

---

## LA TÉCNICA DE LAS ESCALAS

---

6.1. LA OPERACIONALIZACIÓN EMPÍRICA DE CONCEPTOS COMPLEJOS	210
6.2. PREGUNTAS DE RESPUESTA GRADUADA: LA AUTONOMÍA SEMÁNTICA DE LAS RESPUESTAS	213
6.3. LA ESCALA DE LIKERT	219
6.4. EL ESCALOGRAMA DE GUTTMAN	227
6.5. EL DIFERENCIAL SEMÁNTICO	233
6.6. EL TEST SOCIOMÉTRICO	238
6.7. ESCALAS UNIDIMENSIONALES Y ESCALAS MULTIDIMENSIONALES	243
6.8. SÍNTESIS DEL CAPÍTULO 6	245
6.9. LECTURAS COMPLEMENTARIAS	246

En este capítulo abordamos uno de los problemas más difíciles de la investigación social. El problema de la «medición», y en particular la manera de transformar en operaciones empíricas los conceptos complejos de las ciencias sociales, como los estados emocionales, los rasgos psicológicos, las actitudes políticas o las escalas de valores. Tras un examen teórico inicial y un breve paréntesis histórico, presentamos algunas de las técnicas de escalas más sencillas y conocidas; por último, en la parte final se presenta una técnica para estudiar las interacciones entre individuos en grupos pequeños.

## 6.1. LA OPERACIONALIZACIÓN EMPÍRICA DE CONCEPTOS COMPLEJOS

La expresión «técnica de las escalas» (en inglés *scaling*), hace referencia a un conjunto de procedimientos creados por la investigación social para «medir» determinadas variables en el hombre y en la sociedad. Ya hemos hablado sobre la complejidad de los conceptos utilizados en las ciencias sociales y sobre la posibilidad o imposibilidad de observar directamente muchos de ellos. Optimismo, depresión, prejuicios raciales, autoritarismo, religiosidad, inteligencia, integración social, conflictividad, etc., son conceptos típicos de las ciencias sociales, que caracterizan al ser humano, pero que no son fácilmente traducibles al lenguaje de la investigación empírica. Podemos pensar que muchos de los comportamientos y las opiniones de un individuo dependen de su conservadurismo político o su religiosidad, sin embargo, estas propiedades no se pueden observar directamente, lo que podemos observar son sus comportamientos o sus opiniones, que se derivan de la propiedad subyacente y que están influidos por ésta, pero que no coinciden con ella.

Ya tratamos esta cuestión en el capítulo 3, cuando hablamos de conceptos e indicadores. Entonces comentamos que un concepto general como la religiosidad puede operacionalizarse mediante un concepto específico que se denomina indicador, y que está ligado a él por una relación de significado (de este modo podemos operacionalizar la religiosidad con la práctica religiosa, y el conservadurismo político con el partido al que se ha votado).

La técnica de las escalas sirve para conseguir este objetivo de forma más sistemática y formalizada. La diferencia respecto a lo que ya hemos dicho sobre el uso de los indicadores radica en el hecho de que ahora no se trata sólo de sustituir un concepto con uno o más indicadores, sino con un conjunto coherente y orgánico de indicadores, elaborando también criterios intersubjetivos para controlar la coherencia efectiva entre indicadores y concepto y comprobar que el procedimiento se ha completado. Podemos decir, por tanto, que una *escala es un conjunto coherente de elementos que se consideran indicadores de un concepto más general*.

El *elemento* es el componente individual (afirmación, pregunta, comportamiento, respuesta a un test, cualidad), y la *escala* es el conjunto de los elementos. El concepto subyacente al elemento tiene distintas denominaciones, según la disciplina: los psicólogos hablan de «rasgo» o «constructo», los sociólogos emplean fre-

cuentemente el término «variable latente».

Una prueba de habilidad matemática es un ejemplo de escala: se establece la hipótesis de que existe un concepto general, como la «habilidad matemática», que se puede registrar mediante una serie de pruebas específicas (resolución de ecuaciones, problemas, pruebas lógicas, etc.), cuyos resultados se sintetizan en una puntuación. La tabla 5.1 (pregunta 1) del capítulo anterior es otro ejemplo de escala, que contiene una serie de trastornos psicofísicos extraídos del cuadro de depresión de Carroll: en este caso los síntomas son los elementos de la escala, la depresión es el concepto general, y el resultado final de la escala es una puntuación, que se basa en las respuestas individuales y se asigna a cada uno de los individuos que realizan la encuesta. Esta puntuación se considera una medida del grado de depresión.

La aplicación más frecuente de la técnica de las escalas en el campo de la sociología y la psicología social es la denominada *medida de las actitudes*, en la que la unidad de análisis es el individuo, el concepto general es una actitud y los conceptos específicos son opiniones. Por «actitud» entendemos «el conjunto de tendencias y sentimientos, prejuicios e ideas preconcebidas, nociones, temores, inquietudes y convicciones de una persona hacia una determinada cuestión» [Thurstone y Chave, 1929, 67]. La actitud es, pues, una creencia que no se puede registrar directamente, mientras que la opinión es una de las formas en que se manifiesta la actitud, es decir, una expresión de ella que se puede registrar empíricamente.

Podríamos decir que la actitud es el concepto general y las opiniones son sus indicadores. El procedimiento para registrar las actitudes consiste en presentar a los sujetos estudiados una serie de afirmaciones y pedirles que expresen su opinión al respecto. Si combinamos adecuadamente las respuestas podemos llegar a una puntuación individual que estima la posición de cada sujeto sobre la actitud en cuestión.

En la tabla 6.4 presentamos una de estas escalas, en la que la actitud es el «sentimiento de eficacia política», un aspecto de la participación política que se registra mediante la opinión expresada en afirmaciones individuales sobre la política, los diputados, etc. Si puntuamos cada una de las respuestas según un baremo que va de 1 a 4, (1 para el mínimo de eficacia y 4 para el máximo) y sumamos los puntos, podemos atribuir a cada entrevistado una puntuación total de eficacia política, que va desde el valor mínimo, 9 (el total de preguntas del ejemplo es nueve), hasta el valor máximo, 36. La escala está formada por la suma de las puntuaciones obtenidas en cada una de las afirmaciones, que son sus «elementos».

Desde las primeras aplicaciones, en los años treinta, los sociólogos y los psicólogos sociales han elaborado centenares de escalas para «medir» las actitudes y, más en general, para registrar varias dimensiones de la personalidad humana: estados emocionales (ansiedad, depresión, resentimiento), rasgos psicológicos (autoestima, introversión), necesidades (autorrealización, poder), relaciones sociales (estatus social, integración familiar), tendencias políticas (izquierda-derecha, alienación política)<sup>1</sup>.

<sup>1</sup> Algunas de las escalas de actitud más utilizadas se pueden encontrar en Miller [1991], que contiene las escalas utilizadas en la «American Sociological Review» (*Inventory of Measures Utilized in the American Sociological Review, 1981-87*), y en las dos obras de Robinson, Shaver y Wrightsman [1991] sobre actitudes psicológicas y [1999] sobre actitudes políticas.

En el ámbito de las ciencias de la educación se han elaborado escalas para medir habilidades y capacidades (manuales, mentales), o la facilidad o dificultad para el aprendizaje (test de evaluación escolar, profesional).

Aunque la «medida de las actitudes» constituye el campo de aplicación más importante de la técnica de las escalas, no es el único. En primer lugar, la técnica se puede utilizar no sólo para registrar las propiedades de los individuos a través de sus respuestas a una serie de estímulos (los elementos de la escala), sino también para atribuir una puntuación a los estímulos en función de las respuestas de los individuos. Por ejemplo, podemos atribuir a las profesiones una puntuación de prestigio social, o bien establecer una jerarquía de popularidad entre los políticos, a partir de las respuestas y las opiniones expresadas por los entrevistados.

Además, la técnica de las escalas puede emplearse para registrar las propiedades de otras unidades que no sean individuos, por ejemplo, para juzgar la eficiencia de las instituciones (gobiernos, empresas, organismos públicos, etc.), para asignar puntuaciones de cohesión social a las comunidades, para determinar el poder de una serie de cargos profesionales a partir de las relaciones de autoridad-sumisión que mantienen en la interacción cotidiana en el trabajo, etc.

¿Qué podemos decir sobre el tipo de variables que genera la técnica de las escalas? ¿Se trata de variables nominales, ordinales o cardinales? La dimensión subyacente se suele concebir como una propiedad continua: el grado de prejuicio racial de una persona, o la religiosidad, varían de forma gradual en los individuos, o al menos así se cree. Que estas propiedades continuas no se traduzcan en variables cardinales se debe sólo a nuestra incapacidad para medirlas, es decir, para establecer una unidad de medida de las mismas. La técnica de las escalas representa la forma en que las ciencias sociales han hecho frente a este problema, proponiendo procedimientos que conduzcan a puntuaciones cardinales para evaluar las propiedades en cuestión.

¿Se ha alcanzado o se puede alcanzar este objetivo? La técnica de las escalas tradicionales ha producido variables que hemos denominado «cuasicardinales» (ver el capítulo 3, apartado 6), y ese «cuasi» indica precisamente la imposibilidad de cumplir el objetivo de atribuir un significado numérico pleno a las puntuaciones de las escalas<sup>2</sup>. Sin embargo, los avances teóricos, junto con los nuevos recursos de cálculo derivados de la informática, han permitido recientemente elaborar modelos estadísticos capaces de producir variables para las que se realiza esa «igualdad de intervalos» que equivale a la existencia de una unidad de medida. Se trata de aplicaciones complejas y de difusión limitada (las escalas de Rasch), por lo que podemos decir que el cuerpo central de las técnicas de investigación social no va más allá de las escalas cuasicardinales.

Recordemos brevemente que los primeros intentos de registrar las actitudes mediante el uso de escalas se remontan a mediados de los años veinte, con las

<sup>2</sup> Ni siquiera la escala más estudiada y perfeccionada, el cociente intelectual, tiene una verdadera unidad de medida, y sigue siendo dudoso que los intervalos numéricos se correspondan realmente con la propiedad latente (no sabemos si la diferencia entre las puntuaciones de 95 y 100 es efectivamente igual a la diferencia entre las puntuaciones de 100 y 105).

propuestas de Allport y Hartman, Bogardus y Thurstone. Fue Thurstone [1927; 1928; 1931] quien hizo una primera sistematización en este campo, al formular tres propuestas diferentes (*Paired Comparison, Rank Order, Equal Appearing Intervals*), de notable interés metodológico (sobre todo la tercera), pero que hoy se consideran superadas, sobre todo porque su aplicación es muy laboriosa. Por lo contrario, es decir, por la gran sencillez de realización, tuvo un gran éxito (que aún perdura) la propuesta expuesta por Likert en 1923. Más adelante, tendría una gran relevancia la aportación de Guttman [1944; 1950]. Los manuales de metodología y la práctica de la investigación social se han centrado durante años en los tres modelos de las escalas de Thurstone, Likert y Guttman. Recientemente, gracias a las nuevas posibilidades de cálculo que ofrece la informática, ha surgido una nueva aproximación de tipo probabilístico para el registro de propiedades continuas, que ha encontrado vías de aplicación adecuadas; esta aproximación fue presentada inicialmente por Lazarsfeld [1950], y vio sus primeras aplicaciones en las propuestas de Rasch [1960] y Mokken [1971]. No entraremos en los detalles de la misma, ya que su complejidad supera los límites de este texto.

## 6.2. PREGUNTAS DE RESPUESTA GRADUADA: LA AUTONOMÍA SEMÁNTICA DE LAS RESPUESTAS

Una escala está formada por varios elementos, que en la mayoría de los casos constan de varias preguntas; usando una noción que presentamos en el capítulo 3, podemos decir que, por lo general, una *escala está constituida por una batería de preguntas*.

Antes de estudiar en detalle la cuestión de la construcción de las escalas, prestaremos atención al formato de las preguntas. En una pregunta cerrada, cuando las opciones de respuesta que se presentan al entrevistado están ordenadas (de manera creciente o decreciente), se pueden presentar de tres formas.

La primera consiste en presentar unas posibilidades de respuesta que son semánticamente autónomas, es decir, cada una tiene un significado intrínseco pleno que, para ser comprendido, no necesita relacionarse con el significado de las otras opciones presentes en la escala. En las preguntas y las correspondientes respuestas de la tabla 6.1 tenemos un ejemplo de esta forma: para responder que su nivel de estudios es universitario un entrevistado no necesita conocer las otras opciones de respuesta.

En la segunda forma las opciones de respuesta tienen una *autonomía semántica parcial*. El caso más común son las respuestas del tipo «mucho», «bastante», «poco», «nada», o similares (véanse los ejemplos de la tabla 6.2). En este caso el significado de cada opción de respuesta es sólo parcialmente autónomo con respecto a las otras opciones. No queda claro qué quiere decir exactamente que la política nos interese «bastante». En el contexto de la serie que va desde «mucho» hasta «nada», donde la opción «bastante» está situada después de «mucho» y antes de «poco», resulta más fácil atribuirle un significado.

