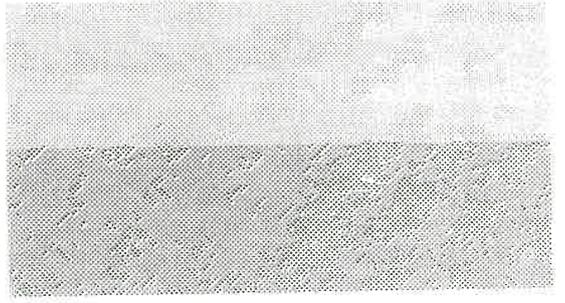


**parte
segunda:
métodos**



Meliá, J.L. (1991). Métodos de Escalamiento Unidimensional. Valencia: CSV. www.uv.es/psicometria

Método de las Comparaciones Apareadas de Thurstone.

El método de las comparaciones apareadas de Thurstone es uno de los procedimientos de escalamiento de más tradición dentro de la medición psicológica. Recibe este nombre porque utiliza como método de obtención de información la comparación de estímulos por pares, en una tarea orientada a juicio. Después las elecciones se convierten en valores de escala mediante la conversión de los porcentajes de preferencia en puntuaciones típicas normalizadas.

El método fue muy utilizado durante mucho tiempo, y todavía actualmente es corriente utilizarlo cuando se está realizando un estudio riguroso contrastando los resultados de diversos métodos de escalamiento. Sin embargo, los costos de esfuerzo y tiempo que supone la comparación de estímulos por pares influyen en que frecuentemente se prefieran otros métodos más sencillos y rápidos.

Método.

El primer paso consiste en aplicar el método de *comparación de estímulos por pares con orientación de juicio (III.A)*.

Si los objetos o estímulos a escalar son items, entonces, previamente, se habrá elaborado la lista de los mismos y se habrán sometido a los controles racionales descritos en el capítulo 3.

Aplicar el método de comparación de estímulos por pares, con orientación de juicio, implica:

-Decidir en qué dimensión se van a escalar los estímulos. De esta decisión dependerá el tipo de comparación de juicio que se pida al sujeto. ¿En qué cualidad, criterio o dimensión se pide a los sujetos que comparen los estímulos?

-Formar todos los pares posibles con los estímulos a escalar. (Combinaciones de n estímulos tomados de 2 en 2). Por razones prácticas el número de estímulos a escalar, n , no puede ser muy grande.

-Presentar todos los pares a una muestra de sujetos que actúan como jueces.

-Cada "juez" tiene que escoger en cada par aquel objeto que tenga "más" (también puede ser "menos") de la dimensión en que se van a escalar los estímulos. Cada sujeto emite un juicio en cada par.

El segundo paso consiste en *obtener los valores de escala de los objetos*. Para ello se sigue el siguiente proceso:

1.- Se forma una matriz de "objetos por objetos" en la que cada celdilla contiene la proporción de sujetos que han dicho que el objeto de la fila es mayor que el objeto de la columna en la dimensión.

Si todos los sujetos han contestado a todos los pares con una elección, entonces la suma de las celdillas simétricas respecto a la diagonal principal debe dar 1. En las celdillas de la diagonal principal se considera, por convención, que hay un 0'5; es decir, que un objeto no se considera ni mayor ni menor a sí mismo en la dimensión. (Por supuesto a los sujetos no se les da a juzgar ningún par con los dos elementos iguales).

2.- Las *proporciones se transforman en puntuaciones típicas* de una distribución normal perfecta. Simplemente, utilizando una tabla de distribución normal, o una calculadora o programa que permita este cálculo, se obtiene qué puntuación z' correspondería en la distribución normal a un valor que deja por debajo de sí la proporción indicada en la celdilla. Al final de este paso cada celdilla ha sido convertida en una z' .

En la diagonal principal, correspondiendo a una proporción de 0'5, quedan puntuaciones z' igual a 0'00.

3. *Se calculan las medias de las columnas.* Es decir, la puntuación típica z' media que corresponde a las proporciones de preferencia de cada estímulo. (Si un estímulo ha sido elegido más veces que los otros con los que ha formado pares, habrá tenido proporciones de elección más altas, y, por tanto, puntuaciones z' más altas, y por tanto un promedio de z' más alto.)

4. Por último, basta restar a todas las z' promedio así obtenidas la z' promedio del estímulo que la tenga más baja. El resultado de estas restas es ya el valor de escala de los estímulos.

Con esta resta el estímulo con z' promedio más baja queda siempre en una puntuación de escala igual a 0, y las puntuaciones de todos los demás expresan la distancia a ese estímulo "0".

Los trabajos originales de Thurstone, los manuales sobre escalamiento (p.e. Dunn-Rankin, 1983) y otros libros que describen el método (p.e., Ponsoda, 1986), suelen dar algunos pasos más antes de llegar al valor de escala. Esos pasos tradicionales no son necesarios y pueden resumirse con el mismo resultado en los que aquí se presentan.

