
INVESTIGACIÓN DEL COMPORTAMIENTO

Innovaciones metodológicas y estrategias de docencia



Compiladores

Vicente Manzano Arrondo
Manuel Sánchez García

Instituto Psicosociológico Andaluz de Investigaciones

Compiladores:

Vicente Manzano Arrondo. *Universidad de Sevilla*
Manuel Sánchez García. *Universidad de Huelva*

Autores:

Carlos Arias Martín. *Universidad de Sevilla*
Carlos Camacho Martínez Vara de Rey. *Universidad de Sevilla*
Raquel Cruz del Pino. *Universidad de Almería*
Salvador Chacón Moscoso. *Universidad de Sevilla*
Hassan Fazelli Khalili. *Universidad de Sevilla*
Juan Sebastian Fernández Prados. *Universidad de Almería*
X. Gabriel Vázquez. *Universidad de La Coruña*
Alicia García Pereira. *Universidad de Sevilla*
Olivia Giménez Fernández. *Universidad de Almería*
Andrés González Gómez. *Universidad de Granada*
Vicente Manzano Arrondo. *Universidad de Sevilla*
Victoria Márquez de la Plata y Cuevas. *Universidad de Sevilla*
Rafael J. Martínez Cervantes. *Universidad de Sevilla*
Regla Martínez Jiménez. *Universidad de Sevilla*
José Antonio Mayor Gallego. *Universidad de Sevilla*
Rafael Moreno Rodríguez. *Universidad de Sevilla*
José Luis Padilla García. *Universidad de Granada*
Cristino Pérez Meléndez. *Universidad de Granada*
Francisco Javier Pérez Santamaría. *Universidad de Sevilla*
Alicia Puerto Martínez. *Universidad de Sevilla*
Antonio Rial Boubeta. *Universidad de Santiago de Compostela*
Antonio José Rojas Tejada. *Universidad de Almería*
Eleuterio Francisco Sánchez García. *Universidad de Sevilla*
Manuel Sánchez García. *Universidad de Huelva*
Santiago Varela Mallou. *Universidad de Santiago de Compostela*

Esta obra no puede ser reproducida en parte o en su totalidad, sin consentimiento expreso del Instituto Psicosociológico Andaluz de Investigaciones. La reproducción comprende fotocopia, copia impresa, magnética o electrónica, utilizando cualquier medio de difusión, para uso individual o colectivo.

El Instituto Psicosociológico Andaluz de Investigaciones es una asociación sin ánimo de lucro cuyo objetivo es favorecer la realización de estudios sobre el comportamiento. En esta línea se encuentra el fomento de las publicaciones que pueden servir de guía para el fundamento metodológico o el desarrollo de aplicaciones sobre el comportamiento.

© Instituto Psicosociológico Andaluz de Investigaciones

Servicio de Publicaciones

Avd. Pablo Iglesias, 57, 4º D; 04003 - Almería

e-mail: arojas@ualm.es

tel: 950 215162 Fax: 950 215420

Depósito Legal: SE 2503-1998

ISBN: 84-930-2470-8

Las contribuciones que contiene esta obra son fruto de las ponencias que diversos especialistas presentaron en el Primer Congreso de Docencia e Investigación del Comportamiento que tuvo lugar en Sevilla, en Octubre de 1998, organizado por el Instituto Psicosociológico Andaluz de Investigaciones.

El comportamiento es, precisamente, el objeto de estudio de la Psicología. No obstante, en varios sentidos, comparte elementos relacionados con su objeto, con otras disciplinas científicas. Así, algunos aspectos del comportamiento son interés de la Pedagogía (comportamiento relacionado con la educación), de la Economía (comportamiento del comprador, de los mercados, de los trabajadores, etc...) de la Sociología (comportamiento colectivo) de la Medicina (disfunciones generadas por el comportamiento o con huella en él) o de la Biología (bases biológicas del comportamiento). A su vez, para abordar su objeto de estudio, la Psicología recurre a herramientas que son, precisamente, el interés principal de otros campos del saber, como la Estadística, la Investigación Operativa o la Informática. Así pues, la investigación del comportamiento es de interés multidisciplinar y, por ello, tiene sentido que profesionales con formaciones y perspectivas dispares aporten sus opiniones e innovaciones sobre el tema.

Gracias al auspicio de IPAI, esta publicación cuenta con aportaciones desde la Psicología, Sociología, Economía y Estadística. Es cierto que su peso es desigual, pero la intención original es utilizar esta pieza como el primer ladrillo de un edificio sólido, donde las aportaciones de diferentes perspectivas terminen siendo habituales. Sucesivas convocatorias del Congreso de Docencia e Investigación del Comportamiento acercarán aún más los hechos al objetivo final de aunar esfuerzos y conocimientos en la mejor comprensión del comportamiento y en el desarrollo de la metodología de investigación.

ÍNDICE

| | |
|--|----|
| 1. Revisión crítica de las categorías sociodemográficas: la Situación Religiosa | 7 |
| <i>Antonio José Rojas Tejada y Juan Sebastian Fernández Prados</i> | |
| 2. Reflexiones críticas sobre aspectos relacionados con las investigaciones mediante encuestas: de la teoría a la práctica | 17 |
| <i>Olivia Giménez Fernández y Raquel Cruz del Pino</i> | |
| 3. Alternativas cuantitativas a las medidas individuales de posición: los masiles | 27 |
| <i>Vicente Manzano Arrondo</i> | |
| 4. Estudio de la estructura poblacional mediante funciones de distribución y dispersión | 37 |
| <i>José Antonio Mayor Gallego</i> | |
| 5. Formatos de respuesta y niveles de medida: una experiencia didáctica | 45 |
| <i>José Luis Padilla García, Andrés González Gómez y Cristino Pérez Meléndez</i> | |
| 6. Una aproximación al Bootstrap: Aplicaciones y limitaciones | 57 |
| <i>Carlos Arias Martín y Victoria Márquez de la Plata y Cuevas</i> | |
| 7. Efectos de la redacción de los items en las respuestas de los encuestados: una aplicación con escalas tipo Likert | 65 |
| <i>Alicia García Pereira, Regla Martínez Jiménez y Alicia Puerto Martínez</i> | |
| 8. Una propuesta de estructuración de la Metodología | 77 |
| <i>Rafael Moreno Rodríguez, Rafael J. Martínez Cervantes y Salvador Chacón Moscoso</i> | |

| | |
|--|-----|
| 9. La validez en las investigaciones mediante encuestas | 85 |
| <i>Andrés González Gómez, José Luis Padilla García y Cristino Pérez Meléndez</i> | |
| 10. Los modelos de estructuras de covarianza como método de validación de constructo | 101 |
| <i>Eleuterio Francisco Sánchez García y Manuel Sánchez García</i> | |
| 11. Combinación del escalamiento multidimensional y el análisis de conglomerados para el estudio de las representaciones mentales del consumidor. Un caso práctico | 113 |
| <i>Jesús Varela Mallou, X. Gabril Vázquez y Antonio Rial Boubeta</i> | |
| 12. Criterios en la validación de modelos. ¿Rechazar o mantener hipótesis? | 123 |
| <i>Carlos Camacho Martínez Vara de Rey</i> | |
| 13. Influencia del mensaje a través de porteros automáticos en la tasa de respuesta | 129 |
| <i>Encarnación León Martínez, Marisol Ortiz Hernández, Cristina Peñafiel Balbuena e Irene Prieto Reina</i> | |
| 14. Evaluación empírica de un doble proceso de aprendizaje de los conceptos de decisión estadística: desarrollo tradicional <i>versus</i> animación por ordenador | 143 |
| <i>Fco. Javier Pérez Santamaría, Vicente Manzano Arrondo y Hassan Fazeli Khalili</i> | |
| Referencias bibliográficas | 155 |

REVISIÓN CRÍTICA DE LAS CATEGORÍAS SOCIDEMOGRÁFICAS: LA SITUACIÓN RELIGIOSA

Antonio José Rojas Tejada
Juan Sebastián Fernández Prados
Universidad de Almería

RESUMEN

Las variables sociodemográficas son incluidas en la mayoría de los cuestionarios. Sin embargo, los investigadores suelen utilizar estas variables sin preguntarse sobre la pertinencia de la pregunta o de las opciones de respuesta. Un ejemplo de esta situación lo encontramos al analizar las preguntas relacionadas con la posición religiosa, donde se da una gran disparidad en las opciones de respuesta utilizadas e, incluso, en el significado atribuido a dichas opciones. El objetivo de este trabajo es hacer una revisión crítica de las categorías usualmente aceptadas para conocer la posición religiosa. El análisis de los datos permite concluir que: a) existe una escasa diferencia entre las categorías “indiferente” y “agnóstico”, lo que da idea de la inutilidad de su separación; b) debemos hacer un esfuerzo por unificar conceptos, debido a la multitud de significados que hay detrás de cada opción de respuesta; y c) para llegar a ese consenso se hace necesaria la colaboración de profesionales de distintas áreas del saber.

INTRODUCCIÓN

Los cuestionarios destinados a la recogida de datos en encuestas de opinión, de actitud o cualquier otro contenido, contienen normalmente una batería de preguntas llamadas socio-demográficas. Estas cuestiones, también llamadas clasificatorias o simplemente demográficas, nos facilitan información de los encuestados sobre variables tales como la edad, género, lugar de residencia, estado civil, situación laboral, ingresos medios, antecedentes étnicos, posición ideológica, religiosa, etc. La utilización de estas variables tiene como fin conseguir posteriores diferencias, clasificaciones u otro tipo de análisis en función de los objetivos de la investigación, aunque en muchas ocasiones el recurso a estos ítems sea pura inercia del investigador. Las dificultades de estas preguntas son múltiples: la elección de las más apropiadas, la redacción de los interrogantes del modo más correcto, la presentación que evite susceptibilidades de posibles violaciones del anonimato, la ubicación y orden en el cuestionario para evitar bajas tasas de respuestas, el cierre de las opciones de respuesta más conveniente y relevante... (Padilla, 1998).

Esta última cuestión es la que ha atraído nuestra atención, debido a que ha sido poco reflexionada en la literatura de las ciencias sociales, porque en algunas ocasiones las preguntas socio-demográficas han sido utilizadas de modo gratuito y frívolo por los investigadores, como se decía anteriormente, y porque en este campo la metodología tiene mucho que decir y profundizar. Las opciones de respuestas ofrecidas en las preguntas socio-demográficas responden a una categorización social, que en muchas ocasiones es la fiel traslación de los modelos culturales que evolucionan continuamente; el análisis y la investigación hasta hoy, prácticamente, es un campo virgen, y es por esta razón por la que posee un futuro prometedor y permanente. Lo que queremos decir es que cerrar las alternativas de respuestas con unas determinadas categorías nunca puede ser definitivo por la realidad socio-cultural siempre en constante cambio, lo que implica que deben ser permanentemente revisadas y redefinidas: por ejemplo, en la categorización de la edad a principios de los noventa se amplió el agregado de la juventud de 15-25 años a 15-30 años, en la escala ideológica izquierda-derecha se ha empezado a utilizar apolítico como una opción más aparte, en la religiosidad ser agnóstico se está considerando como una postura más, etc. ¿Cuáles son los criterios que fundamentan la selección de las opciones de respuesta en las preguntas socio-demográficas?, ¿la intuición del investigador?, ¿las experiencias de otros estudios previos?; y después de seleccionar las alternativas,

¿sabemos que son las mejores simplemente porque nos dan bajas tasas de no respuesta o porque hay pocos “no sabe/ no contesta”?

Tabla 1. Encuestas y las opciones de respuesta de la pregunta sobre religiosidad.

| FUENTE | AÑO | ÁMBITO | OPCIONES DE RESPUESTA |
|-----------------------------|---------|----------------|--|
| CIRES | 1995 | Estado Español | - Católico - Otra ¿Cuál? ----- - Ninguna |
| DATA | 1992 | Estado Español | - Católico - Indiferente - No creyente, ateo - Creyente en otra religión |
| DATA | 1994 | Estado Español | - Católico - Indiferente - Agnóstico - No creyente, ateo - Creyente en otra religión |
| TABULA V | 1992 | Estado Español | - Católico - Indiferente - Agnóstico - Ateo - Creyente de otra religión |
| TABULA V | 1995 | Estado Español | - Católico - Indiferente - Agnóstico - No creyente, ateo - Creyente en otra religión - Creyente sin religión concreta |
| EUROBARÓMETRO | 1989 | Europeo | - Persona religiosa - Persona no religiosa - Agnóstico - Ateo |
| ENCUESTA MUNDIAL DE VALORES | 1995-96 | Mundial | - Una Persona religiosa - Una persona no religiosa - Un ateo convencido |

LAS OPCIONES DE RESPUESTAS EN LAS PREGUNTAS DE POSICIÓN RELIGIOSA O RELIGIOSIDAD

Las encuestas más importantes que se realizan en España y en el mundo contienen en sus cuestionarios preguntas relacionadas con la posición religiosa o la religiosidad de las personas interrogadas. Como se puede apreciar en los ejemplos que hemos propuesto en la tabla 1, encontramos algunas coincidencias y diferencias significativas en las opciones de respuestas utilizadas en la cuestión de la religiosidad dependiendo de la empresa o entidad coordinadora que realiza la investigación, el año y el ámbito. Las diferencias más importantes las encontramos entre los trabajos realizados dentro y fuera de España; los primeros muestran siempre la opción “católico” mientras que los estudios llevados a cabo en Europa o en el Mundo reparan más entre las opciones de religiosidad o no (Del Pino Artacho y Bericat, 1998). Otra diferencia la encontramos en los propios trabajos españoles que han sufrido algunas variaciones a lo largo del tiempo, que incluyen una nueva posibilidad que en los años ochenta y principios de los noventa no se contemplaba: “agnóstico” (Andrés Orizo, 1996, p 203; De Miguel, 1994, 1997). Por último, las semejanzas en las investigaciones españolas coinciden básicamente en proponer: “*católico*”, “*indiferente*”, “*agnóstico*” y “*ateo*”; además de la consabida “no sabe/no contesta”, preguntas complementarias o sobre el grado de práctica religiosa (“católica muy practicante”; “católica practicante”; “católico poco practicante”...):

DEFINICIONES DE LAS ALTERNATIVAS DE RESPUESTAS

A continuación se presentan las definiciones obtenidas del Diccionario de Filosofía (Ferrater Mora, 1984) que nos hace caer en la cuenta de la multitud de acepciones (en tanto que estemos hablando de ontología, teología, gnoseología, ética, etc...) que tienen cada uno de estos términos y la variedad de usos que han tenido en función del autor o pensador presentado. Aquí mostramos un breve resumen:

Católico: Es un ejemplo de teísmo que defiende la existencia de Dios, en este caso un Dios personal, histórico y encarnado tal y como lo define la Iglesia Católica y Romana (a partir de la definición de *Teísmo* de Ferrater Mora, 1984, p. 3205).

Indiferente: “La indiferencia puede considerarse como un temple de ánimo que cubre todas las cosas con un velo que las hace aparecer iguales. Según Heidegger, este tipo de indiferencia emerge en los estados de auténtico y profundo aburrimiento” (Ferrater Mora, 1984, p. 1657).

Agnóstico: “En sus orígenes, el agnosticismo estaba relacionado con la renuncia a saber nada de Dios [...]. Según, Tierno Galván, el no echar de menos a Dios -lo que equivale a no necesitar más que vivir en la finitud o, si se quiere, en este mundo” (Ferrater Mora, 1984, p. 68-69).

Ateo: “Ser ateo, o abrazar el ateísmo, es negar que haya Dios, o negar que haya dioses, o negar que haya alguna realidad que pueda llamarse divina, o todas estas cosas a un tiempo” (Ferrater Mora, 1984, p. 238).

Otras religiones: Es un ejemplo de teísmo que defiende la existencia de Dios, en este caso cualquier otro Dios distinto al que defiende la Iglesia Católica y Romana (a partir de la definición de *Teísmo* de Ferrater Mora, 1984, p. 3205).

Pero nuestro trabajo no posee una finalidad terminológica, aunque no cabe duda que habría que contar con ella, sino más bien metodológica. Aún así, ambas perspectivas necesariamente tendrán que desembocar en una revisión y selección de las expresiones más adecuadas, en las posibilidades de contestar a una pregunta, como la de auto-consideración religiosa que estamos tratando.

En definitiva, el objetivo que perseguimos con el presente trabajo es comprobar el grado de diferencia existente en actitud religiosa que está detrás de cada categoría sociodemográfica. De esta forma queremos revisar de manera crítica las categorías que generalmente se proponen para conocer la situación o posición religiosa.

MÉTODO

Sujetos

En el estudio participaron un total de 1365 sujetos (20.1% de hombres y 79.9% de mujeres), todos estudiantes pertenecientes a las universidades de Almería (48.4%), Granada (39.9%) y Sevilla (11.8%), con edades comprendidas entre 17 y 46 años (con media en 21.19 y desviación típica en 3.41).

Instrumentos

Se utilizó la ‘*Escala de Actitudes Religiosas (R-1)*’ desarrollada por Morales (1988), en la versión de 22 ítems empleada por Rojas (1998). Este test pretende medir la ‘*tónica religiosa o la cercanía a la fe*’. La definición del constructo es muy genérica, y no pretende aludir a ninguna religión o práctica religiosa concreta, tal y como lo conciben Hood (1970) y Pargament et al. (1988).

Todos los ítems tienen un formato tipo Likert de 5 puntos (*muy de acuerdo, de acuerdo, ni de acuerdo ni en desacuerdo, en desacuerdo, muy en desacuerdo*). La puntuación máxima en el test es de 110 puntos y la mínima de 22.

Además de esta escala se incluyó una serie de preguntas sociodemográficas, donde una de ellas interrogaba por el tipo de situación religiosa de cada sujeto (*indiferente, agnóstico, ateo, católico y otras*).

Procedimiento

El cuestionario fue administrado de forma colectiva. La recogida de datos tuvo lugar durante los meses de noviembre y diciembre de 1997, en las instalaciones de las universidades de Almería, Granada y Sevilla.

Análisis realizados

A partir de las respuestas de los sujetos al test se obtuvo la puntuación total de cada uno de ellos. Con estos datos y con las respuestas respecto al tipo de religión al que pertenecían, se llevaron a cabo dos tipos de análisis:

- En primer lugar, un análisis descriptivo de los valores obtenidos en el test en función del tipo de situación religiosa.
- En segundo lugar, un ANOVA de los valores de puntuación de los sujetos en el test en función del tipo de situación religiosa.

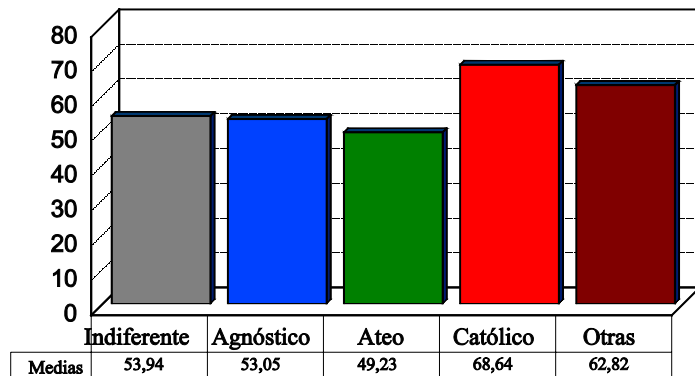
Para realizar todos los análisis se utilizó el programa SPSS para Windows en su versión 7.5.

RESULTADOS

Descriptivos

Realizados los cálculos numéricos (tabla 2) y la representación gráfica de las medias de las puntuaciones de los sujetos en función del tipo de situación religiosa (figura 1) podemos observar que:

Figura 1. Puntuaciones medias en el test de religiosidad, según la categoría religiosa manifestada



- Más de la mitad de la muestra manifiesta ser católico (52.53%). Le siguen el 17.73% de indiferentes, el 14.14% de agnósticos y el 7.40% de ateos. El 6.45% manifestó tener por situación religiosa otra diferente a las citadas.
- Las puntuaciones medias más altas en el test fueron para el grupo de Católicos (68.64). Las puntuaciones más bajas en el test la obtuvo el grupo de ateos (49.23).

Tabla 2. Descriptivos.

| | Indiferente | Agnóstico | Ateo | Católico | Otras | Total | Val.Perd. |
|-------|-------------|-----------|-------|----------|-------|-------|-----------|
| n | 242 | 193 | 101 | 717 | 88 | 1365 | 24 |
| % | 17.73 | 14.14 | 7.40 | 52.53 | 6.45 | 100 | 1.76 |
| Media | 53.94 | 53.05 | 49.23 | 68.64 | 62.82 | 61.90 | ----- |
| D.T. | 9.67 | 9.38 | 9.57 | 11.66 | 12.27 | 13.34 | ----- |

ANOVA

En la tabla 3 se presenta la tabla resultante del ANOVA de las puntuaciones de los sujetos en función de su situación religiosa.

Tabla 3. ANOVA.

| | S.C. | g.l. | M.C. | F | sig. |
|--------------|------------|------|-----------|---------|-------|
| Inter-grupos | 79335.540 | 4 | 19833.885 | 166.620 | 0.000 |
| Intra-grupos | 159032.667 | 1336 | 119.036 | | |
| Total | 238368.207 | 1340 | | | |

Los resultados obtenidos señalan la existencia de diferencias estadísticamente significativas ($p < 0.001$) entre las medias del test en función de la citada situación religiosa de los sujetos.

Realizadas las pruebas de comparaciones a posteriori, resultaron estadísticamente significativas las diferencias entre las puntuaciones medias del grupo de ateos con todas las demás situaciones religiosas (tal y como se muestra en la tabla 4). Asimismo, resultaron estadísticamente significativas las diferencias entre las puntuaciones medias del grupo de católicos con todas las demás situaciones religiosas. Las únicas diferencias que no resultaron significativas fueron las de comparar las puntuaciones de agnósticos e indiferentes.

Tabla 4. Comparaciones Múltiples. Significaciones estadísticas (p).

| | Indiferente | Agnóstico | Ateo | Católico | Otras |
|-------------|-------------|-----------|-------|----------|-------|
| Indiferente | | 0.916 | 0.002 | 0.000 | 0.000 |
| Agnóstico | | | 0.035 | 0.000 | 0.000 |
| Ateo | | | | 0.000 | 0.000 |
| Católico | | | | | 0.000 |
| Otras | | | | | |

CONCLUSIONES

La primera conclusión que hemos encontrado se refiere al hecho particular de la escasa o nula diferencia entre “indiferente” y “agnóstico”, lo cual nos hace pensar en la inutilidad del uso de estas dos categorías por separado en los estudios de investigación, al menos en los estudiantes universitarios. Además, creemos que es generalizable la revisión a partir de escalas como la utilizada de religiosidad, a otro tipo de variables socio-demográficas como la posición política, la posición nacionalista, etc. El objetivo sería similar al expuesto en este trabajo: examinar las alternativas de respuesta empleadas y determinar si son significativas o no respecto a lo que se mide, ya que es posible que sean inservibles una o incluso hallemos la falta de otras nuevas, realizando estudios en todo tipo de poblaciones.

La necesidad de la revisión continua de las alternativas utilizadas en las preguntas socio-demográficas es consecuencia de la dinámica social que impone claves, valores y criterios culturales en diferentes contextos y en distintos momentos históricos. Por ejemplo, como hemos encontrado en la revisión bibliográfica, hasta no hace muchos años no aparecía “agnóstico” en las opciones de respuesta, o también las diferencias que hemos observado dependiendo del ámbito social o cultural que tratemos (europeo, mundial...).

Otra cuestión que aparece fruto de este trabajo es la obligatoriedad de llevar a cabo un esfuerzo filológico y lingüístico que aporte los elementos semánticos y de comprensión de los términos utilizados en una determinada comunidad. Las acepciones, los sentidos propios de cada colectivo humano y social se reflejan también en la elección de una

determinada alternativa. Superar esta traba en este tipo de preguntas es imprescindible debido a la multitud de significados que hemos descubierto que se esconden detrás de cada opción en este tipo de preguntas.

El empleo del mismo test o escala que nos ayude a medir la religiosidad o cualquier otra variable que estemos estudiando, debe ser el mismo siempre y en todo lugar. Solamente utilizando el mismo test nos asegurará que las opciones que escojamos definitivamente para la cuestión concreta (p.e. la situación religiosa) puedan ser comparables en el tiempo. Esto implica un trabajo coordinado, un centro de colaboración multidisciplinar en el que converjan estudiosos e investigadores de distintas especialidades, intentado llegar a un acuerdo conjunto.

La anterior conclusión es consecuencia de la propia experiencia productiva, sabrosa y definitivamente universitaria (en su sentido más original) de dos profesores e investigadores provenientes de dos especialidades y áreas del saber distintas (Metodología y Sociología). Las ciencias sociales, y las ciencias en general, solo avanzarán con esfuerzos conjuntos y con experiencias compartidas de intereses comunes, que el conocimiento siempre puede hallar.

REFLEXIONES CRÍTICAS SOBRE ALGUNOS ASPECTOS RELACIONADOS CON LAS INVESTIGACIONES MEDIANTE ENCUESTAS: DE LA TEORÍA A LA PRÁCTICA

Olivia Giménez Fernández
Raquel Cruz del Pino
Universidad de Almería

RESUMEN

La realización de encuestas personales, cara a cara, depende en gran medida del comportamiento de los encuestadores. Se insiste, por un lado, en su papel importantísimo a desempeñar durante la fase de recogida de datos; por otro lado, el encuestador ideal es considerado como una máquina que no debe tomar decisiones más allá que las que se derivan de los procedimientos que se le han suministrado. En la práctica, la naturaleza no mecánica del encuestador, junto con las características del contexto, investigador y entrevistado, confluyen de tal forma que se observan comportamientos no ideales en los encuestadores. En este trabajo presentamos brevemente las estrategias llevadas a cabo por un grupo de encuestadoras para realizar su trabajo, así como las consecuencias previsibles en la validez del estudio y algunas sugerencias de solución.

INTRODUCCIÓN

La encuesta es considerada como una técnica de investigación científica donde se utiliza un conjunto de procedimientos estandarizados para la obtención de datos (García Ferrando, Ibañez y Alvira, 1992). En la mayoría de las encuestas (personales y telefónicas) están al menos implicados dos factores. Por un lado nos encontramos el elemento teórico, donde se planifican los pasos que conlleva la realización de una encuesta y donde todos los procedimientos deben estar estandarizados y con todas las garantías de calidad; por otro lado estaría el elemento práctico, el trabajo de campo, en el que el protagonismo recae sobre los encuestadores. Para que no existan diferencias entre ambos planos, en el caso ideal, y para garantizar que el encuestador sigue fielmente todos los procedimientos estandarizados en las distintas etapas del trabajo de campo, el encuestador debe convertirse en un mero ejecutor sin poder de decisión (Manzano y González, 1998).

A raíz de nuestra formación y trabajo como encuestadoras y de consideraciones realizadas entre los compañeros acerca de sus experiencias, surge la idea de profundizar y dar a conocer una serie de reflexiones entorno al paso existente entre los aspectos teórico y práctico de la investigación mediante encuestas. Aunque estas diferencias entre los aspectos teóricos y prácticos son bien sabidas por los profesionales dedicados a la investigación mediante encuestas, es probable que se obvian, siendo al final el encuestador el que dé cuenta de ello, improvisando y creando determinadas estrategias que le sean útiles para conseguir llevar la teoría a la práctica.

Por tanto, consideramos decisivo conocer la experiencia de los encuestadores, y es nuestro objetivo acercar, a partir de la experiencia de varias encuestadoras, de una manera informal, la realidad con la que nos encontramos los encuestadores para que sea tenida en cuenta por aquellos que quieren contribuir al conocimiento científico y han elegido la investigación mediante encuestas como técnica idónea.

MÉTODO

Para este trabajo hemos querido reflejar no sólo la experiencia de las autoras, sino también la de siete encuestadoras que han trabajado en la provincia de Almería.

Se ha realizado una entrevista en profundidad de la que se obtuvo información acerca de las dificultades encontradas, durante todo el proceso del trabajo de campo para llevar el plano teórico al práctico, y otras cuestiones como las estrategias desarrolladas en determinados momentos, sugerencias, etc.

Como ejemplos de procedimientos teóricos estandarizados de realización del trabajo de campo vamos a basarnos en el *Manual para encuestadores* (Manzano, Rojas y Fernández, 1996) y en las Normas del IESA (Instituto de Estudios Sociales de Andalucía, 1997) de manera que nos sirvan de referente de cómo llevar a cabo las distintas etapas que siguen las encuestadoras en su trabajo de campo.

RESULTADOS

A continuación vamos a indicar algunos aspectos del proceso de encuesta que reflejan cómo la teoría es llevada a la práctica por la gran mayoría de las encuestadoras entrevistadas y algunas dificultades encontradas.

1. La TABLA DE NÚMEROS ALEATORIOS ayuda a tomar decisiones en el proceso de selección del encuestado. En este proceso se busca que el comportamiento del encuestador sea aleatorio, para lo que hay que eliminar, en la medida de lo posible, todas las características personales del encuestador que puedan influir en la toma de decisiones: prejuicios, preferencias, gustos, manías (Manzano, Rojas y Fernández, 1996).

No obstante, en la práctica, la tabla se usa solamente en las primeras encuestas, siendo sustituida por elecciones personales del encuestador: llamar a todas las puertas por orden de planta y puerta; tocar en una puerta sin ninguna preferencia expresa; preguntar al portero, vecino por alguien que cumpla las características requeridas por la encuesta.

Dado que el uso de la tabla de números aleatorios ha sido calificado de fácil, pueden ser otros factores los que lleven a optar por no usarla. El cansancio, la fatiga (por lo repetitivo y lento del proceso) y lo poco reforzante del método aleatorio, son posibles variables que llevan a que el encuestador “olvide” la importancia del hecho de que no debe participar en la toma de decisiones. Durante el trabajo de campo prima que la encuesta esté rellena, que lo haga alguien de la edad y sexo buscado y no el hecho de que todos los miembros de la población seleccionada tengan las mismas oportunidades, y por tanto sea el encuestador quien rompa el principio de equiprobabilidad.

