

IMPACTO DE LAS ETIQUETAS EN LA INTERPRETACIÓN DE LA ESCALA DE LIKERT

Nuria Martínez¹, Daniel Gómez², Javier Montero¹

¹ Facultad de Matemáticas, Universidad Complutense de Madrid, 28040, Madrid

² Escuela de Estadística, Universidad Complutense de Madrid, 28040, Madrid

nuriamartinez@gmail.com, dagomez@estad.ucm.es, monty@mat.ucm.es

Resumen

En este trabajo se comparan las distribuciones de las respuestas dadas a un cuestionario formulado en dos versiones, una con etiquetas representando el grado de acuerdo, frente a otra con etiquetas expresando la frecuencia de realización de una acción. Se examina la relación entre la percepción de cada etiqueta lingüística y las etiquetas frecuentistas mediante números borrosos. Se analiza la equivalencia entre ambos métodos, estudiando en detalle la equidistancia en la percepción de las etiquetas extremas. Se evalúan además las diferencias en función de la importancia del tema sobre el que se interroga.

Palabras Clave: Lógica difusa, escala de Likert, frecuencia, etiqueta, proceso de decisión de compra, acciones sostenibles

1 INTRODUCCIÓN

En la investigación social es frecuentemente necesaria la medición de actitudes, opiniones o intensidades de un concepto. Para ello, es habitual el uso de diferentes escalas que generan niveles ordinales del atributo medido.

La escala de Likert es muy utilizada por su rápida construcción, aplicación e interpretación y por la suposición de que es posible aumentar la precisión de la respuesta incluyendo más categorías y suponiendo la escala continua [17]. Esta escala consiste en que para cada ítem el encuestado asigna una puntuación de acuerdo a su actitud respecto al ítem propuesto. La puntuación se selecciona de un rango de valores cerrado, a menudo de 1 a 5, aunque pueden utilizarse rangos más amplios (1 a 7, 1 a 10... etc).

La escala de Likert presenta algunos problemas [11] como la supuesta unidimensionalidad de las actitudes, la localización en un espacio continuo y la suposición de la equidistancia entre las alternativas expuestas.

Son varios los autores que han estudiado en los últimos años la influencia de la escala utilizada, del número de

opciones y de la formulación de las respuestas incluidas en los cuestionarios [13], [17], planteando cuestiones abiertas como si el contenido del cuestionario tendrá influencia sobre las respuestas extremas [17].

En los cuestionarios de marketing es frecuente disponer de datos de muy diferente naturaleza, encontrando en el mismo cuestionario preguntas de tipo difuso, así como preguntas de carácter más frecuentista. Un ejemplo de pregunta de naturaleza difusa sería: *Expresar el grado de acuerdo con la siguiente frase: "Compro productos con ofertas especiales"*, mientras que una formulación frecuentista alternativa sería: *"¿Con qué frecuencia compra productos con ofertas especiales?"*.

Este trabajo plantea la cuestión de si las respuestas y conclusiones finales de un estudio varían dependiendo de la formulación de las opciones de respuesta. Para ello, se realizan dos versiones de un mismo cuestionario, formulando las preguntas de modo que la primera versión admita respuestas sobre el grado de acuerdo y la segunda versión sobre la frecuencia de realización.

La utilización de la lógica borrosa es amplia en muchas disciplinas, y en particular en los campos sociales. Concretamente se utiliza a menudo en estudios de evaluación o satisfacción, existiendo diversas investigaciones relacionadas con experimentos en línea con lo aquí presentado. Existen artículos describiendo experimentos tales como asignar de forma libre etiquetas lingüísticas y valores numéricos a las mismas para estudiar su relación [13],[16], presentar diferentes escalas lingüísticas y completar con la nota numérica para analizar la equivalencia y distancia entre ellas [11] u otros en los que se ofrecen escalas numéricas y verbales y se pide que se relacionen ambas [9].

Vemos por tanto que son muchos los autores que han abordado el problema sobre la dimensión numérica de variables lingüísticas. Nuestro objetivo es analizar la equivalencia, dentro de una encuesta construida sobre una escala subyacente de Likert, de una formulación de preguntas expresando el grado de acuerdo con una frase que

conlleve una acción y la formulación de la misma pregunta indicando la frecuencia con la que se realiza esta acción. Encontramos, por lo expuesto anteriormente, que es de gran interés profundizar en el tipo de relación existente entre etiquetas lingüísticas y etiquetas frecuentistas que a menudo parecen considerarse equivalentes, dada la mezcla que existe en la práctica entre ambas formulaciones.