Los valores de escala obtenidos pueden utilizarse de diversas formas para *medir sujetos* (si lo que se había escalado eran items para medir después sujetos). Quizá el procedimiento más sencillo para obtener información consista en utilizar una tarea de *estímulo simple con orientación de respuesta (I.B)*.

Después, para *obtener la puntuación de los sujetos* se utiliza un método de media de los valores de escala asignados a los items acertados o aceptados. Es decir, un método *H2; I:1; II:2; III:2*.

No obstante, el método de medición de los sujetos por el que se opte no es parte, estrictamente hablando, del método de comparaciones apareadas.

Ejemplos.

El mejor modo de comprender cómo trabaja en la práctica un método de escalamiento es verlo operar sobre algunos ejemplos concretos.

Ejemplo 1. Estamos realizando un estudio sobre el valor que conceden los sujetos a su tiempo de ocio, y a los diversos tipos de ocio, dentro de su jornada cotidiana. Como un procedimiento para estudiar esta cuestión, hemos decidido escalar diversas actividades cotidianas en función de la importancia que éstas tienen en la vida diaria de una muestra de sujetos. Hemos decidido utilizar el método de las comparaciones apareadas de Thurstone. La dimensión a estudiar es, pues, "importancia que se le concede en la vida diaria a cada actividad". Para volverla operativa y fácilmente manejable, la pregunta que hacemos a los sujetos ante cada par es: "¿A cuál de estas dos actividades dedica Ud. más tiempo cotidianamente?"

Las actividades a escalar son:

- A. Tiempo para practicar Deporte
- B. Tiempo para ver Televisión
- C. Tiempo para estar con la Familia
- D. Tiempo para estar con los Amigos.
- E. Tiempo para Lectura.

Se forman los 10 pares posibles y se presentan a los 100 sujetos de la muestra.

Los resultados obtenidos los resumimos en la siguiente tabla:

	A	B	C	D	E
A		90	70	50	35
B	10		55	10	15
C	30	45		40	10
D	50	90	60		10
E	65	85	90	90	

N=100.

Las celdillas de la tabla recogen cuantos sujetos han dicho que dedican más tiempo a la actividad de la columna que a la de la fila.

Por ejemplo, 90 sujetos han dicho que dedican más tiempo a la TV que al deporte, mientras que los 10 restantes dicen dedicar más tiempo al deporte que a la TV.

Por ello, la suma de dos celdillas simétricas respecto a la diagonal vacía es siempre N, en este caso 100.

El cuadro anterior se puede expresar como proporciones:

	A	B	C	D	E
A		.90	.70	.50	.35
B	.10		.55	.10	.15
C	.30	.45		.40	.10
D	.50	.90	.60		.10
E	.65	.85	.90	.90	

N=100.

Para obtener las proporciones anteriores (que se han representado suprimiendo el 0 a la derecha del punto decimal), simplemente hemos dividido cada frecuencia por N.

Ahora, con la ayuda de una tabla de la curva normal, o de una calculadora adecuada, buscamos que valores z' corresponden a esas proporciones. (Es decir, lo que buscamos es que valores z' en la curva normal dejan por debajo de sí esa proporción).

Los resultados son:

	A	B	C	D	E
A	0	1'2816	0'5244	0	-0'3853
B	-1'2816	0	0'1257	-1'2816	-1'0364
C	-0'5244	-0'1257	0	-0'2533	-1'2816
D	0	1'2816	0'2533	0	-1'2816
E	0'3853	1'0364	1'2816	1'2816	0

A veces un valor concreto no se encuentra en una tabla, con lo que solemos tomar, convencionalmente, como aproximación, la z' cuya proporción acumulada se acerca más a la que buscamos. (Los valores de una tabla de curva normal pueden por ello diferir ligeramente de los que aquí se muestran ya que estos no han sido obtenidos de una tabla sino calculados).

Obsérvese que, al ser simétrica la distribución normal, cada proporción tiene la misma z' que su complementaria a 1, pero con el signo opuesto.

Así, la z' para 0'10 es -1'2816, mientras que para 0'90 es 1'2816. Por eso, no es necesario buscar todos los valores z' . Basta con buscar los que son mayores que 0'5; a sus complementarios les corresponde la misma z' con signo negativo. La z' para 0'5 es 0, porque, como se sabe, la media de la distribución normal está en $z'=0$, y la

distribución normal acumula un 50% de los casos hasta su media, al ser simétrica.

Ahora simplemente sumamos las columnas y dividimos por el número de estímulos. (Es decir, hallamos la media de las z' de cada columna).