2. Avanzando en el proceso de la encuesta, una vez hecho el recuento de unidades muestrales, puede que la vivienda seleccionada se encuentre en un edificio y el encuestador deba entrar en éste. Las encuestadoras entrevistadas consideran que una de las mayores dificultades para realizar con éxito una encuesta es la ENTRADA EN UN EDIFICIO cerrado. Para todas este momento es decisivo por la dificultad de que los vecinos abran la puerta a un encuestador. La motivación por continuar depende mucho de este momento y, en general, las encuestadoras usan estrategias para conseguir su objetivo (una vez dentro es más fácil llegar al encuestado), siendo la más utilizada decir que se pertenece a una empresa de publicidad o la espera a que alguien entre o salga del portal. Cuando se les habló de la posibilidad de decir la verdadera identidad, aquellas que lo habían hecho en alguna ocasión obtuvieron escasos resultados, por lo que decidieron cambiar de estrategia.

3. En poblaciones de mayor número de habitantes, en las que los edificios agrupan la mayoría de las viviendas, encontramos ciertas características que llevan a la dificultad de introducirnos en las viviendas y, por tanto, de encuestar. En CIUDADES Y PUEBLOS GRANDES, y dentro de éstas, en barrios de clase social alta, la esperanza de colaboración es más baja que en otras zonas. El potencial encuestado es poco receptivo a recibir a extraños, a poner en peligro su anonimato (aún cuando se asegura lo contrario) y a perder su tiempo colaborando gratuitamente en una encuesta, ya se trate o no de una entidad de mayor o menor prestigio. Y, por el conocimiento de las poblaciones en las que se ha trabajado, no pensamos que se trate tanto de inseguridad ciudadana como de falta de receptividad de ciertas comunidades a la presencia de un extraño y al método de encuestas (en este caso a las personales). El impedir el acceso al portal es una forma de control y es una de las grandes dificultades con las que se encuentra el encuestador. Este aspecto resulta aún más claro cuando el trabajo de campo se realiza en un pueblo pequeño, comunidad de vecinos de clase social media-baja, donde encontramos viviendas unifamiliares, casas de planta baja, La mayoría de estos vecinos no plantean tantos problemas para colaborar, abren inmediatamente las puertas de su casas y son, en general, hospitalarios. Podemos decir que son más receptivos a que llegue un extraño y haga una encuesta, siendo quizás el único inconveniente que plantean la creencia de no saber responder. Si contrastamos los dos tipos de poblaciones con las que nos encontramos, la motivación para seguir realizando el trabajo de encuestar es diferente en una y otra. Nuestras encuestadoras realizan mejor su trabajo, más eficazmente, sin introducir tantos sesgos y siguiendo fielmente los pasos del método en ambientes más agradables, más colaboradores, donde son más valoradas y por tanto mejor atendidas. Así, no podemos obviar la motivación del encuestador en relación al trato que recibe en su trabajo como

factor importante para que las diferencias entre los aspectos teóricos y prácticos aumenten o disminuyan durante el proceso.

4. Otra de las grandes dificultades con las que ha de enfrentarse la encuestadora es que LA PERSONA SELECCIONADA ACCEDA A CONTESTAR LA ENCUESTA.

Este es el momento decisivo, todas las encuestadoras usaban su tarjeta de identificación, apelaban al anonimato de la encuesta, a la importancia de haber sido seleccionado entre todo el edificio y de formar parte de una investigación, a la facilidad y rapidez de contestación de la encuesta, y a representar a un organismo o institución importante. Estas estrategias son importantes y dependen de las habilidades del encuestador para que sean efectivas.

Un factor que han resaltado como importante para mantener al encuestado es el tiempo que se tarda en contestar una encuesta. Todas las encuestadoras mienten en el tiempo real que puede tardarse en realizar la encuesta, por ser ello decisivo para que la persona acceda a participar, no obstante este factor también les ha influido en que la persona deje la encuesta a medias. En general cuanto mayor es el tiempo mayor es el riesgo de que dejen la encuesta a medias.

Las características del medio que se han relatado anteriormente (receptividad del medio, colaboración en las encuestas, etc.) son incluso más aplicables en esta fase del proceso en el que el sujeto seleccionado pasa a ser encuestado, pues el contacto directo (cualquiera que sea la estrategia de entrada) es el decisivo para la aceptación o el rechazo hacia el encuestador; sin dejar en segundo plano las habilidades que el encuestador ponga en marcha. Estas habilidades y estrategias van a ser más efectivas con la práctica: así, cuantas más encuestas realice y cuantas más veces se haya enfrentado con encuestados, mejor va a desempeñar su trabajo.

5. ¿Cuál es la reacción del encuestador ante la negativa a colaborar por parte del individuo seleccionado?. Pese al despliegue de recursos por parte del encuestador, se puede dar la negativa a la participación por lo que el encuestador puede optar por relajar la rigurosidad del proceso, (llegando incluso a buscar al encuestado de una manera no aleatoria). Incluso, llega a la manipulación de datos y, en consecuencia, la participación del encuestador pasa a ser muy perjudicial en la investigación. Como se ha mencionado ya, corremos el riesgo de que se habitúen al uso de estas trampas en vez de perfeccionar las estrategias y mejorar las habilidades. Estos casos se dan con frecuencia en nuestras encuestadoras y creemos que

seguirán ocurriendo, independientemente de que la formación sea excelente y del convencimiento de que se debe seguir el proceso sin que influya el encuestador. Es por esto que la mejor recomendación es paralizar el trabajo del día cuando comiencen los problemas, comunicar al coordinador del trabajo de campo y llegar entonces a un acuerdo en el tiempo necesario para tener realizadas todas las encuestas, así como la revisión sistemática del proceso, adecuando éste a la información que se ha recogido en la práctica. A menudo, la presión del tiempo que una empresa o institución da a sus encuestadoras para terminar el trabajo de campo lleva a que no se puedan considerar otras alternativas, por ejemplo, abandonar el trabajo por un día cuando no va todo lo bien que se puede esperar, por la fatiga, la receptividad del medio, la dificultad de las encuestas (por sus características a la hora de la entrevista), y, en definitiva, por la motivación que ese día está moviendo al encuestador.

6. La REVISITA es otro punto del trabajo de campo que todo encuestador tiene siempre presente por ser un punto teórico cuya importancia en el proceso de muestreo se suele resaltar; sin embargo, ninguna encuestadora hizo la revisita, esto es, ninguna volvió a la vivienda en la que había seleccionado a un sujeto que no estaba en ese momento.

Las encuestadoras entrevistadas suelen, en estos casos, buscar otro encuestado de similares características, en vez de volver a una casa donde no es seguro que se vaya a efectuar una encuesta, con la pérdida de tiempo, esfuerzo y la gran desmotivación que puede suponer la negativa. Para que esto no ocurra y se cuente con la certeza de que va a ser hecha la encuesta, recomendamos pedir el teléfono de la casa (si lo hubiera) y llamar antes de iniciar el recorrido, así como resaltar a quien nos ha atendido la importancia de que sea la persona seleccionada quien responda al cuestionario, e incluso se puede barajar la posibilidad de la encuesta telefónica.

7. En relación al desarrollo de la encuesta, los encuestadores nunca han de dar su OPINIÓN ACERCA DEL CONTENIDO DE LAS PREGUNTAS Y RESPUESTAS, pero esto no ocurre siempre así.

Aunque no todas las encuestadoras opinaban acerca de las preguntas y respuestas del cuestionario buena parte de ellas sí lo hacían. La razón fundamental que dieron para ello es asegurar la finalización del cuestionario, esto es, que el clima sea agradable, que la entrevista se dé con normalidad, etc., para lo que se veían forzadas a opinar en determinadas situaciones (por ejemplo, ante una petición expresa del encuestado). Lo contrario se percibiría como una oportunidad para finalizar la entrevista sin completar el

cuestionario. Quizá la mejor alternativa es advertir que esto ocurre y que es mejor no dar pie a conversar desde el primer momento, no contestar a menos que se pregunte y en ese caso resaltar la importancia de que el encuestador no puede responder nada pues es la opinión del encuestado la que importa.

El sesgo del encuestador que se detecta cuando el entrevistador explica una pregunta o interpreta un fenómeno, está presente en las encuestadoras con las que hemos contactado. Ello presenta un gran inconveniente en las encuestas personales aún cuando se intenta prevenir profesionalizando a las encuestadoras. Por tanto, no consideramos que se subsane el problema apelando a la importancia de ser neutral durante los cursos de formación, puesto que la realidad es que la tendencia es a ser más "humano", menos neutral, más participante con el encuestado, para que el clima sea tan agradable como éste lo necesite y, de esta manera, la encuesta se desarrolle sin problemas.

8. Un fenómeno nos ha llamado la atención pues se da en todas las encuestadoras que hemos entrevistado ante la pregunta de "si han seguido alguna vez su criterio a la hora de ANOTAR LA RESPUESTA DEL ENCUESTADO", contestaron que no, pero al profundizar un poco en la cuestión observamos que ante repuestas con un abanico de posibilidades que resulta ambiguo para el encuestado dada su avanzada edad, nivel cultural, etc. (p.e. Muy de acuerdo, De acuerdo, En desacuerdo, Muy en desacuerdo) y respondiendo éste de manera dicotómica (como si solo hubieran dos alternativas de respuesta), las encuestadoras suelen contestar en función de la valoración que se han hecho de las opiniones del encuestado. La existencia de muchas preguntas seguidas con escalas tipo Likert o similares, puede llevar a ser demasiado monótono tanto para el encuestado como para el encuestador. Ello nos ha llevado a reflexionar si para determinadas poblaciones es correcto diseñar una encuesta que no refleja el modo de respuesta de aquellos que van a ser encuestados, con el sesgo que introduce el encuestador, y si éste es subsanable con alternativas de respuestas más adaptadas. Este tipo de interpretación o manipulación de las respuestas de los encuestados sólo lo hemos encontrado por problemas en las alternativas de respuesta, pues en casos en que la encuestadora detectaba (o se decía explícitamente) que se estaba mintiendo, usaba las anotaciones para aclarar la alternativa de respuesta (ejemplo, ante preguntas acerca de la cantidad de dinero que entra mensualmente en una casa, las encuestadoras señalan la opción que el encuestado quiere que aparezca, aún cuando éste dice ser otra, y en este caso la encuestadora anota a modo de observación el comentario realizado, para que el investigador decida). Otro ejemplo es aquel en que el encuestado pide que no aparezcan los datos de su vivienda (dirección), la

encuestadora procede indicando el interés del encuestado pero no puede obviar dicha información.

9. Por último, ante la pregunta de si consideran fácil llevar la TEORÍA A LA PRÁCTICA en el trabajo de campo, las encuestadoras dicen que sí, a pesar de los sesgos que introducen en el proceso, y en ello ven decisivo: el cansancio, el número de encuestas, el barrio, la desconfianza hacia el encuestador, las características del encuestado, la remuneración, la dificultad de la encuesta, el rendimiento conseguido, el estado de ánimo,... Por tanto, no es tan problemático el proceso de encuestas como todas las variables que le rodean, íntimamente relacionadas con la realidad en que se desarrolla la encuesta. El trabajo de encuestar se les presenta tan complicado y poco gratificante que la posibilidad de que una encuestadora vuelva a realizar un trabajo de campo depende: del tiempo que pase desde su último trabajo (deben trascuir algunos meses para que estén dispuestas), de que su nivel económico lo demande, de que su nivel de actividad en ese momento no sea muy alto y del tipo de encuesta y población con que deban trabajar; tenidos en cuenta estos factores independientemente de que la remuneración por encuesta ha de ser aceptable (mínimo 1000 pesetas por encuesta). Aunque esto depende también de la longitud y complejidad de la encuesta.

CONCLUSIONES

En este trabajo hemos intentado mostrar algunas reflexiones acerca de la dificultad del encuestador a la hora de llevar la teoría que implica su tarea a la práctica en el trabajo de campo. La intención fue la de identificar las estrategias que las encuestadoras entrevistadas llevan a cabo para garantizar el trasvase de lo teórico a lo práctico. En su lugar, hemos encontrado un conjunto de comportamientos elaborados por los encuestadores, con el fin de hacer posible o facilitar su labor de entrevista y cumplir con los objetivos formales transmitidos por los investigadores. Tales estrategias han resultado contraproducentes con respecto al objetivo general de obtener una muestra representativa y unos datos de calidad. A saber:

- Toma de decisiones subjetivas, en la selección de viviendas y encuestados.
- Interferencia de los encuestadores en la comprensión de las preguntas y la emisión de las respuestas

Creemos que la técnica de investigación mediante encuestas no concede suficiente importancia a las variables personales del encuestador, al medio social y a las

características propias de la encuesta concreta, siendo además relevante la influencia de estas variables en la calidad de la información recogida.

El papel teórico de un encuestador como mero ejecutor de un proceso y sin poder de toma de decisiones, debe ser sustituido por el de un encuestador con nombre y apellidos, con unas cualidades determinadas, una realidad a la que enfrentarse, que toma decisiones y con un trabajo difícil, cuya imagen es poco atractiva y además, poco gratificante. Una vez en esta posición podremos encontrar más estrategias que transmitir en la formación teórica y que puedan aplicar en la labor práctica, y así disminuir las diferencias existentes entre los planos teórico y práctico de la investigación mediante encuestas, en lo que al encuestador se refiere.

Por último, hay que decir que esperamos que esta información sea aliciente para el establecimiento de futuras investigaciones que acaben de solventar los problemas y dificultades a los que se ven sometidos los encuestadores, aumentando con ello la validez de las investigaciones mediante encuestas.

ALTERNATIVAS A LAS MEDIDAS INDIVIDUALES DE POSICIÓN: LOS MASILES

Vicente Manzano Arrondo
Universidad de Sevilla

RESUMEN

Las medidas individuales de posición más utilizadas para la interpretación de puntuaciones son los cuantiles. Gracias a su uso, el valor concreto de un dato puede ser interpretado según las características de la distribución en la que se inserta. No obstante, la insensibilidad de los cuantiles a la masa o peso de los valores, presenta algunas desventajas. En este trabajo, se proponen los “masiles” como alternativa para la interpretación de puntuaciones individuales, considerando la masa de los valores. Se evalúan sus propiedades, aislando ventajas e inconvenientes.

INTRODUCCIÓN

Entre los abundantes productos que genera la Estadística, orientados al conocimiento de un conjunto de datos, se encuentran las medidas llamadas “de posición”. Su objetivo es ayudar al investigador a interpretar valores concretos, gracias a su situación dentro de la distribución. Dos de estos productos son, sin duda, los más conocidos y utilizados: los cuantiles y las puntuaciones típicas.

Los cuantiles surgen ante la necesidad de interpretar puntuaciones, sin la consideración de la cuantía, masa o peso de los datos. Lo único que interesa en el proceso de cálculo es el rango u orden de los datos en la distribución, según su valor.

Por ejemplo, un valor concreto de una variable concreta como un $x_i = 13$ en la variable $X = \text{Ancho de lomo}$, no dice nada por sí mismo, salvo que conozcamos lo suficiente la distribución donde $x_i = 13$ se encuentra inmerso. Si los valores se reparten uniformemente en el intervalo (1, 14), podemos interpretar $x_i = 13$ como un valor alto (un dato situado en la parte alta de la distribución). Si, por el contrario, la distribución se sitúa uniformemente en el intervalo (10, 100), $x_i = 13$ debería ser evaluado como un valor bajo (un dato situado en la parte baja de la distribución). El valor es el mismo en ambos casos, pero la distribución (el contexto que rodea al dato) no, por lo que la interpretación debe adecuarse y variar.

Para calcular cuantiles, se dispone el conjunto de datos de forma ordenada según su valor, del más pequeño al mayor. Acto seguido, se divide a la distribución resultante en k partes, mediante $k-1$ puntos de corte. Cada parte tiene la misma cantidad de datos. Las puntuaciones individuales se interpretan, entonces, por la parte en la que se sitúan. Así, encontramos diferentes cuantiles, en función del número de partes que se generan, principalmente: cuartiles (cuatro partes), deciles (diez) y centiles o percentiles (cien).

Existen varios procedimientos alternativos de cálculo para estas mismas medidas, según se trate de variables discretas, continuas, objetivos de la interpretación, etc... Dado que tal circunstancia excede los objetivos de este trabajo, para nuestra discusión utilizaremos un procedimiento de fácil comprensión, según el cual los percentiles coinciden con los porcentajes acumulados. En Psicología, estos productos se utilizan, con gran frecuencia, para la baremación de tests, gracias a lo cual, la puntuación de un entrevistado puede ser

situada en un colectivo más amplio. Por ejemplo: la puntuación del sujeto X (sea cual fuere) deja por debajo de sí al 65% de la población.

No obstante su amplia utilización, los cuantiles cuentan con algunas desventajas, de las que nos ocupamos aquí de una: la pérdida de información derivada de sustituir las cuantías de los datos originales por su rango o posición en la distribución ordenada. Esta pérdida de información facilita realizar algunas interpretaciones no muy acertadas. Veamos un ejemplo.

En un examen a un grupo reducido de alumnos, se obtienen las siguientes puntuaciones:

1 2 2 3 3 3 4 4 5 5 5 6 6 10

Ordenados los datos y dispuestos en una tabla de frecuencias, se obtiene:

| Valor | Frecuencia | Propor | P. Acum | %Acum |
|-------|------------|--------|---------|-------|
| 1 | 1 | 0.071 | 0.071 | 7 |
| 2 | 2 | 0.143 | 0.214 | 21 |
| 3 | 3 | 0.214 | 0.429 | 43 |
| 4 | 2 | 0.143 | 0.571 | 57 |
| 5 | 3 | 0.214 | 0.786 | 79 |
| 6 | 2 | 0.143 | 0.929 | 93 |
| 10 | 1 | 0.071 | 1.000 | 100 |
| Total | 14 | 1.000 | | |

Obsérvese que la distribución de notas cuenta con un vacío entre las puntuaciones 6 y 10. Sin embargo, este evento no es considerado en absoluto en la puntuación centil. Lo que importa es que 10 es un valor superior a 6 y, por tanto, queda expuesto después. Se habrían obtenido los mismos centiles si en lugar del valor 10 se observara un 6,1.

El principal inconveniente de no considerar las masas o cuantías de los datos más allá que lo suficiente para establecer el orden, es que los centiles no reflejan las distancias o

intervalos entre los valores. Un alumno que ha obtenido una nota de valor 6 se encuentra a escasas centésimas (0,071) del alumno que ha obtenido un 10. Sería deseable contar con otras medidas individuales que fueran sensibles a estas distancias o intervalos.

LOS MASILES

Una alternativa para el cálculo de medidas individuales que ayuden a interpretar valores concretos, es considerar la masa de los valores en el proceso. La idea básica de los masiles es operar en términos relativos, al igual que las proporciones, pero no con las frecuencias, sino con las masas totales de los valores.

Denominamos “masa” de un valor al producto de su cuantía absoluta por su frecuencia. Esta medida refleja su importancia en la distribución de datos:

$$m_i = f_i |x_i|$$

donde:

- x_i = valor que ocupa la posición i-ésima en la tabla
- m_i = masa del valor x_i
- f_i = frecuencia del valor x_i

La masa m_i es utilizada en cuantas expresiones de cálculo se basen en los datos agrupados por sus frecuencias.

Definimos ahora la “masa relativa” como la masa expresadas en tantos por uno a utilizar el referente de la masa total de la distribución:

$$mr_i = \frac{f_i |x_i|}{\sum_{j=1}^k f_j |x_j|} = \frac{m_i}{\sum_{j=1}^k m_j}$$

donde:

- k = número de valores de la distribución
- mr_i = masa relativa del valor x_i

Por último, definimos el “masil” asociado al valor x_i como la masa relativa acumulada hasta el valor x_i :

$$M_i = \frac{\sum_{j=1}^i m_j}{\sum_{j=1}^k m_j} = \frac{\sum_{j=1}^i f_j |x_j|}{\sum_{j=1}^k f_j |x_j|}$$

donde:

M_i = masil asociado al valor x_i

Los masiles cuentan con algunas propiedades que les hacen preferibles a los cuantiles en algunas ocasiones. Observaremos su comportamiento en varios ejemplos que se exponen a continuación. En las tablas que siguen, se ha utilizado la expresión “fractil” para hacer referencia a la proporción acumulada. De esta forma, la comparación con el masil se establece en la misma escala.

| Valor | Frecuencia | Propor | M. rel | Fractil | Masil |
|-------|------------|--------|--------|---------|-------|
| 1 | 1 | 0.071 | 0.017 | 0.071 | 0.017 |
| 2 | 2 | 0.143 | 0.068 | 0.214 | 0.085 |
| 3 | 3 | 0.214 | 0.153 | 0.429 | 0.237 |
| 4 | 2 | 0.143 | 0.136 | 0.571 | 0.373 |
| 5 | 3 | 0.214 | 0.254 | 0.786 | 0.627 |
| 6 | 2 | 0.143 | 0.203 | 0.929 | 0.831 |
| 10 | 1 | 0.071 | 0.169 | 1.000 | 1.000 |

El ejemplo es el mismo que en la exposición de los porcentajes acumulados. Esta vez, se ha añadido la información referida a los masiles. Se observa que la distancia entre 6 y 10 ha aumentado (de 0,071 a 0,169). Las puntuaciones más importantes (debido a su peso) sobresalen más con los masiles. Veamos ahora algunas propiedades.

1. Los masiles adoptan valores comprendidos en el intervalo)0 , 1)

Como $i \neq k$, entonces

$$\sum_{j=1}^i m_j \neq \sum_{j=1}^k m_j$$

Por otro lado, como $0 < f_j$ y también $0 \neq |x_j|$, entonces:

$$0 < \sum_{j=1}^i m_j \neq \sum_{j=1}^k m_j$$

Al dividir todo entre $\sum_{j=1}^k m_j$, se obtiene que:

$$0 < M_j \neq 1$$

2. Los masiles son independientes de la escala de medida

En otros términos, al multiplicar a todos los valores de la distribución por una constante, los masiles permanecen inmutables:

Definimos la variable Y donde cada uno de sus valores surge de multiplicar los de la variable X por la constante v , de esta forma, $y_i = v x_i$. Con ello:

$$f_i |y_i| = v f_i |x_i|$$

Luego:

$$M_i(y) = \frac{\sum_{j=1}^i f_j |y_j|}{\sum_{j=1}^k f_j |y_j|} = \frac{\sum_{j=1}^i v f_j |x_j|}{\sum_{j=1}^k v f_j |x_j|} = \frac{v \sum_{j=1}^i f_j |x_j|}{v \sum_{j=1}^k f_j |x_j|} = \frac{\sum_{j=1}^i f_j |x_j|}{\sum_{j=1}^k f_j |x_j|} = M_i(x)$$

En el ejemplo, si se multiplica por 10 a cada valor, masiles y fractiles no varían:

| Valor | Frecuencia | Propor | M. rel | Fractil | Masil |
|-------|------------|--------|--------|---------|-------|
| 10 | 1 | 0.071 | 0.017 | 0.071 | 0.017 |
| 20 | 2 | 0.143 | 0.068 | 0.214 | 0.085 |
| 30 | 3 | 0.214 | 0.153 | 0.429 | 0.237 |
| 40 | 2 | 0.143 | 0.136 | 0.571 | 0.373 |
| 50 | 3 | 0.214 | 0.254 | 0.786 | 0.627 |
| 60 | 2 | 0.143 | 0.203 | 0.929 | 0.831 |
| 100 | 1 | 0.071 | 0.169 | 1.000 | 1.000 |

3. Los masiles son sensibles a la posición de la distribución

Esta propiedad ofrece una ventaja frente a los fractiles o cuantiles, puesto que es coherente con las cercanías relativas de los valores. La diferencia entre 1 y 2 es la misma que entre 14356 y 14357. No obstante, en este segundo caso, la diferencia en términos absolutos no debe tener la misma interpretación, pues en términos relativos es menos importante.

Una forma de operativizar la posición de la distribución es sumar una constante a todos los datos. Al hacerlo, la ordenación de éstos no varían, por lo que los fractiles permanecen inmutables. No obstante, al acercarse las distancias (ante la suma de una constante positiva), éstas son menos importantes y así lo registran los masiles.

$$M_i(y) = \frac{\sum_{j=1}^i f_j |y_j|}{\sum_{j=1}^k f_j |y_j|} = \frac{\sum_{j=1}^i f_j |x_j|}{\sum_{j=1}^k f_j |x_j|} \cdot \frac{\sum_{j=1}^i v}{\sum_{j=1}^k v} = \frac{\sum_{j=1}^i f_j |x_j|}{\sum_{j=1}^k f_j |x_j|} \cdot \frac{\sum_{j=1}^i v}{\sum_{j=1}^k v} = M_i(x) \cdot \frac{\sum_{j=1}^i v}{\sum_{j=1}^k v}$$

Conforme v sea mayor, las distancias relativas entre los valores disminuyen y los masiles terminan siendo sensibles únicamente a las frecuencias, con lo que coincidirán con los fractiles. En el siguiente ejemplo, se muestra la misma distribución de datos, con sumas de constantes cada vez mayores y sus efectos en los masiles (obsérvese el acercamiento de éstos a los fractiles conforme $v \div 4$)

| Frecuen | Fractl | Valor | Masil | Valor | Masil | Valor | Masil |
|---------|--------|-------|-------|-------|-------|-------|-------|
| 1 | 0.071 | 1 | 0.017 | 11 | 0.055 | 5001 | 0.071 |
| 2 | 0.214 | 2 | 0.085 | 12 | 0.176 | 5002 | 0.214 |
| 3 | 0.429 | 3 | 0.237 | 13 | 0.372 | 5003 | 0.428 |
| 2 | 0.214 | 4 | 0.373 | 14 | 0.513 | 5004 | 0.571 |
| 3 | 0.429 | 5 | 0.627 | 15 | 0.739 | 5005 | 0.786 |
| 2 | 0.214 | 6 | 0.821 | 16 | 0.899 | 5006 | 0.928 |
| 1 | 0.071 | 10 | 1.000 | 20 | 1.000 | 5010 | 1.000 |

4. Los masiles desuniformizan la distribución de frecuencias

Los masiles registran la masa de cada valor. Ésta se compone de dos elementos, tal y como hemos visto: la cuantía absoluta del valor ($|x_i|$) y su frecuencia (f_i). Luego, si las frecuencias coinciden, las proporciones también lo harán, pero las masas relativas no. Este efecto podría ser conceptualizado como una desuniformización: una distribución uniforme de frecuencias pasa a una función de masas relativas con más apariencia de exponencial, pues los valores imponen una aceleración positiva. En el siguiente ejemplo, la distribución de base es uniforme:

| Valor | Frecuencia | Propor | M. rel | Fractil | Masil |
|-------|------------|--------|--------|---------|-------|
| 11 | 2 | 0.111 | 0.022 | 0.111 | 0.022 |
| 22 | 2 | 0.111 | 0.044 | 0.222 | 0.067 |
| 33 | 2 | 0.111 | 0.067 | 0.333 | 0.133 |
| 44 | 2 | 0.111 | 0.089 | 0.444 | 0.222 |
| 55 | 2 | 0.111 | 0.111 | 0.556 | 0.333 |
| 66 | 2 | 0.111 | 0.133 | 0.667 | 0.467 |
| 77 | 2 | 0.111 | 0.156 | 0.778 | 0.622 |
| 88 | 2 | 0.111 | 0.178 | 0.889 | 0.800 |
| 99 | 2 | 0.111 | 0.200 | 1.000 | 1.000 |

Así, por ejemplo, si la variable es discreta y se cumple que el intervalo constante entre valores es

$$x_i - x_{i-1} = x_1 \quad \div \quad x_i = x_{i-1} + x_1$$

al ser una función suma, ocurrirá que:

$$mr_i = mr_{i-1} + mr_1 \quad \div \quad mr_i = i \cdot mr_1$$

y:

$$mr_1 = \frac{x_1 h}{k} = \frac{x_1}{k} \quad \text{ó} \quad mr_i = \frac{i \cdot x_1}{k} = \frac{i \cdot x_1}{k}$$

(donde h es la frecuencia de cualquier x_i). Por otro lado, con respecto a la suma total:

$$\sum_{j=1}^k x_j = x_1 + 2x_1 + 3x_1 + \dots + kx_1 = x_1 \frac{k(k+1)}{2}$$

Por lo que, finalmente:

$$mr_i = \frac{i \cdot x_1}{k} = \frac{i \cdot x_1}{x_1 \frac{k(k+1)}{2}} = \frac{2i}{k(k+1)}$$

Dado que tanto 2 como k son constantes, se observa que la función “masa relativa” aumenta conforme lo hace la posición original de x_i , es decir, su valor.

DISCUSIÓN

Hemos visto que los masiles permiten registrar las cuantías de los valores y, con ello, utilizar más información que sólo su orden. Su utilización permite distinguir casos individuales, por su posición en la distribución, contemplando su distancia con respecto al resto de los datos del conjunto. Conforme las distancias entre valores se hace mayor, mayor es también la diferencia entre masiles, mientras que los percentiles son inmutables.

No obstante, el objeto de este trabajo no es proponer un cambio en la utilización de medidas de posición, sustituyendo los fractiles por masiles. Existen algunas razones que no aconsejan esta medida. Por un lado, los masiles sólo pueden ser calculados cuando se cuenta con una variable cuantitativa. Ante una característica nominal u ordinal burda, no tiene sentido considerar la cuantía o masa concreta y específica del valor, puesto que las cantidades utilizadas en una variable medida según una escala ordinal sólo son indicativas de las propiedades “inferior”, “superior”, “mayor” o “menor”.

Por otro lado, la información que suministran los fractiles o sus medidas derivadas, son directamente útiles para multitud de usuarios, habituados ya a este tipo de medidas. La interpretación de las masas relativas es más complicada.

La intención del trabajo es señalar otra estrategia para la interpretación de puntuaciones individuales sin los inconvenientes de los fractiles (sensibles sólo a las frecuencias) ni a las puntuaciones estandarizadas tradicionales que, por carecer de acotación, son difícilmente interpretables.