En la sección 2 de este documento se describen el método y los materiales de esta investigación. En la sección 3 se presentan los resultados. En concreto, se recogerá la incertidumbre generada en torno a cada etiqueta lingüística mediante números borrosos. A continuación se analizará si ambos métodos son equivalentes o si existe algún tipo de tendencia o patrón en las respuestas. Se abordará también la cuestión encontrada en la literatura consultada [17] sobre si el tema interrogado puede ser un factor influyente en las diferencias de los resultados. Para ello, el cuestionario incluirá dos temas de distinta relevancia en la vida del encuestado. Finalmente, teniendo en cuenta los problemas mencionados sobre equidistancia, se estudiará un comportamiento encontrado en los gráficos por el que parece que la dispersión de las respuestas en torno al valor extremo “Muy de acuerdo” es mayor que la dada en torno al valor opuesto “Muy en desacuerdo”. Se querrá detectar si estos conceptos lingüísticos corresponden a un rango de frecuencias diferente, no siendo por tanto equidistantes (ver Figura 1). Finalmente, en la sección 4 se exponen posibles líneas futuras de trabajo.

2 MATERIALES Y MÉTODO

El instrumento utilizado en este estudio es un cuestionario que interroga sobre dos temas. El primero consta de 13 preguntas (C1, C2... C13) relacionadas con las acciones llevadas a cabo durante el proceso de decisión de compra de un champú y el segundo se compone de 14 preguntas (S1, S2... S14) sobre la consideración de la sostenibilidad medioambiental en las acciones diarias del entrevistado.

El cuestionario se formula en dos versiones, atendiendo a la naturaleza de las respuestas:

- **Cuestionario difuso:** las respuestas a cada pregunta son definidas mediante etiquetas lingüísticas

- D1. Muy de acuerdo
- D2. Bastante de acuerdo
- D3. Ni de acuerdo, ni en desacuerdo
- D4. Bastante en desacuerdo
- D5. Muy en desacuerdo

- **Cuestionario frecuentista:** las respuestas a cada pregunta son definidas mediante frecuencia de realización

- F1. Alrededor del 100% de las veces
- F2. Alrededor del 75% de las veces

F3. Alrededor del 50% de las veces

F4. Alrededor del 25% de las veces

F5. Alrededor del 0% de las veces

Previo al lanzamiento del campo, se llevó a cabo una fase de testeo de ambas versiones del cuestionario con el objeto de asegurar la correcta expresión y comprensión de cada pregunta, así como la equivalencia semántica de ambos cuestionarios. Con el análisis de los resultados de esta fase de test y los comentarios recibidos se redactaron los cuestionarios finales.

La primera parte del cuestionario se construye en torno a los factores de decisión de compra de un champú, conteniendo preguntas sobre las diferentes etapas del proceso habitual de compra de un producto (ver [7], [8], [14]). La segunda parte trata diferentes acciones sostenibles de los consumidores descritas en manuales de organizaciones públicas (ver [2], [10]) susceptibles de ser formuladas con etiquetas lingüísticas y con etiquetas frecuentistas. Se incluye una pregunta de filtro para asegurar que los encuestados han comprado champú en los últimos doce meses y tienen por tanto criterio para responder a la primera parte del cuestionario.

Se define un público objetivo que permita obtener una muestra suficiente para el análisis por lo que se decide llevar a cabo el estudio en el ámbito de la Universidad. La captación se realiza mediante contactos en colegios mayores, anuncios en diferentes facultades de la Universidad Complutense de Madrid y la publicación del estudio en la página de Facebook “Doctorado Matemáticas”

Se recoge una muestra que, tras eliminar las encuestas incompletas y depurar los datos, consta de 95 estudiantes universitarios que completaron ambas fases del estudio. Esta muestra se compone de un 58% de estudiantes con perfil de ciencias y un 42% con perfil de letras, habiendo un 17% de hombres y 83% de mujeres. Este último desbalanceo es probablemente debido a la pregunta de filtro sobre la compra de champú en los últimos doce meses.