	A	B	C	D	E
A	0	1'2816	0'5244	0	-0'3853
B	-1'2816	0	0'1257	-1'2816	-1'0364
C	-0'5244	-0'1257	0	-0'2533	-1'2816
D	0	1'2816	0'2533	0	-1'2816
E	0'3853	1'0364	1'2816	1'2816	0

Σ -1'4207 3'4739 2'1850 -0'2533 -3'9849

Media -0'2841 0'6948 0'4370 -0'0507 -0'7970

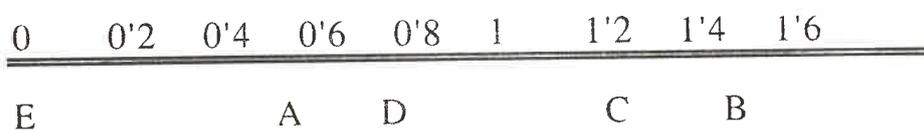
Podríamos tomar esos sumatorios o esas medias como los valores de escala de los cinco estímulos. Sin embargo, para que la escala tenga como punto "0" arbitrario el estímulo menor, lo que se hace es restar a la media de cada estímulo el valor de la media del estímulo menor.

En este caso concreto la actividad "E. Tiempo para lectura" ha quedado clasificada como la menor en la dimensión "tiempo dedicada a la misma", con una z' media de -0'7970. Ahora se resta -0'7970 a todas las z' medias para conseguir que el estímulo E quede ubicado en 0 y los valores de todos los demás expresen la distancia a E en puntuaciones z' medias.

Estímulos:	Valor z' media:	Menos z' menor:	Valor de escala:
A	-0'2841	$-(-0'7970)=$	0'5129
B	0'6948	$-(-0'7970)=$	1'4918
C	0'4370	$-(-0'7970)=$	1'2340
D	-0'0507	$-(-0'7970)=$	0'7463
E	-0'7970	$-(-0'7970)=$	0

Ahora todos los estímulos tienen un valor de escala positivo y el estímulo E se ha convertido en el origen relativo "0"

Podemos representar gráficamente la escala:



De acuerdo con estos resultados ficticios, la actividad a la que los sujetos dicen dedicar más tiempo es a ver televisión, seguida a cierta distancia del tiempo para la familia. En un nivel intermedio se sitúa el tiempo para estar con los amigos y algo por debajo el tiempo para practicar deporte. En último lugar, a una distancia considerable, aparece el tiempo para la lectura.

Obsérvese que en este caso se ha aplicado el método de escalamiento únicamente para escalar los estímulos sobre una dimensión, sin que estos estímulos escalados vayan a ser usados para medir sujetos.

Ejemplo 2. Estamos interesados en escalar 4 enunciados acerca de la actitud ante el divorcio, para después poder utilizarlos en un estudio psicosocial en el que se necesita cuantificar la actitud ante el divorcio que presentan los sujetos. Hemos decidido utilizar el método de las comparaciones apareadas de Thurstone. Los cuatro enunciados son los siguientes:

- A. El matrimonio es indisoluble y no debería permitirse el divorcio bajo ninguna circunstancia.
- B. El divorcio es un remedio muy extremo que sólo los jueces deberían aplicar y de modo muy restrictivo.
- C. El divorcio debe utilizarse únicamente después de ciertas garantías, tales como un año de separación previa, mutuo consentimiento o decisión judicial.
- D. Las personas deben ser libres de establecer vínculos matrimoniales y de romperlos, por tanto el divorcio debería ser libre y sin restricciones burocráticas.

Estamos interesados en determinar el valor de escala en la dimensión "desfavorable-favorable al divorcio", en consecuencia le presentamos a cada sujeto cada uno de los 6 pares posibles y le preguntamos "¿Cuál de estos dos enunciados es más favorable al divorcio?" (Tarea de juicio).

Los resultados, para una muestra de 50 sujetos, los sintetizamos en la siguiente tabla:

	A	B	C	D
A		35	40	45
B	15		35	40
C	10	15		35
D	5	10	15	

$N=50$.

Cada celdilla representa el número de sujetos que han dicho que el enunciado de la columna es más favorable al divorcio que el de la fila. La suma de dos celdillas simétricas respecto a la diagonal vacía debe sumar $N=50$.

Ahora convertimos las frecuencias en proporciones dividiendo cada celdilla por el valor de N , en este caso por 50.

	A	B	C	D
A		0'7	0'8	0'9
B	0'3		0'7	0'8
C	0'2	0'3		0'7
D	0'1	0'2	0'3	

Convertimos las proporciones en puntuaciones z' y calculamos el promedio de las columnas.

	A	B	C	D
A	0	0'5244	0'8416	1'2816
B	-0'5244	0	0'5244	0'8416
C	-0'8416	-0'5244	0	0'5244
D	-1'2816	-0'8416	-0'5244	0

Σ -2'6476 -0'8416 0'8416 2'6476

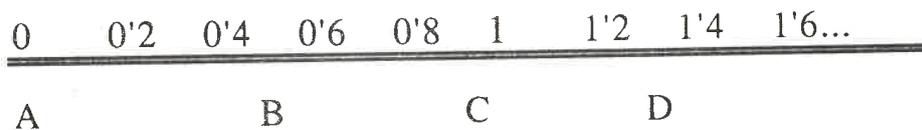
Media -0'6619 -0'2104 0'2104 0'6619

Ahora restamos el valor z' media menor a todos los valores z' media:

<u>Estímulos:</u>	<u>Valor z' media:</u>	<u>Menos z' menor:</u>	<u>Valor de escala:</u>
A	-0'6619	$-(-0'6619)=$	0
B	-0'2104	$-(-0'6619)=$	0'4515
C	0'2104	$-(-0'6619)=$	0'8723
D	0'6619	$-(-0'6619)=$	1'3238

Como resultado todos los estímulos tienen un valor de escala positivo, y el estímulo A se ha convertido en el origen relativo "0".

