



Escalas cualitativas: Aplicación de un nuevo enfoque basado en proximidades

Raquel González del Pozo
PRESAD, IMUVA
Dep. de Economía Aplicada
Universidad de Valladolid
 Valladolid
 raquel.gonzalez.pozo@uva.es

José Luis García Lapresta
PRESAD, BORDA, IMUVA
Dep. de Economía Aplicada
Universidad de Valladolid
 Valladolid
 lapresta@eco.uva.es

Resumen—La asignación de números a los términos lingüísticos de las escalas cualitativas es muy habitual en el ámbito de las Ciencias Sociales. Sin embargo, esta asignación carece de sentido cuando las escalas son no uniformes, es decir, cuando los individuos aprecian proximidades distintas entre los términos consecutivos de las mismas.

En este trabajo se propone un procedimiento ordinal para ordenar un conjunto de alternativas a partir de las valoraciones mostradas por los individuos mediante una escala cualitativa no necesariamente uniforme. Dicho procedimiento evita el uso de números en las valoraciones cualitativas y se basa en las proximidades ordinales entre los términos lingüísticos de las escalas. El procedimiento propuesto se ilustra con un ejemplo tomado de los Barómetros del Centro de Investigaciones Sociológicas.

Palabras clave—escalas cualitativas, encuestas, CIS

I. INTRODUCCIÓN

El creciente interés por conocer las preferencias, opiniones y necesidades de la población ante determinadas políticas económicas y sociales hace que se utilicen cada vez más escalas cualitativas formadas por términos lingüísticos en los cuestionarios y encuestas.

El uso de escalas cualitativas formadas por palabras permite a los seres humanos expresarse de una forma sencilla ante situaciones de vaguedad e imprecisión. Autores como Zimmer [10] señalan que la mayoría de las personas prefiere comunicar sus opiniones inciertas a través de las palabras que con valores numéricos, ya que las expresiones verbales se ajustan mejor a sus pensamientos que los números.

No obstante, y a pesar de la preferencia de los seres humanos por las palabras, la gran mayoría de encuestas y cuestionarios miden las actitudes y el grado de acuerdo de los encuestados ante determinadas cuestiones mediante escalas tipo Likert [8], a cuyos términos lingüísticos se les asignan valores numéricos.

Estas asignaciones de valores numéricos (habitualmente equidistantes) carecen de sentido en el contexto de las escalas cualitativas no uniformes, es decir, cuando los individuos aprecian proximidades distintas entre los términos consecutivos de la escala, pudiendo generar diferentes resultados en función de cómo se codifique numéricamente la escala (véanse Roberts [9] y Franceschini et al. [4] entre otros).

Por otra parte, el Centro de Investigaciones Sociológicas (CIS) y otros organismos nacionales e internacionales presentan los resultados de sus investigaciones agrupando determinados términos lingüísticos. En la metodología de los anuarios del CIS [3] puede comprobarse que términos como “Mucho” y “Bastante”, utilizados en algunas preguntas de batería para conocer el grado de acuerdo de los encuestados ante determinadas cuestiones, han sido agrupados, descartándose el resto de términos lingüísticos de la escala (“Poco” y “Nada”). Este tratamiento supone una pérdida importante de la información cualitativa de las escalas, además de no considerar la no uniformidad de las mismas.

En este trabajo se plantea cómo ordenar un conjunto de alternativas valoradas cualitativamente a través del procedimiento puramente ordinal propuesto recientemente por García-Lapresta y Pérez-Román [7]. Este procedimiento evita el uso de números en las valoraciones cualitativas y considera únicamente las proximidades entre los términos de las escalas, por lo que puede ser aplicado en el contexto de escalas cualitativas no uniformes. En el presente trabajo también se incluye una comparación entre el procedimiento puramente ordinal antes mencionado y la metodología correspondiente a la agrupación de varios términos lingüísticos utilizando datos del Barómetro del CIS de mayo de 2011.

II. PROCEDIMIENTO

Se considerará un conjunto de agentes $A = \{1, \dots, m\}$, con $m \geq 2$, que evalúan un conjunto de alternativas $X = \{x_1, \dots, x_n\}$, con $n \geq 2$, a través una escala cualitativa ordenada $\mathcal{L} = \{l_1, \dots, l_g\}$, con $g \geq 3$ y $l_1 < l_2 < \dots < l_g$.