La versión difusa del cuestionario se llevó a cabo mediante encuestas auto-administradas online desarrolladas con el software SurveyMonkey durante el mes de Febrero de 2011. Dos semanas después se lanzó el cuestionario frecuentista, prolongando el campo hasta finales de Abril.

El análisis de este estudio requiere identificar las respuestas de un mismo entrevistado. Para ello, a cada individuo se le pidió que incluyese las iniciales de su nombre y primer apellido así como su fecha de nacimiento en ambos cuestionarios de forma que se pudiese generar un identificador único manteniendo el anonimato.

Para poder alcanzar a los mismos individuos en el segundo cuestionario, se pidió en la primera fase que incluyesen su dirección de correo electrónico. Se utilizó además la mencionada página de Facebook “Doctorado Matemáticas” para que aquellas personas que fuesen reacias a proporcionar este dato, pudiesen hacerse “fan” de la misma y ser alertadas de forma anónima en su perfil de Facebook cuando se lanzase el segundo cuestionario.

El análisis estadístico de los datos se llevó a cabo con Excel y con los paquetes estadísticos STATA y R.

3 RESULTADOS Y CONCLUSIONES

Se tienen por tanto 27 preguntas contestadas por 95 individuos aplicando dos métodos distintos (método difuso y método frecuentista).

Para cada pregunta se muestra a continuación un gráfico de burbujas con los datos de la tabla de contingencia asociada, teniendo en el eje horizontal las categorías difusas y en el vertical las frecuentistas:

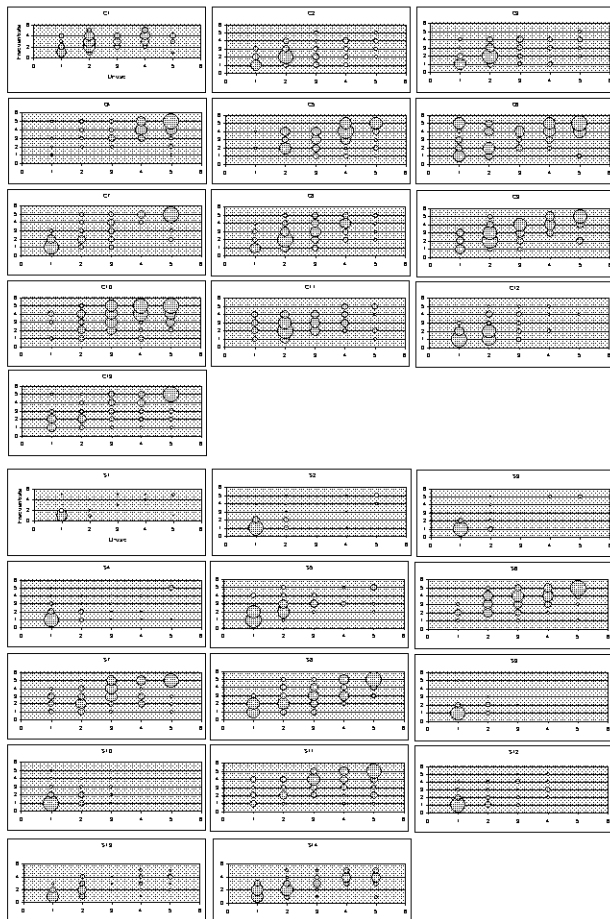


Figura 1: Gráfico de respuestas Difuso vs. Frecuentista, para cada pregunta

Se observa, como era de esperar, que en general las diagonales tienen el mayor número de casos, teniendo el 100% de ellos si ambos métodos fueran equivalentes. Nótese además que parece existir una mayor equivalencia en el segundo bloque de preguntas sobre sostenibilidad que en el primero. La figura anterior muestra también el comportamiento comentado previamente por el que a respuestas difusas “Muy de acuerdo” parece corresponderles una distribución frecuentista más amplia que la equivalente en respuestas “Muy en desacuerdo”.