Podemos representar gráficamente la escala:



De acuerdo con estos datos ficticios, el estímulo D es el más favorable al divorcio, seguido por el C, por el B, y por último, por el A, que se presenta como el enunciado más desfavorable.

Ahora que conocemos el valor de escala de estos enunciados podemos utilizarlos como items de una escala de actitud ante el divorcio. Lo más sencillo es presentar los items en una tarea de estímulo simple con orientación de respuesta. Sencillamente le presentamos cada enunciado a cada sujeto y le preguntamos: ¿Está Ud. de acuerdo con esta afirmación sobre el divorcio? Lo más habitual será calcular la puntuación del sujeto mediante H2, con I:1; II:2 y III:2.

Supongamos que un sujeto J contesta "Sí" a los enunciados B y C, y "No" ante los enunciados A y D. La puntuación total del sujeto será:

$$J = (0'4515 + 0'8723) / 2 = 1'3238 / 2 = 0'6619$$

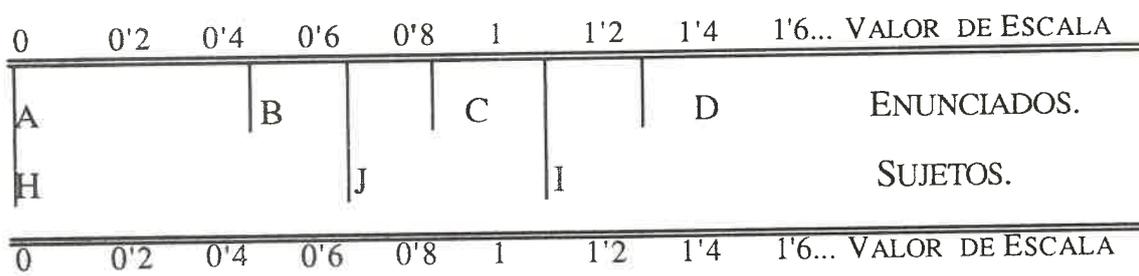
El sujeto J mantiene una actitud ante el divorcio entre la representada por el ítem B y la representada por C.

Supongamos que un sujeto I no da ninguna respuesta (omisión) ante el enunciado A, contesta negativamente al B y afirmativamente a los dos restantes. La puntuación del sujeto I será:

$$I = (0'8723 + 1'3238) / 2 = 2'1961 / 2 = 1'0981$$

Un tercer sujeto H, contesta afirmativamente al enunciado A y negativamente a los otros tres. Su puntuación será igual a 0.

Podemos representar a los tres sujetos sobre la escala:



Este segundo ejemplo muestra una forma de utilizar los valores de escala obtenidos con el método de las comparaciones apareadas de Thurstone para medir sujetos.

Fundamentos.

Thurstone fue pionero en el campo del escalamiento psicológico. Habiendo trabajado inicialmente en escalamiento psicofísico dió el paso hacia el escalamiento psicológico, el escalamiento de dimensiones que no tienen un referente inmediato de medición física. El método de las comparaciones apareadas es precisamente el que evoca de forma más clara un método psicofísico de comparación de estímulos.

Thurstone elaboró tres métodos de escalamiento principales: el método de las comparaciones apareadas, el método de los intervalos sucesivos y el método de los intervalos aparentemente iguales. Los tres se fundamentan en lo que se denomina "*la ley del juicio comparativo*".

Torgerson (1958) ha presentado de un modo sistemático la ley del juicio comparativo de Thurstone. En palabras de este autor, formalmente, "la ley del juicio comparativo es un conjunto de ecuaciones que relacionan la proporción de veces que un estímulo cualquiera k es juzgado mayor en un atributo dado que otro estímulo cualquiera j , con los valores escalares y las dispersiones discriminales de los dos estímulos en el continuo psicológico." Veamos, con un enfoque no matemático, cual es el significado de la ley del juicio comparativo.

Supongamos que tenemos una serie de delitos (homicidio, atraco a un banco, tráfico de influencias, ...) y que deseamos escalarlos sobre una dimensión de "gravedad". Se trata de ubicar estos delitos en una escala hipotética que va desde "Nada grave" hasta "Muy grave". La gravedad asignada a cada delito dependerá de la percepción, conocimientos, experiencia, actitudes, etc. de cada sujeto, de modo que cada sujeto que actúa como juez asignará una posición en la escala de gravedad a cada delito.

Thurstone supone que, con suficiente número de jueces, la distribución de las posiciones en la escala que los sujetos asignan a un mismo estímulo, tendrá forma de curva normal. Por tanto, la media, mediana y moda de esa distribución de puntuaciones coincidirán en un mismo valor de escala; ese valor de escala lo denomina "*proceso discriminial modal*" y constituye la mejor expresión de la posición del estímulo sobre la dimensión.