Con objeto de evitar la asignación de números a los términos lingüísticos de escalas cualitativas no uniformes y manejar su información de una forma puramente ordinal, en este trabajo utilizamos la noción de medida de proximidad ordinal introducida por García-Lapresta y Pérez-Román [6].

Una medida de proximidad ordinal es una función que asigna un grado de proximidad a cada par de términos lingüísticos de una escala cualitativa ordenada \mathcal{L} . Estos grados de proximidad pertenecen a un orden lineal $\Delta = \{\delta_1, \dots, \delta_h\}$, con $\delta_1 \succ \dots \succ \delta_h$, denotando δ_1 y δ_h la máxima y la mínima proximidad, respectivamente. Es importante destacar que los

elementos de Δ no son números, son solo objetos abstractos, que representan diferentes grados de proximidad.

Definición 1: [6] Una *medida de proximidad ordinal* sobre \mathcal{L} con valores en Δ es una función $\pi : \mathcal{L}^2 \rightarrow \Delta$, donde $\pi(l_r, l_s) = \pi_{rs}$ denota el grado de proximidad entre l_r y l_s , que satisface las condiciones siguientes:

- 1) *Exhaustividad:* Para cada $\delta \in \Delta$, existen $l_r, l_s \in \mathcal{L}$ tales que $\delta = \pi_{rs}$.
- 2) *Simetría:* $\pi_{sr} = \pi_{rs}$, para cualesquiera $r, s \in \{1, \dots, g\}$.
- 3) *Máxima proximidad:* $\pi_{rs} = \delta_1 \Leftrightarrow r = s$, para cualesquiera $r, s \in \{1, \dots, g\}$.
- 4) *Monotonía:* $\pi_{rs} \succ \pi_{rt}$ y $\pi_{st} \succ \pi_{rt}$, para cualesquiera $r, s, t \in \{1, \dots, g\}$ tales que $r < s < t$.

Toda medida de proximidad ordinal puede representarse mediante la *matriz de proximidades*

$$\begin{pmatrix} \pi_{11} & \cdots & \pi_{1s} & \cdots & \pi_{1g} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \pi_{r1} & \cdots & \pi_{rs} & \cdots & \pi_{rg} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \pi_{g1} & \cdots & \pi_{gs} & \cdots & \pi_{gg} \end{pmatrix},$$

que es simétrica y de dimensión $g \times g$, con coeficientes en Δ , y cuyos elementos de la diagonal principal son $\pi_{rr} = \delta_1$, $r = 1, \dots, g$, y $\pi_{1g} = \delta_h$ (véase García-Lapresta y Pérez-Román [6, Prop. 2]).

Las valoraciones dadas por los agentes a cada una de las alternativas son recogidas en un *perfil*

$$V = \begin{pmatrix} v_1^1 & \cdots & v_i^1 & \cdots & v_n^1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ v_1^a & \cdots & v_i^a & \cdots & v_n^a \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ v_1^m & \cdots & v_i^m & \cdots & v_n^m \end{pmatrix},$$

donde el elemento $v_i^a \in \mathcal{L}$ es la valoración lingüística otorgada por el agente $a \in A$ a la alternativa $x_i \in X$.

Una vez fijada la medida de proximidad ordinal y conocidas las valoraciones lingüísticas dadas por los agentes a cada una de las alternativas, puede aplicarse el procedimiento introducido por García-Lapresta y Pérez-Román [7] para ordenar las alternativas.

El procedimiento está dividido en los siguientes pasos.