Por último están las denominadas escalas de intervalos de autopoicionamiento [Cantrill y Free, 1962], en las que sólo hay dos categorías extremas dotadas de

Tabla 6.1. Preguntas con opciones de respuesta semánticamente autónomas

Pregunta 1. ¿Ha ido usted a la iglesia en el último año?	Pregunta 2. ¿Cuál es su nivel de estudios?
1. Nunca	1. Sin estudios
2. Dos-tres veces al año	2. Graduado escolar
3. Una vez al mes	3. Secundaria
4. Dos-tres veces al mes	4. Diplomatura
5. Una o más veces a la semana	5. Licenciatura

**Pregunta 3.** *Una pareja puede gestionar su dinero de diversas maneras. ¿Puede indicar, entre las siguientes, cuál se parece más a la que se aplica en su familia?*

1. El marido se queda con su sueldo y la esposa le pide dinero cuando lo necesita.
2. El marido se queda con su sueldo, pero le da a su esposa una cantidad fija de dinero para los gastos de la casa.
3. El marido ingresa su sueldo en una cuenta común de la que los dos cónyuges retiran dinero cada vez que lo necesitan.
4. El marido le da su sueldo a su mujer, que después le devuelve a su marido un porcentaje de ese dinero para sus gastos personales.
5. Otra (especificar).

significado, y entre ellas se coloca una escala, representada por casillas, cifras o un segmento, en el que la entrevistado marca su posición. En la tabla 6.3 reproducimos cuatro ejemplos de ello. El primero refleja la dimensión de orientación ideológica izquierda-derecha, el segundo emplea el llamado «termómetro de los sentimientos», y los dos últimos contienen preguntas sin especificidad propia, para que se comprenda que la técnica se puede aplicar a cualquier pregunta cuyas respuestas se puedan imaginar como un continuo delimitado por dos alternativas contrarias.

Tabla 6.2. Preguntas con respuesta de autonomía semántica parcial

¿Está usted de acuerdo con la siguiente afirmación: «Hoy en España los trabajadores hacen demasiadas huelgas, lo que perjudica al progreso del país»?	¿Le interesa la política?	Indique el tipo de lugar de residencia
1. Totalmente de acuerdo	1. Mucho	1. Urbano
2. Bastante de acuerdo	2. Bastante	2. Cuasiurbano
3. Más bien en desacuerdo	3. Poco	3. Semiurbano
4. Totalmente en desacuerdo	4. Nada	4. Semirural
		5. Cuasirural
		6. Rural

En cuanto al tipo de variable que se obtiene con estos tres procedimientos, podemos decir que en el primero de los casos obtendremos variables ordinales: sólo se asegura el orden de las modalidades, pero se desconoce totalmente la distancia entre ellas. La autonomía semántica de las categorías hace que el entrevistado las escoja por su contenido, al margen de su posición con respecto a las otras.

**Tabla 6.3.** Escalas de intervalos de autoposicionamiento

**Pregunta 1.** En política suele hablarse de izquierda y de derecha. Sirviéndose de esta ficha, ¿dónde situaría su orientación política?




Izquierda					Derecha					No sé
1	2	3	4	5	6	7	8	9	10	99

**Pregunta 2.** Ahora queremos pedirle su opinión sobre algunas instituciones, hechos o partidos cuyo nombre está escrito en estas fichas (dar las fichas y dejar que las estudien con calma; después presentar la figura con el termómetro). Esta figura representa un termómetro, graduado de 0 a 100 grados. Puede usar este termómetro para darnos una idea de su simpatía o antipatía hacia las instituciones que aparecen en las fichas: si siente mucha simpatía o tiene usted una opinión muy positiva sobre ella, coloque la ficha a 100, o a 95-90 grados; si siente mucha antipatía o tiene una opinión muy negativa sobre ella, coloque la ficha a 0 o a 5-10 grados; si tiene usted una opinión intermedia, coloque la ficha entorno a los 50 o los 40-60 grados. Coloque con calma todas las fichas y desplácelas todas las veces que quiera.

**Pregunta 3.** Hay quien sostiene que los desempleados deberían aceptar cualquier trabajo disponible o perder la prestación por desempleo. En cambio otros piensan que los desempleados deberían tener derecho a rechazar un empleo que no desean. Ayudándose de esta ficha, ¿dónde situaría su posición al respecto?

Aceptar cualquier empleo o perder la prestación					Tener derecho a rechazar un empleo que no desea					No sé
1	2	3	4	5	6	7	8	9	10	99

**Pregunta 4.** ¿Cuál es su grado de satisfacción con respecto a los siguientes aspectos de su vida? (Marque una cruz en la barra horizontal: si está totalmente insatisfecho, ponga la cruz cerca del extremo izquierdo; si está muy satisfecho póngala cerca del extremo derecho, de manera que su grado de satisfacción sea proporcional al segmento que queda a la izquierda del signo trazado).

	Insatisfecho	Satisfecho	No sé
Situación económica			
Relación con su mujer			
Relación con sus hijos, etc.			

Tampoco en el caso de las respuestas con autonomía semántica parcial estamos en condiciones de afirmar que las distintas modalidades de respuesta son equidistantes entre sí: para un entrevistado, «bastante» podría tener una connotación muy positiva y estar más cercano a «mucho» y más alejado de «poco», mientras que para otro podría suceder lo contrario. Sin embargo, dado que las respuestas individuales no tienen significado autónomo, el entrevistado toma como referencia las otras opciones para interpretarlas, y es probable que en este proceso intervenga un mecanismo de comparación cuantitativa, especialmente cuando el número de opciones que se ofrecen es bastante elevado.

Este proceso de autoevaluación de la distancia entre las distintas opciones de respuesta es más probable que se produzca con las respuestas de intervalos de auto-posicionamiento. Al atribuir un significado a las categorías intermedias, el entrevistado pone en marcha de manera automática un proceso mental de subdivisión graduada del espacio semántico entre los dos extremos, estableciendo correspondencias entre los estados sobre la propiedad y las posiciones en dicho espacio semántico. De este modo, si está muy satisfecho con una determinada situación o está totalmente de acuerdo con una determinada afirmación, se situará en un extremo, si está un poco menos de acuerdo se dirigirá hacia la zona central, si está equidistante entre los dos extremos se colocará exactamente en el centro, y pasará a la otra mitad de la escala si su opinión se orienta hacia el polo opuesto. Es probable que la subdivisión que el sujeto realiza mentalmente sobre la escala consista en una subdivisión de la misma en espacios iguales, por lo que esta técnica debe garantizar la equidistancia entre las categorías.

El hecho de que sea el mismo sujeto quien implícitamente establezca la unidad de medida de la escala, hace que este procedimiento sea subjetivo: en otras palabras, no disponemos de una unidad de medida externa, válida para todos los sujetos estudiados y para el investigador. Precisamente por ello no hablamos de verdaderas variables cardinales, sino de variables *cuasicardinales*.

Todos los investigadores están de acuerdo en que, a la hora de realizar el análisis estadístico, estas variables se traten como si fueran cardinales. En la práctica de la investigación social es bastante común tratar como variables cardinales también las variables del segundo y del primer tipo, y por lo general se asigna a las categorías valores numéricos en secuencia simple (1, 2, 3...). Sin embargo, esta práctica es controvertida, y ha suscitado un amplio debate entre los investigadores sociales.

Para concluir este punto, trataremos algunas cuestiones de carácter técnico relativas a la formulación de las preguntas. Los resúmenes 6.1 y 6.2 tratan sobre los temas de la respuesta neutra y el número de opciones de respuesta.

La expansión de la entrevista telefónica, que no permite utilizar soportes visuales (por ejemplo, fichas donde están escritas las siete opciones de respuesta que abarcan desde «totalmente de acuerdo» hasta «totalmente en desacuerdo»), cuya finalidad es facilitar la comprensión de las opciones de respuesta y la elección entre las distintas posibilidades, ha contribuido al uso de preguntas formuladas mediante una sucesión de elecciones binarias. Se le puede preguntar al entrevistado si está de acuerdo o en desacuerdo con una determinada afirmación; después, en función de lo que haya contestado, se le pregunta si «mucho» o sólo «bastante»; y así sucesiva-



### Resumen 6.1. LA OPCIÓN CENTRAL NEUTRA

Cuando la variable es «de autonomía semántica parcial» se plantea el problema de si es conveniente ofrecer de manera explícita al entrevistado una opción central neutra (por ejemplo, «totalmente de acuerdo, bastante de acuerdo, *ni de acuerdo ni en desacuerdo*, bastante en desacuerdo, totalmente en desacuerdo»), o si es mejor no hacerlo, obligando al entrevistado a adoptar una postura clara. Los expertos no se ponen de acuerdo sobre cuál de las dos formulaciones es más adecuada: la presencia de una opción neutra puede favorecer la fidelidad del registro en los casos en que existan efectivamente posiciones intermedias, pero también puede ofrecer una escapatoria a quienes prefieren no exponer claramente lo que piensan<sup>3</sup>. En general, parece oportuno ofrecer una opción neutra, a menos que se quiera forzar al individuo a que elija una opción. Pero es muy recomendable, como ya se ha dicho en el capítulo 5, apartado 4, ofrecer siempre la posibilidad de la opción «no sé», para evitar lo que hemos llamado «seudoopiniones»<sup>4</sup>.

mente<sup>5</sup>. Parece que esta formulación lleva a resultados similares a los que se obtienen con una única pregunta que contenga desde el principio todas las alternativas graduadas [Groves, 1989, 467].

En cuanto a las gradaciones de los intervalos de autopoicionamiento, en los ejemplos antes citados hemos presentado tres soluciones distintas para la secuencia de las respuestas: la de casillas vacías, la de la secuencia con cifras (a menudo de 1 a 7, o de 1 a 10; o bien, con el termómetro de los sentimientos, de 1 a 100), y la que presenta una línea continua. Entre la primera y la segunda técnica no hay grandes diferencias. La formulación de la línea continua ha sido propuesta por las ciencias sociales hace poco, adoptando así modalidades de encuesta que se han aplicado con éxito durante los últimos treinta años en el campo de las sensaciones físicas

<sup>3</sup> Para una reseña de las investigaciones que se han realizado sobre la cuestión de la opción neutra remitimos a la obra de Schuman y Presser [1981, capítulo 6]. Más en general, sobre la formulación de las preguntas para las escalas, véase De Vellis [1991].

<sup>4</sup> El tratamiento de las respuestas «no sé» en la fase de análisis de los datos es una cuestión bastante delicada. Existe una costumbre bastante común que consiste en atribuirles una puntuación intermedia entre los dos extremos, tratándolas como si fueran neutras. Pero esto es un error. Por ejemplo, ante la alternativa entre mayores impuestos y menores servicios sociales, quien elige una posición intermedia, interpretable como una posición de equidistancia entre los dos extremos, no es lo mismo que quien elige «no sé» porque no entiende la pregunta o porque nunca se ha planteado el problema.

<sup>5</sup> Este procedimiento se utiliza por ejemplo para la escala de posicionamiento político izquierda-derecha. Se le pregunta al entrevistado si es «de izquierdas, de centro o de derechas». Si dice que «de izquierdas (derechas)», se le pregunta: «¿Pero es usted exactamente de izquierdas (derechas), o bien de centro-izquierda (derecha)?». Si dice que «de centro», se le pregunta: «¿Pero es usted exactamente de centro-centro, o bien es un poco más de centro-izquierda o de centro-derecha?». De este modo, al final se obtienen siete posiciones.

### Resumen 6.2. EL NÚMERO DE OPCIONES DE RESPUESTA

Otra cuestión controvertida es la del número de opciones de respuesta que se ofrecen al entrevistado: cinco, como en el ejemplo anterior, o siete (muy de acuerdo, más bien de acuerdo, en parte de acuerdo, etc.), o tres, etc. En general, podemos afirmar que si la escala contiene pocas preguntas (o directamente se basa en una sola pregunta), conviene ofrecer al entrevistado una amplia gama de posibilidades de respuesta (por ejemplo, cinco o siete opciones). Y si la escala contiene muchas preguntas, éstas podrían plantearse como una elección binaria (Sí/No). El objetivo de la escala es captar la diversidad entre los entrevistados con relación a la propiedad que se registra. Si una escala está compuesta por una sola pregunta y ofrece diez opciones posibles, podemos distribuir a los sujetos en diez posiciones. Si una escala está compuesta por diez preguntas, todas binarias, también obtendremos (sumando las respuestas de las diez preguntas) una puntuación final a diez posiciones.

Téngase en cuenta, por otra parte, que cuanto mayor sea el número de opciones, más difícil resultará la elección (es más fácil responder con un «de acuerdo/en desacuerdo» que establecer el propio grado de acuerdo). En cualquier caso, el problema ha encontrado soluciones específicas dentro de las diversas técnicas (por ejemplo, en las escalas de Likert las preguntas suelen presentar cinco o siete opciones, en las de Guttman las respuestas son dicotómicas, etc.). Por ello volveremos sobre el tema en la presentación de estas técnicas.

(luminosidad, intensidad de los sonidos, sabores, olores, estimaciones de peso, etc. [Lodge, 1981]).

Normalmente se le pide al entrevistado que trace un segmento de longitud proporcional al grado de aprobación de la afirmación propuesta, de simpatía hacia un determinado personaje, de confianza en una cierta institución, etc. La longitud se mide después en la fase de codificación y se transforma en un número. Este procedimiento de tipo gráfico se defiende argumentando que las convenciones relativas a las distancias físicas tienen un carácter de intersubjetividad mayor que las relativas a las distancias semánticas.

