enviar a las Olimpiadas un decatleta. Es sabido que las pruebas del decatión incluyen tres lanzamientos, tres saltos y cuatro carreras. En este caso, no basta con normalizar los puntajes antes de sumarlos. Esto porque en los saltos y lanzamientos cuanto más alta es la cifra, mayor es el mérito; para las carreras, vale obviamente lo contrario. Por eso, se debe invertir la dirección de los puntajes relativos a las carreras —y esto es exactamente lo que hace la federación mundial de atletismo en su tabla de puntajes para el decatión.

El mismo problema semántico se presenta con el índice de progresismo. Si también se quiere insertar en el índice los puntajes de agrado asignados con el termómetro a líderes y símbolos conservadores, y las respuestas a frases conservadoras en una batería de escalas de Likert, puede hacerse sólo después de invertir la dirección de los puntajes.

La segunda condición semántica de la cual se hablaba es más compleja. Hasta ahora se presentaron operaciones (estandarización, normalización, in-

⁴⁷ La estandarización es sólo una forma sofisticada de normalización.

⁴⁸ Sin embargo, hay quien lo hizo —por ejemplo, el politólogo norteamericano Smith (1969)— sin darse cuenta de que de esta forma estaba considerando cada termómetro veinte veces más importante que cada escala Likert.

versión de dirección semántica) que tenían el fin de hacer efectivamente equivalentes los indicadores antes de combinarlos en un índice, en el supuesto de que el investigador los considerara todos igualmente válidos.⁴⁹ Pero puede pasar que un investigador considere algunos más válidos que otros, por motivos que pueden radicar en su conocimiento directo del problema y/o del ámbito investigado, en sus experiencias previas de investigación, en la literatura sobre el tema. Si tiene buenos motivos para creer que algunos indicadores son más válidos que otros, tiene que incorporar esta convicción en la formación del índice, sin temor a no ser "objetivo". Más allá del hecho de que la elección de indicadores no puede de ningún modo ser "objetiva", se debe agregar que considerar igualmente válidos todos los indicadores es un supuesto, exactamente como lo es lo contrario.

Para ponderar la contribución de cada indicador a la suma que va a proporcionar a cada caso su puntaje en el índice, el investigador tiene dos caminos. Puede guiarse por sus valoraciones semánticas, multiplicando los puntajes en los indicadores que considera más válidos por un coeficiente mayor de 1 (puede ser 1,2 o 1,5 o algo parecido) y los indicadores que considera menos válidos por un coeficiente menor de 1 (puede ser 0,7 o 0,8 o algo parecido). O puede aprovechar las posibilidades que le ofrece la matriz de los datos, sometiendo la canasta de las variables que operativizan posibles indicadores del concepto que interesa a un análisis de componentes principales.⁵⁰

Ahora basta con decir que el análisis de componentes principales tiene tres funciones, a las que corresponden tres fases en el procedimiento:

- 1) seleccionar en la canasta inicial de posibles indicadores los que resultan tener una estricta relación empírica con los otros, y por lo tanto se puede suponer que tengan una apreciable relación semántica con el concepto que interesa;
- 2) estimar la fuerza de esta relación semántica (es decir la validez) entre el concepto y cada uno de los indicadores seleccionados;
- 3) basándose en esta estimación, proporcionar un coeficiente para ponderar los puntajes de cada indicador teniendo en cuenta las posibles superposiciones semánticas, es decir el hecho de que dos o más indicadores pueden representar aspectos muy cercanos de la intensión del concepto.

En este caso, el índice final no es otra cosa que la suma de los puntajes de cada individuo en los indicadores seleccionados en la fase (1); antes de sumarlos, los puntajes son multiplicados por el correspondiente coeficiente calculado en la fase (3).

⁴⁹ Sobre el concepto de validez, véase el apartado 9.4.

⁵⁰ El análisis de componentes principales es preferible al más antiguo y más difundido análisis factorial, porque no realiza manipulaciones arbitrarias de las matrices de los datos antes de someterlas a tratamiento con el álgebra matricial. Una reseña de otras técnicas antiguas para ponderar indicadores basándose en sus relaciones con otras variables en la matriz está en Edgerton y Kolbe (1936).