ORIENTACIONES BIBLIOGRÁFICAS

Cualquier libro de análisis de datos que contenga aspectos relacionados con la Estadística Descriptiva, trata el cálculo de los cuantiles (por ejemplo, Fazeli y otros, 1998). Para acceder a diferentes estrategias de cálculo se puede ver, por ejemplo, Manzano (1993), donde se explicitan las variantes que suministra el programa SPSS.

ESTUDIO DE LA ESTRUCTURA POBLACIONAL MEDIANTE FUNCIONES DE DISTRIBUCIÓN Y DISPERSIÓN

José Antonio Mayor Gallego
Universidad de Sevilla

RESUMEN

El estudio de poblaciones finitas con un enfoque más global que el que proporcionan ciertos parámetros puntuales, como la media o la varianza, puede llevarse a cabo satisfactoriamente mediante ciertas funciones poblacionales como la función de distribución y la función de dispersión. No obstante, debido a las características especiales que presentan estas funciones, los métodos clásicos de estimación de parámetros sobre poblaciones finitas no suelen proporcionar resultados satisfactorios. En este trabajo, presentamos una nueva metodología basada en el estudio global de la función de varianza mediante normas, que permite la obtención de diseños muestrales apropiados para dicha estimación. La comparación de los resultados, con los obtenidos por procedimientos clásicos nos permite calificar esta metodología como prometedora.

INTRODUCCIÓN

Gran parte de la teoría del muestreo en poblaciones finitas se ha centrado tradicionalmente en la estimación puntual de ciertos parámetros poblacionales como medias y totales de una variable, varianzas y razones, que resumen en una única cantidad numérica información sobre las características de la población en estudio. No obstante, podemos considerar el problema de la estimación de otros objetos poblacionales que proporcionen información de naturaleza más global sobre el comportamiento de la población en relación a una o más variables.

Nosotros vamos a considerar el problema de la estimación de la función de distribución poblacional de una determinada característica numérica, y de otra función, íntimamente relacionada con la anterior, denominada función de dispersión.

Con relación a la estimación de la función de distribución, diremos que este importante problema así como otros relacionados, relativos a la estimación de la mediana y en general, la inferencia sobre cuantiles poblacionales, han sido estudiados por varios autores, con diferentes enfoques.

Citemos en primer lugar el trabajo inicial de Woodruff (1952), donde sugiere cómo construir un intervalo de confianza para la mediana y otras medidas de posición, basado en muestreo aleatorio simple. Sendransk y Meyer (1978) han estudiado el mismo problema bajo un enfoque de estadísticos ordenados, para muestreo aleatorio simple y estratificado, Hill (1968) mediante un enfoque bayesiano y Kuk y Mak (1989), usando variables auxiliares.

En relación a la estimación de la función de distribución propiamente dicha, podemos encontrar en la bibliografía diferentes enfoques. Así Chambers y Dunstan (1986) proponen un enfoque de modelo de superpoblación para desarrollar un procedimiento de estimación. Kuk (1988) estudia varios estimadores de la función de distribución bajo muestreo con probabilidades desiguales, y Rao, Kovar y Mantel (1990) mediante información auxiliar.

A continuación vamos a exponer un nuevo enfoque basado en la minimización en norma de la varianza de la función de distribución, lo que nos proporcionará las probabilidades de inclusión de primer y segundo orden más adecuadas.

Para ello, sea U una población finita, $U = \{1, 2, \dots, N\}$, y sea Y una variable numérica de estudio, cuyos valores sobre U son (Y_1, Y_2, \dots, Y_N) , y que sin pérdida de generalidad podemos suponer que estos valores están ordenados de menor a mayor, es decir, $0 < Y_1 < Y_2 < \dots < Y_N$. Nuestro objetivo, como ya se ha mencionado, es la estimación de la función de distribución poblacional de la variable Y ,

$$F(t) = \frac{1}{N} \text{CARD}(\{i \in U \mid Y_i \leq t\})$$

Esta función presenta interés, no sólo por sí misma, ya que expresa el comportamiento global de la variable, sino también por estar relacionada con otros parámetros interesantes. Así, recordemos que la solución o soluciones de la ecuación $F(t) = 1/2$ es la mediana o medianas, y que ecuaciones similares definen los cuantiles poblacionales.

Si empleamos la función indicadora del intervalo $[Y_i, Y_{i+1})$

$$I_{[Y_i, Y_{i+1})}(t) = \begin{cases} 1 & \text{si } t \in [Y_i, Y_{i+1}) \\ 0 & \text{si } t \notin [Y_i, Y_{i+1}) \end{cases}$$

podemos expresar $F(t)$ como,

$$F(t) = \frac{1}{N} \sum_{i \in \mathcal{O}} I_{[Y_i, Y_{i+1})}(t)$$

es decir, una forma lineal, apropiada para poder ser estimada mediante el estimador Horvitz-Thompson.

Supongamos pues que m es una muestra obtenida a partir de U mediante un diseño muestral $d = (M, p(\cdot))$, sin reemplazamiento y con matriz de diseño $\Pi = \{\pi_{ij}\}_{1 \leq i, j \leq N}$ donde, como es usual, suponemos

$\pi_{ii} = \pi_i$. Podemos entonces estimar $F(t)$ mediante el estimador de Horvitz-Thompson,

$$\hat{F}(t) = \frac{1}{N} \sum_{i \in \mathcal{O}_m} \frac{I_{[Y_i, Y_{i+1})}(t)}{\pi_i}$$

Como se sabe, dicho estimador es insesgado, y su varianza se puede expresar de la forma,

$$V[\hat{F}(t)] = \frac{1}{N^2} \sum_{i, j \in \mathcal{O}} (\pi_{ij} - \pi_i \pi_j) \frac{I_{[Y_i, Y_{i+1})}(t)}{\pi_i} \frac{I_{[Y_j, Y_{j+1})}(t)}{\pi_j}$$

de manera que para cada valor de t , dicha varianza proporciona información sobre la acuracidad de la estimación.

ESTUDIO DE LA VARIANZA COMO DE NORMA $\|\cdot\|_1$

Podemos definir una medida de la eficiencia global de la estimación empleando la siguiente distancia basada en la norma $\|\cdot\|_1$,

$$d(V[\hat{F}(t)], 0) = \|V[\hat{F}(t)]\|_1 = \int_{S(\delta', \delta)} V[\hat{F}(t)] dt$$

donde $S(\delta', \delta)$ denota el soporte de la variable Y , adecuadamente extendido, es decir, $S(\delta', \delta) = [Y_1 - \delta', Y_N + \delta]$, $\delta, \delta' \geq 0$. Mediante un cálculo directo, obtenemos,

$$\begin{aligned} \|V[\hat{F}(t)]\|_1 & \cdot \frac{1}{N^2} \int_{i,j \in \Omega} \frac{\pi_{ij} \& \pi_i \pi_j}{\pi_i \pi_j} (Y_N \% \delta \& \max\{Y_i, Y_j\}) \\ & \cdot \frac{1}{N^2} \int_{i,j \in \Omega} \frac{\pi_{ij}}{\pi_i \pi_j} (Y_N \% \delta \& \max\{Y_i, Y_j\}) \\ & \& \frac{1}{N^2} \int_{i,j \in \Omega} (Y_N \% \delta \& \max\{Y_i, Y_j\}) \end{aligned}$$

lo que nos lleva a plantear el siguiente problema de minimización,

$$\min_{\Pi} \cdot \frac{1}{N^2} \int_{i,j \in \Omega} \frac{\pi_{ij}}{\pi_i \pi_j} (Y_N \% \delta \& \max\{Y_i, Y_j\})$$

Para el tratamiento de este problema, podemos seguir dos líneas distintas, aunque ciertamente relacionadas. Una de ellas, buscar procedimientos de muestreo que produzcan diseños muestrales con matrices de diseño que tiendan a hacer pequeña la expresión a minimizar, tomando como base el estudio cualitativo de la expresión a minimizar. Véase Mayor (1997).

La otra posibilidad es la aplicación de técnicas de programación matemática para intentar minimizar la anterior función objetivo. De esta forma, resolviendo el programa de programación fraccional anterior, encontraremos las probabilidades de inclusión de primer y segundo orden, óptimas, siendo posible encontrar incluso el diseño muestral que se ha de aplicar para obtener la muestra, siguiendo la metodología general expuesta en Fernández y Mayor (1996).

En cualquiera de las dos opciones, y ya que la variable de estudio no es conocida, será necesario introducir un modelo de superpoblación adecuado, que relacione dicha variable, Y , con una variable auxiliar, completamente conocida, X . El modelo que hemos supuesto, de gran generalidad, es el siguiente,

$$Y_i \cdot \alpha \% \beta X_i \% g_i, \quad \beta > 0, \quad E_s[g_i] \cdot 0$$

Bajo este modelo, y mediante un cálculo directo se obtiene la siguiente expresión para la esperanza de la varianza,

$$E_s \left[\int_{i,j \in \Omega} \frac{\pi_{ij}}{\pi_i \pi_j} (Y_N \% \delta \& \max\{Y_i, Y_j\}) \right] \cdot \int_{i,j \in \Omega} \frac{\pi_{ij}}{\pi_i \pi_j} (X_N \% \delta \& X_{\max\{i,j\}})$$

expresión a la que sí podemos aplicar las técnicas indicadas.

LA FUNCIÓN DE DISPERSIÓN Y SU ESTIMACIÓN

A continuación presentamos una nueva herramienta para el análisis de poblaciones finitas, en lo que se refiere a la dispersión de determinada variable. Esta herramienta es una adaptación de la función de dispersión para una variable aleatoria, X , definida como $D(t) = E[|X-t|]$. Esta función ha sido ampliamente estudiada en la literatura, véase por ejemplo Muñoz (1990) y Muñoz y Sánchez (1990).

Tomando como referencia el concepto anterior, se define en Mayor (1995) la función de dispersión de una variable Y , sobre una población finita, $U = \{1, 2, \dots, N\}$, en la forma,

$$D_Y(t) \cdot \frac{1}{N} \int_{i \in \Omega} |Y_i \& t| \quad \text{œ } 0 \text{ ú}$$

Los autores citados han demostrado una serie de propiedades de la función de dispersión genéricas, $D(t)$, que se vuelven a reproducir en la versión para la población finita, siendo algunas de ellas las que se exponen a continuación.

- Es continua en todo \hat{U} y además es convexa por ser combinación lineal, con coeficientes positivos, de funciones convexas del tipo $|Y_i - t|$, luego solamente tiene mínimos globales en \hat{U}
- Su gráfica es una poligonal, rectilínea a trozos, siendo los puntos de separación de los diferentes trozos del soporte, $\mathcal{Y} = (Y_1, Y_2, \dots, Y_N)$. Se tiene pues que es una función derivable en $\hat{U} - \mathcal{Y}$, verificando además,

$$\begin{aligned} D_Y(t) &' \bar{Y} \& t & t \neq Y_{(1)} \\ D_Y(t) &' t \& \bar{Y} & t \neq Y_{(N)} \end{aligned}$$

donde $Y_{(1)}$ e $Y_{(N)}$ son, respectivamente, el menor y el mayor valor de la variable Y .

- $D_Y(0) = \bar{Y}$, es decir, la ordenada en el origen coincide con la media poblacional de la variable de estudio.
- Entre la función de dispersión y la función de distribución,

$$F_Y(t) = \frac{1}{N} \text{CARD}\{i | Y_i \leq t\} \quad t \in \hat{U}$$

existe la siguiente relación,

$$D_Y(t) = 2F_Y(t) - 1 \quad t \in \hat{U} \& y$$

- Alcanza el mínimo en la mediana, $M(Y)$, o en todos los puntos del intervalo mediano, $IM(Y)$, según los casos.

Las propiedades anteriores indican que la gráfica de la función de dispersión puede resultar muy ilustrativa para analizar globalmente una población, y también para realizar estudios comparativos entre varias poblaciones.

En lo que se refiere a la estimación de la función de dispersión a partir de una muestra, m , obtenida mediante un determinado diseño muestral, $(M, p(\cdot))$, definido sobre U . Para un valor $t \in \hat{U}$, podemos estimar $D_Y(t)$ mediante el estimador Horvitz-Thompson,

$$\hat{D}_Y(t) = \frac{1}{N} \sum_{i \in \mathcal{O}_m} \frac{|Y_i - t|}{\pi_i}$$

siendo esta estimación insesgada para cualquier valor de t .

Observemos que $\hat{D}_Y(t)$ sigue siendo una función convexa pero su valor mínimo ya no tiene por qué obtenerse en la mediana (o valores medianos) de las observaciones muestrales, Y_i , $i \in \mathcal{O}_m$, a menos que el diseño sea uniforme en las probabilidades de inclusión de primer orden.

Con objeto de encontrar un diseño muestral apropiado para la estimación de la función de dispersión, vamos a considerar la varianza, suponiendo tamaño muestral fijo, dada por la fórmula de Yates-Grundy-Sen (véase Fernández y Mayor (1994)),

$$V[\hat{D}_Y(t)] = \frac{1}{2N^2} \sum_{i,j \in \mathcal{O}_m} (\pi_{ij} - \pi_i \pi_j) \left(\frac{|Y_i - t|}{\pi_i} - \frac{|Y_j - t|}{\pi_j} \right)^2$$

que depende de t y que para un diseño uniforme verificando $\pi_{ij} = \pi_i \pi_j$, toma valor máximo si $t \in Y_{(N)}$ o $t \in Y_{(1)}$, siendo dicho máximo con $\pi_i = \alpha$, $\alpha \in \hat{O}_U$,

$$V_{max} = \frac{1}{2N^2} \sum_{i,j} \frac{\pi_{ij} \alpha^2}{\alpha^2} (Y_i \& Y_j)^2$$

lo que inmediatamente sugiere el uso de estrategias de muestreo que produzcan probabilidades de inclusión de segundo orden mayores para pares de elementos más distantes.

FORMATOS DE RESPUESTA Y NIVELES DE MEDIDA: UNA EXPERIENCIA DIDÁCTICA

José Luis Padilla García
Andrés González Gómez
Cristino Pérez Meléndez
Universidad de Granada

RESUMEN

La Teoría de la Medida forma parte del curriculum de los estudiantes de Ciencias Sociales. Sin embargo, la presentación de sus contenidos suele abordar sólo la clasificación de los niveles de medida propuesta por S. S. Stevens (e. g., Stevens, 1946). La presentación de este apartado en los manuales de referencia está limitada a la enumeración de los niveles con un ejemplo banal o con poco contenido psicológico. Este trabajo presenta una propuesta de experiencia didáctica para facilitar la comprensión por los alumnos de los niveles de medida. La lógica de la experiencia didáctica consiste en vincular el objetivo de la medición con el nivel de medida y mostrar como diferentes formatos de respuesta pueden conducir a distintas interpretaciones para las mismas personas. Los alumnos participan en la experiencia utilizando la información de cada formato para responder al objetivo de la medición.

INTRODUCCIÓN

La teoría de la medición es uno de los contenidos presentes en el currículum del estudiante de Psicología y, cada vez más, del resto de las Ciencias Sociales. Sin embargo, el tratamiento habitual del tema suele ser superficial y fragmentario, quedando limitado en la mayoría de los casos a la presentación, apresurada y fuera de contexto, de la clasificación de los niveles de medida elaborada por S. S. Stevens (e.g., Stevens, 1946). Esta situación contribuye a dificultar el desarrollo de la medición psicológica sobre bases más sólidas que las actuales.

La propia enseñanza de la clasificación de los niveles de medida sufre esta situación. Basta repasar los manuales más utilizados en Psicometría (Crocker y Algina, 1986; Martínez, 1995; Nunnally y Berstein, 1995) para, sin tener que contactar con el "ser omnisciente" de la medición psicológica, adivinar la forma en que se aborda el tema en clase. La presentación se limita a enumerar los diferentes niveles junto con un "enjundioso" ejemplo para cada uno: nominal, los números de las camisetas de los jugadores de diferentes deportes (según las latitudes, fútbol o baloncesto); ordinal, el nivel socioeconómico o las calificaciones en las asignaturas; de intervalos, la medición de la temperatura; y, de razón, el tiempo. Esta presentación descontextualizada y con ejemplos banales o no claramente relacionados con contenidos psicológicos, provoca que los alumnos no capten la importancia que puede tener el tema para su futura práctica profesional y que sea estudiado, en el mejor de los casos, con la "caducidad" marcada por la fecha del examen.

Durante nuestra práctica docente hemos detectado que se consigue un mejor aprendizaje del apartado de los niveles de medida, si su presentación se articula en torno a dos ideas: 1) la interpretación que se desea hacer de los resultados de la medición; y 2) el papel que desempeña el formato de respuesta en el nivel de medida de los números que se obtienen con su aplicación.

El objetivo del presente trabajo es presentar una propuesta de experiencia didáctica diseñada para facilitar el aprendizaje de los niveles de medida, mostrando su relación con el objetivo para el que se realiza la medición y con los formatos de respuesta disponibles para llevarla a cabo. La lógica de la experiencia consiste en mostrar cómo diferentes formatos de respuesta pueden conducir a conclusiones diferentes sobre las mismas personas.

EXPERIENCIA DIDACTICA

Introducción

La presentación de los niveles de medida es parte de las clases dedicadas al tema de la teoría de la medida. Los profesores, antes de abordar este apartado, han presentado ejemplos para destacar la importancia de la medición para la práctica profesional del psicólogo, comentado las grandes etapas en la evolución histórica de la teoría de la medida y abordado los principales contenidos de la versión sobre la medida dominante hoy día en psicología: la teoría representacional. Los profesores han mostrado también a los alumnos un esquema que relaciona el objeto general de la medición psicológica (los constructos), con los instrumentos de medida (tests, cuestionarios, preguntas de una encuesta) y la medición. La Figura 1 presenta el esquema que sirve de referencia para la presentación de los niveles de medida.

Por medio del esquema representado en la Figura 1, los profesores destacan el papel del instrumento de medida como medio para obtener las muestras de comportamiento que una vez "cuantificadas" son la base de la interpretación de los valores del constructo que presentan los sujetos de la medida (personas o estímulos).

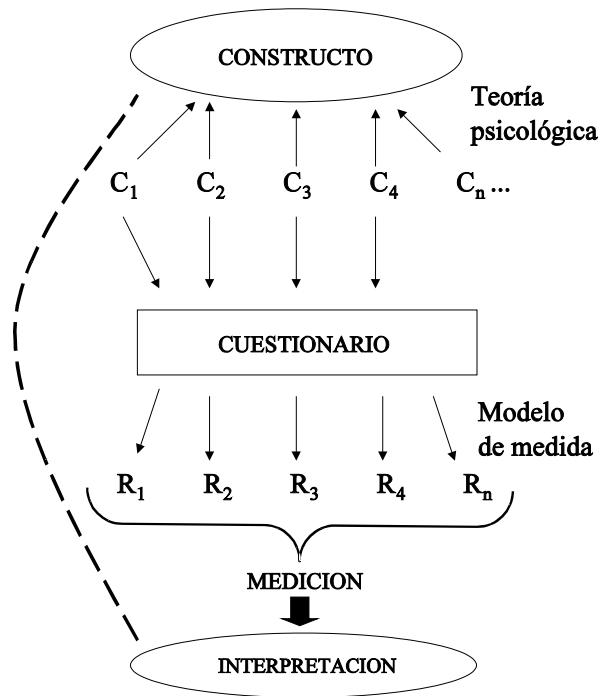
Hemos estructurado la presentación de la experiencia didáctica en una serie de pasos. En cada uno indicamos la actuación del profesor y las tareas que deben realizar los alumnos para lograr el objetivo educativo previsto.

El profesor comienza la clase presentando el objetivo educativo que se pretende alcanzar a través de este contenido: identificar los niveles de medida y relacionarlos con la interpretación deseada de los resultados de la medición; después lo relaciona con el resto del tema y termina la presentación describiendo las actividades que se realizarán en la clase.

Paso 1.

El profesor define los niveles de medida por el tipo de relación que representan los números obtenidos en el proceso de medida. Los niveles de medida se diferencian por representar distintos tipos de relación:

Figura 1. Esquema del proceso de medición



- Nivel nominal: Relaciones de "diferencia o igualdad".
- Nivel ordinal: Relaciones de "más o menos" respecto a la variable.
- Nivel de intervalos: Relaciones de "distancia".
- Nivel de razón: Relaciones de proporción o razón entre los valores.

A continuación, el profesor señala la importancia de identificar el nivel para determinar si se corresponde con el que requiere la utilización que se pretende hacer de la medición. Para aclarar esta relación propone a los alumnos la realización de una experiencia que intenta simular una situación de medida.

Paso 2.

El profesor comenta a los alumnos que va a presentarles un supuesto de medida sobre el que ellos tendrán que realizar una serie de tareas dirigidas a facilitar la comprensión del apartado sobre los niveles de medida. El profesor comienza a describir el supuesto en los términos siguientes:

"Una agencia matrimonial intenta encontrar la "pareja ideal" para los clientes que solicitan sus servicios. Los expertos de la agencia ponen en contacto a personas que tengan una visión de la vida lo más parecida posible..."

En este momento, el profesor debe señalar que relacionar hombres y mujeres con "una visión de la vida lo más parecida posible...", puede requerir no sólo identificar las actitudes, opiniones o valores de estas personas, sino también poder "medir" la distancia que hay entre sus "visiones de la vida". Por tanto, este objetivo de la medición requiere que los números estén al menos en un nivel de intervalos.

Paso 3.

El profesor sigue desarrollando el supuesto:

"Para conocer la visión de la vida de sus clientes, los expertos de la agencia piden a sus clientes que les digan la importancia que estos conceden a una serie de valores a la hora de tomar decisiones sobre aspectos importantes de su vida..."

El profesor interrumpe la descripción del supuesto para indicar a los alumnos que la información demandada por los expertos de la agencia se puede obtener por diversos procedimientos, y que, a continuación, presentaremos algunos de estos procedimientos para analizar la información que proporcionan y compararlos en términos del objetivo de la medición: poner en contacto a personas con visiones de la vida lo más parecidas posibles.

Paso 4.

El profesor comienza a describir la tarea que deberán realizar los alumnos. Los alumnos deben adoptar el papel de expertos de la agencia matrimonial cuya misión es relacionar a

tres clientes. La tarea de los alumnos será recomendar que se relacionen dos de esos tres clientes. Obviamente, deben ser las dos personas con visiones de la vida más parecidas. La decisión ha de tomarse a partir de sus respuestas a un apartado del formulario de admisión. La agencia dispone de tres procedimientos para obtener esta información de los clientes: Procedimiento 1, 2 y 3. Los expertos de la agencia deciden comparar la información que proporcionan cada uno de los tres procedimientos con estos tres cliente a los que conocen desde hace tiempo. La introducción del Procedimiento 1 en el formulario de acogida describe la tarea a la que tienen que responder los clientes en los siguientes términos:

Nos gustaría saber que valores guían sus decisiones sobre aspectos importantes de su vida. Para ello, lea la siguiente lista de valores y escriba al lado de cada uno si para Ud. es "Muy importante", "Regular" o "Poco importante".

Los valores son: Independencia, Bienestar económico, Solidaridad y Vida familiar.

Presentada la tarea, el profesor muestra en una transparencia una tabla con las respuestas de los tres clientes a los que, por la confianza que hay en la agencia, vamos a llamar por sus nombres: Lola, Vicente, y José Luis.

La Tabla 1 presenta las respuestas de Vicente, Lola y José Luis a cada uno de los cuatro valores:

Tabla 1. Respuestas de los tres clientes al Procedimiento 1

| VALORES | CLIENTES | | |
|---------------------|-----------------|-----------------|----------------|
| | Lola | Vicente | José Luis |
| Independencia | Muy importante | Regular | Muy importante |
| Bienestar económico | Muy importante | Poco importante | Muy importante |
| Solidaridad | Muy importante | Poco importante | Muy importante |
| Vida familiar | Poco importante | Poco importante | Regular |

El profesor plantea de nuevo la tarea. Recuerda a los alumnos que el objetivo de la medición es "relacionar" a las dos personas con visiones de la vida más parecidas y que

deben utilizar las respuestas que han dado las personas sobre la importancia que cada valor tiene para ellas. La tarea de los alumnos es recomendar qué dos personas se deberían conocer por tener visiones semejantes de la vida.

Paso 5.

Tras dejar un tiempo prudencial, el profesor retoma el análisis del supuesto para orientar la respuesta a la pregunta. El profesor debe indicar a los alumnos que la respuesta más razonable sería recomendar que se conocieran Lola y José Luis. Si revisamos sus respuestas veremos que Lola tiene una visión de la vida más parecida a la de José Luis (los dos dan la respuesta "Muy importante a tres de los cuatro valores), que a la de Vicente con quién sólo coincide en considerar la "Vida familiar"... poco importante.

A continuación, el profesor debe recordar a los alumnos que los expertos de la agencia disponían de otros dos procedimientos para obtener información sobre la visión de la vida sus clientes. El interrogante ahora sería: ¿qué hubiera ocurrido si respondieran al Procedimiento 2?

Paso 6.

El profesor informa a los alumnos de que van a seguir analizando el supuesto de medida pero imaginando que se hubiera utilizado el Procedimiento 2 de medida. La introducción de este procedimiento en el formulario de acogida dice lo siguiente.

Para conocer mejor su visión de la vida, nos gustaría que ordenara los siguientes valores. Debe escribir el número 1 al lado del valor más importante para Ud., el número 2 al lado del segundo valor más importante para Ud.; el número 3, al lado del tercer valor; y por último, el número 4, al lado del valor menos importante para Ud. Antes de escribir los números revise todos los valores de la serie.

Presentada la tarea del Procedimiento 2, el profesor muestra una transparencia con las respuestas de Lola, José Luis y Vicente.

La Tabla 2 muestra las respuestas de los tres clientes a la tarea planteada en el Procedimiento 2.

El profesor vuelve a recordar la tarea que deben realizar los alumnos. El objetivo de la medición es el mismo que en la situación anterior: "poner en contacto" a las dos personas con visiones de la vida más parecidas. El dato del que disponemos ahora es la ordenación de los valores en función de la importancia que tienen para cada persona. La semejanza en las visiones de la vida se deberá concretar ahora en ordenaciones similares de los valores. De nuevo, la pregunta que deben responder los alumnos es: ¿Qué dos personas recomendarías que se pusieran en contacto?

Tabla 2. Respuestas de los tres clientes al Procedimiento 2

| VALORES | CLIENTES | | |
|---------------------|----------|---------|-----------|
| | Lola | Vicente | José Luis |
| Independencia | 1 | 1 | 2 |
| Bienestar económico | 3 | 3 | 3 |
| Solidaridad | 2 | 2 | 1 |
| Vida familiar | 4 | 4 | 4 |

Paso 7.

Tras dejar el tiempo necesario para responder a la pregunta, el profesor debe retomar el análisis del supuesto de medida para indicar que, a partir de la información aportada por el Procedimiento 2, la recomendación debería ser que se conocieran Lola y Vicente, al haber dado ordenaciones idénticas. El profesor debe señalar que las respuestas dadas por los tres clientes al Procedimiento 2 son *coherentes* con las que dieron al Procedimiento 1.

Tras ello, el profesor debe recordar que los expertos de la agencia disponen aún de un tercer procedimiento. Los expertos, tan perplejos como los alumnos por las diferencias en los resultados de los Procedimientos 1 y 2, deciden analizar las respuestas al Procedimiento 3.

Paso 8.

La introducción del Procedimiento 3 describe en estos términos la tarea que deben realizar las personas:

Nos gustaría confirmar qué valores utiliza como referencia cuando tiene que tomar decisiones sobre aspectos importantes de su vida. Para ello, nos gustaría que leyese la siguiente lista de valores y que les asigne a cada uno un valor de 0 a 100 puntos en función de la importancia que tengan para Ud.

Lola, Vicente y José Luis, presumiblemente mosqueados por la reiteración en el tema, realizan el reparto de puntos que el profesor muestra a los alumnos en una transparencia.

La Tabla 3 muestra los puntos asignados por los tres clientes a cada uno de los valores:

Tabla 3. Respuestas de los tres clientes al Procedimiento 2

| VALORES | CLIENTES | | |
|---------------------|----------|---------|-----------|
| | Lola | Vicente | José Luis |
| Independencia | 100 | 25 | 75 |
| Bienestar económico | 50 | 15 | 55 |
| Solidaridad | 80 | 20 | 85 |
| Vida familiar | 20 | 5 | 25 |

De nuevo, la tarea de los alumnos es recomendar que se relacionen las dos personas con las visiones de la vida más parecidas. El dato ahora para tomar esa decisión son las estimaciones que han dado cada uno sobre la importancia que esos valores tienen para ellos.

Paso 9.

El profesor debe indicar a los alumnos que la respuesta más razonable sería volver a poner en contacto a Lola con José Luis. Si repasamos sus respuestas se puede constatar que tienen las "visiones de la vida" más similares, por haber dado las estimaciones numéricas más próximas entre sí, por ejemplo: el valor "Independencia" recibe de parte de Lola 100 puntos y de José Luis 75, mientras que Vicente le asigna 25; el valor Solidaridad, recibe 80, 85 y 20 puntos, respectivamente, etc. El profesor debe señalar a los alumnos que si revisamos los puntos asignados a cada uno de los valores podemos constatar que las distancias entre las asignaciones de Lola y José Luis son más pequeñas, que las que hay entre las asignaciones de Lola y Vicente. Por tanto, la recomendación razonable es la de que Lola y José Luis se conozcan...

Paso 10... y último

El profesor presenta ahora la conclusión final de la experiencia didáctica. Primero, debe llamar la atención de los alumnos sobre dos hechos que revela el ejercicio: 1) las respuestas de las tres personas a los tres procedimientos son "coherentes" entre sí; y 2) las diferencias entre los procedimientos conducen a respuestas e interpretaciones diferentes.

Parece razonable esperar que los alumnos que hayan seguido las vicisitudes de los "tres corazones solitarios" se planteen al menos dos interrogantes: 1) ¿Por qué hemos alcanzado recomendaciones diferentes para las mismas tres personas? 2) ¿Cuál es la más adecuada para el objetivo de la medición?.