En este apartado hemos estudiado la estructura de las preguntas. Pero una escala suele estar compuesta por varias preguntas o, para ser más exactos, varios elementos. Nunnally [1978, 66-68] cita tres razones por las que las escalas de varios elementos son preferibles a las de un único elemento. En primer lugar, la complejidad de los conceptos que se pretende registrar hace difícil abarcarlos con un único indicador (recordemos lo dicho en el capítulo 3, apartado 7, sobre las distintas dimen-

siones de los conceptos complejos). En segundo lugar, un único registro no ofrece la precisión necesaria, puesto que no consigue discriminar entre las distintas posiciones de los sujetos con respecto a la propiedad considerada. Por ejemplo, en el caso muy frecuente de las preguntas dicotómicas (por ejemplo, a favor/en contra de una determinada afirmación), se consigue dividir a los sujetos en dos grupos, y no es posible hacer distinciones más sutiles. Por último, cada pregunta está más expuesta a los errores accidentales, que tienen menor peso cuando la escala se basa en un número mayor de respuestas.

En conclusión: las escalas compuestas por un único elemento son menos válidas, menos precisas y menos fiables. Por este motivo, en los apartados que siguen presentaremos las técnicas de escalas de varios elementos más difundidas.

### 6.3. LA ESCALA DE LIKERT

El nombre de esta técnica proviene del psicométrico Renis Likert, que la propuso por primera vez a principios de los años treinta [Likert, 1932]. La denominación de «escala de Likert» engloba una amplia variedad de escalas, también llamadas *summated rating scales* en inglés, término que se puede traducir como *escalas aditivas*. Esta técnica sigue siendo actualmente el procedimiento más utilizado para el estudio de las actitudes y se ha desarrollado y enriquecido notablemente en comparación con las propuestas iniciales de Likert. Al ilustrarlas introduciremos también algunos elementos generales (sobre la redacción de las preguntas, la validación de la escala, etc.), que se aplican a todas las técnicas de construcción de las escalas.

El procedimiento que subyace a las escalas de Likert es simple e intuitivo. Es la primera operación que se le ocurre a cualquiera para atribuir una puntuación total basándose en las puntuaciones de pruebas individuales: hacer la suma de los puntos de las pruebas individuales. Así, en un test escolar compuesto por 30 preguntas, podemos atribuir un punto a cada pregunta exacta y sumar los distintos puntos: el estudiante que no ha cometido errores obtendrá un 30. De manera análoga, en el campeonato de fútbol la puntuación de cada equipo en la clasificación general se obtiene sumando los puntos conseguidos en cada partido.

Tradicionalmente el formato de las preguntas individuales de las escalas de Likert está representado por una *serie de afirmaciones* para cada una de las cuales el entrevistado debe decir si está de acuerdo y en qué medida. En la versión inicial del mismo Likert se proponían siete alternativas: totalmente de acuerdo, de acuerdo, en parte de acuerdo, dudoso, en parte en desacuerdo, en desacuerdo, totalmente en desacuerdo. Después se han reducido a cinco, a veces incluso a cuatro, con la eliminación de la categoría intermedia. Remontándonos a la clasificación del apartado anterior, se trata de preguntas con respuestas de autonomía semántica parcial.

La construcción de la escala consta de cuatro fases: *a)* formulación de las preguntas, *b)* aplicación de las preguntas a una muestra de sujetos, *c)* selección de las preguntas y determinación del grado de coherencia de la escala, y *d)* control de la validez y la unidimensionalidad de la escala. A continuación pasamos a comentar

cada una de estas fases.

En la *primera fase* se determinan las dimensiones de la actitud estudiada y se formulan afirmaciones que reflejen los diversos aspectos del concepto general que se quiere destacar. Por ejemplo, Adorno y sus colegas identificaron uno de los componentes del autoritarismo en una «una actitud de impaciencia y oposición hacia los individuos de carácter cariñoso; los individuos autoritarios tienen miedo de los sentimientos verdaderos, pues las emociones podrían escapar a su control» [Adorno *et al.*, 1950, 235]. A partir de esta consideración incluyeron en su escala F de fascismo-autoritarismo algunas preguntas destinadas expresamente al reconocimiento de este componente particular de la personalidad.

Hay que subrayar la importancia de esta fase teórica. Aunque la teoría se puede aplicar a un conjunto cualquiera de afirmaciones centradas en un único tema, confiando luego a operaciones estadístico-matemáticas la tarea de seleccionar las afirmaciones pertinentes y descartar las poco congruentes, la escala resultante será más válida, es decir, más adecuada para registrar efectivamente la propiedad subyacente para la que ha sido diseñada, si las afirmaciones que contiene se formulan después de identificar en la teoría los distintos aspectos de los conceptos que se quiere estudiar.

Esos conceptos suelen ser complejos y contienen múltiples *dimensiones* que la escala debe cubrir (recordemos la relación entre conceptos e indicadores comentada en capítulo 3, apartado 7). Volviendo al que es uno de los ejemplos más conocidos de aplicación de esta escala, *La personalidad autoritaria*, de Adorno y varios autores más, éstos identificaron, a partir de investigaciones anteriores, de estudios psicológicos, de la literatura general sobre el antisemitismo y el fascismo, nueve dimensiones de la personalidad autoritaria, en torno a las cuales construyeron cada una de las preguntas-afirmaciones de la escala.

Para concluir la reflexión sobre este punto con una afirmación sintética (que no sólo se aplica a las escalas de Likert, sino también al resto de escalas), podemos decir que es muy recomendable adoptar un enfoque deductivo, más que inductivo.

A continuación, en la *segunda fase*, la escala se aplica a una muestra de entrevistados. Lo único que merece la pena subrayar de esta fase es el hecho de que en general esta técnica exige un alto nivel de formación de los entrevistados. Afirmaciones del tipo «sería preferible aumentar los impuestos antes que reducir las prestaciones sociales que ofrece el Estado», o bien «es preferible renunciar a la utilización de las centrales nucleares para la producción de energía eléctrica, dado que las ventajas económicas que ofrecen quedan anuladas por los riesgos que conllevan para la población», pueden ser difíciles de comprender para un número importante de entrevistados.

Es necesario recordar también lo que ya hemos dicho sobre las baterías de preguntas: que son particularmente sensibles a los errores provocados por las respuestas emitidas al azar (pseudo-opiniones) o de manera mecánica y sistemática para todas las preguntas de la batería. Por ello es importante que se incluyan las respuestas «no sé», y que no todas las afirmaciones tengan el mismo sentido, a fin de obligar a los entrevistados a reflexionar, puesto que para que sus respuestas sean coherentes deberán expresar acuerdo en unas ocasiones y desacuerdo en otras.

En la tabla 6.4 hemos reproducido una escala del «sentimiento de eficacia política». Este concepto fue introducido en 1954 por el equipo de investigación del Survey Research Center de la Universidad de Michigan para indicar la impresión de del ciudadano de que está integrado en el sistema político, en el sentido de que percibe que puede ejercer influencia en los procesos políticos. En la investigación de la que se ha extraído este ejemplo, la escala contenía nueve afirmaciones, cinco sobre política general y cuatro sobre política municipal.

Una vez que se ha aplicado la escala a los sujetos, corresponde valorar la capacidad efectiva de la escala para cumplir el objetivo para el que ha sido creada. Como sabemos, la premisa para ello es que todos los elementos que componen la escala estén correlacionados con un mismo concepto subyacente, ya que en la primera fase los elementos se han elegido partiendo de esta perspectiva. Lo que no sabemos es si esta elección, que el investigador ha realizado partiendo de una reflexión sobre el concepto a estudiar, ha sido correcta.

Por lo tanto, tras la aplicación de la escala a los sujetos, es necesario definir algún criterio empírico para asegurarnos de que esta dimensión común a todos los elementos sea efectivamente reconocible incluso para los entrevistados. De hecho, es posible que algunos de los elementos de la escala no sean coherentes con el resto, es decir, que estén semánticamente relacionados con otros conceptos distintos, y que, por tanto, se deban eliminar. Dicho de manera más rigurosa: debemos asegurarnos de que la escala es *unidimensional*.

Esta comprobación se realiza en la *tercera fase*. Para ello se utilizan dos instrumentos básicos: la correlación *elemento-escala* (o correlación «parte-todo»), que sirve para identificar los elementos de la escala que no son coherentes con los demás, y el *coeficiente alfa*, que sirve para juzgar el grado total de coherencia interna de la escala (ver también los resúmenes 6.3 y 6.4).

Al eliminar los elementos inadecuados de la escala, el investigador tendrá en cuenta tanto las correlaciones elemento-escala como el coeficiente alfa. Eliminará los elementos con correlación elemento-escala demasiado baja, y de este modo hará que ante la correlación de un elemento con el resto de elementos, es decir, se eliminarán los elementos que presenten una correlación elemento-escala menor, siempre y cuando esta operación produzca un aumento en el valor de alfa.

En la tabla 6.4 hemos reproducido el coeficiente alfa de la escala, los coeficientes de correlación elemento-escala y, para cada elemento, el valor de alfa que quedaría si ese elemento se eliminara. Vemos que la escala presenta un valor de alfa muy elevado ( $\alpha = 0,88$ ), por lo que en principio podríamos decir que es satisfactorio. Los valores de las correlaciones elemento-escala son elevados para todos los elementos. El que presenta la correlación más baja es el número 2 ( $r = 0,47$ ), pero su eliminación no produce un aumento de alfa (que sigue siendo igual a 0,88); por otro lado, era la única afirmación de carácter cognitivo de la escala, por lo que los investigadores decidieron mantenerla.

A continuación pasamos a la *cuarta fase*, en la que corresponde controlar la validez de la escala y su carácter unidimensional. Para controlar la validez debemos considerar la diferencia entre la aplicación clásica de la técnica de las escalas la aceptabilidad de la escala (siguiendo el procedimiento descrito en la fase ante-

**Tabla 6.4.** Escala de eficacia política

*Pregunta: A continuación voy a leer algunas afirmaciones que se suelen hacer con frecuencia. Querría saber si está usted de acuerdo o no con cada una de estas afirmaciones. Le ruego que a medida que yo lea estas afirmaciones, me vaya diciendo si está totalmente de acuerdo, de acuerdo, en desacuerdo o totalmente en desacuerdo con cada una de ellas. Diga exactamente lo que piensa; no hay respuestas correctas ni incorrectas. Si no sabe qué responder a alguna de estas preguntas o no tiene las ideas claras, diga «no sé».*

	% totalmente de acuerdo + de acuerdo	correlación elemento- escala <sup>a</sup>	valor
1. La gente como yo no tiene ninguna influencia en lo que hace el gobierno	62,9	0,58	0,87
2. A veces la política parece tan complicada que no es posible entender lo que está pasando	79,6	0,47	0,88
3. No creo que los políticos se interesen demasiado por lo que piensa la gente como yo	69,6	0,62	0,87
4. Normalmente las personas elegidas para el Parlamento pierden pronto el contacto con los electores	82,5	0,60	0,87
5. A los partidos sólo les interesan los votos de la gente, no sus opiniones	67,7	0,64	0,87
6. La gente como yo no tiene ninguna influencia sobre lo que hace el gobierno de los Ayuntamientos	50,1	0,70	0,86
7. Normalmente las personas elegidas para el gobierno municipal de mi ciudad pierden muy pronto el contacto con los electores	47,4	0,71	0,86
8. La gente como yo no tiene ninguna influencia sobre lo que hace un concejal de distrito	46,8	0,67	0,87
9. Normalmente las personas elegidas como concejales de distrito pierden pronto el contacto con los electores	41,5	0,68	0,87

Alfa = 0,88

<sup>a</sup> Correlación «correcta»: correlación entre las puntuaciones sobre ese elemento y las puntuaciones sobre el índice de los elementos restantes.

### Resumen 6.3. LA CORRELACIÓN ELEMENTO-ESCALA

Para calcular la correlación elemento-escala, se calcula la puntuación de cada sujeto sobre toda la escala y el coeficiente de correlación entre esta puntuación y la puntuación sobre cada elemento individual <sup>6</sup>. El coeficiente de correlación es una medida que cuantifica el grado de relación existente entre dos variables cardinales. Si las dos variables covarían (es decir, al variar una, varía también la otra), el coeficiente de correlación toma un valor elevado (hasta el máximo de 1 si la correlación es positiva y de -1 si la correlación es negativa) <sup>7</sup>; si las variables no están correlacionadas, el coeficiente tendrá un valor bajo, que en caso de ausencia total de correlación será igual a cero. En nuestro caso, el coeficiente de correlación nos dice si la puntuación de cada elemento va en la misma dirección que la puntuación global que tiene en cuenta el resto de los elementos.

Por ejemplo, si estamos midiendo el autoritarismo, tendrá que ocurrir que si a una persona se le atribuye una puntuación global elevada sumando las respuestas a todas las preguntas (es decir, resulta «autoritaria»), esa persona tendrá que tener una puntuación bastante elevada (respuestas de tipo «autoritario») también sobre cada uno de los elementos de la escala. Es posible encontrar excepciones individuales, pero si hay una pregunta que de manera sistemática (es decir, para un número muy elevado de individuos) presenta valores contradictorios con la puntuación global de cada uno de los individuos, habrá que deducir que esa pregunta específica presenta algún problema. Puede ser que sea ambigua, que sea malinterpretada debido a una formulación inadecuada, o que haga hincapié en un aspecto distinto del autoritarismo. En todo caso, está claro que esa pregunta no es congruente con el resto de la escala y por tanto se deberá eliminar.

y una aplicación simplificada. En el planteamiento tradicional, la construcción de la escala es una operación con autonomía propia, es decir, cuando se construye una escala para medir variables como la autoestima, la ansiedad, la religiosidad, etc., se busca crear un instrumento que sea aplicable también a poblaciones distintas de aquella para la que se creó inicialmente la escala.