- 1) Para cada alternativa $x_i \in X$ se seleccionan las valoraciones de x_i otorgadas por todos los agentes: $v_i^1, \dots, v_i^m \in \mathcal{L}$ (columna i de V).
- 2) Para cada alternativa $x_i \in X$ se calculan los grados de proximidad entre las valoraciones dadas por los agentes y la máxima valoración posible de la escala, l_g :

$$\pi(v_i^1, l_g), \dots, \pi(v_i^m, l_g) \in \Delta.$$

- 3) Para cada alternativa $x_i \in X$ se ordenan de forma decreciente los grados de proximidad obtenidos. Una vez ordenados, con objeto de evitar la pérdida de información, se selecciona el par de medianas de grados de proximidad, M_i , de la siguiente forma:

- a) Si el número de valoraciones es par, entonces se tienen en cuenta las dos medianas $M_i = (\delta_r, \delta_s)$, con $r, s \in \{1, \dots, h\}$ tales que $r \leq s$.
- b) Si el número de valoraciones es impar, se duplica la mediana. Por tanto, $M_i = (\delta_r, \delta_r)$ para algún $r \in \{1, \dots, h\}$.

Por consiguiente, $M_i \in \Delta_2$, donde Δ_2 es el *conjunto de medianas factibles*:

$$\Delta_2 = \{(\delta_r, \delta_s) \in \Delta^2 \mid r \leq s\}.$$

- 4) Se ordenan los pares de medianas de grados de proximidad obtenidos previamente para cada una de las alternativas, considerando el orden lineal \succeq sobre Δ_2 definido como

$$\begin{aligned} (\delta_r, \delta_s) \succeq (\delta_t, \delta_u) &\Leftrightarrow \\ \begin{cases} r + s < t + u \\ \text{o} \\ r + s = t + u \text{ y } s - r \leq u - t, \end{cases} & \quad (1) \end{aligned}$$

para cualesquiera $(\delta_r, \delta_s), (\delta_t, \delta_u) \in \Delta_2$.

- 5) Se ordenan las alternativas a través del orden débil \succcurlyeq sobre X definido como $x_i \succcurlyeq x_j \Leftrightarrow M_i \succeq M_j$, donde \succeq es el orden lineal definido en (1).

Al ordenar las alternativas es posible que algunas de ellas estén empatadas por tener el mismo par de medianas. En estos casos puede utilizarse un proceso secuencial de desempate basado en Balinski y Laraki [1]. Este procedimiento consiste en eliminar las dos medianas de las alternativas que están empatadas, seleccionar el nuevo par de medianas de entre los grados de proximidad restantes y aplicar el procedimiento del paso 4 hasta romper los empates y ordenar todas las alternativas.

III. APLICACIÓN A LAS ESCALAS CUALITATIVAS DEL CENTRO DE INVESTIGACIONES SOCIOLÓGICAS

En esta sección se presenta una comparación entre la metodología de agrupación de varios términos lingüísticos y el procedimiento puramente ordinal de la sección II.

Para ello se han tomado los datos de la pregunta 20 del Barómetro del CIS de mayo de 2011 [2], en la cual se pide a los individuos que manifiesten, a través de la escala cualitativa de la Tabla I, hasta qué punto confiarían en la ayuda que les puedan proporcionar las nueve alternativas contenidas en la Tabla II ante una situación de catástrofe o desastre.

La muestra utilizada asciende a 2.219 individuos, de los cuales 1.114 son hombres y 1.105 son mujeres.

Para la aplicación del procedimiento ordinal introducido en la sección II es necesario utilizar una medida de proximidad



ordinal adecuada. Con objeto de establecer esta medida, se realizó una encuesta acerca de las proximidades entre los términos de la escala de la Tabla I dirigida a estudiantes de cuarto curso del grado en Economía de la Universidad de Valladolid. Una vez conocidas sus respuestas, se obtuvieron las medidas de proximidad ordinal para cada uno de los estudiantes y se aplicó el procedimiento de agregación introducido por García-Lapresta et al. [5] para determinar la medida de proximidad ordinal que mejor representaba las opiniones de los encuestados.

La medida de proximidad ordinal resultante es la asociada a la matriz

$$A_{232} = \begin{pmatrix} \delta_1 & \delta_2 & \delta_4 & \delta_5 \\ & \delta_1 & \delta_3 & \delta_4 \\ & & \delta_1 & \delta_2 \\ & & & \delta_1 \end{pmatrix}$$

representada en la Fig. 1.



Fig. 1. Medida de proximidad ordinal asociada a la matriz A_{232} .