Para responder al primero es necesario volver al esquema con que se introdujeron las nociones de constructo, procedimiento de medida y medición. La teoría psicológica sobre la variable objeto de la medición debe proporcionar la relación de indicadores que la ponen de manifiesto (en nuestro caso, la importancia asignada a una serie de valores como indicador de la visión de la vida de las personas); y también, proponer el procedimiento observacional (Procedimiento 1, 2 ó 3), para recoger las muestras de conducta con las que inferir el valor de la variable que presentan las personas. El procedimiento de medida no es independiente de la definición del objeto de medida que maneje la teoría. Luego, no debemos sorprendernos de que formatos de respuesta diferentes nos lleven a resultados distintos.

Respecto a qué interpretación es la más adecuada, debemos volver de nuevo a la teoría psicológica para obtener la respuesta. La teoría debe recoger la relación entre el constructo y otras variables. No disponemos de otra estrategia que la de relacionar las mediciones obtenidas con cada uno de los procedimientos con esas otras variables, hasta encontrar una evaluación indirecta del procedimiento por la vía de cuál muestre la relación prevista por la teoría. De todas formas, siempre podremos confiar en que Lola, ante la incertidumbre de los expertos, decida conocer a los dos y que la vida nos ofrezca el criterio con el que validar los tres procedimientos...

CONCLUSIONES

Las conclusiones están limitadas a recordar la finalidad de la propuesta y los objetivos educativos a los que responde. La intención de los autores es realizar un estudio para comparar los resultados de la aplicación de esta experiencia frente a la presentación

"tradicional" de los niveles de medida. Los resultados de dicho estudio contribuirán a valorar la calidad de la propuesta.

La finalidad de la experiencia didáctica es contribuir a mejorar la enseñanza de uno de los contenidos peor tratados en el curriculum de Psicología: la Teoría de la Medida. A pesar de que los contenidos de la teoría son de aplicación a todas las disciplinas psicológicas, la situación más extendida es que estén incluidos dentro de los programas de Psicología Matemática, Análisis de Datos o Psicometría. Esta ubicación, junto con una presentación que no la pone en relación con el resto de los contenidos psicológicos, dificulta que el alumno valore la importancia que pueda tener para su formación.

Los objetivos educativos a los que pretende contribuir la experiencia didáctica se pueden resumir en los siguiente puntos:

- Incidir en la relación entre el objetivo de la medición y los niveles de medida que resultan de la aplicación de los procedimientos de medida.
- Mostrar cómo el formato de respuesta condiciona la interpretación de los resultados de la medición.
- Señalar la importancia de que la teoría psicológica ofrezca la relación de conductas indicadoras de la variable y el procedimiento adecuado para medirla.

La práctica docente irá diciendo el grado de acierto de nuestras intuiciones y propuestas para mejorar la enseñanza de la Teoría de la Medida.

UNA APROXIMACIÓN AL BOOTSTRAP. APLICACIONES Y LIMITACIONES

Carlos Arias Martín
Victoria M. Marquez de la Plata y Cuevas
Universidad de Sevilla

RESUMEN

Los métodos estadísticos por ordenador, entre los que se encuadra el Bootstrap, son un ejemplo del impacto que tienen las nuevas tecnologías en todas las ciencias, no sólo en lo que se refiere al almacenamiento de datos o al tratamiento de la información, sino en los cambios que se realizan en la forma de investigar la realidad que nos rodea. Esta realidad cambiante, y abundante en información, requiere un esfuerzo constante a los investigadores que tratan de conocer más, y mejor, determinados aspectos de la misma. Con este objetivo se realizan aportaciones como los métodos bootstrap, que buscan conseguir mejores resultados con costes más bajos, es decir, mejores estimadores sin incrementar el tamaño de las muestras, y sin necesidad de realizar hipótesis restrictivas sobre el comportamiento de la población.

INTRODUCCIÓN

El primero de los Métodos Estadísticos Intensivos por Ordenador, como los dieron en llamar Diaconis y Efron, que apareció fue el *Jackknife*, ideado por Maurice Quenoille en 1949, y bautizado y reformado por John W. Tukey un año más tarde. Partiendo de una muestra, va suprimiendo una observación cada vez, y estudia, en cada una de las diversas muestras que resultan, lo que en principio estudiaría en una sola: la muestra inicial o base.

El Bootstrap debe el nombre a su funcionamiento, y hace referencia al movimiento de "izarse a uno mismo" tirándose de los cordones de las botas. Surgió como un intento de reducir la varianza de un estimador cualquiera, intentando eludir la dificultad matemática de algunas expresiones, cuando el estimador de interés resultaba ser distinto de la media muestral. No existe un único método bootstrap, sino tres, y de ellos múltiples variantes.

Quizás sea necesario prescindir, hasta cierto punto de las distintas clasificaciones de esta metodología y centrarnos en describir las dos variantes esenciales del Bootstrap: Bootstrap No Paramétrico Con Reemplazo (BCR) y el Bootstrap No Paramétrico Sin Reemplazo (BSR).

El primero de ellos, Bootstrap No Paramétrico Con Reemplazo, funciona siguiendo las siguientes etapas. De una población: $X=(X_1, X_2, \dots, X_N)$, se extrae una muestra con reemplazo de tamaño n : $x=(x_1, x_2, \dots, x_n)$ que llamamos muestra inicial o base, de ella se extraen B muestras, todas del mismo tamaño y diseño que la muestra inicial. Cada una de las muestras así extraídas denominada "muestra bootstrap". En cada una de éstas se calcula la estimación en la que estamos interesados, obteniendo entonces una colección de B estimaciones, las B "replicaciones bootstrap"; es decir, un vector aleatorio de componentes que son variables aleatorias independientes e idénticamente distribuidas; finalmente se toma como estimación del valor del parámetro la media de las estimaciones de todas las muestras bootstrap. Para medir su precisión podemos calcular la varianza del estimador bootstrap. Disponemos de un conjunto de replicaciones bootstrap $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \dots, \hat{\theta}^{(B)}$, un vector formado por los estimadores de cada muestra bootstrap, que por ser variables aleatorias independientes e idénticamente distribuidas, todas tienen la misma media y la misma varianza. Si se define ahora el estimador bootstrap para las B muestras bootstrap, tenemos:

$$\hat{\theta}^{(c)} = (1/B) \sum_{b=1}^B \hat{\theta}^{(b)}$$

donde $\hat{\theta}^{(b)}$, es la replicación bootstrap en la muestra bootstrap: x^{*b} , $b=1,2, \dots, B$. Es fácil comprobar que la varianza del estimador es:

$$Var(\hat{\theta}^{(c)}) = (1/B)^2 Var\left(\sum_{b=1}^B \hat{\theta}^{(b)}\right) = \frac{1}{B} Var(\hat{\theta}^{(b)})$$

es decir, la varianza del estimador de B muestras será menor que el que ofrece una única muestra. Como a su vez, las muestras bootstrap y la inicial deben ser idénticamente distribuidas puesto que se extraen bajo el mismo diseño muestral, también será menor que la varianza del estimador que se obtenga para la muestra inicial. Por lo tanto cabe esperar que el bootstrap, en esta primera versión, disminuya la varianza del estimador.

Otra versión del bootstrap que trata de mejorar la estimación bootstrap sin incrementar el tamaño de la muestra inicial es la que se conoce como el Bootstrap Sin Reemplazo, ya que para éste la extracción de los elementos que forman las muestras bootstrap se realiza sin reemplazo a partir de la muestra inicial. El primer autor que sugirió esta alternativa fue Gross (1980), posteriormente Bickel Y Freedman en 1984 perfeccionaron el método que propuso Gross, y entre unos y otros, Diaconis y Efron en 1983, sugirieron otra versión. Nos interesamos ahora por esta última, donde se propone el siguiente proceso que exponemos a continuación.

1) Sea una muestra inicial $x = (x_1, x_2, \dots, x_n)$ extraída de una población $X = (X_1, X_2, \dots, X_N)$, que puede considerarse como un vector de N componentes desconocidas, con función de distribución de probabilidad F . Se considera una característica de interés, $\theta = f(F; n)$, es decir, el parámetro se considera función de la distribución F desconocida, y del tamaño de la muestra n .

2) Se repite cada elemento de la muestra inicial un número grande de veces, que llamamos k (los autores utilizan hasta 10^9) mezclándolos concienzudamente, y se extrae una muestra, de tamaño n , de este conjunto que se denomina *población artificial*. Para que sea una

muestra aleatoria simple debe cumplir que todos los elementos que la forman tengan la misma probabilidad de formar parte de ella; en este caso, dicha probabilidad será $1/nk$. A esta muestra la denotamos por: x^{*1} , y será la primera muestra bootstrap¹.

En los diversos métodos bootstrap sin reemplazamiento para poblaciones finitas, el número de replicaciones es un elemento esencial de las diferencias que tienen entre sí, concretamente las versiones de Gross, Bickel y Freedman así como la de Sitter, tienen en común que el número de replicaciones es tal que el vector con los elementos ya copiados tiene una dimensión igual o muy similar a la de la población, es decir, la población artificial tiene el mismo tamaño que la población en la que estamos interesados.

3) Se repiten los dos primeros pasos, de forma independiente, un número B de veces. Se obtienen B muestras bootstrap: $x^{*1}, x^{*2}, \dots, x^{*B}$, es decir, tenemos una colección de B muestras aleatorias de tamaño n , que se obtienen sin reemplazamiento de los elementos de la población artificial.

4) Se calcula para cada muestra bootstrap, la estimación del parámetro de interés, es decir, $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \dots, \hat{\theta}^{(B)}$, “versiones bootstrap”² de $\hat{\theta} = f(x_1, x_2, \dots, x_n; \hat{F}, n)$, donde la estimación bootstrap $\hat{\theta}^{(b)}, \hat{\theta}(x^{(b)}; \hat{F}, n)$, depende de la muestra bootstrap x^{*b} para $b=1, 2, \dots, B$; y de la función de distribución empírica \hat{F} , que se construye asignando la probabilidad de $1/n$ por cada vez que se observe que x_i (para cualquier $i, i = 1, 2, \dots, n$) aparece en la muestra. Llamando s al número de veces que un elemento aparece en la muestra (s variará desde 0 hasta n) si x_i aparece en s ocasiones, la probabilidad a asignar será s/n .

¹ Se han descrito formas para las que emplear el bootstrap con reemplazamiento, introduciendo el factor de corrección, y se engloban en los “bootstrap con re-escala”. Rao y Wu (1988), págs. 620-630.

² Tal y como se indica en la tesis doctoral *Simulaciones bootstrap: correcciones a su variabilidad muestral*, de Muñoz y Reyes (1994), en la literatura se consideran dentro del método 2, todas aquellas versiones del bootstrap donde se proponga como método de aproximación a la verdadera función de distribución, la función de distribución empírica, \hat{F} distribución bootstrap.

5) Finalmente se representa el histograma, o cualquier otra estimación de la distribución de $\hat{\theta}^{(b)}$ con $b = 1, 2, \dots, B$; y se toma como estimación del parámetro la media de las estimaciones que se obtienen para todas las muestras bootstrap, es decir:

$$\hat{\theta}^{(\cdot)} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^{(b)}$$

En este último método, se espera la disminución de la varianza, puesto que al extraerse los elementos de las muestras bootstrap sin reemplazo, la varianza del estimador bootstrap será menor que la del estimador clásico.

En ninguno de los métodos bootstrap se ha impuesto restricción funcional alguna sobre \hat{F} , por eso a los dos métodos aquí descritos se les conoce como *Bootstrap No Paramétrico*.

RESULTADOS OBTENIDOS

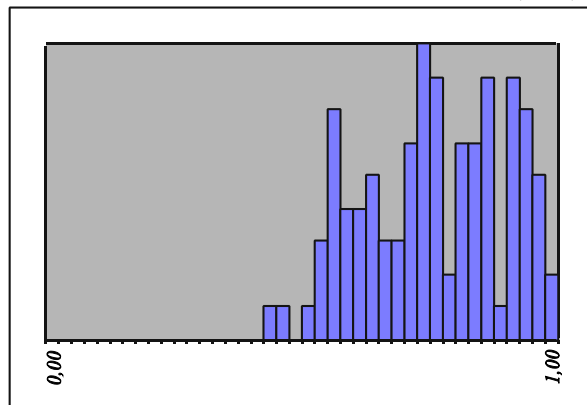
Se ha realizado un estudio de simulación, utilizando el mismo ejemplo que Efron, pero para someter el método a condiciones más adversas, se han considerado otras muestras precedentes de la misma población que al del autor, además de la de Efron. El ejemplo es en realidad de otro autor, concretamente de D.B. Rubin, y aparece recogido en su obra *Using empirical Bayes techniques in the Law School Validity Studies Law School Admission Council* (1977). La población está formada por pares de puntos ordenados según las calificaciones de dos pruebas de conocimiento realizadas a 82 alumnos que ingresaron en 1973 en facultades de derecho americanas; y el parámetro por el que estamos interesados es la correlación lineal entre las calificaciones obtenidas en la prueba de acceso a la universidad (LSAT); y el test de ingreso en las facultades de derecho (GPA). El valor de dicho parámetro es de 0.759997856. Se han extraído de la misma población seis muestras, todas del mismo tamaño que la de Efron con $n=15$, que hemos llamado Muy Buena, Buena, Regular, Mala, Muy Mala 1 y Muy Mala 2, según el criterio del sesgo relativo, es decir:

$$\frac{|\text{estimación \& parámetro}|}{\text{parámetro}}$$

Tal y como se define este cociente, si se aleja de cero la “calidad” de la muestra empeora. Hemos comparado los resultados con los del autor, estableciendo una ordenación para las muestras, y se observa que existen ciertas coincidencias, sobre todo en las tres peores muestras; de forma que cuando la muestra inicial se aleja del parámetro las estimaciones bootstrap, con o sin reemplazo, no son mejores que la clásica. En el siguiente cuadro se ofrece la ordenación de las muestras seleccionando la mejor estimación de todo el conjunto de estimaciones realizadas (B=25, 50, 100, 200, 400, 800, 1.600, 3.200, 10.000, 100.000, 1.000.000 y 10.000.000; y k=100, 1000, 10.000, 100.000, 1.000.000).

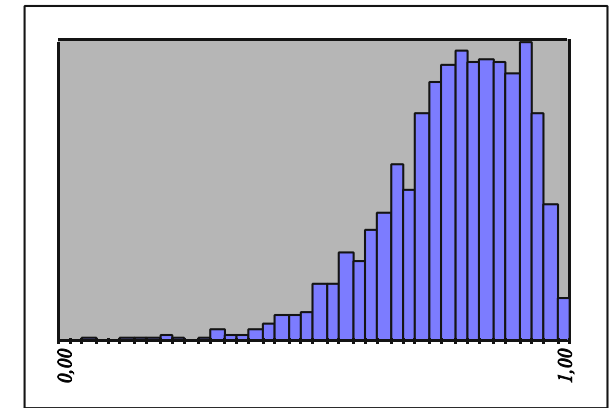
| Ordenación de las muestras según el sesgo relativo de las estimaciones clásicas | Ordenación de las muestras según el sesgo relativo de las estimaciones <i>jackknife</i> | Ordenación de las muestras según el sesgo relativo de las estimaciones bootstrap (BCR) | Ordenación de las muestras según el sesgo relativo de las estimaciones bootstrap (BSR) |
|---|---|--|--|
| Muy Buena Efron | Muy Buena Efron | Regular | Muy Buena Efron |
| Buena | Regular | Muy Buena Efron | Regular |
| Regular | Muy Mala 2 | Buena | Buena |
| Mala | Mala | Mala | Mala |
| Muy Mala 2 | Buena | Muy Mala 2 | Muy Mala 2 |
| Muy Mala 1 | Muy Mala 1 | Muy Mala 1 | Muy Mala 1 |

Figura 1. Histograma de las replications bootstrap de la correlación, en la muestra de Efron, con B=100 (BCR)



Por otra parte, se ha comprobado la similitud que existe entre el histograma construido con 8000 muestras extraídas de la población, es decir, la distribución empírica del estimador, y el construido con 3200 replications bootstrap sobre la muestra de Efron. Este parecido no se da cuando el número de muestras bootstrap es de 100, tal y como se muestra en las siguientes figuras:

Figura 2. Histograma de las replications bootstrap de la correlación, en la muestra de Efron, con B=3200 (BCR)



Otro aspecto importante en la estimación de la correlación, es que la distribución de ésta resulta ser asimétrica a la izquierda, puesto que la correlación es positiva, y creemos que como consecuencia de este hecho, para las muestras que subestiman el parámetro, el bootstrap no consigue acercarnos al valor del parámetro, como en los casos de las muestras Mala y muy Mala 1.

Tal y como anuncian los autores, un número de 25 ó 50 muestras bootstrap ya es suficiente para tener una idea de la estimación de la correlación, aunque para el bootstrap con reemplazo los mejores resultados se han obtenido para B=100 y B=400, no obstante, se han realizado simulaciones con B=10⁷, sin que el coste de la realización de estos cálculos se haya compensado con un incremento en la precisión del estimador. En el caso del bootstrap sin reemplazo, fijando K se han obtenido los siguientes resultados:

| Muestras | B que hace mínimo el sesgo relativo con K=100 | B que hace mínimo el sesgo relativo con K=1000 | B que hace mínimo el sesgo relativo con K=10000 | B que hace mínimo el sesgo relativo con K=100000 |
|------------|---|--|---|--|
| Muy Buena | 100 | 400 | 400 | 100 |
| Efron | 400 | 400 | 400 | 100 |
| Buena | 200 | 100 | 200 | 100 |
| Regular | 200 | 3200 | 200 | 3200 |
| Mala | 3200 | 100 | 100 | 200 |
| Muy Mala 2 | 100 | 400 | 200 | 200 |
| Muy Mala 1 | 3200 | 200 | 200 | 200 |

Pero tal y como se hacía notar al principio, la ventaja de éste método de estimación está en que en ningún momento se ha necesitado realizar hipótesis sobre la distribución que sigue el estimador, y la limitación que tiene, común a la inferencia clásica, está en la calidad de la muestra inicial.

EFFECTOS DE LA REDACCIÓN DE LOS ÍTEMS EN LAS RESPUESTAS DE LOS ENCUESTADOS: UNA APLICACIÓN CON ESCALAS TIPO LIKERT.

Alicia García Pereira
 María José Martín Torrejón
 Regla Martínez Jiménez
 Alicia Puerto Martínez
Universidad de Sevilla

RESUMEN

Con objeto de observar los efectos de la redacción de los ítems en las respuestas de los sujetos, se llevó a cabo una investigación mediante encuestas en la cual fueron seleccionados 200 alumnos de la Facultad de Psicología de Sevilla. Mediante la aplicación de un cuestionario, se recogieron sus respuestas, en relación con sus opiniones sobre el plan de estudios del centro. En algunos ítems escogidos, fueron introducidas algunas variaciones. Como era de esperar, las modificaciones realizadas en estos ítems (una negación, una variación en el cuantificador o un cambio en el sentido), produjeron cambios en las respuestas de los sujetos.

INTRODUCCIÓN

Los procedimientos de investigación mediante encuestas establecen reglas que nos permiten acceder de forma científica a lo que las personas opinan. Uno de los principales objetivos de estas reglas es que un segundo investigador pueda repetir el proceso siguiendo los mismos pasos, es decir, debe ser sistemático.

Aunque decidamos seguir una vía científica, la forma más habitual de abordar lo que la gente piensa u opina es preguntándole. Y si bien el acto de preguntar e indagar en la opinión o pensamiento de otros, es común a multitud de ámbitos (incluyendo la vida cotidiana), la generación y gestión de las preguntas recibe un tratamiento especial en las ciencias sociales y del comportamiento. Así, la construcción de un cuestionario, como documento que recoge de forma organizada las preguntas sobre el objetivo de la investigación, sigue un proceso sistemático y del que se espera que cubra un mínimo de rigor metodológico.

Así pues, hay una preocupación comprensible y extendida acerca de cómo deben realizarse las preguntas (o ítems, en sentido más amplio y técnico) dentro de un cuestionario.

En un cuestionario, la totalidad de las preguntas están determinadas. Además, con el objeto de facilitar la codificación posterior, en la mayoría de los casos la/s respuesta/s a cada pregunta sólo puede/n ser alguna/s fijada/s en el texto. Por ello, este tipo de preguntas recibe la denominación de *preguntas cerradas*.

Dentro de este grupo se puede incluir el formato de respuesta tipo Likert, en el que se presenta a un grupo de personas un conjunto de ítems, formados tanto por ítems favorables como desfavorables, con respecto a una actitud que se pretende medir.

Se les pide a los sujetos que respondan en función de su grado de acuerdo / desacuerdo. Cuando más favorable (o desfavorable) sea la actitud mayor (o menor, respectivamente) será su puntuación en el ítem.

La escala tipo Likert es uno de los formatos más utilizados por investigadores y encuestadores en el ámbito de las ciencias psicosociales. Es por ello que consideramos importante profundizar en este tema. Así, un aspecto de gran trascendencia es la redacción

de los ítems en dichas escalas de medida, ya que su redacción puede condicionar el comportamiento del sujeto. Por ejemplo, se debe evitar las denominadas tendencias de respuesta del encuestado o entrevistado, como es el caso de la inclusión de frases negativas, que puede alterar la forma de responder la escala.

Partiendo de todo lo expresado anteriormente, se propuso realizar una investigación mediante una encuesta, con el procedimiento general utilizado en este campo. El objetivo fue observar si la redacción de un ítem, en una escala tipo Likert, afecta a la tendencia de respuesta de los sujetos.

MÉTODO

Sujetos

Se utilizó una muestra de 200 sujetos, alumnos de la Facultad de Psicología de la Universidad de Sevilla, con edades comprendidas entre los 17 y 27 años y de ambos sexos.

La encuesta se aplicó en las aulas de dicha facultad. Para aumentar las garantías de representatividad de la muestra se estratificó la variable *curso* con afijación constante.

Por otro lado, dos grupos de sujetos diferían en un aspecto: los alumnos de 5º curso pertenecen al plan 79 y los de los restantes cursos al plan 94, aunque excepcionalmente podíamos encontrarnos con personas adaptadas a este último plan. Considerando que el objetivo de medida del cuestionario era recoger información sobre la actitud de los alumnos frente a los planes de estudio, la existencia de las dos subpoblaciones era deseable.

Los cuestionarios empleados estaban formados por 31 ítems, de los cuales se modificaron 6 de ellos, quedando finalmente 3 tipos de cuestionarios distintos que fueron repartidos aleatoriamente a los alumnos que formaron parte de esta investigación. (Ver anexo).

Instrumento / material

El instrumento utilizado en la investigación fue el cuestionario, el cual fue estructurado de la siguiente forma: por un lado se podía encontrar una serie de preguntas cerradas a las que debían contestar señalando una opción, así como preguntas abiertas a las que se respondían con dígitos. Por otro lado, había preguntas que constituían una escala tipo Likert, con 5 valores de respuesta que oscilaban entre muy de acuerdo – muy en desacuerdo. También

se utilizaron preguntas-llave que permitían que todos los sujetos respondieran a la primera parte del cuestionario (ítems 1 – 15); los que iniciaron sus estudios en el plan 94 responderían hasta el ítem 18; y los sujetos adaptados (del plan 79 al 94) contestaban hasta el ítem 27. Además, todos los sujetos eran remitidos a la pregunta 28 donde rellenaban algunos datos identificativos (sexo, edad, curso y año inicio carrera).

Esta estructura fue la misma en todos los ejemplares pero, en función de los objetivos de la investigación, se cambiaron algunos aspectos del contenido de las preguntas, por lo que se les aplicó tres tipos de cuestionarios que variaban en los ítems: 10, 11, 12, 13, 21 y 23. La tabla 1 muestra las redacciones concretas originales (grupo 1) y sus variaciones (grupos 2 y 3).

Tabla 1: Relación de las preguntas que varían entre grupos de sujetos.

| | |
|--------------------|---|
| <i>Pregunta 10</i> | Grupo 1: Tu horario te permite llevar a cabo actividades extraescolares Grupo 2,3: Tu horario te impide llevar a cabo actividades extraescolares |
| <i>Pregunta 11</i> | Grupo 1: Tus estudios requieren que hagas uso de la biblioteca Grupo 2,3: Tus estudios no requieren que hagas uso de la biblioteca |
| <i>Pregunta 12</i> | Grupo 1: Tus estudios requieren que hagas uso de la hemeroteca Grupo 2,3: Tus estudios no requieren que hagas uso de la hemeroteca |
| <i>Pregunta 13</i> | Grupo 1: Existen grandes diferencias entre el plan 1979 y el plan 1994 Grupo 2,3: Existen grandes similitudes entre el plan 1979 y el plan 1994 |
| <i>Pregunta 21</i> | Grupo 1: Te ha costado mucho adaptarte Grupo 2: Te ha costado adaptarte Grupo 3: Te ha costado poco adaptarte |
| <i>Pregunta 23</i> | Grupo 1: La nueva organización de la fecha de los exámenes es más acertada que la anterior Grupo 2,3: La anterior organización de la fecha de los exámenes es más acertada que la actual |

Procedimiento

Se utilizó un total de 200 cuestionarios. Para poder contar con el mayor número de sujetos posible, se seleccionaron asignaturas troncales de cada curso y, bajo previa cita con el

profesor responsable, se procedió a la aplicación de los cuestionarios, repartiéndose 40 ejemplares en cada curso. Dentro de los 40 se encontraban los tres tipos de cuestionarios mezclados, los cuales se repartieron aleatoriamente a los alumnos.

Cabe destacar que los cuestionarios fueron rellenados de forma anónima. El tiempo que los sujetos tardaron en cumplimentarlos osciló entre 15 y 20 minutos.

Análisis de los datos

Como se indicó en el apartado anterior, dado que se modificó el sentido de algunos ítems, hubo que tener en cuenta dicho cambio a la hora de analizar las respuestas de los sujetos. Para ello, se realizó la operación

$$p' = 6-p$$

donde p representa la puntuación directa en el ítem invertido y p' la puntuación que se utilizó finalmente en el análisis de los datos.

Con los datos obtenidos se realizó, en primer lugar, un análisis descriptivo. En el proceso de inferencia, preocupó medir el efecto de los cambios en los ítems mencionados, mediante la aplicación de la U de Mann-Whitney en el caso de la comparación entre dos grupos (ítem original - ítem invertido) y la H de Kruskal-Wallis para la medida del efecto con tres grupos (ítem original y dos variaciones en el cuantificador).

RESULTADOS

Se presenta a continuación una tabla resumen de las comparaciones realizadas en aquéllos ítems que contaban con dos redacciones diferentes. La prueba de comparación empleada en estos casos ha sido la U de Mann-Whitney.

Se aprecian diferencias significativas ($\alpha = 0.05$) en dos de las comparaciones realizadas mediante el estadístico U de Mann-Whitney. La tabla 1 muestra las redacciones de los ítems implicados. Se observa que los ítems 11 y 12 son los únicos cuya variación ha implicado incluir un “no” en la redacción. El grupo 2, que contiene la negación, muestra una media de rangos significativamente mayor (mayor grado de acuerdo). Recuérdese que estos ítems fueron invertidos para el análisis. Luego, en el cuestionario original, la tendencia del grupo 2 a estar en desacuerdo ha sido significativamente mayor que la tendencia del grupo 1 a estar de acuerdo.

En el ítem 21 se redactaron tres enunciados diferentes, por lo que la prueba empleada ha sido la H de Kruskal-Wallis.

COMPARACIONES U de MANN-WHITNEY

| | Media de Rangos | | Valor U | Grado de significación |
|----|-----------------|---------|---------|------------------------|
| | Grupo 1 | Grupo 2 | | |
| 10 | 93,39 | 101,61 | 4305,5 | 0,2883 |
| 11 | 85,43 | 109,57 | 3534,0 | 0,0009 |
| 12 | 82,93 | 110,92 | 3305,5 | 0,0002 |
| 13 | 96,65 | 88,16 | 3839,5 | 0,2568 |
| 23 | 12,17 | 10,07 | 42,5 | 0,4652 |

PRUEBA H DE KRUSKAL-WALLIS

| | Media de Rangos | | | H | Grado de significación |
|----|-----------------|---------|---------|--------|------------------------|
| | Grupo 1 | Grupo 2 | Grupo 3 | | |
| 21 | 13,50 | 4,00 | 14,38 | 7,1423 | 0,0281 |

Para obtener una mayor capacidad explicativa de las diferencias constatadas, se realizaron diversas pruebas de U de Mann-Whitney entre los tres grupos, seleccionados dos a dos, que mostraron, como se indica en la siguiente tabla, diferencias significativas ($\alpha = 0,05$) tan sólo al comparar el primer y segundo grupo de sujetos, obteniendo éste un promedio de rangos significativamente inferior al de aquél (es decir, puntuaciones menores en su respuesta).

| | Grupo 1 | Grupo 2 | Grupo 1 | Grupo 3 | Grupo 2 | Grupo 3 |
|------------------------|---------|---------|---------|---------|---------|---------|
| Media de Rangos | 11,77 | 3,38 | 9,73 | 11,00 | 3,13 | 5,88 |
| U de Mann-Whitney | 3,5 | | 26,0 | | 2,5 | |
| Grado de significación | 0,0068 | | 0,6786 | | 0,0885 | |

Los grados de acuerdo con el enunciado se respetaron en el caso de los grupos 1 y 2. Pero en el grupo 3, la variable fue recodificada por el mismo procedimiento especificado $p' = 6-p$, al considerar que el cuantificador “poco” invierte el sentido de la sentencia, frente a “mucho”. En el análisis global de los tres grupos, se observa que la media de rangos es similar en las redacciones con cuantificador. Luego y en contra de lo que pudiera dictaminar el sentido común, parece que el cuantificador “mucho” aumenta la tendencia al acuerdo (al polo positivo), con respecto a no incluir ningún cuantificador. Por otro lado, “poco” parece conseguir el mismo efecto, si bien hacia el polo opuesto (no obstante, al invertir la respuesta, las puntuaciones finales coinciden aparentemente con el cuantificador opuesto).