Se parte de un elevado número de elementos (que puede llegar hasta los cincuenta), que se aplican a una muestra limitada de sujetos (por lo general 100-200); a partir de esta primera prueba se eliminan los elementos incoherentes y se establece

<sup>6</sup> El procedimiento más correcto consiste en calcular para cada elemento la correlación entre su puntuación y la que resulta de sumar (o calcular la media) de todos los elementos *restantes* (correlación «correcta» elemento-escala).

<sup>7</sup> Correlación positiva: al aumentar una variable también aumenta la otra; correlación o negativa: al aumentar una variable disminuye la otra.

#### Resumen 6.4. LA COHERENCIA INTERNA DE LA ESCALA: EL ALFA DE CRONBACH

Se han elaborado algunos criterios para valorar *la coherencia interna global de una escala*. Se trata de índices que resumen esta coherencia, y entre los más conocidos se encuentra el alfa de *Cronbach*, que se basa en la matriz de relación entre todos los elementos y su número. Su fórmula es:

$$\alpha = \frac{nr}{1 + r(n - 1)}$$

donde  $n$  es el número de elementos de la escala, y  $r$  es su correlación media.

Aunque lo parezca, alfa no es un coeficiente de correlación; suele tomar un valor positivo<sup>8</sup>, entre 0 y 1; cuanto más alto es su valor, mayor es la coherencia interna de la escala. Nunnally [1978, 245] sugiere un valor mínimo de 0,70 como criterio de aceptabilidad de la escala. Un alfa inferior a este valor significa que los elementos de la escala tienen poco en común, o que su número es demasiado bajo. Por ejemplo, si los elementos de la escala son 10 y la media de las 45 correlaciones entre los elementos es 0,30, el valor de  $\alpha$  será 0,81 (suficiente). Si los elementos son cinco y la correlación media es la misma, el valor de  $\alpha$  será igual a 0,68 (insuficiente). Alfa aumenta a medida que aumentan el número de elementos de la escala y su correlación media<sup>9</sup>.

rior). La escala se aplica a diversas investigaciones, y en el curso de su utilización se van acumulando pruebas de validez sucesivas (validez predictiva, validez concomitante, validez por grupos conocidos, validez por frase), que permiten comprobar que efectivamente registra la propiedad para la cual ha sido creada<sup>10</sup>.

Este modo de proceder es común en la investigación psicológica, cuando las propiedades que son objeto del estudio están representadas por rasgos de la personalidad, conceptos psicológicos muy delimitados y situados en lo profundo de la psi-

<sup>8</sup> Si el valor de alfa es negativo, quiere decir que hay elementos que están correlacionados negativamente, debido a polaridades no coherentes (por ejemplo, en un caso la puntuación máxima indica un autoritarismo alto y en otro caso indica un autoritarismo bajo). Si todas las puntuaciones van en la misma dirección, alfa tendrá sólo valores positivos.

<sup>9</sup> Esta particularidad matemática lleva a muchos psicométricos a construir y proveer escalas con demasiados elementos para aumentar la probabilidad de que alfa tenga un valor alto. Sin embargo esto aumenta también la probabilidad de que el entrevistado se canse y acabe respondiendo al azar.

<sup>10</sup> Recordemos la definición de «validez» que presentamos en el apartado 9 del capítulo 3 («la validez es la medida en que una determinada forma de traducción de un concepto en variable registra efectivamente el concepto que se pretende registrar»); ver dicho apartado para consultar las formas de control de la validez.



que, que requieren escalas bastante elaboradas, generalmente constituidas por varias decenas de elementos<sup>11</sup>. Los sociólogos y politólogos han adoptado un punto de vista más simplificado en el uso de la técnica de las escalas, que en su caso suelen tener una decena de elementos aproximadamente; además, en el mismo instrumento de registro (que suele ser un cuestionario) pueden coexistir varias escalas; la escala es sólo un aspecto de la investigación.

Un ejemplo de ello sería una investigación sobre actitudes hacia la política compuesta por una escala de «sentimiento de eficacia política», otra de «participación política» y otra de posicionamiento «izquierda-derecha», junto con otras muchas preguntas sobre socialización familiar, educación, profesión, religiosidad, etc. En este caso la escala se construye con una sola operación, sin separación entre «construcción» y «aplicación»; los controles de validez son insuficientes y no sistemáticos, y el ámbito de aplicación de la escala queda limitado a la población específica estudiada.

Pasamos ahora al problema del *control del carácter unidimensional* de la escala. Aunque el procedimiento de *análisis de los elementos* que hemos presentado en la tercera fase tenga como fin establecer que todos los elementos son indicadores de la misma propiedad, éste no es suficiente para garantizar el carácter unidimensional de la escala. Es posible, por ejemplo, que los elementos impliquen dos propiedades distintas, incluso muy distintas (por ejemplo, autoritarismo e ideología de clases), pero correlacionadas entre sí, lo que haría que los valores de las correlaciones elemento-escala y el coeficiente alfa fueran satisfactorios, aunque la escala no tuviera carácter unidimensional.

Una técnica muy eficaz para el control del carácter unidimensional de la escala es el *análisis factorial*. Este procedimiento es demasiado complejo para abordarlo en este manual, pero podemos entender su aplicación para el control de la unidimensionalidad de las escalas sin necesidad de conocerlo en profundidad (véase el resumen 6.5, en la página siguiente).

Para concluir este apartado, diremos que las escalas de Likert son la técnica más difundida en las ciencias sociales para el registro de las propiedades continuas, en particular las actitudes. Su popularidad se atribuye a la simplicidad de su estructura teórica y su facilidad de aplicación, que ni siquiera requiere, en principio, los recursos de cálculo de los ordenadores modernos, lo que explica su enorme éxito en los años cincuenta-setenta.

Esta técnica presenta, sin embargo, algunos inconvenientes. El primero guarda relación con la puntuación que se asigna a cada uno de los elementos. Cada elemento es una variable ordinal, por lo general «con autonomía semántica parcial», que suele presentar cinco opciones de respuesta (desde «totalmente de acuerdo» hasta «totalmente en desacuerdo», y similares), a las que se asignan de manera arbitraria, las puntuaciones en escala simple 1, 2, ..., 5, que se tratan luego como escalas

<sup>11</sup> Los psicométricos trabajan con escalas individuales muy largas (para elevar el valor de alfa) y en el laboratorio; los sociólogos y politólogos incluyen escalas (necesariamente más breves) en cuestionarios que aplican a estudios de campo.

cardinales (por ejemplo, en la correlación elemento-escala). El mismo Likert, consciente de este problema, propuso inicialmente una técnica más elaborada en la que las puntuaciones de las respuestas se asignaban a partir de los resultados de la encuesta («método sigma», Likert, 1932), pero la abandonó, ya que las puntuaciones más simples basadas en la secuencia de los números naturales eran muy parecidas a las calculadas mediante el criterio más sofisticado.

El segundo inconveniente guarda relación con la imposibilidad de *reproducir* la escala, es decir, el hecho de que la puntuación final de la escala es global y no hay manera de conocer a qué combinación de respuestas corresponde una puntuación determinada. Por este motivo puede suceder que dos puntuaciones idénticas corres-

### Resumen 6.5. EL ANÁLISIS FACTORIAL

El objetivo del análisis factorial es reducir una serie de variables relacionadas entre sí a un número inferior de variables hipotéticas (factores o variables latentes) independientes entre sí<sup>12</sup>. El punto de partida para el análisis es la matriz de correlación entre las variables observadas, y el objetivo es explicar estas correlaciones mediante la existencia de factores subyacentes. Por ejemplo, la correlación entre las notas en álgebra y las notas en geometría de un determinado grupo de estudiantes (es decir, el hecho de que, por lo general, si un estudiante obtiene buenas notas en álgebra, también obtiene buenas notas en geometría, y si es mediocre en una de las dos materias, también lo es en la otra), se podría reinterpretar deduciendo que existe un factor subyacente que influye en ambas notas, que en este caso sería la capacidad de abstracción matemática del estudiante.

En la tabla 6.5 hemos reproducido los resultados del análisis factorial realizado sobre los elementos de la escala de la tabla 6.4. Las «saturaciones factoriales» indicadas en la tabla se pueden interpretar como coeficientes de correlación entre el elemento de la escala y el factor subyacente. El análisis, realizado según el procedimiento estándar, muestra la existencia de dos factores: el primero saturado por las cinco primeras preguntas y el segundo por las cuatro últimas. Por tanto, la escala es bidimensional, ya que las propiedades subyacentes son dos. La interpretación de los dos factores es bastante inmediata. El primero está relacionado con las preguntas relativas a la política nacional, el segundo con las relativas a los representantes políticos del Ayuntamiento y del distrito municipal. Por ello, los investigadores decidieron dividir las preguntas en dos escalas de «sentimiento de eficacia política»: uno referido a la política nacional y otro a la política local.

<sup>12</sup> Los factores también pueden estar relacionados entre sí («rotación oblicua»), pero aquí nos ceñiremos a la situación más sencilla.

**Tabla 6.5.** Análisis factorial sobre la serie de indicadores de eficacia política

Elemento n°	Saturaciones	
	Factor n° 1	Factor n° 2
1	0,661	0,295
2	0,702	0,104
3	0,764	0,245
4	0,752	0,224
5	0,679	0,360
6	0,346	0,761
7	0,300	0,811
8	0,210	0,858
9	0,204	0,875

pondan a respuestas muy distintas.

Por último, la puntuación final de la escala no representa una variable cardinal: no disponemos de ningún elemento para poder afirmar, por ejemplo, que la distancia entre las puntuaciones 14 y 18 sea la misma que entre las puntuaciones 8 y 12.

#### 6.4. EL ESCALOGRAMA DE GUTTMAN

La propuesta de Guttman [1944; 1950] nace con el objetivo de ofrecer una solución para el problema del carácter unidimensional de la escala, que es uno de los puntos débiles de la técnica de Likert. La escala de Guttman consiste en una sucesión de elementos que presentan una dificultad creciente, de manera que si se ha respondido afirmativamente a una pregunta, se debe haber respondido también afirmativamente a todas las que la preceden en grado de dificultad. De ahí el nombre de escalograma o escala acumulativa (por contraste con el nombre de «escala aditiva», atribuido a la técnica de Likert, en la que todos los elementos de la escala tienen el mismo grado de dificultad).

Para ilustrar este carácter acumulativo de los elementos utilizaremos como ejemplo la «escala de distancia social» propuesta por Bogardus [1925] en los años veinte. En esta escala, para determinar los prejuicios de los entrevistados con respecto a ciertas minorías étnicas, se planteaba una secuencia de preguntas del tipo: ¿Estaría dispuesto a aceptar a un individuo de color como visitante de su país?, ¿estaría dispuesto a aceptar a un individuo de color como vecino de su casa?, ¿estaría dispuesto a aceptar a un individuo de color como amigo?, ¿estaría dispuesto a casarse con un individuo de color?

Es evidente que quien está dispuesto a casarse con un individuo de color, probablemente también estará dispuesto a tenerlo como amigo, como vecino, etc., mientras que quien no esté dispuesto a aceptarlo como amigo, probablemente tampoco esté dispuesto a casarse con él. Si los elementos de la escala están perfectamente

escalonados, sólo son posibles ciertas secuencias de respuestas. Si asignamos la puntuación 1 a la respuesta positiva y 0 a la negativa, en principio sólo deberían ser posibles las secuencias: 1111, 1110, 1100, 1000, 0000 (ver la tabla 6.6), mientras que, por ejemplo, la secuencia 1011 no debería producirse (sería una persona que está dispuesta a casarse con un individuo de color y a tenerlo como amigo, pero no como vecino). Como puede verse en la tabla 6.6, las respuestas posibles producen una matriz dividida en dos triángulos, uno de los cuales contiene sólo respuestas positivas y el otro sólo respuestas negativas.

Asignando la puntuación 1 a la respuesta afirmativa y 0 a la respuesta negativa y sumando las puntuaciones de cada individuo sobre todos los elementos, obtenemos la puntuación total sobre la escala para cada sujeto. Nótese que a partir de esta puntuación individual se pueden deducir las respuestas dadas por ese individuo a cada uno de los elementos de la escala: por ejemplo, quien haya obtenido una puntuación de 2 en la escala de la tabla 6.6 habrá dado la secuencia de respuestas 1100. Por tanto podemos saber, no sólo a cuántas preguntas, sino también a qué preguntas concretas ha dado una respuesta afirmativa. Esta posibilidad de deducir las respuestas a cada una de las preguntas a partir de la puntuación de la escala es una característica típica del escalograma de Guttman, que también se designa con el término *reproducibilidad* (a partir de la puntuación se pueden «reproducir» las respuestas a cada pregunta). Nótese también que esta técnica sólo utiliza elementos dicotómicos, es decir, preguntas con respuestas del tipo «sí/no», «de acuerdo/en desacuerdo», «a favor de/en contra de» (por contraste con la técnica de Likert, en la que la respuesta se graduaba sobre una escala de intensidad).

*Es obvio que sólo una escala ideal proporciona las secuencias de respuesta que aparecen en la tabla 6.6. En cada aplicación real habrá también respuestas que no estén contempladas en las secuencias previstas en el modelo, y que se considerarán «errores». Se trata de establecer cuántos errores puede tolerar una escala para ser aceptable y cuáles son los procedimientos a seguir para minimizar dichos errores.*

Igual que en el caso de la escala de Likert, en la construcción de la escala de Guttman podemos distinguir tres o cuatro fases: la formulación de las preguntas, su aplicación a una muestra de sujetos, el análisis de los resultados con la eliminación de los elementos que indican demasiados errores, la definición de un índice global de aceptación de la escala, y una eventual fase de validación de la misma.