La desviación típica de esa distribución normal significa el grado de dispersión o discrepancia entre los jueces respecto a la posición en la escala de un estímulo; la denomina "*dispersión discriminial modal*". Cuanto mayor sea la dispersión, menor es el acuerdo de los "jueces" en el valor de escala de un estímulo.

En nuestro ejemplo, supongamos que los sujetos pudieran expresar directamente la gravedad del delito "atracó a un banco" sobre una escala de 0 (nada grave) a 100 (muy grave). Supongamos que disponemos de una muestra de 200 jueces. Thurstone no espera que los 200 estén de acuerdo en ubicar al delito en una misma posición, por ejemplo 60. En lugar de ello, se espera que se produzca una distribución normal de puntuaciones en torno a un valor medio. Por ejemplo, puede esperarse que 40 jueces digan que el valor de escala de ese estímulo es 60; que 32 digan que el valor de escala es 61 y que otros 32 digan que es 59; que 20 sujetos digan que el valor de escala es 62 y otros 20 que es 58, etc., formando una curva normal más o menos apuntada. La media de esos valores, -en el ejemplo supuesto sería 60,- se toma como valor de escala del delito, como la "gravedad" del mismo.

De ese modo, supone Thurstone, los juicios de los sujetos sobre el valor de escala de un estímulo, sobre su ubicación sobre la dimensión considerada, adoptan la forma de una distribución normal. Esta suposición está avalada por cierta evidencia empírica en algunos casos, pero más allá de esto es un supuesto formal del método, necesario para sostener algunos razonamientos posteriores.

Hasta aquí se ha expuesto una concepción de los juicios de los sujetos de naturaleza probabilista y la forma que se supone que adopta la distribución de esos juicios, si los sujetos juzgaran directamente cuál es la posición de un estímulo en la escala. Sin embargo, en el método de las comparaciones apareadas los sujetos juzgan únicamente cuál de dos estímulos es mayor en una dimensión; por ejemplo, cuál de dos estímulos es más grave. ¿Cómo se relaciona la concepción de la distribución de los juicios de los sujetos como una curva normal con las elecciones que los sujetos producen en los pares?

El razonamiento es más o menos el siguiente. Si dos estímulos están muy próximos en su valor de escala, sus distribuciones de puntuaciones estarán muy solapadas, con las medias próximas. Ello supone que para un número de jueces un estímulo será mayor que el otro, y para los otros al revés; ello supondrá que el porcentaje de sujetos que dicen que el primero es mayor que el segundo estará próximo al 50% y que el porcentaje de sujetos que dicen que el segundo es mayor que el primero estará también próximo al 50%.

Si, por el contrario, dos estímulos están muy lejanos en su valor de escala, son muy diferentes en el atributo que se juzga, las medias que representan esos valores de escala estarán muy alejadas la una de la otra. Es decir, las distribuciones normales en torno a esas medias estarán muy poco solapadas, y por tanto habrá un alto porcentaje de sujetos que dirá que un estímulo es mayor que el otro (y sólo un pequeño porcentaje que sostendrá lo contrario). El solapamiento de las distribuciones de dos estímulos A y B se utiliza como explicación de los porcentajes de elecciones en el par formado por ambos. Si la media de A por ejemplo es mayor, y además esta muy lejos de la media de B, entonces casi todos los sujetos dirán que $A > B$ en el par, y sólo muy pocos lo contrario.

La clave de este razonamiento reside en que muestra cómo el porcentaje de sujetos que dice que A es mayor que B es proporcional a la distancia entre A y B.

Supongamos que efectivamente A es mayor que B (por ejemplo que el delito A es más grave que el B) pero que ambos están muy próximos; entonces, habrá más porcentaje de sujetos que digan que, efectivamente, A es mayor que B que que digan lo contrario; sin embargo, para muchos sujetos esa "diferencia" no será clara y podrán "confundirse" afirmando que B es mayor que A.

Por ejemplo, si A es mayor que B pero ambos están muy próximos podríamos esperar algo así como que el 55% de los sujetos dijera que A es mayor que B y el 45% restante que B es mayor que A.

Ahora supongamos que A es mayor que B pero que la diferencia entre ellos es muy grande (por ejemplo que el delito A es clarísimamente más grave que el B), entonces podemos esperar que la diferencia entre los porcentajes sea patente; por ejemplo quizás podemos esperar que el 90% de los sujetos diga que A es mayor que B y que sólo el 10% restante afirme lo contrario.

La diferencia entre el porcentaje que produce una elección y el porcentaje complementario que produce la elección inversa es muy grande cuando la diferencia en los valores de escala de los estímulos es muy grande, y muy pequeña cuando la diferencia en los valores de escala es muy pequeña. Es decir, esa diferencia es proporcional a la diferencia en valores de escala.

Hasta aquí queda claro que existe una proporcionalidad entre los porcentajes que se producen en las elecciones y las distancias entre los estímulos

implicados, ahora bien ¿cómo podría formularse esa proporcionalidad de una forma más precisa?