DISCUSIÓN

Las diferencias en las respuestas de los sujetos se observaron al incluir la negación en los ítems. También al modificar el gradiente poco/ mucho. Por lo tanto, estos cambios alteraron las tendencias de respuesta de los sujetos en esta investigación. Así:

1. La inclusión de la negación explícita “no”, favorece el que los encuestados manifiesten mayor desacuerdo que el acuerdo que mostrarían si el enunciado se redactase en términos afirmativos.
2. La existencia de un cuantificador, por sí mismo, parece generar diferencias con respecto a su ausencia. Así, en el ítem 21 se ha observado que la tendencia de respuesta ha aumentado hacia el grado de acuerdo cuando se ha incluido un cuantificador.

11. Tus estudios requieren que hagas uso de la biblioteca.

12. Tus estudios requieren que hagas uso de la hemeroteca.

13. Existen grandes diferencias entre el plan 1979 y el plan 1994.

14. ¿A qué plan de estudios perteneces en la actualidad? 1979 1994

15. ¿Te cambiarías al plan de estudios contrario al que estás realizando?

Sí No

SI PERTENECES AL PLAN DE ESTUDIOS 1979, PASA A LA PREGUNTA N°28.

16. ¿Cuántos créditos de libre configuración tienes hechos?

17. Si no tienes hecho ningún crédito de libre configuración ¿cuándo piensas realizarlos?

En los últimos años de carrera.

Mientras realizas el prácticum.

Cuando termines la carrera.

¿Has pertenecido al plan de estudios 1979? Sí No

SI HAS RESPONDIDO QUE “NO”, PASA A LA PREGUNTA N° 28

19. ¿Por qué te cambiaste de plan?

Creo que recibiré una mejor formación.

1. Muy en desacuerdo 2. En desacuerdo 3. Ni de acuerdo ni en desacuerdo 4. De acuerdo 5. Muy de acuerdo

No quería continuar en un plan en extinción.

Dispongo de clases de todas las asignaturas.

No me corren convocatorias si no las utilizo.

Otras (indicar a continuación):

20. Hay un cambio importante respecto al plan 1979 en cuanto a:

Número de asignaturas.

1. Muy en desacuerdo 2. En desacuerdo 3. Ni de acuerdo ni en desacuerdo 4. De acuerdo 5. Muy de acuerdo

Horario.

Método de enseñanza.

Fechas de exámenes.

21. Te ha costado mucho adaptarte.

1. Muy en desacuerdo 2. En desacuerdo 3. Ni de acuerdo ni en desacuerdo 4. De acuerdo 5. Muy de acuerdo

22. Has cambiado tu planificación a la hora de estudiar.

23. La nueva organización de la fecha de los exámenes es más acertada que la anterior.

24. El no tener clase durante la realización de los exámenes te parece una buena decisión.

25. ¿Cuántas asignaturas aprobaste en tu primer año de adaptación?

26. ¿Cuántas asignaturas aprobaste en tu primer año de carrera?

27. ¿Volverías al plan 1979? Sí No

28. Sexo V H

29. Edad

30. Curso

31. Año inicio carrera

Fin del Cuestionario

UNA PROPUESTA DE ESTRUCTURACIÓN DE LA METODOLOGÍA

Rafael Moreno Rodríguez
Rafael J. Martínez Cervantes
Salvador Chacón Moscoso
Universidad de Sevilla.

RESUMEN

Se presenta una estructuración de un amplio conjunto de nociones metodológicas, provenientes de muchas y variadas fuentes de intereses y objetivos, considerándolas como especificaciones de otras nociones más generales, que se proponen en este trabajo y que han sido extraídas de la práctica investigadora. Un grupo de estas nociones está constituido por determinadas dimensiones de los conceptos implicados en el trabajo de la ciencia, y otro grupo por determinadas categorías descriptoras de dicho trabajo. Con ambos grupos y sus combinaciones, se obtiene una determinada estructuración que, además de implicar la pretendida integración de nociones metodológicas, facilitaría un mayor conocimiento de las tareas previas al análisis de datos -como, por ejemplo, la generación y delimitación de objetivos de investigación-, y potenciaría un mayor aprovechamiento de los logros del análisis de datos, poniéndolos al servicio de objetivos y diseños de investigación mejor delimitados.

INTRODUCCIÓN

En cuanto descripción y prescripción del trabajo científico, la metodología puede ser considerada -utilizando el símil de la medición- como una escala cuyo sistema numérico formal, muy desarrollado, está constituido por un amplio bagaje de modelos y técnicas para el análisis de datos. Ahora bien, como señala la teoría representacional de la medida (p. ej. Berka, 1983; Suppes, Krantz, Luce y Tversky, 1989) todo sistema formal sólo puede adquirir mayor validez y utilidad en la medida en que se ajuste a un sistema empírico, también ampliamente desarrollado, que sin embargo en el caso de la metodología ha recibido poca atención.

Por esa escasa atención prestada a los conceptos que constituirían el sistema empírico de la metodología, existen en la bibliografía inconvenientes como los siguientes: lagunas e imprecisiones de determinadas nociones metodológicas, como por ejemplo las de constructo, valor, o la dimensión cualitativo-cuantitativo; problemas de exhaustividad y/o exclusividad de algunas clasificaciones, como algunas de métodos y de técnicas de recogida de datos; generalidad limitada de determinadas definiciones, como la de confundido planteada casi exclusivamente entre variables independientes y extrañas, y además sólo para estudios relacionales; polisemias de ciertos términos técnicos, como los de control, descripción o metodologías cualitativa y cuantitativa; relaciones innecesarias entre algunas nociones y contextos metodológicos como ocurre con la noción de sistema de categorías y la metodología observacional; y también falta de otras asociaciones posibles, como por ejemplo entre categorización y experimentación, o entre ciertos muestreos y determinadas metodologías específicas.

En nuestra opinión, las insuficiencias señaladas y otras adicionales podrían solventarse en la medida en que se lograra un conjunto de nociones generales integradoras de las nociones metodológicas existentes, que mostrara a éstas como concreciones de las más generales pretendidas. Esas nociones generales permitirían estructurar en mayor medida lo que hoy es más bien una acumulación de nociones metodológicas, aportadas a lo largo de muchos años por diferentes orientaciones teóricas y preocupaciones tecnológicas. Esa estructuración facilitaría además un mayor conocimiento de las tareas previas al análisis de datos, como por ejemplo la generación y delimitación de los objetivos de investigación (McGuire, 1994; 1997); también potenciaría un mayor aprovechamiento de los logros del

análisis de datos poniéndolos al servicio de objetivos y diseños de investigación mejor delimitados (Cook, Levinton y Shadish, 1985; Maxwell y Delaney, 1990; Micceri, 1989).

UNA PROPUESTA DE ESTRUCTURACIÓN.

Tras sucesivos análisis de la bibliografía y ensayos diversos, entendemos (Chacón, Martínez y Moreno, en prensa; Moreno, Martínez y Chacón, en preparación) que las nociones generales integradoras pueden ser las siguientes.

Un primer grupo de nociones serían las referidas a las tareas de la ciencia, agrupadas de acuerdo con propuestas de Campbell (1988), Popper (1963) y Toulmin (1979), en una sucesión de **variaciones** conceptuales y **selecciones válidas** de algunas de esas variaciones. Entendemos las *variaciones* como cambios desde unos conceptos de partida a otros diferentes, o a otros similares planteados en contextos diferentes. Entendemos la *selección válida* como el ajuste o ausencia de variaciones entre el concepto en estudio surgido como variación y los correspondientes datos *recogidos* y *analizados* con procedimientos existentes en la bibliografía, considerados ambos términos en contextos similares. En otras palabras, la variación supone una correspondencia entre un concepto y contexto de partida y otros diferentes, mientras que la selección válida supone también una correspondencia pero de semejanza entre un concepto surgido como variación y planteado ahora como referencia y los datos obtenidos en un contexto similar al planteado (ver cuadro 1).

Cuadro 1: Nociones de variación y selección válida

| |
|---|
| <p>Ciencia como sucesión de: Variación 6 Selección Válida 6 Variación...</p> <p>siendo</p> <p>Variación: {Concepto y contexto de partida à Otros conceptos en contextos}</p> <p>Selección válida: {Concepto y contexto planteados• Datos en contexto}</p> |
|---|

Un segundo grupo de nociones generales estaría constituido por seis **dimensiones** o facetas que entendemos identificables en todo concepto de la ciencia: a) **contenido**

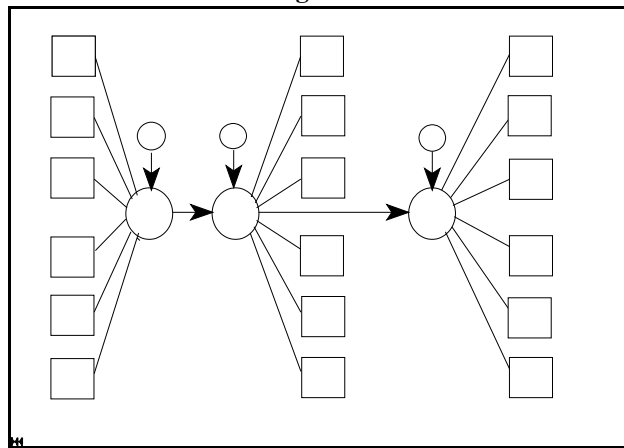
temático de cada concepto, que puede ser **explicitado** o delimitado: b) con una **composición** o estructura de variables y/o de relaciones de diversa complejidad, c) a **nivel empírico** de constructo y/o de indicadores empíricos, d) con **tipos** o modos consistentes en comprensión y/o extensión, e) en un número y tipo de **valores** determinados, y f) según una o más de las **formas de expresión** o códigos, verbal, numérico-formal y gráfico (ver cuadro 2)

Cuadro 2: Dimensiones en los conceptos de la ciencia

| DIMENSIONES | CASOS |
|---------------------------------|--|
| Contenido | *Características de sujeto *Características de medio *Características de estudio |
| Composición (Tipo y número) | *Carácter o variable *sencilla *múltiple (multivariable) *Relación * sencilla * multivariable *de relaciones (interacción) |
| Nivel empírico de explicitación | *Indicador *Constructo |
| Tipo de explicitación | *Comprensión *Extensión |
| Valores (Tipo y número) | *Cualitativos *Ordinales *Dicotómicos y politómicos *Cuantitativos *Contínuos y discretos |
| Formas de expresión | *Verbal *Numérico-formal *Gráfica |

Si combinamos los dos grupos mencionados de nociones, resulta que cada una de las tareas de variación y selección válida, y toda sucesión entre ellas, puede ser considerada detalladamente en términos de una o más de las dimensiones o facetas de los conceptos implicados en dichas tareas. En la figura 1 representamos mediante los círculos situados en la fila central la sucesión de las correspondencias definitorias de una variación y una selección, o de dos variaciones o dos selecciones consecutivas. Cada concepto implicado en dichos procesos puede ser detallado en términos de sus dimensiones, representadas por los respectivos cuadrados. El contexto correspondiente a cada concepto está indicado por los círculos de menor tamaño.

Figura 1



Sobre la base de tal estructuración, las nociones de la bibliografía metodológica podrían ser entendidas como concreciones de una o más de las: a) dimensiones conceptuales, b) tareas mencionadas, o c) combinaciones entre esas dimensiones y tareas. Pongamos algunos ejemplos de las tres posibilidades recién mencionadas.

a) Cada uno de los términos metodológicos que aparecen en la columna derecha del cuadro 2 corresponden a casos o concreciones de tan sólo una determinada dimensión conceptual. En cambio, otros términos y nociones responden a más de una dimensión. Por ejemplo, la noción metodológica de *distribución de frecuencias* corresponde a la siguiente conjunción de dimensiones: explicitación de algún contenido de composición usualmente no relacional, realizada normalmente a nivel empírico de indicadores -con datos-, por extensión y de forma numérica; por su parte, los *polígonos de frecuencias* implican

normalmente una expresión por extensión, numérica y gráfica de algún contenido con estructura relacional, mientras que *diagramas* como los usados en los modelos de ecuaciones estructurales -o el mostrado en la figura 1- suelen ser una expresión gráfica por extensión de determinados contenidos de composición relacional y no relacional.

b) Por su parte, *problemas* y *soluciones* son nociones metodológicas que pueden ser consideradas como referencias respectivamente de la carencia y logro de variaciones realizadas a partir de algún concepto tomado como partida. Asimismo, cuando esas soluciones son consideradas no seleccionadas empíricamente aún, son consideradas *hipótesis*. Por su parte, la identificación o recogida de datos es la tarea primaria de la selección de variaciones.

A su vez y como ejemplos de los casos señalados en c), es decir de combinaciones entre dimensiones y tareas, cabe mencionar a nociones como las de *intuiciones*, *inducción* y *deducción*, el carácter *confirmatorio* o *exploratorio* de las recogidas de datos, así como algunos tipos de validez (Ver cuadro 3). Asimismo los problemas e hipótesis, como integrantes de la tarea de variación, pueden ser considerados en términos de las diferentes dimensiones o facetas antes mencionadas. De tal modo, los *problemas abiertos* y *cerrados* parecen corresponder a su consideración en términos de la menor o mayor explicitación con los que sean planteados sus componentes. Por su parte, si la identificación o *recogida de datos* es el establecimiento de la correspondencia entre un concepto de referencia y datos necesaria para la selección válida, la *categorización* y la *medición* son casos particulares de esa recogida entendidos en términos de que el concepto tomado como referencia incluya al menos las formas de expresión verbal y numérica respectivamente. A su vez, los conceptos utilizados como referencia para las recogidas de datos son denominados *sistema de categorías* o *clasificación* si están expresados verbalmente, y *escala de medida* si lo están numéricamente.

Más ejemplos son posibles. Los métodos *observacional*, *selectivo de encuesta* y *experimental* pueden entenderse como diferentes especificaciones de la misma tarea de recogida de datos, cuando es considerada en términos de la influencia que los factores de estudio -investigadores e instrumentos- tengan o no sobre los contenidos de sujeto y medio que son objetos de estudio y sus contextos. De modo similar, las numerosas nociones metodológicas referidas a las *técnicas de control y muestreo* -operaciones necesarias para la validez de las selecciones-, al *diseño* de recogida de datos y a su *análisis*, pueden ser consideradas como determinadas combinaciones de las tareas y dimensiones planteadas (Moreno, Martínez y Chacón, en preparación).

Cuadro 3: Casos de cada tarea atendiendo a dimensiones

| DIMENSIONES | Tema 3 VARIACIONES O CAMBIOS... | Tema 4 IDENTIFICACIONES DE DATOS... | Tema 5 VALIDEZ... |
|---------------------|---|--|---|
| Explicitación | entre tipos y/o niveles de explicitación empírica (denominadas intuiciones al faltar explicitación suficiente) | confirmatorias o exploratorias , de referencias explicitadas como constructo y/o indicador, por extensión y/o comprensión. | * de las identificaciones confirmatorias, o al menos de las dimensiones explicitadas. *de contenido y de constructo , según nivel. |
| Contenido | entre categorías de sujeto, medio y estudio, o entre características de alguna de esas categorías. | de características específicas de las categorías de sujeto, medio, y/o estudio. *También simulación de sujetos y/o medio | * de población de sujetos * ecológica * de tratamiento |
| Composición | entre variables y relaciones, o sus tipos: * inducción * deducción | de variables y relaciones de cualquier tipo. | * de variables y de relaciones |
| Valores | en el conjunto de ellos, y/o en su tipo. | de conceptos cualitativos, ordinales y cuantitativos. | * con los diferentes tipos |
| Formas de expresión | de sistemas de códigos de un tipo -verbal, numérico o gráfico-, o entre tipos. | verbalmente, con códigos numéricos y formales, o mediante gráficos. | * con los diferentes códigos. |

Una anotación adicional para completar esta presentación de la estructuración. Los términos señalados en cursiva en los párrafos anteriores son ejemplos de algunas de las concreciones de tareas y dimensiones que corresponden a nociones recogidas por la bibliografía metodológica. Sin embargo, ésta no recoge otras posibles derivables también

de la estructuración que planteamos. Por ello habrá que evaluar en un futuro la relevancia de casos como éstos no contemplados usualmente; pero su mero señalamiento o reconocimiento supone una aportación adicional derivada de la estructuración o modelo que venimos presentando.

CONCLUSIONES

La estructuración presentada pretende mostrar a las nociones metodológicas provenientes de muchas y diversas aportaciones e intereses durante años, como concreciones o especificaciones de otras más generales. Sin duda las nociones generales aquí consideradas -es decir, las dimensiones conceptuales y tareas científicas señaladas- no son las únicas posibles y pueden ser mejoradas; pero en todo caso ellas parecen mostrar abundantes posibilidades para cubrir el objetivo previsto. Si esa estructuración es adecuada, la enseñanza de la metodología y su práctica se puede ver favorecida como en otros trabajos hemos empezado a encontrar (p. ej. Moreno, Martínez y Chacón, en prensa).

LA VALIDEZ EN LAS INVESTIGACIONES MEDIANTE ENCUESTAS

Andrés González Gómez
José Luís Padilla García
Cristino Pérez Meléndez.
Universidad de Granada.

RESUMEN

En este trabajo presentamos una aproximación al estudio de la validez en las investigaciones mediante encuestas, que conjuga dos perspectivas previas. Por un lado, clasificaciones de los errores en la encuesta como la aportada por Groves. Por otro, la conceptualización de la validez de la investigación como un proceso escalonado de inferencia, tal como lo conciben Cook y Campbel para las investigaciones cuasi-experimentales. De forma añadida, ampliamos el concepto de validez de estos autores para incluir, tanto la relevancia de la asignación numérica como los nuevos desarrollos de la validez que se han producido en las últimas décadas.

INTRODUCCIÓN

La investigación mediante encuestas está presente en la sociedad actual, no sólo en el ámbito científico, sino en la vida cotidiana en forma de encuesta de voto, de consumo, de actitudes, etc. Las encuestas, en cualquiera de sus modalidades, suelen tener como objetivo obtener información que ayude a tomar algún tipo de decisión. Dado que no sería admisible una decisión apoyada en los resultados de una encuesta sin alguna prueba de la calidad de ésta, la información sobre los resultados de una encuesta suele ir acompañada de una serie de datos que intentan demostrar su calidad.

En la mayoría de los casos, esta información se refiere casi exclusivamente al error muestral, tamaño de la muestra seleccionada y procedimiento de muestreo. Podemos encontrar un ejemplo en las fichas técnicas que acompañan las encuestas realizadas por el Centro de Investigaciones Sociológicas. (Ver Tabla 1).

Podría pensarse que si los datos presentados en esa tabla son satisfactorios (es decir, si el error estimado es suficientemente pequeño) las interpretaciones en base a los resultados de la encuesta estarán justificadas. Esto sería así si la única inferencia a partir de los datos obtenidos en la investigación se refiriese a la correspondencia entre estos datos y la población objetivo. Es decir, el grado en que la muestra es representativa de la población. Sin embargo, no es ésta la única inferencia que se realiza en una investigación mediante encuesta. El investigador debería plantearse, no ya si obtendría los mismos datos con otra muestra de sujetos, sino si obtendría los mismos datos caso de volver a encuestar a la misma muestra de sujetos. También, debería interrogarse, por ejemplo, sobre el verdadero significado de los datos obtenidos, cuestionándose si se hubiesen obtenido los mismos resultados realizando las preguntas de otra manera. Es más, se podría plantear que la correspondencia entre los datos muestrales y la población objetivo ni siquiera es la interpretación más importante. Como apuntábamos al principio, la encuesta proporciona los argumentos para tomar algún tipo de decisión: predecir un comportamiento, obtener el efecto de una política, etc. Por tanto, no parece recomendable que la información aportada para juzgar la calidad de la encuesta se limite al error estimado. En definitiva, durante una investigación mediante encuesta se realizan una serie de inferencias (de forma explícita o implícita) que es necesario justificar para asegurar la validez de la investigación.

Tabla 1. Ficha técnica del estudio (barómetro julio C.I.S.)

| |
|---|
| URL: http://www.cis.es |
| Ámbito: Nacional. Se incluyen las provincias insulares y se excluyen Ceuta y Melilla |
| Universo: Población española de ambos sexos de 18 años y más |
| Tamaño de la muestra: Diseñada: 2.500 entrevistas. Realizada: 2.486 entrevistas |
| Afijación: Proporcional. |
| Ponderación: No procede. |
| Puntos de Muestreo: 168 municipios y 46 provincias. |
| Procedimiento de muestreo: Polietápico, estratificado por conglomerados, con selección de las unidades primarias de muestreo (municipios) y de las unidades secundarias (secciones) de forma aleatoria proporcional, y de las unidades últimas (individuos) por rutas aleatorias y cuotas de sexo y edad. Los estratos se han formado por el cruce de las 17 comunidades autónomas con el tamaño de hábitat, dividido en 7 categorías: menor o igual a 2.000 habitantes; de 2.001 a 10.000; de 10.001 a 50.000; de 50.001 a 100.000; de 100.001 a 400.000; de 400.001 a 1.000.000, y más de 1.000.000 de habitantes. Los cuestionarios se han aplicado mediante entrevista personal en los domicilios. |
| Error muestral: Para un nivel de confianza del 95,5% (dos sigmas), y $P = Q$, el error es de ± 2 para el conjunto de la muestra y en el supuesto de muestreo aleatorio simple. |
| Fecha de realización: Del 9 al 13 de julio de 1998. |

El objetivo del presente trabajo es mostrar las inferencias -explícitas o implícitas- que se realizan durante una investigación mediante encuestas, mostrando las distintas fuentes de error que pueden estar amenazando una interpretación válida de los resultados. Para ello, unificamos el esquema de los errores de la encuesta propuesto por Groves (1989) con el de las amenazas a la validez de las investigaciones cuasi-experimentales planteado por Cook y Campbell (1979).

LOS ERRORES EN LA ENCUESTA

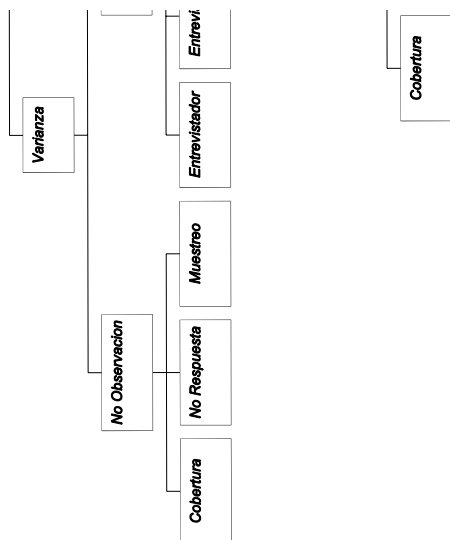
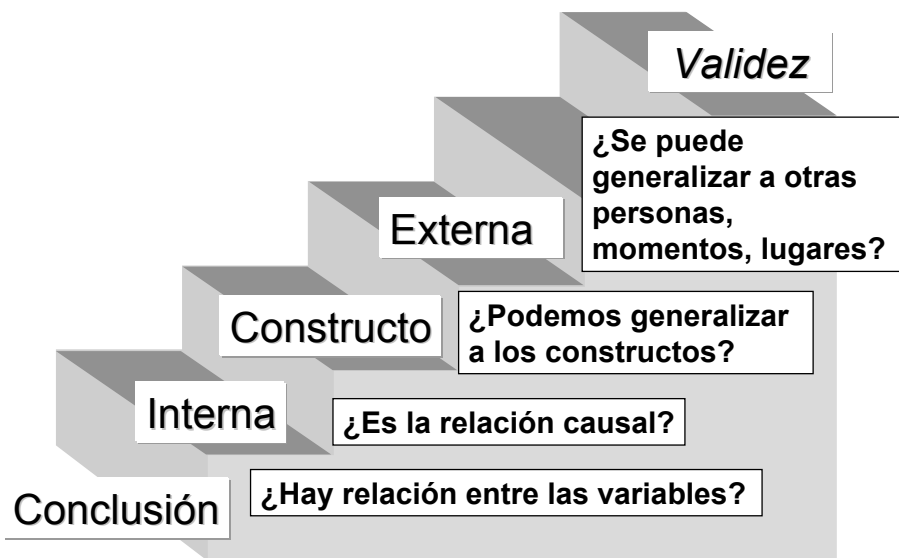
Son numerosos los manuales sobre métodos y técnicas de investigación que incluyen un apartado específico a la investigación mediante encuestas (ej. Anguera y col, 1995) y los dedicados específicamente a este procedimiento (ej. Fowler, 1993, Rojas, Fernández y Pérez, 1998). Incluso, es posible encontrar monografías sobre la calidad de encuesta (ej. Groves, 1989). En muchos de estos textos se señala que el error que puede estar presente en una encuesta no es sólo el error muestral, sino que una investigación mediante encuesta puede estar contaminada por muchos otros factores. Una de las clasificaciones más exhaustivas es la presentada por Groves (1989, op. Cit), que divide las fuentes de error en errores de no observación y errores de observación. En los primeros, se incluirían los errores de cobertura, no respuesta y muestreo. En los segundos, los debidos al entrevistador, al entrevistado, al instrumento utilizado y al modo concreto en que se realizó la encuesta. A su vez, cada una de estas fuentes aparece dos veces en función de que el error cometido sea constante (sesgo) o variable (varianza). Puede verse un esquema resumen en la figura 1.

Muchas otras clasificaciones propuestas por otros autores podrían adaptarse y quedar incluidas en la propuesta por Groves. Es indudable la utilidad de esta clasificación para identificar la distintas fuentes de error que han de controlarse para garantizar la calidad de una investigación mediante encuesta. Sin embargo, quedando claro que esos factores pueden introducir error en la investigación, no está tan claro que consecuencias tendrá en cada caso el error introducido.

EL PROCESO DE INFERENCIA

El objetivo de toda investigación es generar algún tipo de conocimiento. Si se siguen los pasos del método científico, el investigador infiere ese conocimiento a partir de los datos que ha obtenido de la realidad. Cook y Campbell (1979) señalan una serie de inferencias escalonadas que se realizan en el contexto de las investigaciones cuasi-experimentales. En la figura 2 presentamos un esquema de estas inferencias tal como lo presenta Trochim (1997). En esa figura observamos una serie de interrogantes que el investigador se va planteando de forma escalonada. La validez de la interpretación final- de la respuesta a la pregunta que originó la investigación- dependerá de la validez de las respuestas -inferencias- a las preguntas previas. Cada una de estas inferencias estará justificada si no se ha cometido ningún error que la invalide. Cook y Campbell ofrecen un listado de posibles fuentes de error bajo el rótulo de *amenazas a la validez*. Como veremos más adelante, este esquema es extrapolable al contexto de la investigación mediante encuesta.

Figura 2: Esquema de la validez según Trochim



Validez de las encuestas

Figura 1. Esquema de los errores de la encuesta de Groves

ión numérica.

Aunque Cook y Campbell no recogen este tipo de inferencia, nosotros pensamos que es necesario reflejarla puesto que representa algo distinto y previo a la inferencia de conclusión estadística que se presenta posteriormente. La poca atención prestada a este tipo de validez es reveladora de uno de los supuestos implícitos comunes a la investigación social: la creencia de que los números utilizados para cuantificar las variables representan fielmente las relaciones entre los valores de las entidades “empíricas” objeto de la medición (Mitchell, 1990). El supuesto está presente cuando, por ejemplo, la cuantificación que realizan los encuestados de la credibilidad de los líderes políticos es utilizada para “ordenarlos” en una dimensión de credibilidad; cuando las opiniones sobre los supuestos de despenalización del aborto son empleados para indicar qué colectivo demográfico es “más o menos” pro-abortista que otro, etc.

El problema de la “representación numérica” ha tenido su reflejo más popular en la cuestión de los niveles de medida (Stevens, 1946). Sin embargo, pensamos que su análisis debería abarcar también interrogantes como: ¿La variable a medir es cuantitativa?; ¿Permite el modelo de medida que estamos usando -aunque en la mayoría de los casos el investigador no sepa cual es- la localización en una única dimensión de personas e ítems?; ¿Disponemos de una unidad de medida significativa?, etc. Sería ingenuo pensar que investigadores y usuarios en general de la metodología de encuestas se van a detener en estas cuestiones de teoría de la medida, pero es nuestra responsabilidad hacerles al menos conscientes de que hay un supuesto cuyo no cumplimiento puede amenazar la validez de la encuesta. En definitiva, el supuesto de que las relaciones entre los números representan de manera adecuada las relaciones entre los valores de la variable objeto de la medición.

Validez de conclusión estadística.

El abanico de posibles objetivos de una encuesta es amplio. Así, podemos encontrarnos con encuestas cuya única finalidad es conocer el valor medio en la población de una única variable, o situaciones donde interesa información sobre distintas variables y su posible relación. En cualquier caso, el investigador se encuentra con un conjunto de datos sobre los que realiza algún tipo de cálculo estadístico. El significado del resultado obtenido constituye la primera inferencia que debe realizar el investigador. Esta puede ser del tipo: “¿Es significativamente distinto de X el valor medio obtenido en la muestra?”, “¿Son diferentes los valores medios de X e Y?” o bien, “¿Hay diferencias en la media de la variable entre los subgrupos A y B de la muestra?”. Se puede considerar las respuestas a

estas preguntas como conclusiones estadísticas, conclusiones que es necesario validar y que se encuentran amenazadas por una serie de factores apuntados ya en gran medida por Cook y Campbell. Estos son:

Baja potencia estadística. Como es sabido la potencia de un estadístico está directamente asociada al tamaño de la muestra. Por tanto, en la mayoría de investigaciones mediante encuesta éste no será un grave problema. Sin embargo, es un aspecto que hay que tener en cuenta en situaciones especiales en las que se cuenta con muestras relativamente pequeñas o en las que se utilizan estadísticos que requieren un elevado número de sujetos.

Violación de los supuestos de los tests estadísticos. Esta amenaza actúa de igual forma que en las investigaciones cuasi-experimentales. Quizá adquiera especial importancia el nivel de medida de las puntuaciones. En muchos casos el estadístico requiere que sea de intervalos. Sin embargo, en investigación mediante encuestas, es habitual el empleo de instrumentos en los que nada se ha demostrado sobre el nivel de medida de la escala resultante. No debe confundirse este problema con el anteriormente presentado acerca de la validez de la asignación numérica. Es posible que un conjunto de números represente adecuadamente la relación “meramente ordinal” que se establece entre los objetos de medida. Esta adecuada representación, no permitiría el empleo de técnicas estadísticas que supusiesen una relación establecida a nivel de intervalos.