Para la fase de formulación de las preguntas (dicho de forma más general, la identificación de los elementos de la escala), recordamos lo ya dicho al hablar de la escala de Likert sobre la importancia de la reflexión conceptual. Para la escala de Guttman podemos añadir dos observaciones específicas al respecto. En primer lugar, el hecho de que las respuestas deben darse de forma binaria. En segundo lugar, el hecho de que las preguntas se deben proyectar teniendo en cuenta la forma final de la escala, un conjunto de elementos de intensidad creciente.

A diferencia de la escala de Likert, en la que todas las preguntas podían tener más o menos la misma intensidad, en el caso de las escalas de Guttman éstas deben cubrir toda la extensión del continuo de la actitud subyacente, a fin de obtener una batería de preguntas con dificultad creciente. Por ejemplo, para una escala sobre la actitud progresista-conservadora en política, el investigador deberá formular

**Tabla 6.6.** Respuestas a la pregunta: ¿Estaría dispuesto a aceptar a un individuo de color como visitante de su país, como vecino de casa, como amigo o como cónyuge?

Como visitante	Como vecino	Como amigo	Me casaría con él	Puntuación
1	1	1	1	4
1	1	1	0	3
1	1	0	0	2
1	0	0	0	1
0	0	0	0	0

una serie de afirmaciones variando el campo de aplicación (de la política fiscal, políticas de empleo, valoración de hechos históricos como el fascismo, etc.), y también proponer afirmaciones graduales desde el extremo revolucionario al extremo reaccionario, pasando por todas las posiciones intermedias.

En cuanto a la fase de aplicación a una muestra de sujetos, la escala de Guttman no plantea problemas muy distintos de los que planteaba la escala de Likert. La diferencia más notable reside en el carácter binario de las opciones de respuesta. La verdadera especificidad de las escalas de Guttman se manifiesta en el momento de hacer el análisis de los resultados. Igual que para las escalas de Likert, el objetivo de esta fase es evaluar la escalabilidad de todos los elementos, descartar los menos coherentes con el modelo, establecer un índice de escalabilidad de la escala y establecer su aceptación o no.

El procedimiento consiste, en primer lugar, en identificar los llamados «errores» de la escala, o sea, las respuestas que no se integran en las secuencias previstas en el modelo. Existen dos técnicas distintas para su identificación: una es la que adoptó originalmente Guttman, y la otra se propuso después [Edwards, 1957]. Presentaremos sólo esta última, puesto que se considera la más adecuada.

El procedimiento se desarrolla como sigue. Supongamos que la matriz de los datos inicial sea la que representamos en la tabla 6.7a (para simplificar imaginemos que sólo hemos aplicado la escala a ocho sujetos). Las columnas y las filas de la matriz se reordenan, de manera que las preguntas de la escala estén ordenadas de izquierda a derecha según el número de respuestas positivas recibidas, y los casos estén ordenados de arriba a abajo según la puntuación total recibida. De esta forma se visualiza la estructura de dos triángulos de la matriz (uno compuesto predominantemente por 1 y el otro por 0, como se muestra en la tabla 6.7b).

La tabla 6.7b sugiere que hay dos secuencias de error: el caso D y el caso G. Los errores se identifican comparando la secuencia observada en cada caso con la secuencia teórica correcta que debería obtenerse *con la puntuación total de ese caso concreto*. Consideremos el caso D: la secuencia observada es 10110, con una puntuación total de tres. La secuencia correcta para esa puntuación total debería ser 11100, por lo tanto, hay dos errores (la segunda posición debería ser un 1 y la

cuarta debería ser un 0).

Una vez identificados de este modo los «errores» (que hemos marcado con un asterisco en la tabla 6.7b), se trata de decidir si son lo bastante irrelevantes como para garantizar la «escalabilidad» de las preguntas de la escala. Guttman propuso un índice, que llamó coeficiente de reproducibilidad, que mide el grado de desviación de la escala observada respecto a la escala perfecta. Su fórmula es la siguiente:

$$C_r = 1 - \frac{\text{Número de errores}}{\text{Número total de respuestas}} = 1 - \frac{\text{Número de errores}}{\text{Número de elementos} \times \text{número de casos}} = \frac{\text{Número de respuestas correctas}}{\text{Número total de respuestas}}$$

En nuestro ejemplo:

$$C_r = 1 - \frac{4}{5,8} = 0,90$$

Este índice también se puede interpretar como la proporción de respuestas «correctas» (correspondientes a las secuencias teóricas) sobre el número total de respuestas. Guttman sugirió un valor de  $C_r = 0,90$  (es decir, errores iguales o inferiores al 10 % de las respuestas) para la aceptación de la escala <sup>13</sup>. Si la escala presenta un coeficiente de reproducibilidad inferior a 0,90, se van eliminando los elementos con mayor número de errores, y se va recalculando el índice  $C_r$ .

El coeficiente de reproducibilidad de la escala es la media de los coeficientes de reproducibilidad de cada uno de los elementos (proporción de respuestas correctas / respuestas totales de cada elemento). A partir de ahí se puede demostrar que el coeficiente de reproducibilidad de cada elemento no puede ser inferior a su proporción de respuestas en la categoría modal <sup>14</sup>. Esto implica que el coeficiente puede estar inflado debido a una distribución de las respuestas muy asimétrica.

Por ejemplo, la respuesta modal del elemento 1 de la tabla 6.7 es «no» y su proporción es 0,75 (seis «no» de un total de ocho respuestas), por lo que el coeficiente de reproducibilidad del elemento número 1 no puede descender por debajo de 0,75. Por lo tanto, una escala de Guttman compuesta íntegramente por afirmaciones extremas que provocan respuestas como un 90 % de «sí» frente a un 10 % de «no» (o a la inversa) obtiene automáticamente un  $C_r$  en torno al 90 %.

Una vez eliminados los elementos no escalables y establecida la aceptación de la

<sup>13</sup> El número de errores depende de la forma en que se calculan éstos, por lo que el umbral de aceptación de  $C_r = 0,90$  resulta ser un criterio más severo si el para calcular estos errores utilizamos el procedimiento de Goodenough-Edwards.

<sup>14</sup> La categoría modal (o «moda») de una variable nominal es la que presenta mayor frecuencia. Si una variable presenta el 40 % de respuestas «sí» y el 60 % de respuestas «no», la moda es la categoría «no» y su proporción es de 0,60; si las respuestas han sido del 90% y el 10%, la proporción de la categoría modal es 0,90.

Tabla 6.7. Identificación de los errores de una escala de Guttman

Casos	Elementos					Puntuación
	n 1	n 2	n 3	n 4	n 5	
a) matriz originaria						
A	1	1	1	1	1	5
B	0	0	0	0	1	1
C	0	0	1	0	1	2
D	0	1	0	1	1	3
E	0	1	1	0	1	3
F	0	0	0	0	0	0
G	1	1	1	0	0	2
H	0	1	1	1	1	4
N.º de respuestas afirmativas	2	4	5	3	6	
	n 5	n 3	n 2	n 4	n 1	
b) matriz ordenada						
A	1	1	1	1	1	5
H	1	1	1	1	0	4
E	1	1	1	0	0	3
D	1	0*	1	1*	0	3
C	1	1	0	0	0	2
G	0*	1	0	0	1*	2
B	1	0	0	0	0	1
F	0	0	0	0	0	0
N.º de errores	1	1	0	1	1	

\* Error

escala, se trata de atribuir puntuaciones a los sujetos. Para los sujetos que presentan una secuencia de respuestas acorde con el modelo no hay ningún problema: la puntuación total de la escala para cada sujeto es la suma de las puntuaciones 0/1 obtenidas en cada elemento. Aunque el sujeto presente una secuencia con errores (por ejemplo, 01010), se respeta el mismo criterio si se ha seguido la técnica de determi-

nación de los errores aquí presentada: se considera correcta la secuencia aceptada por el modelo que tenga la misma puntuación final (en el ejemplo: 11000), y por lo tanto la puntuación del sujeto se mantiene igual a la suma de las respuestas positivas que ha dado, aunque presente errores. En este punto, el proceso de construcción de la escala se considera concluido.

Se han realizado diversas sugerencias para mejorar el rendimiento de esta técnica<sup>15</sup>. Aquí vamos a presentar tres de ellas, recomendadas por el mismo Guttman. La primera es evitar preguntas con tasas demasiado elevadas (superiores al 80 %) de aceptación o rechazo, ya que son poco discriminantes y elevan de manera artificial el coeficiente de reproducibilidad de la escala. En segundo lugar, Guttman sugirió

#### Resumen 6.6. MEDIDAS DE REPRODUCIBILIDAD

Edwards [1957] propuso calcular un *índice de reproducibilidad marginal mínima* (*MMR, minimal marginal reproducibility*) con la siguiente fórmula:

$$MMR = \frac{\sum \text{proporción de respuestas en la categoría modal}}{N}$$

donde *N* es el número de elementos de la escala

En nuestro ejemplo de la tabla 6.7:

$$MMR = \frac{0,75 + 0,50 + 0,63 + 0,63 + 0,75}{5} = 0,65$$

Este índice indica el valor mínimo que puede adoptar el coeficiente de reproducibilidad de la escala, cualesquiera que sean las secuencias de las respuestas. Se sugiere, por tanto, comparar el *C*, con el *MMR*: sólo si el primero es superior a 0,90 y también claramente superior al segundo se puede afirmar que la adecuada reproducibilidad de la escala se debe a una escalabilidad efectiva de sus elementos y no a la distribución marginal de las respuestas. En nuestro ejemplo se cumple este requisito.

<sup>15</sup> Señalemos que la técnica de Guttman también puede emplearse con respuestas politómicas, transformándolas en dicotómicas mediante la agregación de las categorías (el punto de corte puede elegirse de tal modo que se maximice la escalabilidad de la escala; esto se consigue, por ejemplo, unificando las respuestas «muy de acuerdo» y «bastante de acuerdo» para ciertos elementos, y contrastando los «muy de acuerdo» con el resto de las respuestas para otros elementos).



emplear un número lo bastante elevado de preguntas, mostrando que una escala con sólo cuatro preguntas puede producir valores elevados de  $C$ , aunque las preguntas sean estadísticamente independientes entre sí. El mismo Guttman recomendó estudiar atentamente las secuencias erróneas, pues la presencia repetida de una secuencia no prevista en el modelo puede ser síntoma de otra dimensión subyacente.

Podemos concluir diciendo que el escalograma de Guttman ha supuesto un avance muy importante en el desarrollo de la técnica de las escalas. En el debate en torno a éste que tuvo lugar durante al menos treinta años, se propusieron numerosas sugerencias para su mejora [cfr. McIver y Carmines, 1981]. No obstante, sigue presentando algunos aspectos problemáticos. En primer lugar, la puntuación final que se obtiene sobre la variable latente sigue siendo fundamentalmente, como sucedía en las escalas de Likert, una variable ordinal, y tampoco en este caso tenemos ningún elemento para afirmar, como sucedía con la participación política, que la distancia entre las puntuaciones 2 y 4 es igual que la distancia entre las puntuaciones 5 y 7.

En segundo lugar, la técnica de Guttman es aplicable cuando estamos ante actitudes (comportamientos, acontecimientos...) bien definidas y escalables. Cuando la actitud se hace compleja resulta difícil escalonarla en secuencias acumulativas, las categorías se pueden superponer, por lo que al final hay demasiados errores.

Por último, el modelo es determinista e inflexible, pero la realidad social sólo puede interpretarse correctamente mediante modelos probabilísticos que contemplen la posibilidad de error y el paso gradual entre las distintas posiciones. Ésta es la crítica más importante al modelo de Guttman. De hecho, su escala se considera ya superada por las escalas de tipo probabilístico, como las escalas de Rasch.

## 6.5. DIFERENCIAL SEMÁNTICO

La técnica del diferencial semántico fue desarrollada en los años cincuenta por el psicólogo estadounidense Charles Osgood y sus colegas [Osgood, 1952; Osgood, Suci y Tannenbaum, 1957]. El objetivo de estos investigadores era crear un instrumento para determinar, con la máxima estandarización, el significado que los conceptos adquieren para los individuos, un instrumento capaz responder a la pregunta: «¿qué significa esto (un concepto, como madre, patria, guerra, etc.) para usted?».

El problema del «significado» que un término, una situación social, un objeto cognitivo en general adquiere para un individuo es complejo, está cargado de elementos filosóficos y psicológicos y tiene una enorme importancia para el estudio de los comportamientos. Como escribe Nunnally, «el comportamiento humano está determinado por el significado de los acontecimientos más que por sus propiedades intrínsecas; el niño reacciona positivamente a la voz de su madre porque para él significa alimento, calor, protección» [1959, 383].

Osgood y sus colegas lo plantearon como sigue: «Cuando queremos descubrir lo que significa para cierta persona una cosa determinada, le pedimos que nos lo diga [...]. [Podemos preguntar]: ¿qué significa para usted la palabra «sofisticado»? La respuesta podría ser: sé lo que significa, aunque me resulte difícil decirlo [...]; significa ser inteligente y conocer las cosas y las personas [...], saber mucho. Es una

forma de ser relajada y refinada, llena de gracia y ligereza [...], equilibrada [...], quiero decir, tener *savoir faire*» [Osgood, Suci y Tannenbaum, 1957, 18].