Una vez determinada la medida de proximidad ordinal, se ordenan las nueve alternativas de la Tabla II mediante aplicación del procedimiento ordinal de la sección II.

Para ello, se calculan los grados de proximidad ordinal entre las valoraciones dadas por los agentes y la máxima valoración posible de la escala (l_4). Seguidamente, estos grados de proximidad se ordenan de mayor a menor proximidad y se selecciona de cada alternativa su par de medianas.

Los pares de medianas obtenidos “considerando la población total” son los siguientes: $M_1 = M_2 = M_3 = M_4 = M_5 = M_6 = M_8 = M_9 = (\delta_2, \delta_2)$ y $M_7 = (\delta_1, \delta_1)$.

Aplicando (1) se obtiene $M_7 \succ M_1 = M_2 = M_3 = M_4 = M_5 = M_6 = M_8 = M_9$, luego la alternativa x_7 (Familiares) ocupa la primera posición del orden.

A continuación, con objeto de ordenar el resto de alternativas, se aplica el procedimiento de desempate descrito en la sección II, obteniéndose las siguientes medianas: $M_1^{(1)} = M_2^{(1)} = M_4^{(1)} = M_5^{(1)} = M_6^{(1)} = M_8^{(1)} = M_9^{(1)} = (\delta_2, \delta_2)$ y $M_3^{(1)} = (\delta_1, \delta_2)$.

Entonces por (1) $M_3^{(1)} \succ M_1^{(1)} = M_2^{(1)} = M_4^{(1)} = M_5^{(1)} = M_6^{(1)} = M_8^{(1)} = M_9^{(1)}$, luego la alternativa x_3 (Bomberos) ocupa la segunda posición del orden.

El procedimiento de desempate se aplica las veces que sean necesarias hasta ordenar las nueve alternativas.

Finalmente, los resultados obtenidos después de la aplicación del procedimiento ordinal aparecen recogidos en la Tabla III. En dicha tabla también se muestran los órdenes obtenidos al agrupar los términos “Mucho” y “Bastante”.

Observando los resultados de la Tabla III cabe señalar, en primer lugar, que no existen diferencias en las posiciones ocupadas por las alternativas vecinos y ONGs, las cuales ocupan

Tabla I
TÉRMINOS LINGÜÍSTICOS

l_1	Nada
l_2	Poco
l_3	Bastante
l_4	Mucho

Tabla II
ALTERNATIVAS

x_1	Guardia Civil
x_2	ONGs
x_3	Bomberos
x_4	Protección Civil
x_5	Policía
x_6	Ejército
x_7	Familiares
x_8	Vecinos
x_9	Amigos

en todos los órdenes de la tabla la séptima y octava posición, respectivamente. En cambio, sí que existen diferencias en la forma de ordenar el resto de alternativas.

Si no se hace distinciones entre género, las diferencias entre los órdenes obtenidos mediante la agrupación de los términos lingüísticos y la aplicación del procedimiento ordinal se encuentran en las alternativas Protección Civil, Ejército, Policía y Guardia Civil. Como puede observarse, Protección Civil y el Ejército, que ocupan la cuarta y la quinta posición, respectivamente, en el orden resultante de agrupar de “Mucho” y “Bastante”, invierten sus posiciones en el caso del orden obtenido tras aplicar el procedimiento ordinal, sucediendo lo mismo en el caso de la Policía y la Guardia Civil.

En cuanto a los hombres, existen diferencias significativas en las primeras posiciones de los órdenes. Los bomberos y los familiares ocupan la primera y la segunda posición, respectivamente, en el orden obtenido tras la agrupación de los términos “Mucho” y “Bastante”. Sin embargo, si se considera el procedimiento ordinal, los familiares se sitúan en la primera posición seguidos por los bomberos. Protección Civil y Ejército también sufren modificaciones, invirtiéndose las posiciones que ocupan en los órdenes resultantes.

En el caso de las mujeres, las diferencias entre los dos órdenes son poco relevantes, variando únicamente las posiciones de la Policía y la Guardia Civil.