Recordemos que los valores de escala de dos estímulos A y B son, en la concepción de Thurstone, las medias de dos distribuciones de juicios. La distancia entre esas medias (o valores de escala) puede expresarse como:

$$\bar{X}_A - \bar{X}_B \quad (1)$$

Esa diferencia puede tipificarse o expresarse en puntuaciones típicas:

$$z_{AB} = \frac{\bar{X}_A - \bar{X}_B}{S_{A-B}} \quad (2)$$

Esa diferencia de medias es bien conocida al utilizarse en el test de la diferencia entre medias de dos distribuciones normales. Desarrollando el denominador puede expresarse a su vez como:

$$z_{AB} = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{S_A^2 + S_B^2 - 2 r_{AB} S_A S_B}} \quad (3)$$

Si despejamos la diferencia entre las medias obtendremos:

$$\bar{X}_A - \bar{X}_B = z_{AB} \sqrt{S_A^2 + S_B^2 - 2 r_{AB} S_A S_B} \quad (4)$$

La ecuación (4) es una expresión matemática de la ley del juicio comparativo. Tal y como está no puede resolverse, dado que no conocemos la mayoría de sus términos. Para hacerla viable es necesario hacer algunas suposiciones adicionales.

En primer lugar, si se supone que los juicios sobre los estímulos no correlacionan entre sí (esto es, $r_{AB}=0$), la ecuación se simplifica:

$$\bar{X}_A - \bar{X}_B = z_{AB} \sqrt{S_A^2 + S_B^2} \quad (5)$$

En segundo lugar, si se asume que la dispersión de la distribución de juicios de A y la de B son iguales, es decir que $S_A^2 = S_B^2$, entonces el valor bajo la raíz se convierte en una constante, y la ecuación vuelve a simplificarse:

$$\bar{X}_A - \bar{X}_B = k \cdot z_{AB} \quad (6)$$

Siendo k una constante que teóricamente puede tomar cualquier valor, pero que usualmente se considera igual a 1 para simplificar. Ahora la ecuación se ha vuelto manejable.

La puntuación z_{AB} se considera la puntuación típica que corresponde a la proporción en que el estímulo A es juzgado mayor que el B. Puede apreciarse ahora que esa puntuación típica es proporcional a la diferencia entre los valores de escala de los estímulos, como antes habíamos mostrado a nivel intuitivo.

Obsérvese que en realidad nunca se sabe exactamente donde, en términos absolutos, está ubicado un estímulo dentro de una dimensión; pero las proporciones de elecciones en comparaciones apareadas pueden utilizarse para obtener puntuaciones típicas que expresan la distancia relativa entre los estímulos.

El recurso de restar a los promedios de esas típicas la típica menor, como hemos visto en el método y ejemplos, es sólo una forma de hacer coincidir el estímulo menor con una posición cero arbitraria.

Supuestos y limitaciones.

La ley del juicio comparativo y el método de las comparaciones apareadas son aportaciones muy importantes a los métodos de escalamiento. Sin embargo, hay algunas limitaciones y dificultades que conviene tener presentes. Entre ellas pueden destacarse los fuertes supuestos que el método implica, las dificultades derivadas de utilizar los sujetos como jueces para comparar estímulos, la cuestión de la dimensionalidad de las escalas y los elevados costes en tiempo y en tarea del método.

Los supuestos del método.

En primer lugar, hay que ser consciente de los supuestos que subyacen al método expuesto. En la medida en que los datos se alejen de esos supuestos el método pierde valor y puede llegar a considerarse inadecuado.

La clave del método reside en que Thurstone llega a poder igualar la diferencia entre dos estímulos, expresada en puntuaciones típicas en una curva normal, a la diferencia entre las posiciones escalares de los estímulos, lo que permite dar el salto entre una proporción de elecciones y una distancia entre estímulos sobre una dimensión. Pues bien, eso es posible, asumiendo que:

- Las respuestas de los sujetos ante los estímulos adoptan una distribución normal sobre la dimensión escalar subyacente.
- Las varianzas de los estímulos sobre esta distribución son iguales.
- Los juicios de los sujetos ante los estímulos no están correlacionados entre sí.

Sin estas tres suposiciones no sería posible simplificar la ecuación (4) en la ecuación (6) y por tanto no sería posible efectuar el escalamiento según el método tal y como se ha descrito.

Las tres suposiciones son muy exigentes acerca del comportamiento de los sujetos que actúan como jueces y en buen número de casos puede cuestionarse seriamente que lleguen a cumplirse.

La ecuación (4) es muy importante. Esa ecuación constituye "la forma completa de la ley del juicio comparativo" (Torgerson, 1.958). En esta ecuación,

$$\bar{X}_A - \bar{X}_B = z_{AB} \sqrt{S_A^2 + S_B^2 - 2 r_{AB} S_A S_B}$$

las medias constituyen los valores de escala de los estímulos, las desviaciones típicas denotan las dispersiones discriminales de los estímulos, la correlación que aparece es la correlación entre las puntuaciones asignadas a los estímulos, y la puntuación típica es la puntuación normal que corresponde a la proporción teórica de veces que el estímulo A se juzga mayor que el B.