Probabilidad del error tipo 1. El error tipo 1 consiste, por ejemplo, en concluir covariación (diferencia) cuando en realidad no la hay. Este problema surge especialmente cuando se realizan múltiples comparaciones. Ésta es una situación habitual en investigación mediante encuestas (ej. Encuestas omnibús) donde se recoge información sobre gran cantidad de variables. Si se realiza una inferencia simultánea, la probabilidad teórica de cometer error tipo 1 en “alguna” de las variables implicadas es mucho mayor que el nivel de significación elegido. Este nivel sólo es correcto cuando se analiza la probabilidad de error en una variable concreta.

Fiabilidad de las medidas. Durante el desarrollo de la encuesta, es normal la utilización de tests y cuestionarios. Es necesario que todos los instrumentos empleados cumplan unos mínimos de fiabilidad o precisión en la medida. Es necesario señalar que el análisis de la fiabilidad es necesario en todos los casos y no sólo en el de tests o cuestionarios complejos. No debemos olvidar que, en todos los casos, la información proviene de los sujetos, y éstos no son infalibles ni siquiera en el caso de preguntas consideradas *a priori* como sencillas. Por ejemplo, es posible que una persona no respondiese de la misma manera si le

preguntamos en dos ocasiones distintas cuantas horas a la semana dedica a ver la televisión. Si suponemos que existe un valor verdadero para esta pregunta, la variabilidad que podría encontrarse en las respuestas estaría indicando un problema de fiabilidad. En el esquema de Groves podría identificarse como el error de observación debido al instrumento.

Fiabilidad en la administración de los tratamientos. Aunque mantenemos esta formulación por ser la original de Cook y Campbell, quizá sería mas adecuado hablar de fiabilidad en la realización de las entrevistas. Nos estamos refiriendo al efecto del entrevistador ya señalado como fuente de error por distintos autores.

Irrelevancias aleatorias en el marco del experimento. o factores aleatorios en el contexto de la entrevista, que pueden alterar el resultado de la misma. Aquí cobran sentido las recomendaciones de muchos autores (p.e. Manzano, Rojas y Fernández, 1996) de realizar las entrevistas en un ambiente tranquilo y controlado.

La propuesta de Cook y Campbell incluye también en esta lista la heterogeneidad aleatoria de los sujetos. No creemos que esta amenaza sea extrapolable al campo de la investigación mediante encuestas, pues en este caso no suele requerirse la homogeneidad de los sujetos. Por otro lado, cabría incluir aquí, como señala Martínez (1995) los errores cometidos durante el procesamiento de los datos (codificación, edición, grabación, etc.).

Validez interna.

En investigaciones cuasi-experimentales, una vez determinada la relación entre variables, cabe preguntarse si dicha relación es causal. Si este interrogante no tiene sentido en las encuestas descriptivas, lo adquiere en las analíticas (Cochran, 1977; Abramson, 1990, citados en Martínez, 1995). En este segundo tipo de encuesta se realizan comparaciones entre grupos. El investigador podría estar tentado de dar una explicación causal de las diferencias encontradas. En este caso, entraríamos de lleno en el terreno de la validez interna, o lo que es lo mismo, de la necesidad de justificar la inferencia realizada sobre la relación causal. Imaginemos que en una encuesta se pide a los sujetos su valoración de distintos líderes políticos y el conocimiento que tengan sobre la actuación política de cada líder. El investigador podría estar tentado de atribuir las diferencias en valoración a los distintos grados de conocimiento, lo que supondría una inferencia causal. Los factores que para Cook y Campbell pueden estar amenazando esta inferencia son: historia, maduración, administración de pruebas, instrumentación, regresión estadística, mortalidad selectiva, interacciones con la selección, ambigüedad en la relación causa-efecto, difusión e imitación

de tratamientos, rivalidad compensatoria entre sujetos y desmoralización de los sujetos. No es nuestra intención volver a explicar cada una de estas amenazas. Para ello, remitimos al lector al texto original. Simplemnte, señalaremos algunos aspectos especialmente relevantes para la investigación mediante encuestas.

Historia y maduración. Sobre todo en el caso de diseños de encuestas longitudinales, es posible que sucedan acontecimientos (historia) o que se produzcan cambios en los sujetos (maduración) que invaliden la inferencia causal que se pretende realizar.

Instrumentación. Sucede cuando las diferencias encontradas se pueden atribuir a los instrumentos utilizados más que al efecto de otra variable. Por ejemplo, en poblaciones con más de una lengua, es posible que la encuesta se desarrolle en distintos idiomas. Las diferencias encontradas entre comunidades pueden deberse al idioma en que se realizó la encuesta. Por otro lado, uno de los "instrumentos" más habituales en la investigación mediante encuestas es el entrevistador. A pesar de que en muchas ocasiones el encuestador debe simplemente leer las preguntas de manera textual, no todos los encuestadores lo hacen exactamente igual. Además, por muy bien redactado que esté el cuestionario y aunque el entrevistador siga las instrucciones al pie de la letra, en ocasiones el encuestado no da una respuesta satisfactoria y es necesario repreguntar. En este momento se exageran las posibles diferencias en los resultados debidas a los entrevistadores. Un adecuado entrenamiento y formación de los entrevistadores son imprescindibles para controlar esta fuente de error (léase, al respecto, el trabajo "Reflexiones críticas sobres aspectos relacionados con las investigaciones mediante encuestas...", en esta misma obra).

Mortalidad selectiva. Podría identificarse, en terminos de las clases de error propuestas por Groves, como un error de no observación (ya sea de cobertura, no respuesta o muestreo) selectivo. Es decir, diferente para distintos subgrupos de la muestra. Básicamente, se refiere a que "no están todos los que son". No obstante, su adecuación, desde los diseños cuasi-experimentales, sólo es enteramente satisfactoria en las encuestas longitudinales, donde se "pierden" encuestados de una aplicación a la siguiente.

Validez de constructo.

En investigaciones cuasi-experimentales, el problema de la validez de constructo se refiere a que lo que un ivestigador interpreta como relación causal entre los constructos A y B, otro experimentador podría considerarlo como relación entre los constructos C y D. Nosotros preferimos redefinir la validez de constructo en términos de lo que realmente

significan las puntuaciones obtenidas. Así, esta idea de validez adquiere significado incluso en investigaciones donde se analiza una única variable.

La validación de constructo es más presumible, en el sentido de que es posible que se haya realizado algún tipo de estudio, en encuestas que utilizan test estandarizados (como ocurre habitualmente en encuestas psicológicas). No es el caso de otro tipo de encuestas, donde parece que se supone directamente la validez de los indicadores utilizados. Así, es necesario justificar que un ítem o test determinado proporciona una medida válida de la inteligencia. Por el contrario, parece aceptarse directamente que una pregunta como "¿A qué partido votará usted en las próximas elecciones?" proporciona una referencia válida de la intención de voto. En nuestra opinión esto no es así. Por sencillo que sea el instrumento de medida, es necesario demostrar la relación entre las puntuaciones obtenidas y el constructo o variable que intentan representar. En otras palabras, es necesario analizar la validez de constructo. Las amenazas a este tipo de validez más relevantes en el contexto de la investigación mediante encuestas son:

Explicación preoperacional inadecuada. Como señalan Crocker y Algina (1986), no existe ninguna aproximación universalmente aceptada a la medición de ningún constructo psicológico. En otras palabras, no tiene por qué existir acuerdo entre distintos investigadores sobre las conductas que pueden representar, por ejemplo, una determinada actitud. Este desacuerdo podría llevar a la elaboración de distintos conjuntos de ítems y, por tanto, a distintas puntuaciones finales y conclusiones. Este hecho, ampliamente aceptado en el terreno de los constructos psicológicos, puede ocurrir también en otro tipo de variables. Por ejemplo, un investigador podría definir el status socioeconómico de un individuo en función de su nivel de ingresos, mientras que otro podría adoptar como indicador su profesión.

Sesgo mono-operación. Una vez que el investigador opta por una definición para cada uno de los constructos relevantes en su investigación, cuenta con un conjunto de operaciones que lo representan. En muchas ocasiones, no incluye todas en el proceso de medida sino que utiliza exclusivamente una. Esto implica que el constructo en general está mal representado. Aunque Cook y Campbell plantean este problema cuando se utiliza exclusivamente una operación para medir el constructo, en nuestra opinión esta amenaza puede estar presente siempre que no se contemplen todas las dimensiones relevantes del constructo. Ésta puede ser una situación habitual en la investigación mediante encuestas, donde las exigencias económicas y temporales tienden a reducir la extensión de los cuestionarios utilizados.

Sesgo mono-método. Aún considerando una única operación o representación de un constructo, las formas de llevar a cabo la medida son varias. Por ejemplo, la información sobre el nivel de ingresos de una persona puede obtenerse mediante una pregunta abierta o cerrada. O mediante una entrevista personal o por teléfono. Como vemos, en la operacionalización del constructo influyen aspectos relacionados tanto con el cuestionario como con el modo concreto en que se realiza la encuesta.

Adivinación de hipótesis. En ocasiones el objetivo con el que se incluye una pregunta en un cuestionario no se identifica claramente. Por ejemplo, el investigador es consciente de que los sujetos no responden con sinceridad cuando se les pregunta sobre su nivel de ingresos. Por ello, decide inferirlo indirectamente a partir de ciertos indicadores, como el número de vehículos que tiene. Aparte de los problemas señalados antes (Explicación preoperacional inadecuada) es posible que el entrevistado averigüe por alguna razón el verdadero sentido de la pregunta y distorsione por tanto su respuesta. En este sentido, es importante cuidar aspectos relativos tanto al cuestionario como al entrevistador. Aspectos relacionados con la disposición y redacción de las preguntas en el primer caso y con una adecuada formación de los entrevistadores, en el segundo.

Recelo de evaluación. O el efecto que la deseabilidad social puede tener sobre las respuestas de los sujetos. Este efecto, muchas veces constatado, puede reducirse asegurando a los encuestados la confidencialidad de sus respuestas. Por otro lado, no todos los modos de realización de encuesta favorecen por igual este fenómeno. Como señalan Arias y Fernández (1998) la deseabilidad social está presente en mayor medida en las encuestas que utilizan entrevistadores.

Expectativas del experimentador. En las encuestas aparece cuando el entrevistador debe interpretar la respuesta del sujeto y cuando debe repreguntar o indagar hasta alcanzar una respuesta satisfactoria. El investigador puede sesgar sus preguntas en función de sus expectativas o de la percepción que tenga del encuestado.

Confusión entre constructos y niveles de constructos. Cook y Campbel lo enmarcan en el problema de la discretización de variables continuas. Este proceso sucede con relativa frecuencia en la investigación mediante encuestas. Por ejemplo, establecer intervalos para la edad, nivel de ingresos, etc. Es posible que los resultados varíen si se estableciesen los intervalos de otra manera. De nuevo, el cuestionario aparece como la fuente de error que puede estar provocando una inferencia equivocada.

Interacción de diferentes tratamientos. Si la encuesta cubre muchos apartados, las respuestas a los últimos items pueden estar afectadas por lo que ocurrió al principio de la entrevista. Sería un amenaza que estaría relacionada con la disposición de las preguntas en el cuestionario.

Interacción entre administración de pruebas y rendimiento. O el efecto que la propia encuesta ejerce sobre las respuestas. En las encuestas es especialmente importante. Si es raro que una persona participe en dos experimentos, es más normal que lo haga en dos encuestas, quizá la segunda vez que se le solicite su participación su predisposición, expectativas, etc, hayan cambiado.

Generalización restringida entre constructos. Los resultados de una encuesta nos permitirán realizar inferencias únicamente a los constructos representados en la encuesta y no a otros. Si el investigador está interesado en generalizar los resultados a otros constructos teóricamente relacionados con alguno de los que aparecen en el cuestionario, deberá obtener información directa sobre éstos para asegurar una adecuada generalización.

Validez externa.

Sin duda la más ampliamente tratada en las investigaciones mediante encuesta. Sin embargo, el terreno en el que se trata en estas investigaciones no agota las posibilidades de generalización que contemplan Cook y Campbell. En este sentido, los procedimientos de muestreo intentan asegurar que los datos obtenidos son representativos de una determinada población, pero nada dicen sobre la pertinencia de otras posibles generalizaciones.

Interacciones entre selección y tratamiento. Esta primera amenaza se reflejaría en la pregunta: ¿A qué categorías de personas pueden generalizarse los resultados obtenidos?. En investigación mediante encuestas no suele formularse así, dado que la población objetivo suele estar identificada a priori. Más que preguntar por *cuál* es la población a la que representan los sujetos de la muestra, suele preguntarse *en qué* medida la muestra es representativa de la población objetivo. Como dijimos al principio, en esta línea se sitúan las fichas técnicas de la mayoría de los estudios. Por otro lado, esta representatividad o este tipo de generalización, estaría amenazado por los errores de cobertura, no respuesta y muestreo señalados por Groves. Y no sólo por este último, como podría pensarse al ser éste el que recibe un estudio específico en muchas investigaciones.

Interacciones entre contexto y tratamiento. ¿Responde de igual forma una persona a una pregunta si se la realizamos en plena calle o si se la planteamos en su domicilio? Cabe la posibilidad de que no. Por tanto, el contexto en el que se realiza la encuesta es un factor a tener en cuenta a la hora de valorar los resultados obtenidos. En la clasificación de errores propuesta por Groves, el modo de realización de la encuesta quizá sea el que más claramente pueda identificarse con esta amenaza concreta.

Interacciones Historia-Tratamiento. La realidad es cambiante en múltiples aspectos y no sólo en aquéllos que en un momento dado puedan ser objetivo de una investigación mediante encuesta. Sin embargo, la generalización temporal de los resultados de una investigación está especialmente amenazada en el caso de las encuestas. Así por ejemplo, si realizamos una encuesta sobre la gravedad percibida de distintos delitos inmediatamente después de que los medios de comunicación den a conocer un caso de violación múltiple, es posible que este tipo de delito sea considerado con más importancia que lo hubiese sido una semana antes o una semana después. Al ser un aspecto ajeno en gran medida al diseño y realización de la encuesta, es difícilmente controlable. Sin embargo, el investigador debe estar atento a posibles acontecimientos que tengan lugar antes o durante la realización de la investigación para poder estimar su efecto.

NUEVOS PROBLEMAS PARA LA VALIDEZ DE LAS ENCUESTAS

La teoría de la validez ha cambiado considerablemente en las últimas décadas (Anastasi, 1986). Aunque las novedades hayan partido de los tests y de la evaluación educativa, pensamos que los argumentos planteados son aplicables a la encuesta como técnica de investigación para fines teóricos o aplicados. De hecho, las nuevas exigencias sociales a las que responden los cambios -validez, credibilidad y responsabilidad- son planteadas con mayor frecuencia en los estudios de investigaciones con encuesta. La ampliación de la teoría de la validez que nos resulta más relevante tiene que ver con los valores y las consecuencias sociales -intencionadas o no- de la realización de una encuesta. Cronbach (1988) propone dos categorías relativas a las consecuencias de la utilización de un procedimiento de medida: la *perspectiva funcional*, que agruparía las reflexiones sobre si la realización de la encuesta tiene consecuencias apropiadas para los individuos e instituciones; y la *perspectiva política* que apela a la responsabilidad de los profesionales de la investigación para que contribuyan a que los agentes sociales valoren la pertinencia de la realización de la encuesta. Messick (1989) plantea la necesidad de analizar los valores

y los contenidos ideológicos implicados en la medición, así como sus consecuencias sociales potenciales y reales.

Las encuestas de intención de voto o los periódicos “barómetros de opinión” son buenos ejemplos de la necesidad de reflexionar sobre las consecuencias sociales y la dimensión valorativa que arrastra la investigación mediante encuestas, así como sobre las cuestiones meramente técnicas, a la hora de juzgar su calidad.

CONCLUSIÓN.

A pesar de que la encuesta es un instrumento potencialmente muy útil y válido para obtener información, padece cierto descrédito, especialmente en algunos contextos. No nos resultarán extrañas expresiones como: *las encuestas no sirven para nada*, o *las encuestas siempre se equivocan*. Este descrédito, que no sufren otras técnicas de investigación, puede deberse a que la encuesta no pasa habitualmente por todos los controles de calidad que se realizan en otros tipos de investigación. Hemos señalado la insuficiencia del error muestral como garante de la calidad de una encuesta. Aunque esto no es nuevo (ver, por ejemplo, Fink, 1995), planteamos un procedimiento para analizar la validez de la encuesta que no se centra exclusivamente en las fuentes de error, sino que las asocia al proceso de inferencia. En definitiva, el objetivo último de cualquier encuesta es una interpretación de la realidad. Sólo en la medida que esta interpretación final (y las posibles interpretaciones o inferencias intermedias) esté justificada, la encuesta resultará provechosa.

LOS MODELOS DE ESTRUCTURAS DE COVARIANZA COMO MÉTODO DE VALIDACIÓN DE CONSTRUCTO

Eleuterio Francisco Sánchez García

Universidad de Sevilla

Manuel Sánchez García

Universidad de Huelva

RESUMEN

La validez de los instrumentos de medida es un requisito necesario para el desarrollo de las Ciencias Sociales. Una de las metodologías cada vez más utilizadas para contrastar la validez de constructo de tests y escalas de medida es la de los modelos de estructuras de covarianza. El objetivo de este trabajo es mostrar la utilidad de dichos modelos en el proceso de validación de constructo. Para ello, se enuncian cuatro tópicos de la validez de constructo y su paralelismo en los modelos de estructuras de covarianza. Estos tópicos son: a) la doble definición (operativa y relacional) de los constructos; b) el concepto de validez como acumulación de evidencias; c) la inclusión de la validez de contenido y la validez referida al criterio en la validez de constructo; y c) la validez de constructo como contraste de teorías. Por último, se describen otras aplicaciones psicométricas de los modelos de estructuras de covarianza.

INTRODUCCIÓN

De todos es bien conocida la importancia que para el desarrollo de la Psicología como ciencia tiene la utilización de instrumentos de medida fiables y, sobre todo, válidos, dado que la validez incluye a la fiabilidad. Es por ello que el tema de la validez se ha convertido en un tópico central en los trabajos de investigación realizados en el ámbito de las Ciencias Sociales, en general, y en el de la Psicología, en particular. Reflejo de ese elevado número de trabajos, y también de la corriente psicológica en boga en cada momento, son las distintas clasificaciones y denominaciones de validez a lo largo de la historia, hasta llegar a la concepción unitaria de la misma imperante en la actualidad. En este enfoque, la validez de constructo se define como integradora del resto de tipos de validez: validez de contenido y validez referida al criterio.

Por otro lado, la mayoría de las variables estudiadas en Psicología son variables no observables directamente, cuya existencia y papel hemos de inferir a partir de ciertas conductas manifiestas que suponemos son indicadores válidos de esa variable latente (no observable ni medible directamente). Inteligencia, ansiedad, neuroticismo, etc. son algunas de las características de las personas que sólo podemos medir indirectamente, a través de sus manifestaciones observables. Estas variables latentes son conocidas como constructos.

Para determinar el grado en el que un test mide el constructo que pretende medir (validación de constructo), partimos de una doble definición del constructo: una definición semántica u operativa, según la cual el constructo queda definido por su relación con variables observables que suponemos indicadores adecuados del mismo; y una definición sintáctica o relacional, a través de la cual ponemos en relación el constructo con otros constructos que suponemos asociados a él dentro de una misma teoría. De este modo lo que hacemos es hipotetizar un modelo de relaciones entre constructos y de los constructos con sus indicadores observables. Validación de constructo es, por lo tanto, validación de hipótesis de trabajo.

En este nuevo enfoque de la validación de tests, se considera que no existe un único coeficiente de validez, sino que la validación de constructo es un proceso de acumulación

de evidencias (ya sean cualitativas o cuantitativas) que permitan justificar las decisiones que posteriormente se tomarán en base a las puntuaciones de los sujetos en el test.

Dentro de este marco, el objetivo de este trabajo es presentar las características y posibilidades de una herramienta estadística de análisis de datos y su adecuación en el proceso de validación de constructo. Esta herramienta estadística está constituida por los modelos de estructuras de covarianza (más conocidos como modelos *causales* o modelos *Lisrel*). A lo largo del texto se argumentará que las cuatro facetas de la validación de constructo previamente desarrolladas se pueden explicar y cuantificar desde los modelos de estructuras de covarianza. Estas cuatro facetas de la validación de constructo a las que hacemos referencia son: a) doble definición (operativa y teórica) del constructo; b) validación como acumulación de evidencias; c) inclusión de la validez de contenido y referida al criterio en la validez de constructo; y d) validación como contraste de teorías. Para finalizar se enumeran otras utilidades de los modelos de estructuras de covarianza.

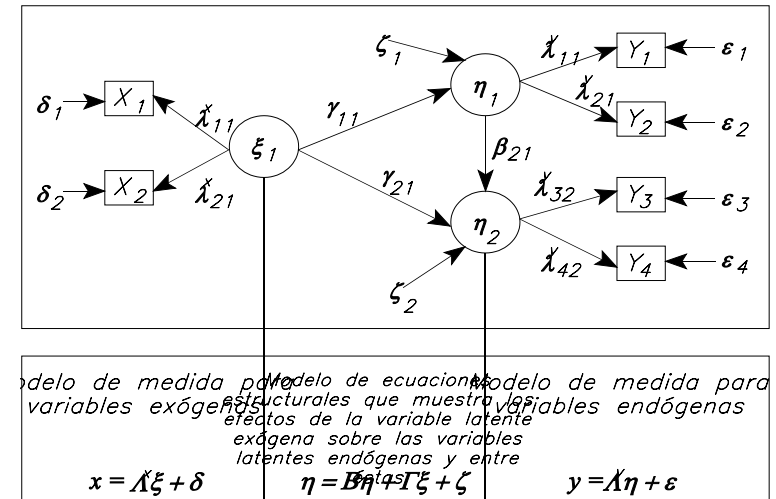
DEFINICIÓN DEL CONSTRUCTO

Como ya se expuso, un constructo se define a través de sus relaciones con otros constructos con los que suponemos relacionado y a través de las relaciones que mantiene con sus indicadores observables.

Esta doble vertiente se recoge en los modelos de estructuras de covarianza, incluso desde sus orígenes. Estos modelos de análisis surgieron de la unión entre los modelos de ecuaciones lineales propios de los econométricos y los modelos de medida (análisis factorial confirmatorio) propios de los psicómetras. Los modelos de ecuaciones lineales desarrollados por los econométricos representan una estructura de relaciones lineales entre variables observadas, partiendo de una hipótesis que preveía dichas relaciones. Sin embargo, en estos modelos, dichas variables se suponen medidas sin error. Gracias a la aportación de los psicómetras, se pudo introducir en los modelos el error de medida, tal y como ellos venían desarrollando a través de los modelos de análisis factorial confirmatorio. En los modelos de medida, las puntuaciones en una variable observada son explicadas por la influencia de un factor o variable latente y el error de medida propio de esa variable. De la unión de ambos surgen los modelos de estructuras de covarianza. En la figura 1 aparece

la representación gráfica de un modelo de estructuras de covarianza con tres variables latentes y dos indicadores observables cada una de ellas.

Figura 1. Modelo de relación entre constructos e indicadores



Las variables latentes que reciben la influencia de otra variable latente incluida en el modelo, son llamadas variables latentes endógenas (η) y las variables latentes que no son afectadas por ninguna de las variables latentes incluidas en el modelo, son variables latentes exógenas (ξ). Por extensión, las variables observadas (X e Y) son llamadas también exógenas o endógenas según sean indicadores observables de una variable latente exógena o endógena, respectivamente.

En la figura 1 se pueden apreciar los dos componentes de un modelo de estructuras de covarianza completo. El primer componente del modelo corresponde a un modelo de medida que relaciona los constructos con cada uno de sus indicadores y, además, incluye el error de medición de los items. El segundo componente del modelo completo corresponde al modelo de relaciones lineales entre variables latentes. Cada uno de los componentes coincide con los conceptos recogidos en la definición de un constructo.

ACUMULACIÓN DE EVIDENCIAS

No existe un único coeficiente de validez, sino que la validación es un proceso de acumulación de evidencias de validez. Los modelos de estructuras de covarianza proporcionan distintos valores que pueden ser utilizados como evidencia de validez. Entre ellos están los índices de fiabilidad de los ítems, la fiabilidad global de la escala o los residuales estandarizados, entre otros.

La exactitud con la que un ítem refleja la medida verdadera de un atributo psicológico se expresa a través de su índice de fiabilidad, de tal modo que cuanto más preciso sea un ítem menor será su error de medida y mayor su índice de fiabilidad. El índice de fiabilidad de un ítem o índice de homogeneidad del mismo es igual al producto de su desviación típica por la correlación del ítem con el test completo:

$$I.F._x = \sigma_x \rho_{xt}$$

Si trabajamos con puntuaciones estandarizadas del ítem y del test completo el índice de fiabilidad del ítem se hace igual a la correlación entre ambos:

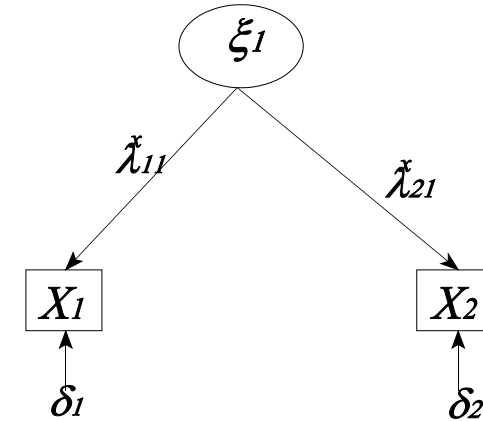
$$I.F._x = \rho_{xt}$$

En un modelo de medida, en el supuesto de que estén incluidos todos los indicadores relevantes de un constructo, podemos considerar las puntuaciones de los sujetos en ese constructo como sus puntuaciones en el test completo. Entonces el índice de fiabilidad de un ítem será igual a su correlación con el factor del que es indicador, es decir, será igual a su parámetro λ . Al estar incorrelacionados (por definición) la variable latente y el error que predicen la puntuación en un ítem, el parámetro de relación de un constructo con sus indicadores es igual a la correlación entre ambos:

$$I.F._x = \rho_{\xi^x} = \lambda^x$$

Es decir, en cualquier modelo de medida (como el de la figura 2), el índice de fiabilidad de cada uno de los ítems será igual a sus respectivos parámetros λ estandarizados.

Figura 2. Modelo de medida de un constructo.



El coeficiente de fiabilidad de cada uno de los ítems, por tanto, será igual a λ^2 o R^2 (porcentaje de varianza del ítem explicada por el factor).

Si todos los índices de fiabilidad de los ítems son suficientemente elevados (iguales o superiores a 0.4), concluiremos que hemos obtenido evidencias necesarias, aunque no suficientes para demostrar la validez de la prueba.

Igualmente, podemos calcular la fiabilidad de dicha prueba a partir de los parámetros de relación entre indicadores y constructo. Así pues, partiendo de la formulación del índice de fiabilidad de un ítem, la fiabilidad de un test se obtiene sumando todas las varianzas de puntuaciones verdaderas (parámetros λ) y dividiendo por la varianza total. Partiendo del ejemplo del apartado anterior, si tenemos una prueba que consta de dos ítems que suponemos indicadores de un mismo constructo (figura 2), podemos calcular la fiabilidad de la prueba a través de la siguiente fórmula:

$$\rho_{xx} = \frac{(\sum \lambda)^2}{(\sum \lambda)^2 + \sum \theta_{ii}}$$

Es decir, la fiabilidad de una escala completa es igual al cuadrado de la suma de los coeficientes de relación (parámetros λ) del constructo con sus indicadores, dividido por la varianza total. A su vez, la varianza total está compuesta por el cuadrado de la suma de los coeficientes de relación del constructo con sus indicadores, más la suma de las varianzas de los errores de cada uno de los indicadores. Cuanto mayor sea la variabilidad debida al constructo respecto a la variabilidad total de las variables observadas (items) mayor será la fiabilidad de la escala.

Por otro lado, al analizar la validez de una prueba a través de modelos estructurales, siempre es aconsejable el examen de la matriz de residuales ($[S - \Sigma(\theta)] =$ diferencia entre la matriz de varianzas-covarianzas muestral y la matriz de varianzas-covarianzas reproducida por el modelo a partir de los parámetros estimados), ya que un ajuste perfecto del modelo daría como resultado una matriz de ceros. Así, por ejemplo, valores residuales significativamente distintos de cero indicarían que las covarianzas entre algunos pares de items quedan mal explicadas por el modelo, evidenciando problemas de especificación del mismo o invalidez de los items utilizados para medir el/los constructos.

LA VALIDEZ DE CONSTRUCTO ENGLOBALA LA VALIDEZ DE CONTENIDO Y LA VALIDEZ REFERIDA AL CRITERIO

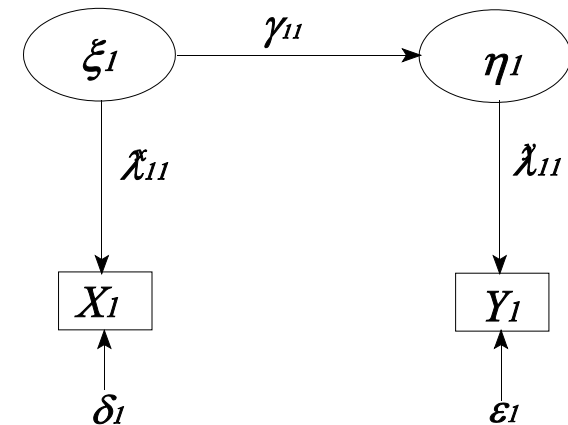
La concepción unitaria de validez implica que todos los demás tipos de validez están incluidos en la validación de constructo. Pues bien, un ejemplo de ello es la evidencia de validez referida al criterio que se obtiene al comprobar la relación existente entre constructos, donde un constructo puede ser utilizado como criterio de validez de otro.