Para analizar las posibles diferencias en la percepción frecuentista de las etiquetas lingüísticas que definen el grado de acuerdo (D1, D2... D5, ver 2 MATERIALES Y MÉTODO), se realizó el cálculo de números borrosos triangulares que tratan de establecer esta relación. Un número borroso triangular es un subconjunto borroso A de X cuya función de pertenencia $\mu_A: U \rightarrow [0,1]$ viene definida por:

$$\mu_A(u) = \begin{cases} \frac{(u-a)}{b-a} & \text{si } a \leq u \leq b \\ \frac{(u-c)}{b-c} & \text{si } b \leq u \leq c \\ 0 & \text{en otro caso} \end{cases}$$

El número borroso A puede ser por tanto representado mediante la tripleta $A = (a, b, c)$ con $a \leq b \leq c$. En este estudio se ha tomado $b = \text{moda}(X)$, $a = b - 2*s$ y $c = b + 2*s$, donde s es la desviación estándar.

Los resultados (ver Figura 2) muestran que las funciones de pertenencia son muy variadas, habiendo un 35% de ellos en los que las modas no se corresponden con la frecuencia esperada de la etiqueta asociada. El 22% muestra un desvío de la distribución frecuentista hacia la derecha del valor que le correspondería. Este comportamiento se encontrará también en el gráfico boxplot de diferencias entre las respuestas (ver Figura 3) y corrobora un patrón que se mencionará posteriormente sobre la equivalencia de los métodos. La percepción de la etiqueta “Ni de acuerdo, ni en desacuerdo” arroja, como era de esperar, el mayor porcentaje de casos de no coincidencia entre la moda de la distribución y la frecuencia “Alrededor del 50% de las veces”, es decir, es donde existe una mayor incertidumbre en la equivalencia entre la etiqueta lingüística y la frecuencia de realización. Además, existen en general desviaciones en torno a la moda que varían hasta alcanzar gran amplitud, siendo ésta mayor en las preguntas relacionadas con el proceso de compra de un champú que en el tema de sostenibilidad.

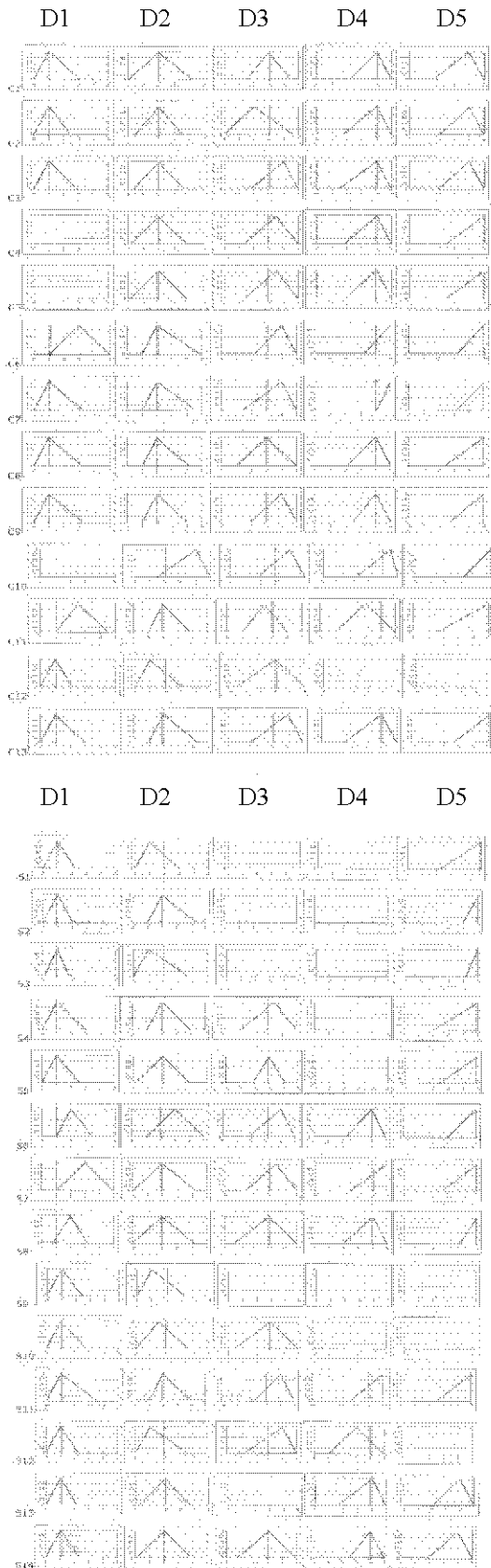


Figura 2: Números borrosos triangulares por pregunta para cada etiqueta lingüística