Si conociésemos todos estos valores la ley del juicio comparativo podría resolverse directamente y no haría falta establecer ningún supuesto. Sin embargo, la ley del juicio comparativo presenta siempre para un número dado de estímulos, más incógnitas que ecuaciones. Por ello es necesario hacer algunos supuestos para resolverla. Thurstone (1927) ensayó varias formas a fin de evitar estas dificultades y resolver la ley del juicio comparativo. Consideró la posible simplificación y solución para diferentes casos en que se cumplieran determinadas asunciones. Los tres supuestos anteriores forman parte de los supuestos simplificadores tratados bajo el "caso V", que es quizás el de mayor resonancia práctica.

Los sujetos como jueces.

En segundo lugar, hay que considerar las objeciones acerca de la calidad de las comparaciones de estímulos que los sujetos efectúan. Esta objeción se

refiere al método de obtención de información para escalar los estímulos, y por tanto no es exclusiva de este método de Thurstone.

En la fase de escalamiento de los items, una muestra de sujetos (muestra I) actúa como jueces respecto a los estímulos. Los sujetos se utilizan como "calibradores" de los estímulos. La cuestión es si los sujetos son buenos calibradores, y por tanto, si deciden correctamente o no en cada comparación.

Este grupo de objeciones tiene dos grandes componentes. Primero, está la cuestión de si los sujetos pueden ser "un buen instrumento de medida". Segundo, está la cuestión de la imparcialidad de los sujetos.

La cuestión de si los sujetos pueden ser un buen instrumento para calibrar los items tiene a su vez varios aspectos. Primero, puede discutirse en el sentido de si los sujetos discriminan adecuadamente entre estímulos realmente ubicados en diferentes puntos del continuo. Segundo, puede discutirse si los sujetos son consistentes en sus juicios consigo mismos, con otros y a través del tiempo.

La cuestión de la imparcialidad de los sujetos hace referencia a la medida en que los sujetos son influidos por *su propia posición en la dimensión medida* al comparar estímulos. Por ejemplo, al escalar dos items A y B que representan actitudes más o menos pacifistas se les pide a los sujetos (muestra I) que digan cuál de los dos es más pacifista que el otro. La tarea implica comparar los estímulos como si los sujetos fueran jueces justos que sólo están influidos en su elección por la magnitud de pacifismo contenido en los items A y B, y no por su posición personal respecto al tema.

La tarea de los sujetos de la muestra I es la de comparar los estímulos por sí mismos, no por la relación que ellos mantienen con los estímulos. Si no fuera de este modo no quedaría patente que los estímulos mismos difieren ubicándose en diferentes puntos del continuo. Quizás sea un supuesto excesivo considerar que la posición de los jueces en el continuo de actitud es independiente de sus juicios sobre los items. Por ejemplo, hay algunas evidencias que señalan que los jueces con actitudes extremistas, positivas o negativas, hacia un fenómeno, no son buenos discriminadores de items

moderados sobre ese continuo. La dificultad reside en que, como Scott (1968) indicó, "el modelo requiere que las diferencias en la localización de los ítems sean aleatorias, sin depender de las características sistemáticas de los jueces."

Hay que considerar que intervienen *dos* procesos distintos en las fases de escalamiento de los ítems (muestra I) y de medición de los sujetos (muestra II). Cuando el objetivo es solamente escalar los estímulos, como en el ejemplo de las actividades de ocio, esta dificultad no aparece. Sin embargo, cuando escalamos un conjunto de ítems para medir después sujetos con ellos, este problema es patente.

A los sujetos que actúan como jueces para escalar los ítems (muestra I) se les pide que comparen los estímulos en razón de la magnitud con que los estímulos mismos contienen una propiedad.

A los sujetos que se desea medir (muestra II) se les pide que se comparen ellos mismos con los estímulos y digan si ellos están ubicados en el mismo punto del continuo de actitud que cada estímulo.

Por supuesto, en algunos casos la muestra I puede ser después usada como muestra II, pero eso no afecta el núcleo del problema. La dificultad reside en que el método asume que las posiciones de los ítems en el continuo serán las mismas para ambos procesos independientemente de la tarea que los sujetos realicen respecto a los mismos. Sin embargo, como McIver y Carmines (1981) señalan, debido a que la tarea demandada a los sujetos a los que se quiere medir es claramente distinta, es posible que las repuestas de estos a los ítems estén determinadas por factores no considerados al elaborar la escala, resultando en una medición no válida.

La cuestión de la dimensionalidad.

El problema de la dimensionalidad es complejo y afecta de diferentes modos a todos los métodos de escalamiento. De un modo elemental puede concretarse en las siguientes preguntas:

Primero, dado un conjunto de estímulos, ¿cuántas dimensiones son necesarias para representarlos de un modo adecuado?