En los modelos de estructuras de covarianza podemos utilizar el componente estructural para mostrar la validez referida al criterio de una prueba de rendimiento. Esto se puede

hacer a través de los parámetros β y γ de relación entre constructos, o más concretamente, a través de la significación o no de dichos parámetros de relación (figura 3).

El parámetro γ_{11} es un coeficiente de validez referida al criterio. Así pues, considerando las puntuaciones de los sujetos en cada uno de los constructos como las puntuaciones en dos tests que suponemos relacionados, obtendremos evidencia de validez referida al criterio cuando el parámetro γ sea estadísticamente significativo. Hemos de hacer constar que el modelo planteado implica una relación unidireccional entre los constructos. Si la relación hipotetizada no fuese causal sino bidireccional, consideraríamos como coeficiente de validez referida al criterio, el coeficiente de correlación entre constructos estimado por el modelo.

Figura 3. Modelo de relación entre constructos



Respecto a la validez de contenido, obtenemos evidencia de la relevancia del contenido a través de los índices de asociación de los indicadores observables con el constructo. Es decir, si la relación entre los indicadores y el constructo es estadísticamente significativa, obtenemos una prueba de que todos los items incluidos son una medida adecuada del constructo (*son todos los que están=relevancia del contenido*). Del mismo modo, a través de la validación de constructo, también obtenemos evidencia de la representatividad del contenido si los constructos definidos se relacionan entre ellos como habíamos hipotetizado. Para demostrar la relación entre dos constructos es necesario que en su

definición operativa estén incluidas todas las conductas que son indicadores observables de los mismos o, al menos, que las conductas incluidas sean una muestra representativa de la población de conductas que son indicadores observables de dichos constructos (*están todos los que son = representatividad del contenido*).

CONTRASTE DE HIPÓTESIS

La validación de constructo implica el contraste de hipótesis. Pues bien, dentro del ámbito de los estudios de validación de constructo, destacaremos finalmente que los modelos de estructuras de covarianza ofrecen también índices de ajuste global entre el modelo teorizado y los datos. Si los índices de ajuste del modelo indican que la teoría de relación entre constructos y de éstos con sus indicadores se ajusta adecuadamente a los datos recogidos a través de las pruebas que intentamos validar, habremos obtenido un índice de validez de constructo de dichas pruebas.

El índice de ajuste del modelo más utilizado es la prueba de χ^2 (una transformación de la función de ajuste, $c = nF[S, \Sigma(\theta)]$, se distribuye como χ^2). La prueba del ajuste del modelo se realiza partiendo de la $H_0: S = \Sigma(\theta)$, así, cuanto mayores sean las diferencias entre las matrices S y $\Sigma(\theta)$, mayor será el valor de χ^2 y tendremos que rechazar la hipótesis nula. Concluiremos, por tanto, que el modelo no ajusta a los datos. Mientras que cuanto menores sean estas diferencias, menor será el valor de χ^2 , mantendremos la hipótesis nula y el ajuste del modelo a los datos.

A la hora de considerar el valor de χ^2 se ha de tener en cuenta que este estadístico es sensible al tamaño de la muestra (n es igual al número de sujetos menos 1), de modo que tamaños muestrales elevados pueden llevarnos a rechazar un modelo cuyo ajuste a los datos sea adecuado y, al contrario, con muestras pequeñas el valor de χ^2 también tiende a serlo y puede llevarnos a aceptar como válido un modelo que no lo es.

Además de χ^2 existen otras medidas de ajuste del modelo, las más utilizadas son (con los umbrales habitualmente aconsejados):

- GFI o índice de bondad de ajuste: es un índice insensible al tamaño muestral, cuyos valores oscilan entre 0 y 1, aunque sólo valores superiores a 0.9 suelen ser aceptables.

- AGFI: es el índice GFI ajustado en base a los grados de libertad del modelo, como en el caso anterior, sus valores oscilan entre 0 y 1 y valores superiores a 0.8 indican un buen ajuste del modelo a los datos.

- RMR o raíz cuadrada de la media residual: es una medida del promedio de las diferencias cuadráticas entre S y $\Sigma(\theta)$. Valores menores que 0.10 indican un buen ajuste.

- RMSEA o raíz cuadrada de la media del error de aproximación. Es una medida de discrepancia que está en función de los grados de libertad del modelo. Un valor de RMSEA igual o menor a 0.05 indica un buen ajuste, valores superiores a 0.08 representan un error de aproximación a los valores poblacionales.

Frecuentemente, sobre todo en modelos muy complejos, se concluye rechazando la H_0 . Ante esta situación, hay algunas alternativas como la validación comparada o validez relativa de los modelos a partir de diferencias de χ^2 . Otra posibilidad es insistir más en la significación estadística de los parámetros o índices de fiabilidad y validez.

OTRAS APLICACIONES DE LOS MODELOS DE ESTRUCTURAS DE COVARIANZA

Los modelos matemáticos descritos pueden ser de gran utilidad en el análisis psicométrico desde distintos puntos de vista. Entre otras aplicaciones, se pueden utilizar en el proceso de depuración de un cuestionario, ya que identifican ítems con baja fiabilidad o permiten descubrir constructos no previstos en el modelo inicial (a través de la matriz de residuales y los índices de modificación, derivados de los primeros), así como ítems que previsiblemente saturan en más de un factor, lo que los convierte en indicadores inválidos del constructo definido.

Los modelos estructurales son también ampliamente utilizados en el análisis de las matrices multirrasgo-multimétodo. La matriz multirrasgo-multimétodo es una matriz de correlaciones de Pearson entre medidas de distintos rasgos psicológicos (constructos) obtenidas a través de distintos métodos. Para obtener una matriz multirrasgo-multimétodo, todos los rasgos han de ser medidos con todos los métodos. En un modelo de estructuras de covarianza, se hipotetiza que las variaciones en cada una de las variables observadas son debidas a los efectos propios del rasgo, más los efectos del método, más su unicidad (representado la unicidad el error de medida aleatorio y los errores sistemáticos no correlacionados ni con los rasgos ni con los métodos):

$$X_{la} = \lambda_{la,l}RI + \lambda_{la,a}MA + \delta_{la}$$

donde:

$\lambda_{la,l}$ = efecto del rasgo *RI* en el test '1a'.

$\lambda_{la,a}$ = efecto del método *MA* en el test '1a'.

δ_{la} = unicidad del test '1a', o errores aleatorios y sistemáticos que afectan al test '1a'.

Si las saturaciones factoriales, parámetros λ , de las variables observadas con los factores de rasgo son elevadas y estadísticamente significativas, obtenemos evidencia de validez convergente. Por el contrario, idealmente, las saturaciones de las variables observadas con los factores de método han de ser bajas y no significativas. Las correlaciones entre factores, parámetros ϕ , deberían ser bajas y no significativas para concluir que el diseño posee validez discriminante. Si las correlaciones entre rasgos son bajas indican que, efectivamente, miden distintos constructos. Por otro lado, bajas correlaciones entre factores de método denotan independencia entre los efectos de método.

Otra posible aplicación de los modelos de estructuras de covarianza es la detección del funcionamiento diferencial de los items. Para ello, se hipotetiza que los items están relacionados con el constructo que miden, con su error de medida y con otro factor que será la variable que introduce sesgo. Este otro factor suele ser la edad, el sexo o cualquier otra variable que denote distintos grupos de referencia. En esencia, la detección del funcionamiento diferencial de los items consiste en la identificación de aquellos factores que hacen que sujetos con el mismo valor en el rasgo obtengan puntuaciones diferentes en el test. Estas diferencias suelen ser explicadas por la pertenencia a determinados grupos. La estimación de los efectos de los factores que producen sesgo y del constructo, permiten determinar el tamaño y significación del sesgo.

ORIENTACIONES BIBLIOGRÁFICAS

Las nociones sobre validez aparecen ampliamente recogidas en el texto de Messick (1989) y en los *Standards* de la APA de 1985. En castellano, se pueden consultar los textos de Martínez (1995) y Paz (1996).

El texto más recomendable para profundizar en el conocimiento de los modelos de estructuras de covarianza es el de Bollen (1989). Gómez Benito (1996) y Batista y Coenders (1997) ejemplifican en sus textos las distintas posibilidades de aplicación de los modelos causales.

En el mercado informático se pueden encontrar distintos programas que permiten el cálculo de los parámetros en modelos estructurales. De todos ellos LISREL y EQS son los más utilizados. Los textos de Jöreskog y Sörbom (1993) y Bentler (1995), respectivamente, pueden ayudar a comprender su funcionamiento.

COMBINACIÓN DEL ESCALAMIENTO MULTIDIMENSIONAL Y EL ANÁLISIS DE CONGLOMERADOS PARA EL ESTUDIO DE LAS REPRESENTACIONES MENTALES DEL CONSUMIDOR. UN CASO PRÁCTICO

Jesús Varela Mallou *

X. Gabriel Vázquez **

Antonio Rial Boubeta *

* *Universidad de Santiago de Compostela*

** *Universidad de A Coruña*

RESUMEN

Las representaciones mentales que los consumidores elaboran frente a la amplia diversidad de productos del mercado, tienen un interés especial para varios campos del saber, como son la Psicología Comercial o el Márketing. Se requiere, además, estrategias para abordar las representaciones mentales como objeto de estudio. Dentro de la metodología aplicada, además de diseños peculiares, es importante mantener una actitud abierta ante las posibilidades de las herramientas de análisis de datos. En este trabajo, se muestra la utilización conjunta del análisis de conglomerados (CLUSTER ANALISYS) y el escalamiento multidimensional (MDS) para estudiar las representaciones mentales de una muestra de consumidores ante varias marcas de cerveza, a partir de un diseño de recogida de datos que permite la emisión de respuestas espontáneas a los sujetos encuestados.

INTRODUCCION

Una de las decisiones clave en toda investigación es, sin duda, la referente a la metodología a utilizar por el investigador para alcanzar los objetivos que se haya planteado. En ese sentido, si lo que queremos es conocer cómo es la representación mental que los individuos tienen de un conjunto de estímulos (o marcas comerciales), podemos optar por la utilización conjunta de dos conocidas técnicas multivariantes: el Escalamiento Multidimensional y el Análisis Cluster. La combinación de ambas técnicas resulta especialmente útil en la investigación comercial (Aaker y Day, 1989), en aquellos casos en que no se conocen a ciencia cierta los atributos o dimensiones en los que se basan las percepciones de los sujetos, en relación a distintos objetos o estímulos.

En el presente trabajo trataremos de ilustrar la utilidad de combinar ambas técnicas, utilizando para ello un ejemplo de la investigación comercial referido precisamente al mercado de la cerveza. Como veremos, mediante una sencilla tarea de ordenamiento podríamos ser capaces de informar de cuáles son las dimensiones o criterios que los consumidores utilizan cuando perciben las distintas marcas de cerveza que a diario les rodean. Más aún, podemos construir un mapa perceptual que simule cómo es la representación mental que los sujetos tienen del mercado, cuántos grupos o tipos de cerveza distinguen, qué caracteriza a cada uno y en qué se fijan los consumidores para clasificarlas. En definitiva, obtendremos información acerca de cómo estructura mentalmente el consumidor la diversidad de la oferta de un mercado concreto, de cara a la planificación estratégica de una acción de marketing como es la gestión de la Imagen de una marca y su Posicionamiento. Sin embargo, antes nos detendremos un instante a describir en qué consiste cada una de estas técnicas.

Habitualmente, en una investigación, disponemos de *Datos de Perfil*, es decir, puntuaciones de sujetos (filas) en variables (columnas), de manera que cada fila viene a representar el *perfil* de un sujeto. Sin embargo, en ocasiones resulta complicado acceder a este tipo de datos, con lo que habría que recurrir a otras estrategias para cubrir los objetivos de la investigación. Una de las alternativas es trabajar con *datos de proximidad*. Éstos se ordenan en una matriz donde tanto las filas como las columnas son estímulos y cada elemento es una medida de proximidad entre un par de estímulos. Dicha proximidad puede medirse en términos de semejanza (o *similitud*) o en términos de desemejanza (o

disimilaridad). Las tareas a través de las cuales se recogen estos datos suelen ser sencillas: bien tareas de ordenamiento de estímulos en una serie de montones, bien comparaciones entre pares de estímulos, etc. Las dos técnicas que centran este trabajo (el Análisis Cluster y el Escalamiento Multidimensional), permiten precisamente utilizar, como entrada para el análisis, este tipo de datos.

En primer lugar, el Análisis Cluster podría ser definido como una amplia variedad de procedimientos que pueden ser usados para generar una clasificación (Aldenderfer y Blashfield, 1991). Dichos procedimientos, a partir de un conjunto de datos, forman "clusters" o grupos de estímulos altamente similares. Dicho de otro modo, se trata de una técnica estadística multivariante que comienza con un conjunto de datos que recogen la información de una serie de estímulos o entidades e intenta reorganizarlos en grupos relativamente homogéneos.

Por su parte, el Escalamiento Multidimensional (MDS) se podría definir como una técnica multivariante adecuada, precisamente, para el análisis de datos de proximidad y preferencia. Esta técnica, además de las ventajas propias del tipo de datos con que trabaja, presenta claras ventajas respecto a otras técnicas multivariantes, tales como el análisis factorial. Entre otras podríamos señalar: que no es necesario el supuesto de linealidad entre variables, no exige un nivel de medida de las variables de intervalo y la solución que ofrece es más parsimoniosa y fácilmente interpretable, ya que cada estímulo se representa por un punto en el espacio. Cabe destacar que el MDS ofrece también algoritmos de tres vías que nos permiten estudiar las diferencias individuales (en este caso, entre segmentos de consumidores), ofreciéndonos una solución dimensional para cada grupo.

Veamos cómo hemos combinado ambas técnicas para resolver un problema concreto de investigación comercial.

OBJETIVO

El objetivo fundamental que perseguíamos era el de conocer cómo representan mentalmente los consumidores el abanico de marcas de cerveza que encuentran en el mercado. Dicho de otro modo, trataremos de dar respuesta a dos cuestiones:

1. Determinar la tipología de las distintas cervezas desde el punto de vista del consumidor.
2. Elaboración de un mapa perceptual que reproduzca (lo más fielmente posible) cómo estructuran mentalmente los consumidores la enorme diversidad de la oferta de cervezas que les rodean. Más concretamente, indagaremos acerca de cuáles son los criterios o dimensiones que el consumidor utiliza para diferenciar entre productos.

DESCRIPCIÓN DE LA TAREA

La tarea tuvo lugar en una sala provista de mesas amplias. A cada sujeto, de forma individual, se le presentaba un total de 16 marcas de cerveza (en botellas de 25cc) y, a continuación, se le pedía que las agrupasen en función de la similaridad que apreciase entre ellas, haciendo para ello tantos montones como creyese conveniente. Finalmente, en una hoja en blanco, se escribirían las marcas de cerveza que había incluido en cada montón. Se tuvo especial cuidado en no proporcionarle al sujeto ningún tipo de pistas o claves en relación a cómo debía agrupar las cervezas, dejando que realizase la tarea de forma espontánea y en función de sus propios criterios. La duración aproximada de la tarea era de unos 10 minutos.

Hay que hacer notar que mediante esta estrategia de recogida de datos, a diferencia de otro tipo de tareas, logramos extraer la información de forma rápida, cómoda y entretenida para el sujeto y, lo más importante, sin establecer ningún tipo de sesgo por parte del investigador. Otra de las ventajas es que se trata de una situación real y no simulada, de ahí la riqueza de la información que nos proporciona.

Sujetos

La tarea fue realizada por 98 sujetos de ambos sexos. Se cuidó que tuviesen relativo hábito en el consumo de cerveza y/o, que estuviesen familiarizados con las marcas que se les iban a presentar. Hay que señalar que en este tipo de trabajos en los que se pretende extraer la dimensionalidad cognitiva de los sujetos a la hora de percibir distintos productos o marcas,

los expertos señalan que no es necesario disponer de grandes muestras, sino que un tamaño de 60 sujetos ya nos permite obtener resultados significativos.

Estímulos

Las marcas utilizadas fueron las siguientes:

| | |
|-------------------------|------------------|
| ESTRELLA GALICIA | HR |
| RIVER | 1906 |
| ESTRELLA DAMM | VOLL-DAMM |
| CRUZCAMPO | MAHOU |
| HEINEKEN | CARLSBERG |
| CORONITA | BUCKLER |
| LAIKER | KALIBER |
| KRONNENBURG | NOSTRUM |

Análisis de datos

Para el análisis de datos se aplicaron las dos técnicas que centran este trabajo:

* CLUSTER ANALYSIS, para conocer los grupos que establecen los sujetos. ¿Cuántos grupos hacen?, ¿cuáles?.

* ESCALAMIENTO MULTIDIMENSIONAL para reconstruir el Mapa Perceptual del mercado de cervezas, comprobando cómo se estructura mentalmente la diversidad de la oferta.

No obstante, antes de llevar a cabo ambos análisis, era necesario transformar las puntuaciones directas de la tarea de agrupación en una matriz triangular de similitudes entre las 16 marcas (de orden 16 x 16), donde cada elemento de la matriz representase la similitud percibida por los sujetos entre dos marcas de cerveza determinadas. Aquellas cervezas que fuesen puestas en un mismo montón por muchos sujetos de la muestra, presentarían en dicha matriz puntuaciones próximas a 1. Por el contrario, aquellas cervezas que en escasas ocasiones fuesen puestas por los sujetos en un mismo montón, presentarían

puntuaciones próximas a 0. Dicha matriz, que iba a ser utilizada como entrada tanto para el Análisis de Cluster como para el Escalamiento Multidimensional, fue obtenida a partir de un programa diseñado expresamente para tal fin.

Resultados

Se procedió, en un primer momento, a ejecutar el Análisis Cluster, que habría de darnos la solución a la primera de nuestras preguntas: *¿Cuántos grupos o tipos de cerveza distinguen los consumidores?*. Para su respuesta, el programa HICLUST, implementado en el paquete PCMDS (Smith, 1990) nos ofrece una representación gráfica que permite informar de la existencia de 5 grupos de cervezas o marcas:

GRUPO 1: Estrella Galicia, Estrella Damm, Mahou y Cruzcampo

GRUPO 2: River, Laiker, Buckler y Kaliber

GRUPO 3: HR y Coronita

GRUPO 4: Heineken, Kronenbourg y Carlsberg

GRUPO 5: Nostrum, 1906 y Voll-Damm

Tratando de interpretar estos resultados podríamos decir que el consumidor distingue un primer grupo de cervezas que podríamos etiquetar como NACIONALES DE SIEMPRE, como Estrella Galicia, Mahou o Cruzcampo, que serían las cervezas con relativa tradición en el mercado y de origen nacional. En un segundo grupo se sitúan las cervezas que podríamos denominar SIN ALCOHOL, como Laiker, Buckler o River. En el tercer grupo distingue únicamente dos cervezas: HR y Coronita, que podríamos etiquetar como SUAVES. En el cuarto grupo incluye las cervezas que podríamos denominar DE IMPORTACIÓN, como Heineken, Kronenbourg y Carlsberg. Y, finalmente, en el quinto y último grupo se incluyen aquellas cervezas que podríamos denominar SELECTAS, con un color más “tostado” y una mayor graduación (Voll-Damm, 1906 y Nostrum).

El segundo de los objetivos perseguidos fue cubierto sometiendo los datos a un análisis de Escalamiento Multidimensional. A la matriz de similaridades, se le aplicó el modelo no métrico de Shepard (1962) y Kruskal (1964a, 1964b), mediante el programa KYST, implementado también en el paquete estadístico PCMDS. Con ello obtuvimos una

representación gráfica en un espacio r-dimensional, donde las distancias entre las marcas se interpretan como las proximidades originales entre ellas. El programa KYST proporciona además, para cada una de las soluciones, una medida de bondad de ajuste o *STRESS*. Cuanto mayor es el ajuste menor será el valor de *STRESS*. La solución más parsimoniosa fue la de TRES DIMENSIONES, con un *STRESS* asociado de 0.072, lo cual puede considerarse muy aceptable en términos estadísticos. En las figuras 1 y 2 se representan las 16 cervezas en las tres dimensiones extraídas, de manera combinada con la información proporcionada por el Análisis Cluster.

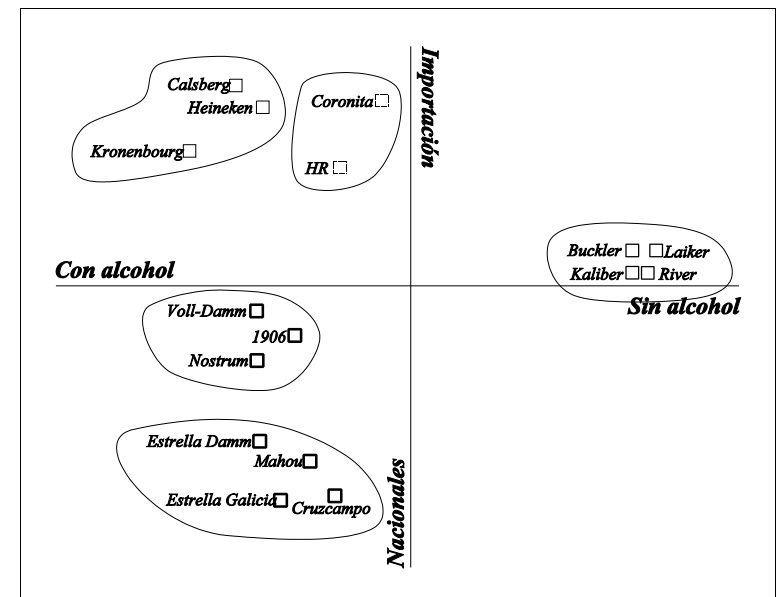


Figura 1: Mapa percentual del mercado de cerveza (dimensiones I y II)

La primera dimensión (eje de abscisas) separa el Cluster formado por las marcas Laiker, Buckler, Kaliber y River, de las restantes cervezas. Ello indica ya un primer criterio que utilizan los sujetos: CERVEZAS SIN ALCOHOL vs. RESTO DE CERVEZAS. La segunda dimensión (eje de ordenadas) separa las marcas nacionales de las de importación. En un extremo se sitúan los cluster de las CERVEZAS DE IMPORTACIÓN (formado por Heineken, Carlsberg Kronenbourg, Coronita y HR) de las NACIONALES (formado por Estrella Galicia, Estrella Damm, Cruzcampo y Mahou).

En la figura 2 se observa que la tercera dimensión (eje de ordenadas) separa las distintas marcas de cerveza en función de su GRADUACIÓN o FUERZA; quedando en el extremo superior Voll-Damm y 1906 y, en el extremo inferior Coronita y HR.

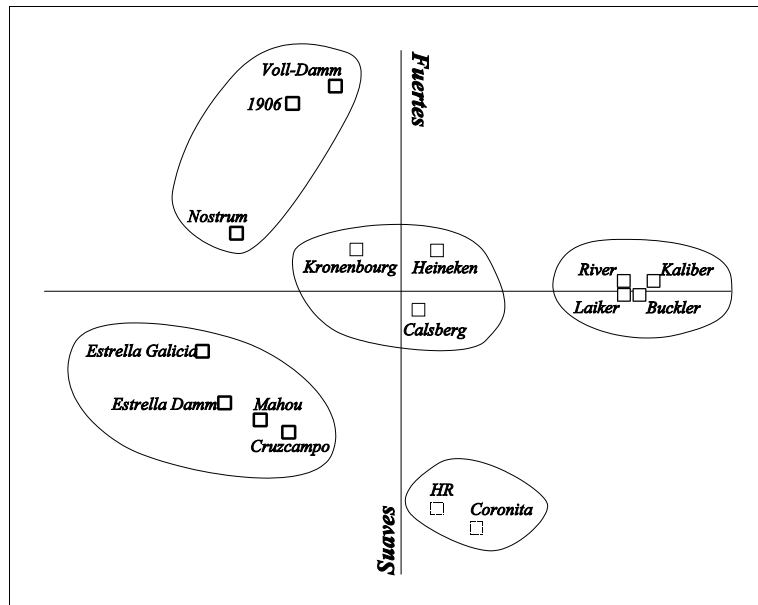


Figura 2: Mapa perceptual del mercado de cerveza (dimensiones I y III)

Podemos observar que el cluster formado por HR y Coronita, aparece como el MÁS SUAVE; el formado por Kronenbourg, Heineken y Carlsberg es percibido como de GRADUACIÓN o FUERZA MEDIA y, lo que resulta más curioso, de mayor fuerza que el cluster de Estrella Galicia, Estrella Damm, Cruzcampo y Mahou. Ello revela la distorsión perceptual de la categoría de IMPORTACIÓN, donde se espera encontrar una cerveza con cuerpo y sabor (atributos asociados con la graduación). La presencia del cluster de las CERVEZAS SIN, tan compacto y a la vez tan distante del resto (eje de abscisas de la figura 1), junto con la posición precisamente intermedia en el eje de ordenadas de la figura 1, pone de manifiesto que los consumidores no perciben esta categoría como cervezas propiamente dichas, sino como una categoría aparte.

En conclusión, el escalamiento multidimensional nos ha permitido responder, al segundo de los objetivos planteados. Los consumidores de cerveza, cognitivamente hablando, utilizan tres criterios a la hora de percibir el mercado de las cervezas: CERVEZAS CON ALCOHOL vs. CERVEZAS SIN ALCOHOL, NACIONALES vs. DE IMPORTACIÓN y FUERTES vs. SUAVES. Esto es, el mapa perceptual que reflejan las figuras 1 y 2 nos ha permitido relacionar los distintos clusters o tipos de cerveza, con los criterios que manejan los consumidores a la hora de percibir las.

La importancia del procedimiento utilizado radica en que a los consumidores no se les ha ofrecido ningún tipo de pistas o indicios, acerca de qué atributos debían utilizar a la hora de clasificar las distintas marcas de cerveza, con lo cual el investigador no estaría introduciendo ningún tipo de sesgos en los resultados. Resulta además especialmente útil en aquellos estudios de mercado en los que no se conocen a ciencia cierta los atributos que determinan las diferencias entre las distintas marcas de un sector (sobre todo en sectores relativamente novedosos o aún sin estudiar), o bien se tiene la sospecha de que los atributos propuestos a nivel teórico no concuerdan plenamente con los que utiliza en realidad el consumidor. En dichos casos, la utilización de esta metodología permite que afloren de manera espontánea los criterios que los consumidores utilizan cuando perciben la diversidad de la oferta que les rodea, y cómo ésta se estructura en función de dichos criterios.

ORIENTACIONES BIBLIOGRÁFICAS

Para obtener más información sobre los aspectos tratados en el texto, el lector interesado puede acudir, además de las referencias utilizadas, a los trabajos de Arce (1993), Smith (1990) y Varela, Rial y Braña (1998).

CRITERIOS EN LA VALIDACIÓN DE MODELOS. ¿RECHAZAR O MANTENER HIPÓTESIS?

Carlos Camacho Martínez Vara de Rey
Universidad de Sevilla

RESUMEN

Es frecuente el uso del término *validez* para referirse a la adecuación de un cierto modelo a la realidad. No obstante, aunque dicho término es el mismo, los procedimientos estadísticos encaminados a tal fin varían según el caso. Se comprueba que en aquellos modelos que podemos denominar clásicos (contrastes de medias, análisis de la varianza, regresión) la validez está montada sobre el rechazo de la hipótesis nula, mientras que en los denominados modelos estructurales es la aceptación o mantenimiento de la hipótesis nula la que determina la validez del modelo.

En el presente trabajo se ofrece una reflexión sobre la lógica implícita en la determinación de la validez en los diferentes tipos de análisis, al mismo tiempo que se intenta ofrecer un criterio unificador del concepto de validez estadística.

INTRODUCCIÓN

Si realizásemos un análisis histórico de la evolución de los contrastes de hipótesis a lo largo de la corta historia de la estadística como ciencia, observaremos que el planteamiento de modelos (normalmente, modelo lineal) es relativamente reciente. Hasta no hace mucho, un contraste de medias, un análisis de la varianza o una regresión múltiple respondían a cuestiones estadísticas bien diferenciadas. La significación de una diferencia, de un cociente o la predicción de una variable en base a otras, eran las cuestiones a plantear. Hoy día todas estas pruebas están dentro de un mismo concepto de modelo.

En la actualidad, el desarrollo de los denominados modelos estructurales han permitido generalizar más aún este concepto, y normalmente cualquier tipo de prueba estadística puede ser reconceptualizada en términos de modelos. De esta forma, un contraste de medias no es más que un modelo bivalente con una variable exógena tipo *dummy* y una variable endógena cuantitativa. Planteamientos equivalentes pueden realizarse con el análisis de la varianza o la regresión múltiple.

No obstante, aún subsiste un cierto lastre a la hora de enfocar el análisis estadístico en estos casos aún cuando son considerados como un mismo tipo de modelo. Tradicionalmente, se aplican pruebas basadas en la *t* de Student o en la *F* de Snedecor, incluso dentro de la perspectiva del modelo lineal. El cambio de tal perspectiva tan sólo afecta a considerar la significación de las pendientes o de los valores de correlación de dichos modelos lineales. Interesa, en el fondo, saber si la fuerza de la relación es o no diferente de cero. Es en estos casos el rechazo de la hipótesis nula lo que otorga credibilidad a estos modelos.

En los modelos estructurales, por el contrario, la validez de un cierto modelo está basada en la adecuación de la estructura definida por dicho modelo con la información que ofrecen los datos de observación, sin que se considere especialmente la intensidad de las relaciones encontradas entre las variables. Aquí, es la aceptación de la hipótesis nula lo que interesa. Puede darse la paradoja de que un mismo modelo, por ejemplo, de regresión múltiple, sea aceptado como válido desde la perspectiva de los modelos estructurales y no desde la misma regresión múltiple.

Veamos más detenidamente la cuestión de la validez en los modelos clásicos y los modelos estructurales.

LA VALIDEZ EN LOS MODELOS CLÁSICOS

Podemos tomar la regresión múltiple como exponente del enfoque clásico en la validación de modelos. Tiene interés plantearlo así por un doble motivo. Por un lado, análisis tales como contraste de medias, análisis de la varianza y regresión simple pueden ser reformulados desde esta perspectiva; y por otro, los modelos de regresión múltiple presentan, igualmente, cierta afinidad con los modelos estructurales, lo que nos permite elegir una plataforma común para ambos tipos de enfoques.