En cuanto a la equivalencia de los métodos, los análisis llevados a cabo no permiten concluir de forma general que exista evidencia para afirmar que los resultados con uno y otro método son significativamente distintos. La extensión de Stuart y Maxwell al test de McNemar para estudiar la homogeneidad marginal [4] permite detectar algunos ítems, 7 de 27, en los que existe evidencia para rechazar que las muestras marginales sean homogéneas:

CHAMPU		SOSTENIBILIDAD	
C1	0,0084	S1	0,5747
C2	0,6824	S2	0,0765
C3	0,0376	S3	0,1556
C4	0,5003	S4	0,4344
C5	0,6297	S5	0,00001
C6	0,0169	S6	0,0388
C7	0,0673	S7	0,0551
C8	0,3446	S8	0,4352
C9	0,1924	S9	0,6518
C10	0,2345	S10	0,0489
C11	0,4463	S11	0,1039
C12	0,0698	S12	0,3904
C13	0,4411	S13	0,0410
		S14	0,0902

Tabla 1: P-valores del contraste de homogeneidad marginal mostrando en rojo aquéllos inferiores a 0,05

Algunas de estas preguntas se muestran también relevantes en otros tests y análisis realizados coincidiendo todas ellas en que tienen una formulación lingüística más compleja que el resto.

Para cada pregunta se estudia el grado de asociación mediante la medida γ de Goodman y Kruskal ([1],[4]) así como el grado de acuerdo mediante el índice Kappa ([1],[3],[4]). En general se observa que existe asociación elevada en las respuestas dadas por cada uno de los métodos. El acuerdo encontrado no es excesivamente alto en muchas de las preguntas, pero, dadas las limitaciones y la sensibilidad del índice Kappa a la distribución de las marginales (ver, por ejemplo, [3], [15], [18]), y el uso meramente descriptivo hecho del índice, no se puede extraer ninguna conclusión directa.

Se aprecia que las preguntas dedicadas a sostenibilidad muestran mayor grado de asociación y acuerdo que las relacionadas con el proceso de compra de un champú.

El análisis de las diferencias entre las respuestas mediante el gráfico boxplot (ver Figura 3) muestra un patrón interesante: salvo valores atípicos, en 21 de las 27 preguntas sucede para al menos la mitad de los individuos que, o bien no hay diferencias ($d_i - f_i = 0$), o bien existe una tendencia por la que las personas declaran una frecuencia de realización de la acción inferior a la etiqueta lingüística seleccionada en el cuestionario difuso ($d_i - f_i = -1 \rightarrow d_i = f_i - 1$).

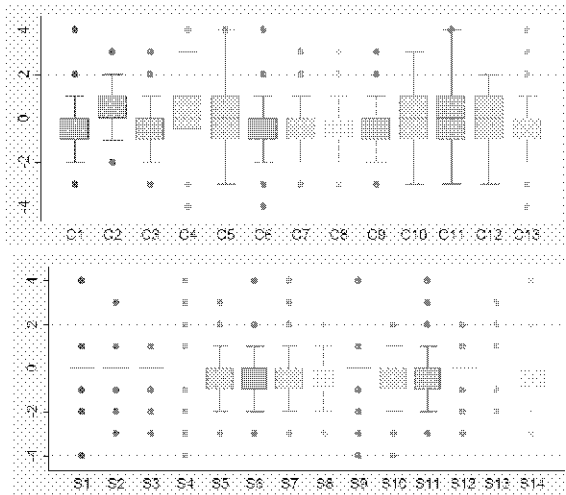


Figura 3: Boxplot de las diferencias de las respuestas dadas por cada individuo

Por ejemplo, una persona podría expresar que está “Bastante de acuerdo” con una pregunta en el cuestionario difuso, y declarar en el cuestionario frecuentista que lo lleva a cabo alrededor de la mitad de las veces (cuando la etiqueta equivalente en frecuentista sería “Alrededor del 75% de las veces”). Es decir, cuando no hay equivalencia directa entre los métodos, parece existir una tendencia a declarar un mayor acuerdo en la pregunta con etiquetas lingüísticas que en la pregunta formulada con etiquetas frecuentistas:

	di = fi - 1		
	di	fi	
Muy de acuerdo	1	2	Alrededor del 75% de las veces
Bastante de acuerdo	2	3	Alrededor del 50% de las veces
Ni de acuerdo, ni en desacuerdo	3	4	Alrededor del 25% de las veces
Bastante en desacuerdo	4	5	Alrededor del 0% de las veces
Muy en desacuerdo	5	-	