Cuando los sujetos estudiados son personas inteligentes, con formación y con excelentes dotes de expresión verbal, esta forma de conocer los significados a partir de la descripción que ellos mismos nos ofrecen, puede ser válida y lo bastante sensible: estas personas son capaces de expresar, mediante una descripción verbal, los elementos distintivos y determinantes de lo que ese concepto significa para ellos. Pero con individuos que no tienen esta capacidad de expresión verbal, este sistema puede ser muy limitado: las descripciones emitidas de manera espontánea pueden resultar demasiado pobres. Además, resulta difícil comparar entre sí todas las descripciones obtenidas con estos procedimientos tan subjetivos.

Con el diferencial semántico Osgood propone una técnica para extraer los significados que determinados conceptos adquieren para los entrevistados, que, en lugar de basarse en la descripción subjetiva y directa del significado por parte del entrevistado, se basa en las asociaciones que establece el entrevistado entre este concepto y otros propuestos de forma estandarizada a todos los entrevistados. En lugar de preguntar: ¿qué significa para usted el término «sofisticado»? se plantea una serie de preguntas como: ¿«sofisticado» es duro o blando?, ¿agradable o desagradable?, ¿rápido o lento? Y así sucesivamente. Y para mejorar la sensibilidad del instrumento se pide que se clasifique la intensidad de cada opinión en una escala de siete puntos. «De esta forma el diferencial semántico es básicamente una combinación de asociaciones controladas y *escalas de intensidad*» [*ibidem*, 20].

Para explicar el formato de la pregunta, consideremos el ejemplo presentado en la figura 6.1, que procede de una investigación realizada en Italia sobre los valores de los jóvenes, en la que se pidió a los entrevistados que evaluaran a cuatro partidos políticos [Sciolla y Ricolfi, 1989]. Esta técnica emplea una serie de escalas de intervalos de autoposicionamiento típicas, en las que las categorías extremas tienen significado autónomo, mientras que el significado (gradual) de las categorías intermedias lo establece el entrevistado por comparación. En la tabla 6.8 hemos reproducido 50 parejas de atributos polares, empleadas por Osgood y sus colegas en sus investigaciones. Ellos aplicaron estas escalas a personas (extranjero, madre, yo mismo, políticos, etc.), objetos concretos (cuchillo, nieve, motor, etc.), objetos abstractos (arte moderno, pecado, liderazgo, etc.), acontecimientos (debate, nacimiento, etc.), e instituciones (Naciones Unidas, hospital, familia, etc.).

Osgood sostenía que la lista de los atributos bipolares no debe tener relación con el objeto evaluado, y por tanto debe ser siempre la misma. Incluso ciertos adjetivos que no tienen nada que ver con el objeto en cuestión resultan útiles para extraer significados más profundos o inconscientes que tiene el objeto para el entrevistado<sup>16</sup>.

El test se puede realizar fácilmente y suele ser bien aceptado por los entrevistados, incluso cuando las parejas de atributos parecen poco pertinentes en relación

<sup>16</sup> Sin embargo el propio Osgood modificó la lista para adaptarla a cada investigación. Además, la técnica se ha aplicado muchas veces utilizando atributos relacionados específicamente con el objeto a evaluar, junto con las parejas tradicionales.

con el objeto al que se aplican; en general, se recomienda al entrevistado que conteste de forma instintiva, sin reflexionar demasiado sobre cada pareja.

En cuanto al número de adjetivos bipolares propuestos para un mismo objeto, la técnica es muy flexible. En los dos ejemplos que hemos presentado abarca desde los 12 en la investigación sobre partidos políticos hasta los 50 propuestos inicialmente por Osgood. De forma análoga, los objetos sometidos a evaluación pueden ser muchos (por ejemplo, una serie de 30 personajes públicos del mundo del espectáculo, de la política, de la economía, etc.), o pocos (por ejemplo, los cuatro partidos políticos que hemos mencionado). La elección dependerá del objetivo del estudio y la relevancia que tenga la técnica en el marco general de la investigación.

En la aplicación a los partidos políticos, el diferencial semántico se incluía dentro de un cuestionario que utilizaba otros instrumentos para revelar las actitudes, desde preguntas tradicionales hasta escalas Likert. De ahí su uso limitado, en cuanto al número de objetos evaluados y el número de parejas de atributos empleadas. Osgood y sus colegas citan otras aplicaciones, en el campo del estudio de la personalidad, basadas totalmente en la técnica del diferencial semántico, y en las que el número de juicios solicitados (que la cifra que resulta de multiplicar el número de atributos bipolares y el número de objetos a evaluar) puede llegar hasta 400 (por ejemplo, 40 parejas polares y diez objetos a evaluar).

Las respuestas obtenidas se pueden someter a diversas operaciones, que aquí sólo citaremos brevemente. Para describir de manera sintética la imagen del objeto

**Pregunta:** «Hablemos ahora de los partidos italianos. Debe decirnos su opinión de cada uno de ellos, tratando de situarlo en cada una de las escalas siguientes» (pregunta aplicada a los partidos Socialista, Comunista, Democracia Cristiana y Verdes [Sciolla y Ricolfi, 1989]).

Partido Socialista	
	1    2    3    4    5    6    7
Viejo	_____
Fuerte	_____
Limpio	_____
Rápido	_____
Cercano	_____
Intolerante	_____
Abierto	_____
Intransigente	_____
Ganador	_____
De izquierdas	_____
Alegre	_____
Ineficiente	_____
Joven	_____
Débil	_____
Sucio	_____
Lento	_____
Lejano	_____
Tolerante	_____
Cerrado	_____
Indulgente	_____
Perdedor	_____
De derechas	_____
Triste	_____
Eficiente	_____

**Notas para el entrevistador:**

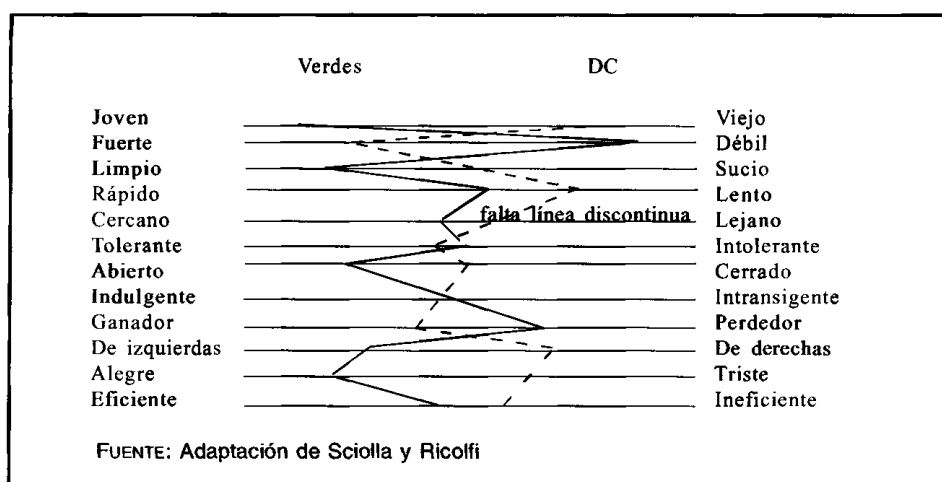
- mostrar la ficha al entrevistado y pedirle que ponga una cruz sobre la posición que se corresponda mejor con su imagen del partido para cada una de las escalas indicadas;
- es mejor que las respuestas se den instintivamente, sin reflexionar demasiado tiempo en cada escala.

**Figura 6.1.** Ejemplo de aplicación del diferencial semántico

sometido a valoración, se puede emplear el «perfil». Este perfil es una representación gráfica que se construye uniendo con una línea la puntuación media obtenida por ese objeto en cada pareja de atributos. En la figura 6.2 hemos reproducido los perfiles de los Verdes y la Democracia Cristiana de la aplicación del diferencial

**Tabla 6.8.** Diferencial semántico. Parejas de atributos empleadas por Osgood, Suci y Tannenbaum [1957]

1. Bueno-malo	18. Agradable-desagradable	35. Simpático-antipático
2. Grande-pequeño	19. Negro-blanco	36. Brillante-apagado
3. Bonito-feo	20. Amargo-dulce	37. Grave-agudo
4. Amarillo-azul	21. Feliz-triste	38. Anguloso-circular
5. Duro-blando	22. Gracioso-soso	39. Perfumado-hediondo
6. Dulce-amargo	23. Vacío-lleno	40. Honesto-deshonesto
7. Fuerte-débil	24. Agresivo-pacífico	41. Activo-pasivo
8. Limpio-sucio	25. Pesado-ligero	42. Rugoso-liso
9. Alto-bajo	26. Mojado-seco	43. Fresco-rancio
10. Tranquilo-inquieto	27. Sagrado-profano	44. Rápido-lento
11. Sabroso-insípido	28. Relajado-tenso	45. Justo-injusto
12. Valioso-sin valor	29. Valiente-cobarde	46. Áspero-suave
13. Rojo-verde	30. Alto-bajo	47. Cercano-lejano
14. Joven-viejo	31. Rico-pobre	48. Agitado-apacible
15. Amable-grosero	32. Luminoso-oscuro	49. Sano-enfermo
16. En voz alta-baja	33. Caliente-frío	50. Ancho-estrecho
17. Profundo-superficial	34. Espeso-ligero	



**Figura 6.2.** Ejemplos de «perfiles» en el diferencial semántico

semántico a los partidos políticos mencionada antes; para facilitar la lectura del gráfico hemos indicado a la izquierda la parte «positiva» de cada pareja de atributos. La imagen de la DC que obtenemos de estos perfiles es la de un partido «viejo, de derechas, ineficiente, sucio, cerrado, lento, triste», mientras que los Verdes son «jóvenes, rápidos, alegres, cercanos, limpios, abiertos, tolerantes, débiles y perdedores»; en la misma investigación el PCI resultó ser «eficiente, de izquierdas, severo y poco tolerante», mientras que el partido socialista se percibía como un partido «indulgente y lejano» [Sciolla y Ricolfi, 1989, 109-110].

En este caso, el perfil está centrado en el concepto y consiste en las valoraciones medias de los entrevistados, pero también se pueden obtener perfiles individuales, es decir, de los sujetos entrevistados, basados en las respuestas de cada uno de los individuos. Por ejemplo, la técnica se ha empleado como instrumento para conocer el nivel de autoestima del entrevistado, pidiéndole que coloque en una batería de atributos bipolares primero el objeto cognitivo «yo, tal como soy» y luego «yo, tal como me gustaría ser». La distancia entre ambos perfiles permite conocer el nivel de autoestima del sujeto y, para citar una aplicación de la técnica, la evolución en el tiempo de esta distancia puede emplearse para seguir el avance de un paciente durante un tratamiento con psicoterapia.

Sin embargo, la aplicación más importante del diferencial semántico es la exploración de la dimensión de los significados. Si pedimos a un entrevistado que valore un determinado objeto cognitivo a partir de 30 parejas de adjetivos, obviamente las distintas valoraciones estarán guiadas por la visión de conjunto (el «significado») que el entrevistado tiene del objeto en cuestión. La condición es que esta visión no sea unidimensional, sino que presente diversos aspectos que contribuyan al significado global de ese objeto. Mediante el análisis factorial se puede determinar cuáles son las dimensiones básicas que subyacen a las valoraciones de una determinada muestra de individuos entrevistados.

Utilizando la serie de 50 adjetivos bipolares (los que aparecen en la tabla 6.8) para una serie desigual de objetos cognitivos y repitiendo la encuesta sobre muestras distintas de entrevistados, Osgood y sus colegas constataron que, en líneas generales, pueden identificarse tres dimensiones básicas que subyacen a las diversas opiniones: la *evaluación* (relacionada con las parejas bueno-malo, positivo-negativo, bello-feo, justo-injusto, valioso-sin valor, etc.), la *potencia* (duro-blando, fuerte-débil, pesado-ligero, etc.), y la *actividad* (activo-pasivo, rápido-lento, caliente-frío, etc.). De estas tres dimensiones la primera es la más relevante, y parece representar lo que en las técnicas de escalas tradicionales se solía denominar la «actitud» (favorable o desfavorable) hacia un determinado objeto cognitivo.

A partir de esta clasificación de las dimensiones del espacio semántico, pueden sumarse<sup>17</sup> los valores para cada atributo bipolar (según la dimensión a la que pertenece el atributo), obteniendo así para cada individuo unas puntuaciones globales sobre las tres dimensiones mencionadas, como sucedería con una escala aditiva de

<sup>17</sup> Se tratará de una suma ponderada, utilizando las puntuaciones factoriales obtenidas del análisis factorial.

tipo Likert<sup>18</sup>. La puntuación sobre la dimensión de «evaluación» es la que mejor se presta a su comparación con las puntuaciones obtenidas con otras técnicas de escalas: Osgood y sus colegas encontraron importantes correlaciones entre el factor «evaluación» del diferencial semántico y las puntuaciones obtenidas con las escalas de Thurstone (ambas técnicas se utilizaron para estudiar las actitudes en relación con la Iglesia, la pena capital y las personas de color), y con las escalas de Guttman.

Osgood y sus colegas afirmaban que las tres dimensiones de evaluación, potencia y actividad rigen el espacio semántico de los individuos en la gran mayoría de los casos, independientemente de la cultura de los sujetos (realizaron estudios en varios países) y de los conceptos sometidos a evaluación. Muchos investigadores no comparten esta afirmación generalizadora, y algunas investigaciones han encontrado dimensiones que no coinciden con ninguna de las tres propuestas por Osgood.