IV. CONCLUSIONES

La representación de las escalas cualitativas a través de valores numéricos exactos es muy común en el ámbito de las Ciencias Sociales debido a las facilidades que proporciona la Estadística a la hora de manejar datos numéricos. Sin embargo, este enfoque cuantitativo de la información cualitativa carece de sentido en el contexto de las escalas cualitativas no uniformes, ya que contradice algunos principios de la teoría de la medición (véase Roberts [9]).

Por otra parte, la práctica utilizada por el CIS y otros organismos consistente en presentar los resultados de sus

Tabla III
 ÓRDENES OBTENIDOS TRAS LA AGRUPACIÓN DE “MUCHO” Y
 “BASTANTE” Y LA APLICACIÓN DEL PROCEDIMIENTO ORDINAL

[10] Zimmer, A.C.: Verbal vs. numerical processing of subjective probabilities. *Decision Making Under Uncertainty* 16, pp. 159-182, 1983.

Agrupación “Mucho” y “Bastante”

Total		Hombres		Mujeres	
1	Familiares	1	Bomberos	1	Familiares
2	Bomberos	2	Familiares	2	Bomberos
3	Amigos	3	Amigos	3	Amigos
4	Protección Civil	4	Protección Civil	4	Protección Civil
5	Ejército	5	Ejército	5	Ejército
6	Policía	6	Guardia Civil	6	Policía
7	Guardia Civil	7	Policía	7	Guardia Civil
8	Vecinos	8	Vecinos	8	Vecinos
9	ONGs	9	ONGs	9	ONGs

Procedimiento ordinal

Total		Hombres		Mujeres	
1	Familiares	1	Familiares	1	Familiares
2	Bomberos	2	Bomberos	2	Bomberos
3	Amigos	3	Amigos	3	Amigos
4	Ejército	4	Ejército	4	Protección Civil
5	Protección Civil	5	Protección Civil	5	Ejército
6	Guardia Civil	6	Guardia Civil	6	Guardia Civil
7	Policía	7	Policía	7	Policía
8	Vecinos	8	Vecinos	8	Vecinos
9	ONGs	9	ONGs	9	ONGs

investigaciones mediante la agrupación de varios términos lingüísticos tampoco está considerando la no uniformidad de las escalas cualitativas. Además supone una pérdida importante de información al descartar parte de las categorías de respuesta de las escalas.

Los resultados obtenidos en este trabajo tras aplicar el procedimiento ordinal introducido recientemente por García-Lapresta y Pérez-Román [7] evidencian la importancia de no omitir información cualitativa de las escalas y considerar su no uniformidad.

AGRADECIMIENTOS

Los autores agradecen al Ministerio de Economía y Competitividad (referencia ECO2016-77900-P) y al Fondo Europeo de Desarrollo Regional (FEDER) la financiación recibida.

REFERENCIAS

- [1] M. Balinski, R. Laraki, A theory of measuring, electing and ranking. *Proceedings of the National Academy of Sciences of the United States of America* 104 (2007) 8720–8725.
- [2] Centro de Investigaciones Sociológicas. Barómetro de mayo de 2011.
- [3] Centro de Investigaciones Sociológicas. Anuario 2016.
- [4] Franceschini, F., Galetto, M., Varetto, M.: Qualitative ordinal scales: the concept of ordinal range. *Quality Engineering* 16, pp. 515-524, 2004.
- [5] García-Lapresta, J.L., González del Pozo, R., Pérez-Román, D.: Metrizable ordinal proximity measures and their aggregation, *Information Sciences*, 448-449, pp. 149-163, 2018.
- [6] García-Lapresta, J.L., Pérez-Román, D.: Ordinal proximity measures in the context of unbalanced qualitative scales and some applications to consensus and clustering. *Applied Soft Computing* 35, pp. 864-872, 2015.
- [7] García-Lapresta, J.L., Pérez-Román, D.: Aggregating opinions in non-uniform ordered qualitative scales. *Applied Soft Computing* 67, pp. 652-657, 2018.
- [8] Likert, R.: A technique for the measurement of attitudes. *Archives of Psychology* 22 (140), pp. 1-55, 1932.
- [9] Roberts, F.S.: *Measurement Theory*. Cambridge University Press., 1979.