Tabla 2: Relación entre las respuestas de una pregunta para el individuo i-ésimo en el caso analizado

A la vista de los resultados obtenidos y los análisis realizados, entre otros el gráfico boxplot de diferencias y los niveles de asociación y acuerdo, podría concluirse que el tema sobre el que se interroga sí es un factor determinante en la posible equivalencia entre las distribuciones de las respuestas difusas o frecuentistas. Así, podría pensarse que temas con cierta relevancia social o simplemente “políticamente correctos”, como es la realización de acciones sostenibles en el día a día, arrojan resultados similares independientemente de las etiquetas con las que se pregunte. En cambio, temas menos importantes en la vida del entrevistado, como son las acciones previas a la decisión de compra de un champú, sí parecen verse más influidos por las opciones de respuesta del cuestionario.

Para analizar la equidistancia de las respuestas frecuentistas en torno a las etiquetas lingüísticas “Muy de acuerdo” (D1) y “Muy en desacuerdo” (D5) se calcula el índice de entropía de la distribución frecuentista condicionada a los

valores D1 ó D5 ([5],[12]). Se define la entropía de la v.a. X condicionada al valor Y = y_j como:

$$H(X|Y=y_j) = - \sum_{i=1}^I p(x_i|y_j) \log_2 p(x_i|y_j)$$

Los resultados (ver Tabla 3) arrojan que existen varias preguntas para las que se confirma la hipótesis de una mayor entropía condicionada a D1 que la resultante condicionando a D5, no siendo éstas suficientes para concluir un resultado de forma general.

Pregunta	CHAMPU		SOSTENIBILIDAD		
	Entropía condicionada a... D1	D5	Pregunta	Entropía condicionada a... D1	D5
C1	1,59	1,85	S1	0,78	0,54
C2	1,35	1,84	S2	0,84	0,95
C3	1,38	1,79	S3	0,87	0,88
			S4	1,27	0,59
			S5	1,28	1,25
C6	2,07	1,19	S6	1,25	1,12
C7	1,20	0,99	S7	1,86	0,81
C8	1,55	1,92	S8	1,23	0,84
C9	1,55	1,38			
C11	1,95	1,84	S11	1,87	1,88
C13	1,83	1,12	S13	0,97	1,66
			S14	1,40	1,67

Tabla 3: Resultados de la entropía condicionada para cada pregunta indicando en rojo aquellos casos en los que se cumple la hipótesis planteada

4 TRABAJOS FUTUROS

Para investigaciones futuras consideramos interesante tratar de llevar a cabo un análisis lingüístico de las preguntas para las que se han detectado diferencias significativas y comportamientos distintos en las respuestas. Esto permitiría extraer patrones acerca de cómo están estipuladas estas preguntas concretas y ver qué diferencias existen con respecto al resto.

Encontramos también de interés realizar un nuevo experimento similar teniendo en cuenta varios temas de diferente relevancia en la vida del encuestado, para estudiar si se repite el comportamiento encontrado en este estudio de a mayor relevancia (individual o social), mayor equivalencia entre los dos métodos.

Otra cuestión que no se ha analizado, pero que convendría tener en cuenta en futuras investigaciones, sería determinar la discrepancia natural en un cuestionario. Es decir, conocer el porcentaje de discrepancia que encontraríamos si repitiésemos dos veces un cuestionario idéntico a la misma muestra de individuos. Esto nos permitiría determinar cuánta de la discrepancia encontrada en el estudio se debe a la nueva formulación de las etiquetas de las respuestas y cuánta sería la discrepancia natural.

Un factor a analizar en el futuro es tratar de discernir el impacto de la percepción que tienen los entrevistados acerca del tipo de escala subyacente.