Aunque estas objeciones puedan contradecir una cierta teorización sobre el espacio semántico que Osgood consideraba tan importante, no restan interés a su técnica como instrumento para revelar la *estructura* de las actitudes. Probablemente la contribución más original del diferencial semántico al reconocimiento de las actitudes sea haber desvelado las múltiples dimensiones de los significados, al introducir la riqueza del punto de vista multidimensional en un campo que hasta entonces había aplicado una perspectiva unidimensional (volveremos sobre esto al final del capítulo).

## 6.6. EL TEST SOCIOMÉTRICO

La técnica que presentamos en este apartado tiene un objetivo bien distinto del que tienen en común las técnicas presentadas hasta ahora. De hecho, no tiene como finalidad revelar variables latentes como la religiosidad o la depresión, ni actitudes con respecto a determinados objetos cognitivos como los partidos políticos o los inmigrantes. Las técnicas sociométricas nacieron con el fin de revelar las relaciones interpersonales existentes dentro de un grupo de individuos. Tal vez convenga recordar que el objeto de este capítulo es presentar las principales técnicas de las escalas, es decir, los procedimientos elaborados por las ciencias del comportamiento, sobre todo entre los años treinta y sesenta, con el objetivo de «medir al hombre y a la sociedad». Y aunque no han cumplido ese objetivo de «medición», han producido técnicas que siguen siendo válidas hoy día para el estudio del hombre y sus comportamientos.

El test sociométrico fue propuesto entre 1940 y 1950 por Jacob Moreno [1953], un académico multidisciplinar, médico, terapeuta y sociólogo, de origen rumano, pero estadounidense de adopción, y nace con el objetivo de estudiar de manera sistemática y con procedimientos estandarizados, las relaciones interpersonales que se

<sup>18</sup> Con la diferencia de que en la escala de Likert las respuestas tienen autonomía semántica parcial (como: totalmente, bastante, poco, en absoluto de acuerdo), mientras que en este caso las respuestas están representadas por gradaciones de intervalos (de autopoicionamiento).

establecen dentro de grupos pequeños (o relativamente pequeños). Su campo de aplicación ideal es una clase escolar <sup>19</sup>. George Bastin comienza su libro sobre esta técnica con la descripción de un experimento muy sencillo. Escogió doce clases de primero de enseñanza primaria de un centro escolar, y tres meses después del inicio del curso entrevistó a cuatro profesores por grupo, de los que daban clase todos los días, pidiéndoles que describieran las relaciones que se habían establecido entre los alumnos: las amistades, los subgrupos, los estudiantes aislados, los líderes, etc. Los profesores no fueron capaces de desentrañar las redes de relaciones existentes entre sus alumnos. Simultáneamente Bastin aplicó un cuestionario sociométrico a las clases y éste reveló con nitidez el mapa de las relaciones entre los chicos, mostrando la existencia de dinámicas que los educadores no habían percibido.

En su formulación más sencilla, el test sociométrico consiste en un cuestionario compuesto por unas cuantas preguntas, que giran en torno al tema de la preferencia o rechazo por el resto de los individuos de un grupo. Siguiendo con el ejemplo escolar, las cuatro preguntas siguientes planteadas a los alumnos de la clase son suficientes para realizar un test sociométrico completo [Bastin, 1961]:

1. ¿De tus compañeros actuales, cuáles te gustaría volver a tener en tu clase el año que viene? Indica todos los que quieras, en orden de preferencia.
2. ¿Y cuáles son los compañeros que no te gustaría volver a tener en tu clase el año que viene? Indica todos los que quieras, ordenados empezando por el que menos te gustaría tener de compañero el año próximo.
3. Adivina cuáles de tus compañeros han dicho que les gustaría estar contigo en clase el año que viene.
4. Adivina cuáles de tus compañeros han dicho que no les gustaría estar contigo en clase el año que viene.

En este ejemplo la encuesta se realizó con motivo del fin del curso escolar y el paso a un nuevo ciclo que da lugar a la formación de nuevas clases, pero las preguntas pueden girar en torno a temas similares: los compañeros con los que te gustaría (no te gustaría) jugar, estudiar, formar un grupo de trabajo, ir al cine, a los que querías (no querías) invitar a tu casa, a tu fiesta de cumpleaños, etc.

La elaboración de los datos es tan sencilla como la aplicación del cuestionario. Se construye una matriz cuadrada, en cuyas partes izquierda y superior se sitúa la lista de los componentes del grupo, y se registran las preferencias o rechazos en las casillas correspondientes a cada pareja.

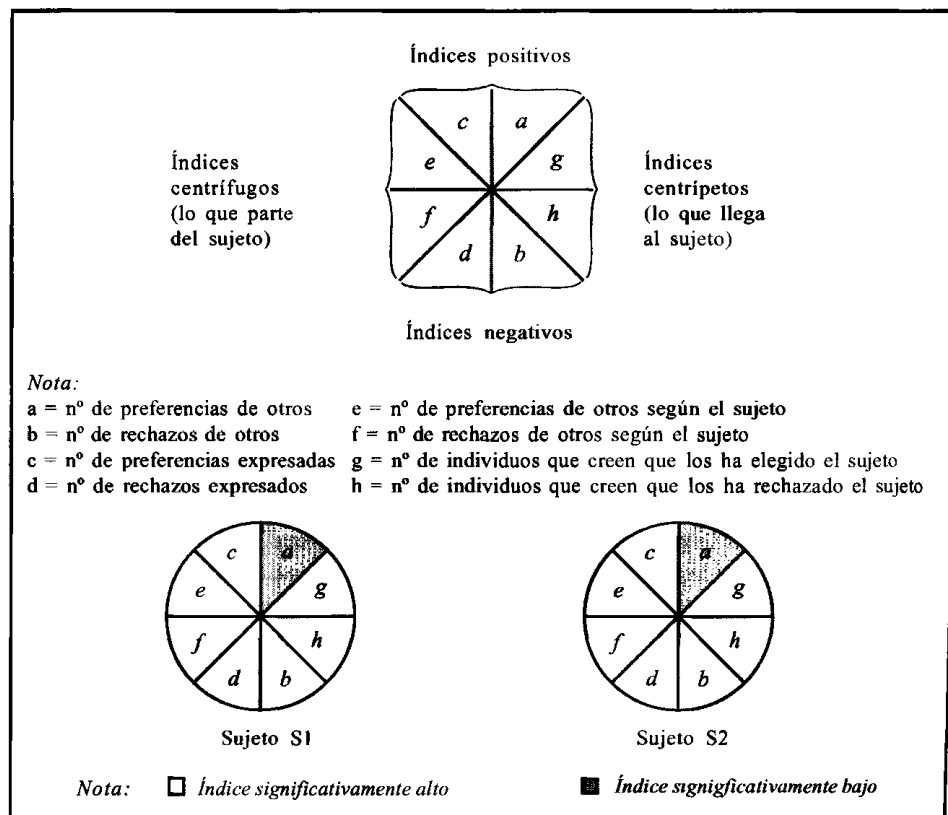
La primera operación que realiza con los datos es el análisis del *estatus socio-*

<sup>19</sup> La figura de Moreno es bastante singular: unió su vocación de investigador a la de profeta y cautivador para fundar un movimiento orientado a la construcción de una sociedad utópica basada en la cooperación, la solidaridad y la libertad. Con el nombre de «sociometría» se suele designar la totalidad de propuestas de Moreno, entre ellas, las técnicas del psicodrama y el sociodrama, orientadas a la liberación de la creatividad y la espontaneidad, con objetivos fundamentalmente terapéuticos. Aquí nos limitaremos a estudiar la técnica propuesta por Moreno para el estudio de las relaciones existentes dentro de los grupos, conocida como «test sociométrico».

*métrico individual*, que consiste en analizar para cada individuo el número de preferencias emitidas y el número de rechazos recibidos. En función de las preguntas arriba mencionadas pueden elaborarse otros índices básicos, que nos ofrecerán un retrato psicosocial del sujeto.

En la figura 6.3 hemos reproducido estos índices y el gráfico para dos sujetos. Hemos tomado estos dos casos aislados, porque hubo muy pocos compañeros que los eligieron (el índice *a* fue muy bajo) y muchos que los rechazaron (el índice *b* fue muy alto); además, también hubo muchos compañeros que se consideraban rechazados por ellos (índice *h* alto). En esta situación, el sujeto *S*<sub>1</sub> percibe su posición de excluido (índice *f* de rechazos que cree haber recibido alto) y reacciona expresando a su vez muchos rechazos (índice *d* alto), mientras que el sujeto *S*<sub>2</sub> parece no ser consciente de esta situación de aislamiento (índice *f* no significativo) y no reacciona ante ello (índice *d* de rechazos no significativo).

Además de los análisis para definir la situación de los individuos en el aspecto

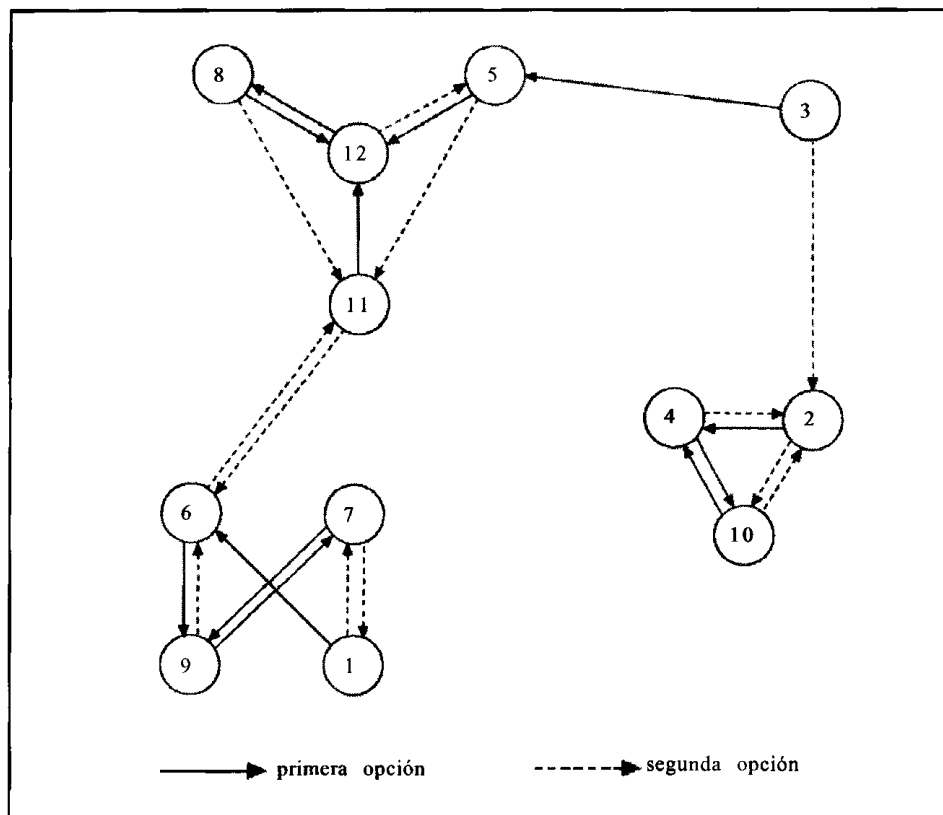


FUENTE: Bastin [1961, 44]

**Figura 6.3.** Índices sociométricos individuales.



de las relaciones con los demás, se realizan operaciones para ilustrar la estructura sociométrica del grupo. La figura 6.4 muestra una sencilla representación o sociograma de las dos primeras preferencias en un grupo de doce estudiantes [Nunnaly, 1959, 392]. Como puede verse en el gráfico, el grupo se organiza en tres subgrupos (todas las primeras opciones de los sujetos y casi todas las segundas se agotan en cada subgrupo), más una persona aislada (3). El grupo formado por los sujetos 5, 8, 11 y 12 presenta un líder (el sujeto 12, que es la primera opción de todos los demás), mientras que los otros dos grupos presentan una estructura más paritaria (aunque las diferencias entre los sujetos son claras: por ejemplo, en el grupo 1, 6, 7 y 9 el sujeto 1 está más aislado que los otros, ya que sólo es una segunda opción). Dos de los tres grupos están conectados entre sí por los sujetos 6 y 11, que funcionan como intermediarios, mientras que el tercero (2, 4 y 10) es un grupo cerrado, ya que todas las preferencias quedan dentro del mismo. El estado de aislamiento y desorientación del sujeto 3 es aun más evidente por el hecho de que, además de no



FUENTE: Nunnaly [1959, 392].

Figura 6.4. Sociograma de las dos primeras preferencias en un grupo de 12 estudiantes.

ser escogido por ningún otro sujeto, él escoge a sujetos pertenecientes a grupos diferentes.

Moreno y los investigadores que han seguido el planteamiento sociométrico propusieron otras muchas formas de representar gráficamente la estructura del grupo y la condición de los individuos, como diagramas de círculos concéntricos en cuyo centro se encuentran los individuos más populares, y hacia el exterior los individuos más aislados. También se pueden construir sociogramas distintos según las preguntas que se consideren (en la figura 6.4 hemos visualizado sólo las preguntas sobre elecciones de sujetos, pero también están los rechazos, las elecciones y rechazos supuestos, las elecciones y los rechazos recíprocos, etc.).