Segundo, dado un conjunto de estímulos representados en un número dado de dimensiones ¿en qué medida están suficientemente bien representados por ellas?

Un modo de entender la cuestión de la dimensionalidad es pensarla en términos de su representación geométrica. Así, un conjunto de estímulos considerados unidimensionalmente puede ser representado sobre una recta, indicando el punto donde se ubica cada estímulo, cuál es el valor escalar de cada estímulo. Un conjunto de estímulos, considerados bidimensionalmente, puede ser ubicado sobre un plano; siendo las proyecciones de cada punto (estímulo) sobre los ejes las que determinan el valor escalar de ese estímulo en cada dimensión. Con la misma lógica se consideran espacios de tres o más dimensiones. La dificultad reside en que comúnmente es posible representar el mismo conjunto de estímulos unidimensional, bidimensionalmente, etc. según el método de escalamiento que se utilice y un conjunto de decisiones que el investigador debe tomar.

Los métodos de Thurstone que consideramos en este libro permiten ubicar los estímulos sobre una dimensión, es decir, escalarlos unidimensionalmente. La limitación del método consiste en que supone que los estímulos pueden escalar unidimensionalmente y que esa representación es adecuada, pero no presenta un procedimiento de contrastación empírica de la unidimensionalidad. Esta situación es compartida de diferentes modos por los diversos métodos unidimensionales.

Costes del escalamiento de Thurstone.

El método de escalamiento de las comparaciones apareadas de Thurstone tiene altos costes temporales y de esfuerzo. En primer lugar, porque presenta separado el proceso de escalamiento de los items y el de medición de los sujetos -aspecto que será común a la mayoría de métodos-. Segundo, y especialmente, porque la tarea solicitada a cada sujeto en la fase de escalamiento de los items puede ser muy costosa. El método depende de una tarea de comparaciones de estímulos por pares que es el procedimiento de obtención de información más costoso en términos de tiempo, esfuerzo y motivación de los sujetos.

En el método de las comparaciones apareadas el número de estímulos a escalar ha de ser necesariamente muy limitado. Como vimos al tratar el método de comparación de estímulos por pares, para n estímulos cada sujeto debe hacer $[n(n - 1)]/2$ juicios comparativos. El número de juicios comparativos a realizar crece muy rápidamente al aumentar el número de estímulos a escalar. Para 10 estímulos bastan 45 juicios comparativos; 20 estímulos requieren 190 juicios comparativos; 30 estímulos requieren 435 juicios comparativos; 40 suponen 780, y 50 nada menos que 1.225. Como punto de comparación puede recordarse que los tests o cuestionarios que pasan de las 200 preguntas pueden considerarse muy largos; y la mayoría de los mismos no suele alcanzar los 100 items.

Cap. 9. METODO DE LAS COMPARACIONES APAREADAS DE THURSTONE.

ELABORACION DE LA ESCALA:

Generación de ítems:

1. Construcción de ítems.

Obtención de información para escalar los ítems:

1. Formar todos los pares posibles de estímulos.
2. Presentar cada par de ítems a cada sujeto de una muestra, solicitándole que elija un estímulo de cada par según algún criterio.
(Tarea comparación de estímulos por pares con orientación de Juicio: III.A).

Obtención de los valores de escala de los ítems:

1. Se forma una matriz de "ítems por ítems" donde cada celdilla representa la proporción de sujetos de la muestra que ha preferido el objeto de la columna al de la fila. (Cada valor representa el número de sujetos que ha preferido el ítem de la columna al de la fila, partido por el número total de sujetos.) (Tabla 1)
2. Las proporciones se convierten en las puntuaciones típicas correspondientes utilizando una tabla de la distribución normal(K). (En la diagonal se considera que hay 50% de elecciones y por tanto se asigna una puntuación típica z' de 0'00.) (Tabla 2)
3. Se calcula la puntuación típica de elección promedio de cada ítem (a).
(Promedios de las columnas de la tabla 2).
4. Se ordenan los ítems en función de su puntuación típica de elección promedio.
(De menos a más, de ítem menos elegido a ítem más elegido.) (Opcional.)
5. Se calculan las distancias de cada estímulo al menor. Considerando al estímulo menor como cero arbitrario, esas distancias expresan los valores de escala de los estímulos.

MEDICION DE SUJETOS:

Obtención de información:

1. Se administra la escala a los sujetos a medir y se les solicita que expresen si cada ítem los representa o no. (P.e., si reflejan o no la actitud del sujeto).
(Tarea de Estímulo simple con orientación de Respuesta: I.B).

Obtención de los valores de escala o puntuaciones de los sujetos:

1. Disposición de la información obtenida:
Se elabora una tabla "ítems por sujetos" en la que cada celdilla contiene el valor (0 ó 1) que representa la respuesta (acuerdo o desacuerdo) de cada sujeto con cada ítem.
2. Puntuaciones de los sujetos:
El promedio de los valores de escala de los ítems que representan a un sujeto se consideran la puntuación del sujeto (su ubicación sobre la dimensión medida).
(Método H2; I:1; II:2; III:2)