Finalmente, otro tema susceptible de estudio sería recoger si existen diferencias en las respuestas en función del perfil de estudios del entrevistado. Es decir, si universitarios de carreras de ciencias (tecnológicas, bio-sanitarias) o de letras (sociales o de humanidades) responden de manera significativamente distinta en un cuestionario de naturaleza difusa o frecuentista.

Agradecimientos

Este trabajo se ha realizado con financiación parcial del proyecto TIN2009-07901 del Plan Nacional de Investigación.

Referencias

- [1] Análisis de datos cualitativos: tablas de contingencia, (2007). Sociedad española de bioquímica clínica y patología molecular (SEQC)
- [2] Consumo sostenible, el equilibrio entre la acción humana y el planeta, (2009). *Guía de consumo responsable*. Consultado en Noviembre de 2010. Donostia-San Sebastián: Servicio Central de Publicaciones del Gobierno Vasco: http://www.ejgv.euskadi.net/r53-2291/es/contenidos/informacion/publi_didactico/es_1302/adjuntos/consumo_sostenible.pdf
- [3] Abaira V. Errores en las mediciones y clasificaciones clínicas: Precisión y validez (n.d.). Hospital Universitario Ramón y Cajal. http://www.hrc.es/bioest/Intro_errores.html
- [4] Aguilera del Pino, A.M. Análisis de tablas de contingencia bidimensionales. Abril 2005, <http://www.ugr.es/~focana/dclasif/aaguilera.pdf>
- [5] Cadena Monroy L.A. (2006). Entropía: Información y Complejidad. *Senderos ambientales. VOL. 1 NO. 1, pp. 81 - 88*
- [6] Gómez D., Biging G. y Montero J.(2008). Accuracy statistics for judging soft classification. *International Journal of Remote Sensing*, Vol 29, pp 693-709
- [7] González Cobo M.A. y Carrero López E. (1997). *Manual de planificación de Medios*. Madrid: ESIC
- [8] Howard J.A. (1993). El comportamiento del Consumidor en la estrategia de Marketing. Madrid : Diaz de Santos.
- [9] Jaffe-Katz A., Budescu D.V., Wallsten T. (1989). Timed magnitude comparisons of numerical and nonnumerical expressions of uncertainty. *Memory&Cognition* Vol. 17 n°3, pp. 249-264
- [10] Jóvenes por el cambio. Manual de educación para un consumo sostenible META (Asociación de los Medios, Ecología y Tecnología), División de Tecnología, Industria y Economía del PNUMA (Programa de Naciones Unidas para el Medio Ambiente) y la División para la Enseñanza Secundaria, Técnica y Profesional de la UNESCO. http://www.oei.es/decada/portadas/YXC_red.pdf
- [11] Lalla M., Facchinetti G. y Mastroleo G. (2004). Ordinal Scales and Fuzzy Set Systems to Measure Agreement: An Application to the Evaluation of Teaching Activity, *Quality&Quantity* 38: 577 – 601, Kluwer Academic Publishers, Netherlands
- [12] Larrañaga P. e Inza I. Teoría de la información estadística. Departamento de Ciencias de la Computación e Inteligencia Artificial. Universidad del País Vasco: <http://www.sc.ehu.es/isc/>
- [13] Martínez García J.A. y Martínez Caro L. (2010). La medición de la satisfacción del consumidor de servicios deportivos a través de la lógica borrosa. *Revista de Psicología del Deporte* Vol. 19, n° 1, pp. 41-58
- [14] Solé Moro M.L. (2003). *Los consumidores del siglo XXI*. Madrid: ESIC.
- [15] Viera A.J., Garret J.M. (2005). Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine* Vol.37, n°5, pp. 360 - 363
- [16] Wallsten T.S., Budescu D.V. y Zwick R.(1993). Comparing the Calibration and Coherence of Numerical and Verbal Probability Judgments. *Management Science* Vol. 39, n°2, pp.176 - 190
- [17] Watkins M.L. (1992). The implications of extreme response style (ERS) for cross-cultural and comparative research in South Africa. *Journal of Industrial Psychology*, 18(1), pp. 13-19.
- [18] Yimprayoon P. y Ruktamatakul S. (2009). Some Statistical Aspects of Measuring Agreement Based on a Modified Kappa. *Kasetsart Journal (Natural Science)*, n°43, pp. 389 - 397