Para un estudio más detallado de este tema se recomienda consultar la literatura especializada. Aquí nos limitaremos a decir que la técnica resulta útil como instrumento de *diagnóstico individual*, con el fin de identificar las relaciones de dominación-dependencia, de aislamiento, y de afinidad (simpatía, amistad) y conflicto (antipatías, enemistades), como instrumento para revelar la *estructura relacional del grupo*, las redes de comunicación, la organización jerárquica informal, las vías por las cuales discurre la información, los comentarios personales, las órdenes, etc., y como herramienta para *estudiar la psicología de los grupos*, identificar los puntos de tensión, las estratificaciones sociales, las barreras étnicas, religiosas, lingüísticas, de género, de edad, etc.

En general, el test sociométrico es adecuado para el estudio de grupos estructurados y menos apropiado para el análisis de grupos informales (los amigos del bar o los jóvenes que juegan juntos en la playa), puesto que el radio de elección del sujeto debe estar limitado. Ya hemos citado su gran utilidad para el estudio de las dinámicas en las clases escolares. Otro campo de aplicación ha sido el laboral (departamentos de fábricas, organización empresarial); también podemos mencionar estudios sobre la organización militar, donde podría analizarse, por ejemplo, si la jerarquía formal se corresponde con la real, si las órdenes siguen el canal de comunicación jerárquico o uno alternativo, etc.

Con respecto a estas últimas aplicaciones, originalmente la técnica se creó para estudiar las relaciones afectivas existentes dentro de los grupos, es decir, criterios de elección de tipo «afectivo», basados en la simpatía, amistad, admiración, con preguntas como: ¿con quién te gustaría jugar?, ¿a quién invitarías a cenar? Posteriormente la técnica se ha aplicado también al estudio de las relaciones funcionales, es decir, criterios de elección de tipo «funcional», basados en la estimación de las capacidades, las competencias y la utilidad, con preguntas como: ¿con quién querías estudiar?, ¿a quién elegirías para tu equipo de fútbol?, ¿a quién querías tener en tu equipo de trabajo? Por último, esta técnica también se ha utilizado, aunque con menos aplicaciones, para estudiar los procesos comunicativos o «diagramas de interacción» dentro de las organizaciones formales.

La técnica propuesta por Moreno fue muy popular en los años cincuenta y sesenta, periodo en el que se aplicaba en múltiples campos. Poco a poco su utilización para la investigación se fue reduciendo, en parte por la disgregación de la construcción teórica de la «sociometría», elaborada por Moreno con un fin de renovación social más que de investigación científica. Más recientemente, el estudio sistemáti-

co de las relaciones sociales ha cobrado vigor en el marco del análisis de redes sociales, que recibió un fuerte impulso gracias al desarrollo de la informática y las nuevas posibilidades que ofrecen los ordenadores para la gestión y el análisis de datos relacionales [cfr. Knoke y Kulinski 1982; Wasserman y Faust 1994].

## 6.7. ESCALAS UNIDIMENSIONALES Y ESCALAS MULTIDIMENSIONALES

En los dos últimos apartados nos hemos alejado en parte del tema central de este capítulo, la técnica de las escalas para la operacionalización empírica de conceptos complejos. En los primeros apartados hemos esbozado el camino que recorrieron las ciencias sociales en sus intentos de «medir» al hombre y su mente, pero hemos hablado sólo de las escalas denominadas «unidimensionales», es decir, de las situaciones de investigación en las que se presume que existe una sola dimensión fundamental subyacente a un conjunto de observaciones.

El razonamiento utilizado ha sido el siguiente: queremos registrar un concepto complejo o latente como la alienación o el autoritarismo, ¿cómo podemos hacerlo? A partir de consideraciones puramente teóricas, los investigadores elegimos unas expresiones explícitas (aunque parciales) de este concepto latente, que formulamos como opiniones, las presentamos a los sujetos entrevistados, verificamos que estas variables también estén relacionadas entre sí en las respuestas de los entrevistados (para comprobar que están determinadas por una única propiedad subyacente), y utilizamos las respuestas de cada individuo para operacionalizar el concepto latente.

El punto de partida en esta secuencia es el concepto general que se quiere registrar, y para operacionalizarlo se identifican unos indicadores. Pero hay otra forma de abordar el problema de los conceptos complejos subyacentes a un conjunto de observaciones, que consiste en partir directamente de las observaciones, preguntándonos *cuántas y qué* dimensiones latentes subyacen a ellas. Ésta es la técnica de las denominadas *escalas multidimensionales*.

Consideremos, por ejemplo, la investigación realizada en 1968 por Weisberg y Rusk [1970] en torno a la opinión de los electores sobre los candidatos a las elecciones presidenciales estadounidenses de ese año. Las preferencias de los entrevistados se registraron con el «termómetro de los sentimientos», aplicado a doce candidatos a las elecciones primarias. Los autores utilizaron la técnica de las escalas multidimensionales para identificar las dimensiones que subyacían a la opinión expresada.

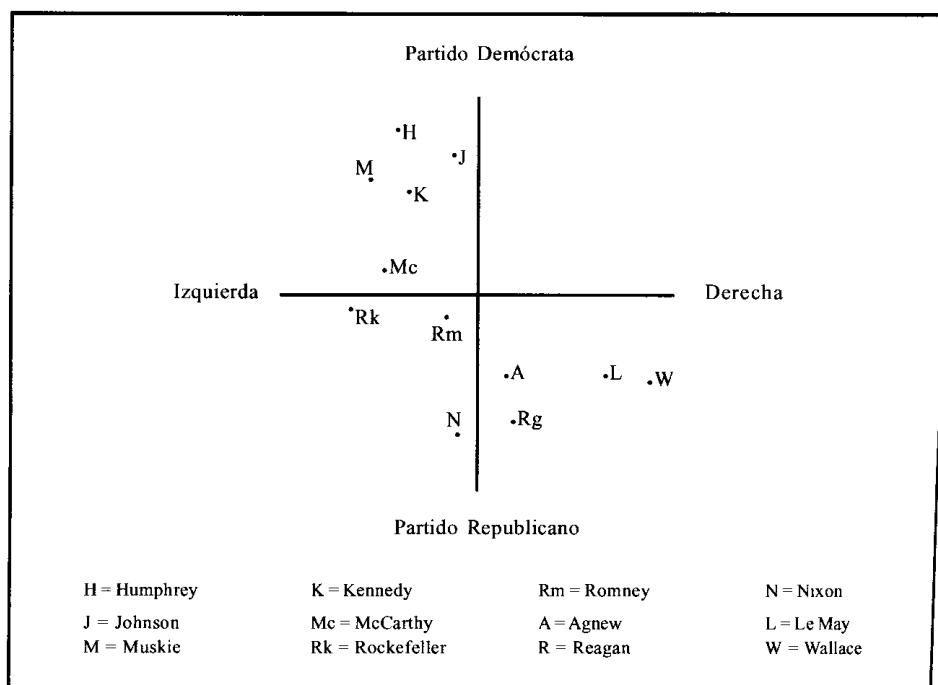
Estas técnicas de análisis interpretan las correlaciones entre los estímulos (o elementos, en nuestra terminología), como medidas de *proximidad*. Si, por ejemplo, existe una alta correlación entre las puntuaciones obtenidas por los candidatos Humphrey y Johnson, eso significa que los dos candidatos están cercanos en el espacio mental de los entrevistados: si un sujeto valora positivamente a uno, también valorará positivamente al otro. Este espacio conceptual puede imaginarse como un espacio físico definido por dimensiones (así como en geometría el plano se define con dos dimensiones, el espacio con tres dimensiones, y se pueden imaginar espacios de  $n$  dimensiones).

Kruskal y Wish [1978, 7] ofrecen un interesante ejemplo de ello. Imaginemos que tenemos un mapa geográfico y que nos piden que construyamos una tabla con las distancias en kilómetros entre las ciudades. La tarea es fácil de realizar midiendo con una regla las distancias sobre el mapa y convirtiéndolas luego en kilómetros. Supongamos que estamos ante el problema contrario: tenemos la tabla con las distancias en kilómetros y queremos reconstruir las posiciones de las ciudades sobre el mapa. Esta tarea es más difícil que la anterior, pero existen procedimientos geométricos para realizarla.

El problema de la técnica de las escalas multidimensionales es similar al proceso de construcción del mapa a partir de las distancias. A partir de las proximidades y distancias entre determinados objetos (en nuestro ejemplo, los candidatos) debemos reconstruir el espacio conceptual en el que se sitúan estos objetos (en la mente de los entrevistados, si se trata de datos de entrevistas).

En primer lugar, se trata de comprender si existe un espacio común a todos los sujetos que han respondido; a continuación, hay que identificar cuántas dimensiones tiene este espacio; por último, es necesario dar un nombre a estas dimensiones (en el caso de las ciudades la tarea era más sencilla, pues se sabía desde el principio que las dimensiones eran dos, y se conocían sus nombres: latitud y longitud).

En la investigación antes citada sobre los candidatos a las presidenciales estado-



**Figura 6.5.** Situación de los candidatos a las elecciones presidenciales de EEUU en 1968 en el espacio definido por las dimensiones «identificación con el partido» e «ideología» [Weisberg y Rusk, 1970].

unidenses, se encontraron dos dimensiones (la figura 6.5 muestra la representación de los doce candidatos en estas dimensiones). Los autores interpretaron estas dimensiones como el eje de identificación con el partido (republicano / demócrata) y el eje de ideología (izquierda / derecha). En otras palabras, al expresar sus preferencias los entrevistados aplicaron de forma más o menos inconsciente dos criterios de evaluación: valoraron a cada uno de los candidatos en función de su pertenencia al partido y su posición progresista-conservadora. Por ejemplo, Humphrey se situó cerca del centro en la dimensión de ideología, y fue identificado claramente con el partido demócrata. Rockefeller, en cambio, se situó en una posición equidistante entre el partido republicano y el demócrata (en realidad era republicano, pero con tendencias liberales), pero con un claro posicionamiento ideológico de izquierdas.

No profundizaremos en las escalas multidimensionales, dado que son técnicas muy complejas desde el punto de vista conceptual y matemático y su estudio supera los objetivos de este manual.

## 7.8. SÍNTESIS DEL CAPÍTULO 6

1. Con el término «técnica de las escalas» designamos una serie de procedimientos elaborados por los investigadores sociales para operacionalizar conceptos complejos. Una escala es un conjunto coherente de elementos que se consideran indicadores de un concepto más general. El *elemento* es una afirmación, pregunta, comportamiento, respuesta a un test o atributo; la *escala* es el conjunto de una serie de elementos. En sociología y en psicología social la aplicación más común de las escalas son las denominadas *escalas de actitudes*. La técnica de las escalas genera variables «cuasicardinales», que presentan propiedades muy similares a las de las variables ordinales.
2. En una pregunta cerrada con respuestas ordenadas las categorías de respuesta pueden ser *semánticamente autónomas*, tener *autonomía semántica parcial*, o ser del tipo denominado *escalas de intervalos* de autoposicionamiento. En los dos primeros casos los resultados son variables ordinales, en el tercero, una variable cuasicardinal.
3. El concepto *escala de Likert* abarca una amplia variedad de escalas, que también se conocen con el nombre de escalas *aditivas*, ya que la puntuación de la escala se deriva de la suma de las puntuaciones de los elementos que la componen. La construcción de una escala de Likert consta de cuatro fases: formulación de las preguntas, aplicación a una muestra de sujetos, análisis de los elementos y del grado de coherencia interna, y control de la validez y el carácter unidimensional de la misma.
4. El *escalograma de Guttman* es una escala *acumulativa*, porque está constituida por elementos de dificultad creciente, y se supone que el sujeto que da una respuesta afirmativa a una determinada pregunta también ha dado una respuesta afirmativa a todas las preguntas que preceden a aquélla en grado de dificultad. Esta característica de escalabilidad se mide con el coeficiente de reproducibilidad.

5. El *diferencial semántico* de Osgood fue creado para estudiar el significado de los conceptos, no a través de la descripción subjetiva de los mismos emitida por los entrevistados, sino mediante las asociaciones entre estos conceptos y otros presentados de forma estandarizada a los sujetos. Con esta técnica, Osgood identificó tres dimensiones fundamentales que subyacen a todas las opiniones de los sujetos sobre los distintos conceptos: *valoración, potencia y actividad*.
6. El *test sociométrico*, creado por Jacob Moreno, tiene por objetivo identificar la red de relaciones interpersonales existentes entre los individuos de un grupo. En su versión básica el test sociométrico consiste en un cuestionario que contiene unas cuantas preguntas para plantear a los sujetos del grupo sobre las preferencias y rechazos que sienten hacia los demás componentes del grupo.

## 6.9. LECTURAS COMPLEMENTARIAS

Una recopilación interesante de las aportaciones clásicas de la tradición empírica estadounidense a las técnicas de medición y las escalas, es la obra clásica de G. M. Maranell *Scaling. A Sourcebook for Behavioral Scientists* (Aldine, 1974, 436 pp.). En D. Krebs P. y Schmidt (eds.) *New Directions in Attitude Measurement* (de Gruyter, 1993, 378 pp.) puede encontrarse una versión actualizada con aportaciones más recientes.

La obra de R. F. DeVellis, *Scale Development: Theory and Applications* (Sage, 1991, 121 pp.) contiene un texto introductorio sobre la medición, que también puede servir como guía práctica para la construcción de las escalas más sencillas.

Una guía muy útil para la construcción de escalas de actitud se encuentra en las recopilaciones elaboradas por J. P. Robinson, P. R. Shaver y L. S. Wrightsman (eds.) *Measures of Personality and Social Psychological Attitudes* (Academic Press, 1991, 753 pp.) y *Measures of Political Attitudes* (Academic Press, 1998, 801 pp.), que contienen una completa exposición sobre los instrumentos disponibles en la actualidad.