En regresión múltiple, los criterios de ajuste se basan en las discrepancias entre las puntuaciones observadas y predichas por el modelo. De esta forma, interesa minimizar la siguiente función:

$$\sum_j (Y_i - \hat{Y}_i)^2$$

Una vez definida la parte explicada por el modelo, se compara con el componente residual, según la conocida prueba de la *F* de Snedecor y se concluye si la parte sistemática queda disuelta en la parte aleatoria o no. Si la parte determinista supera el componente aleatorio, concluimos que se reconoce un cierto modelo tras la distorsión impuesta por el azar. En términos de hipótesis, se rechaza la hipótesis nula -azar-, y se admite, en consecuencia, la validez del modelo. Además, podemos reinterpretar la varianza explicada en términos de R^2 y realizar la comparación con el valor cero de correlación, que establece la hipótesis nula. Concluimos que la correlación observada no procede de una población caracterizada por una correlación de cero, y en consecuencia, suponemos que las variables del modelo están relacionadas.

Como se sabe, el rechazo de la hipótesis nula conlleva ciertas implicaciones. Existe un riesgo α conocido; esto es, sabemos la probabilidad de que la correlación observada

proceda de una población cuya correlación valga cero. Este aspecto conviene resaltarlo para contrastarlo con el basado en las ecuaciones estructurales que veremos a continuación.

Abundando en este tema, digamos que el rechazo de la hipótesis nula nos permite reconocer que existe algo en los datos que es distinto de cero, pero no nos dice, en principio, nada sobre ese algo. Sabemos que no es sólo azar lo que impregna todo lo observado, pero sin embargo, no sabemos exactamente lo que ocurre. Dicho en otros términos, *sabemos lo que no es pero no sabemos lo que es*. Justamente, un intento por saber lo que es, es lo que pretende el enfoque basado en los modelos estructurales.

LA VALIDEZ EN LOS MODELOS ESTRUCTURALES

En los modelos estructurales, lo importante no es el ajuste de las puntuaciones individuales predichas por el modelo, con las observadas en el fenómeno estudiado, sino el ajuste de las varianzas-covarianzas reproducidas por el modelo, con las varianzas-covarianzas observadas en la realidad. En términos matemáticos.

$$j \quad (S_{ij} & \hat{\sigma}_{ij})^2$$

Esta circunstancia marca un matiz diferencial importante. No es el ajuste a casos individuales lo que interesa, sino la estructura especificada por las matrices de varianzas-covarianzas. Normalmente el ajuste se realiza mediante una prueba de Chi-cuadrado. Y en caso de aceptarse la hipótesis nula, concluimos que el modelo reproduce fielmente la matriz de varianzas-covarianzas observada en la realidad.

Las implicaciones que conlleva aceptar la hipótesis nula, como se sabe, son muy diferentes a las de su rechazo. El riesgo ya no es alpha (conocido) sino beta (desconocido) por cuanto no sabemos cuantas posibles poblaciones existen afines a nuestros datos.

Podemos afirmar que con la aceptación de la hipótesis nula sabemos lo que nuestros datos pueden ser. Más exactamente, sabemos qué posible estructura pueden adoptar, pero no

estamos seguros de que sea la que nos proponemos, por cuanto desconocemos posibles estructuras alternativas que no presenten discrepancia con la estructura observada en la realidad. No sabemos lo que es, pero sí lo que puede ser.

Obsérvese que un mismo modelo (por ejemplo, de regresión múltiple) podemos enfocarlo desde la perspectiva clásica o bien desde la lógica de los modelos estructurales. Este mismo modelo si ofrece una R^2 alta, concluiremos, tras aplicarle la prueba F que dicho modelo es válido. Y lo es porque la parte explicada es significativa.

Si aplicamos a este caso, la lógica de los modelos estructurales nos podemos encontrar con una casuística muy variada y (aparentemente) contradictoria con la anterior. Podremos aceptar la hipótesis nula si el valor de Chi-cuadrado no es muy alto y concluir que nuestro modelo es válido, pero por motivos muy diferentes al caso anterior. Aquí es por el ajuste a la estructura y no por el valor de R^2 . Puede darse la paradoja que R^2 no sea significativo y el modelo sea válido desde el punto de vista estructural. O bien, que R^2 sea significativo pero estructuralmente no se reproduzca adecuadamente las varianzas-covarianzas observadas.

CONCLUSIONES

En definitiva, en estas páginas, hemos querido alertar sobre las diferentes lógicas aplicadas en el tema de la validación de modelos. Las pruebas clásicas están basada en el rechazo de la hipótesis nula, y es entonces, cuando se rechaza el valor de cero, cuando reconocemos algo. Por el contrario, los modelos estructurales operan sobre la base de aceptar la hipótesis nula en la discrepancia de las matrices de varianzas-covarianzas reproducida y observadas en la realidad. Cuando la discrepancia es pequeña (aceptación de la hipótesis nula) concluimos que el modelo es válido.

Al ser lógicas distintas puede presentarse una casuística de lo más variada y frecuentemente contradictoria en nuestras conclusiones si no somos conscientes de ello. Hay que decir que las pruebas clásicas sólo hablan de la intensidad de la relación, dada una cierta estructura que se da como buena, mientras que en los modelos estructurales, lo que se chequea fundamentalmente es la bondad en la estructura, aunque es preciso reconocer

que se contemplan en el modelo la fuerza de las relaciones entre las variables, aunque no sea objeto de estudio específico para analizar la validez.

Hay que insistir que la elaboración de modelos es una labor fundamentalmente teórica. Y que han de ser, en primer lugar *correctos* y además *precisos*. La prueba de Chi-cuadrado nos ofrece la corrección de lo especificado y la significación de R^2 la precisión en la predicción de los datos. Ambos aspectos son complementarios y ambos han de considerarse en el análisis estadístico.

INFLUENCIA DEL MENSAJE A TRAVÉS DE PORTEROS AUTOMÁTICOS EN LA TASA DE RESPUESTA

Encarnación León Martínez

Marisol Ortiz Hernández

Cristina Peñafiel Valbuena

Irene Prieto Reina

Universidad de Sevilla

RESUMEN

Mediante el presente trabajo hemos pretendido determinar la efectividad de distintos mensajes emitidos a través de porteros automáticos (operativizada en apertura del portal del edificio) para que potenciales encuestados permitiesen el acceso a su vivienda. Fue seleccionada una muestra de 2800 llamadas a porteros automáticos, siguiendo el procedimiento de muestreo de rutas aleatorias. Nuestra hipótesis de partida preveía que los mensajes con información más específica en su contenido serían más efectivos que aquellos cuyo contenido ofreciese una información de carácter más general. Se postuló además la posible influencia de la zona y de la hora del día en la tasa de respuesta de los inquilinos. Los resultados invirtieron el sentido de nuestra primera hipótesis y confirmaron la influencia de las variables zona y hora.

INTRODUCCIÓN

Hoy día, nuevos procedimientos de encuestas están siendo continuamente desarrollados y algunos de los más tradicionalmente utilizados se revisan y rediseñan para atender a la demanda de la economía y la investigación en ciencias psicológicas y sociales.

Los cambios sociales están influyendo en el desarrollo de la investigación mediante encuestas. La evolución de esta disciplina se ve directamente invadida por la evolución y los últimos avances en la tecnología de nuestro tiempo. Sirve como ejemplo la posibilidad de realizar encuestas vía internet.

En definitiva, el diseño de encuestas y de todo lo que ello conlleva, tiende cada vez más a dar soluciones operativas y novedosas en todo los tipos de encuestas, incluso en los tradicionales “por correo”, “por teléfono” y “cara a cara”.

El procedimiento de encuestas “cara a cara” es uno de los más aceptados en la investigación científica y el más habitual. Aún así, presenta una serie de inconvenientes entre los que se encuentra el “acceso” al encuestado.

Dicho “acceso” puede entenderse como un proceso que constaría, básicamente, de las siguientes fases:

- 1.- Selección de la vivienda (a través de una base de datos o por rutas aleatorias).
- 2.- Apertura del portal (acceso al edificio).
- 3.- Acceso a la vivienda.
- 4.- Selección del encuestado una vez dentro de la vivienda.

Nuestra investigación se centra en la fase 2, de apertura del portal del edificio, con el objetivo de encontrar los mensajes más exitosos para aumentar la probabilidad de apertura.

Cabe hacer una pequeña aclaración ética, en cuanto a los mensajes a utilizar en esta investigación, ya que realizar una encuesta supone, de algún modo, colaborar en un intento por conocer o mejorar determinados aspectos de la sociedad. De ahí que no se deba “mentir”. No obstante, se han probado mensajes dentro del engaño, que permiten comprobar una supuesta ventaja. Es el caso, por ejemplo, del mensaje número dos (buenos días/tardes, publicidad), emitido a través del portero automático cuando no teníamos intención de entregar ningún tipo de publicidad.

MÉTODO

Sujetos

Se utilizó una muestra de 2800 llamadas a viviendas de la ciudad de Sevilla con acceso mediante porteros automáticos procedentes de distintas zonas o barrios de la misma (cuatro zonas).

Materiales

Para la recogida de datos se utilizaron hojas de registro en las que se especificaron los siete mensajes a emitir y diferentes casillas en las que codificar la información obtenida a través del trabajo de campo realizado (ver anexo 1).

Variables

Se estudió la influencia de tres variables fundamentales para los objetivos de la investigación: El mensaje, la zona y el intervalo horario de realización de la llamada.

Los mensajes emitidos fueron de dos tipos:

a) De contenido más general:

1. Buenos días/tardes, correo.
2. Buenos días/tardes, publicidad.

3. Buenos días/tardes, para realizar una encuesta.

b) De contenido más específico:

4. Buenos días/tardes, Universidad de Sevilla, para realizar una encuesta.

5. Buenos días/tardes, Facultad de Psicología, para realizar una encuesta.

6. Buenos días/tardes, Universidad de Sevilla, para realizar unas preguntas.

7. Buenos días/tardes, Facultad de Psicología, para realizar unas preguntas.

En lo referente a las zonas, fue seleccionado un conjunto de cuatro barrios de la ciudad de Sevilla, extraídos al azar de un bombo con los nombres de los barrios de la misma. Los barrios seleccionados fueron Los Remedios, Nervión, Bami y Reina Mercedes.

Se realizaron 100 llamadas por mensaje y 700 llamadas por barrio.

Respecto al intervalo horario, se dividió el tiempo de realización de las llamadas en intervalos de dos horas: de 9 a 11 h., de 12 a 14 h., de 16 a 18 h. y de 19 a 21 h.

Estas variables venían especificadas en las hojas de registro, en las cuales aparecían una serie de columnas encabezadas con diferentes rótulos que cubrían el abanico de posibles respuestas que interesaba estudiar (ver anexo 1).

Otros aspectos tenidos en cuenta fueron los días (no se realizaron llamadas los días festivos) y el piso o la letra de la vivienda (no se llamó de manera consecutiva a dos viviendas de la misma planta de un edificio, aunque al terminar debía cumplirse el objetivo de haber llamado a todas las viviendas del mismo).

Procedimiento

Los portales fueron seleccionados siguiendo el procedimiento de rutas aleatorias, mediante la culminación de las siguientes etapas:

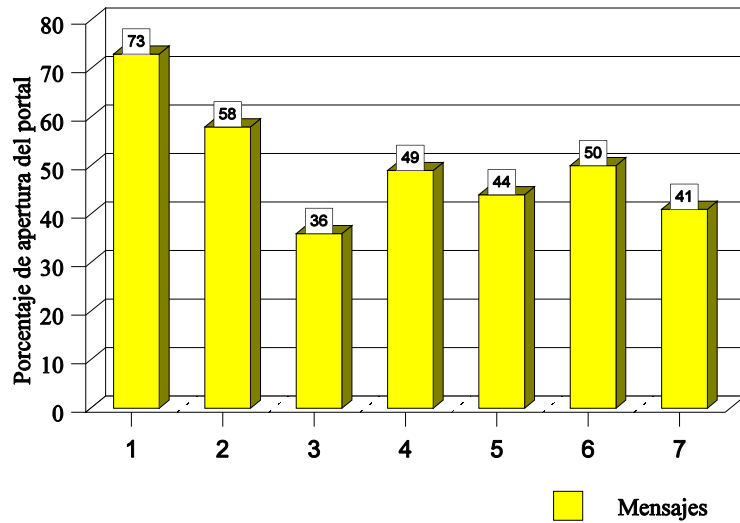
1.- Inicio del itinerario por un punto aleatorio del barrio.

2.- En la construcción de la ruta se aplicó el procedimiento sistemático “izquierda a derecha”, hasta dar con edificios con porteros automáticos, sin videocámara, en los que se realizaron las pertinentes llamadas a todas y cada una de las viviendas. Al terminar la serie de llamadas en el primer edificio se prosiguió con el siguiente de la misma acera, hasta completar 700 llamadas.

RESULTADOS

Cada mensaje fue emitido a través de los porteros automáticos tantas veces como número de descuelgues de telefonillo se obtuvo como respuesta a las llamadas realizadas.

Figura 1. Porcentaje de respuesta en función del mensaje.



Como se indica en la tabla 1A, de un total de 2800 llamadas se obtuvo una tasa de respuesta de 1498 descuelgues de telefonillo, así pues, fueron emitidos un total de 1498 mensajes. La figura 1 muestra el porcentaje de éxito correspondiente a cada mensaje. Por otro lado, la tabla 1B contiene los resultados del análisis realizado comparando todos los mensajes entre sí.

Tabla 1A. Frecuencia de respuesta de los inquilinos en función del mensaje.

| MENSAJE | Nº DESCUELGUE S TELEFONILLO | ABRE DIRECTAMENTE | NO ABRE | ABRE |
|---------|-----------------------------|-------------------|---------|------|
| 1 | 208 | 15 | 42 | 151 |
| 2 | 219 | 7 | 85 | 127 |
| 3 | 230 | 15 | 133 | 82 |
| 4 | 197 | 18 | 82 | 97 |
| 5 | 205 | 17 | 98 | 90 |
| 6 | 212 | 15 | 91 | 106 |
| 7 | 227 | 12 | 121 | 94 |
| TOTAL | 1498 | 99 | 652 | 747 |

Se realizaron un total de 700 llamadas a porteros automáticos por cada barrio y por cada intervalo de dos horas; esto supuso un tamaño muestral de 2800 llamadas.

En la tabla 2A aparece reflejada la tasa de respuesta de los inquilinos en función de la zona donde se realizó la llamada. La figura 2 muestra el porcentaje de descuelgue del telefonillo por barrio. La tabla 2B contiene los resultados del análisis realizado comparando todos los barrios entre sí.

Tabla 1B. Comparaciones entre las frecuencias de respuesta en función del mensaje.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|-------------------------|-------------------------|-------------------------|--------------------|--------------------|--------------------|
| 2 | 3.16 * 0.00 16 | | | | | |
| 3 | 7.74 * 0.00 00 | 4.74 * 0.00 00 | | | | |
| 4 | 4.82 * 0.00 00 | 1.79 0.07 34 | 2.84 * 0.00 46 | | | |
| 5 | 5.91 * 0.00 00 | 2.90 * 0.00 38 | 1.76 0.07 84 | 1.07 0.28 46 | | |
| 6 | 4.75 * 0.00 0 | 1.66 0.09 7 | 3.05 * 0.00 2 | 0.15 0.88 08 | 1.25 0.21 12 | |
| 7 | 6.55 * 0.00 00 | 3.50 * 0.00 02 | 1.26 0.20 76 | 1.62 0.10 52 | 0.52 0.60 30 | 1.81 0.07 04 |

En cada casilla aparece el valor estandarizado y debajo de éste su probabilidad asignada. El asterisco indica significación estadística.

Tabla 2A. Frecuencia de respuesta en función de la zona

| ZONA | Nº NO DESCUELGUES TELEFONILLO | Nº DESCUELGUES TELEFONILLO | TOTAL LLAMADAS A PORTERO |
|----------------|-------------------------------------|----------------------------------|--------------------------------|
| Los Remedios | 419 | 281 | 700 |
| Nervión | 364 | 336 | 700 |
| Bami | 256 | 444 | 700 |
| Reina Mercedes | 263 | 437 | 700 |
| TOTAL | 1302 | 1498 | 2800 |

Figura 2. Porcentaje de respuesta en función de la zona.

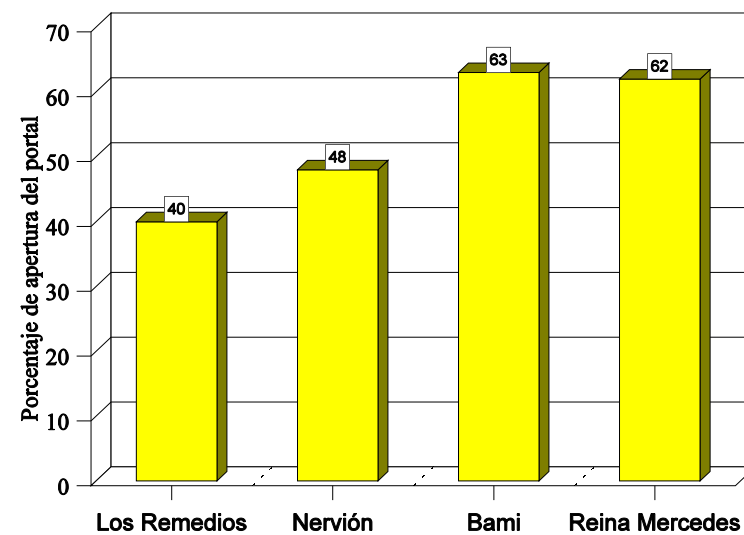


Tabla 2B. Comparaciones entre las frecuencias de respuesta en función de los barrios.

| | LOS REMEDIOS | NERVIÓN | BAMI |
|-----------------|-------------------------|-------------------------|--------------------|
| NERVIÓN | 2.96 * 0.00 30 | | |
| BAMI | 8.72 * 0.00 00 | 5.81 * 0.00 00 | |
| LOS REMEDIOS | 9.92 * 0.00 00 | 7.03 * 0.00 00 | 1.24 0.21 50 |

En cada casilla aparece el valor estandarizado y debajo de éste su probabilidad asignada. El asterisco indica significación estadística.

Tabla 3A. Frecuencia de respuesta de los inquilinos en función de la hora de la llamada.

| HORAS DE LLAMADA | N _§ NO DESCUELQUES TELEFONILLO | N _§ DESCUELQUES TELEFONILLO | TOTAL LLAMADAS A PORTEROS |
|---------------------|---|--|---------------------------------|
| 9 - 11 | 334 | 366 | 700 |
| 12 - 14 | 309 | 391 | 700 |
| 16 - 18 | 355 | 345 | 700 |
| 19 - 21 | 304 | 396 | 700 |

| TOTAL | 1302 | 1498 | 2800 |
|-------|------|------|------|
|-------|------|------|------|

En la tabla 3A se expone la tasa de respuesta de los inquilinos en función de la hora de realización de la llamada. La figura 3 muestra el porcentaje de descuelgues de telefonillo, por tramo horario. Por otro lado, la tabla 3B contiene los resultados del análisis realizado comparando todas los tramos horarios entre sí.

Tabla 3B. Comparaciones entre las frecuencias de respuesta en función de la hora de la llamada.

| | 9 - 11 | 12 - 14 | 16 - 18 |
|---------|--------------------|-------------------------|-------------------------|
| 12 - 14 | 1.34 0.18 02 | | |
| 16 - 18 | 1.12 0.26 28 | 2.46 * 0.01 38 | |
| 19 - 21 | 1.61 0.10 74 | 0.27 0.78 72 | 2.73 * 0.00 64 |

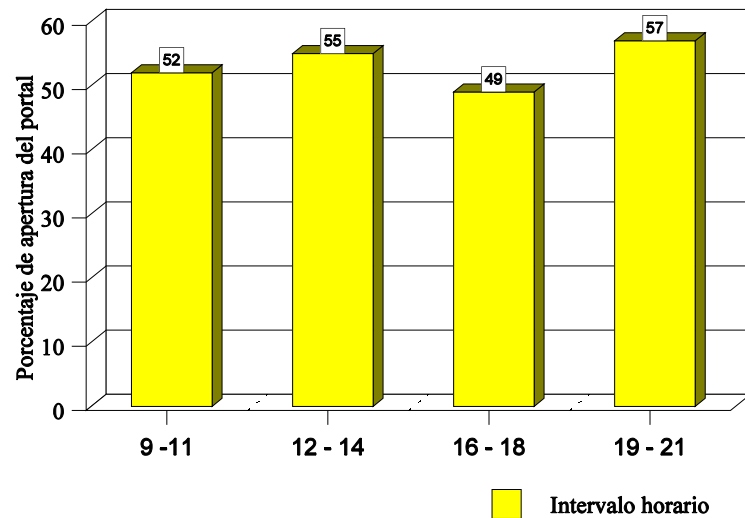
En cada casilla aparece el valor estandarizado y debajo de éste su probabilidad asignada. El asterisco indica significación estadística.

DISCUSIÓN

Los resultados muestran información de interés, tanto en lo que se refiere a la variable principal del estudio (mensajes) como a las que se utilizaron como medio de control (tramo horario y zona de encuesta).

El mensaje más exitoso ha resultado ser “correo”, seguido de “publicidad”, si bien el éxito de éste no difiere significativamente de los mensajes 4 y 6, donde queda incluida la Universidad de Sevilla. Obsérvese que los dos primeros mensajes difieren del resto en la circunstancia de que no requieren más participación que la apertura del portal. Los mensajes 3 a 7 informan de la realización de una encuesta para la que, aunque no se haga explícito, es necesaria la colaboración del encuestado.

Figura 3. Porcentaje de respuesta en función de la hora de realización de la llamada



Este resultado debe ligarse necesariamente con la reflexión de naturaleza ética a que hicimos referencia en la introducción a este trabajo. Ocurre, pues, que los mensajes basados en una mentira han tenido mayor éxito en la práctica. No obstante, el éxito se ha medido únicamente en términos de apertura del portal mediante el portero automático y no en la actitud positiva del encuestado para participar en la investigación. Cabe esperar sorpresa y, posiblemente, resitencia por parte de la persona que ha abierto la puerta a “correo” o “publicidad”, cuando observa que ha sido engañada y que, realmente, se trata de una encuesta.

Dentro del resto de los mensajes (3 a 7), los análisis muestran una clara superioridad de los mensajes que incluyen a la “Universidad de Sevilla” como institución que está detrás de la encuesta, frente a la inexistencia de entidad (“para realizar una encuesta”). Los mensajes que incluían a la “Facultad de Psicología” se encuentran a medio camino en términos de efectividad, aunque esta conclusión se establece en la muestra, no estando acompañada de significación estadística con respecto a ningún otro mensaje (3 a 7).

Luego, puede concluirse con la recomendación de mencionar una institución responsable en el mensaje para la apertura del portal. Igualmente, es necesario conocer o “sospechar” el prestigio asociado con las diferentes instancias implicadas en el estudio, seleccionando para el mensaje aquella que guarde mayor prestigio, aceptación, credibilidad, etc.

En referencia al resto de las variables contempladas en el estudio, los análisis muestran:

1. La peor hora para llamar se sitúa entre las 12 y las 14 horas.
2. Existe diferencia en cuanto a la respuesta por el telefonillo, de unas zonas a otras. Las características peculiares de los cuatro barrios seleccionados hacen sospechar que el éxito en la respuesta pueda tener alguna relación con el nivel socio-económico de la zona. Parece ser que los barrios de nivel más alto implican una mayor dificultad para lograr la apertura del portal.

**EVALUACIÓN EMPÍRICA DE UN DOBLE PROCESO DE
APRENDIZAJE DE LOS CONCEPTOS DE DECISIÓN
ESTADÍSTICA: DESARROLLO TRADICIONAL *VERSUS*
ANIMACIÓN POR ORDENADOR**

Francisco Javier Pérez Santamaría

Vicente Manzano Arrondo

Hassan Fazelli Khalili

Universidad de Sevilla

RESUMEN

El trabajo que aquí se presenta es consecuencia de una experiencia más en la utilización de nuevas tecnologías, con el objetivo de mejorar el aprendizaje en el contexto de la enseñanza universitaria. La creciente importancia dada al uso de las nuevas tecnologías por los psicólogos, pedagogos, sociólogos, etc... ha propiciado desde hace años, la preocupación por contrastar la utilidad de dicho uso. En nuestro caso, se pretende evaluar el aprendizaje de ciertos conceptos estadísticos en los alumnos, utilizando dos formas diferentes de explicación de los mismos. Una de las formas fue la aplicación del método tradicional y la otra, consistente en la exposición de los conceptos mediante animación por ordenador.

INTRODUCCIÓN

Quizás el concepto de innovación pedagógica tenga tanta antigüedad como el término de enseñanza. Desde una visión simplista, esta innovación, en muchas ocasiones, se ha convertido en sólo la incorporación instrumental en el proceso de la enseñanza de los últimos avances tecnológicos. De todos es conocido el ritmo tan vertiginoso de cambios y aparición de nuevas tecnologías, de tal forma que, no en pocas veces, antes de familiarizarnos con la última adquisición se la empieza a considerar desfasada (de forma muy especial todo lo relacionado con el mundo de la informática).

El mundo educativo no es ajeno a la nueva situación, haciendo uso en mayor o menor medida, de las novedades tecnológicas; aunque, con frecuencia, sin algún aval científico que justifique tal uso, guiándose tan sólo por la subjetividad o sencillamente por la moda.

Efectivamente, se conocen múltiples formas de transmitir la información; sin embargo, parecen imponerse aquéllas cuyo soporte es la utilización de las herramientas informáticas. Pero, ¿Podemos estar seguros de que estas herramientas favorecen realmente el aprendizaje? Según Steffens (1996) el problema requiere, en primer lugar, que el proceso de aprendizaje utilizando nuevas tecnologías cuente con criterios claros de evaluación que nos permitan aumentar nuestro conocimiento de los resultados de los citados procesos.

El presente informe es el resultado del análisis comparativo de dos situaciones de aprendizaje: la enseñanza de tipo tradicional, en la que la presentación de los conceptos estadísticos se realiza de forma convencional, (llamémosla “de pizarra”) y una enseñanza en la que la presentación de los mismos conceptos se realiza mediante secuencias animadas por ordenador y proyectadas con un cañón de imágenes de video (llamémosla “de ordenador”). El objetivo se centró en averiguar si realmente con el segundo de los métodos se conseguía un mejor aprendizaje de los conceptos estadísticos impartidos a los alumnos del primer curso de Psicología de la Universidad de Sevilla. Tras la aplicación de los dos métodos se llevó a cabo un análisis pormenorizado de las respuestas de los alumnos, dadas a una serie de preguntas, directamente relacionadas con los contenidos previamente expuestos.

Es cierto que en muchos foros educativos han surgido y surgen constantemente llamadas acerca del gran potencial de las nuevas tecnologías, especialmente informáticas, con propósitos educativos. Creemos, en efecto, que su potencial es enorme, sobre todo porque está basado en aspectos que teóricamente deben favorecer el aprendizaje. Sin embargo, la visión presentada, tan positiva en principio, se oscurece un tanto al comprobar también que existen problemas, sobre todo los que tienen que ver con la utilización directa del ordenador por parte de los alumnos, problemas, como apuntan Fariña y Arce (1993) que tienen que ver con la generación de procesos de ansiedad o de aislamiento del alumno ante la máquina, lo que le desorienta al sentirse atrapado entre las redes de las nuevas tecnologías.

Para evitar o minimizar dichos problemas, autores como Kozma (1991) y Johnsey (1992), entre otros, han propuesto soluciones realizando, por ejemplo, las exposiciones de los contenidos en grupos, pudiendo afirmar que cuando el aprendizaje se realiza en grupo, generalmente dicha exposición afecta positivamente a su ejecución posterior, a su productividad, transferencia y tiempos de prueba de lo aprendido.

MÉTODO

Sujetos

El trabajo fue llevado a cabo sobre un total de 187 alumnos de la Facultad de Psicología de la Universidad de Sevilla. Todos ellos asistían regularmente a las clases de la asignatura “Análisis de Datos en Psicología I” que se imparte en el primer curso.

Inicialmente, los alumnos estaban repartidos en 5 grupos de clase (A, B, C, D y E), siendo subdivididos en 10 subgrupos de prácticas.

Instrumentos

Para llevar a cabo el estudio se utilizó:

- 1.- Un programa para generar las animaciones: Autodesk Animator Pro.

2.- Un ordenador Pentium y tarjeta de video-color VGA de 640 x 480.

3.- Un aula, con el equipo informático conectado a un cañón de imágenes para la proyección de las animaciones en una pantalla del tamaño habitual para la proyección de transparencias.

4.- Un cuestionario, diseñado ex-profeso para la evaluación de los resultados. Se encuentra en el anexo y consta de cinco preguntas de respuestas cerradas (tres con opciones de verdadero y falso y dos de opción múltiple) y tres preguntas de respuestas a desarrollar, que fueron recodificadas en variables, tras el proceso de codificación inicial de las respuestas.

5.- Los análisis de datos fueron realizados con el programa SPSS (Paquete Estadístico para las Ciencias Sociales).

Cabe mencionar que la creación del programa de animación en soporte de cinta de video fue realizado por uno de los autores ante la inexistencia de dicho producto en el mercado, tarea que supuso un gran esfuerzo, tanto en lo intelectual como en cantidad de horas de trabajo.

Método y Procedimiento

Como quedó dicho anteriormente, hemos aplicado a los alumnos dos tipos de tratamiento: el denominado de “pizarra” versus el de “ordenador”, asignando cinco grupos de prácticas (grupos naturales) a cada uno de los tratamientos. Como consecuencia de esta asignación, 108 alumnos recibieron las explicaciones de los conceptos con el método “pizarra” y 79 con el método “ordenador”.

Los controles ejercidos en la experiencia implicaron que el procedimiento de enseñanza de los conceptos de decisión estadística siguieran un mismo guión en ambos grupos, aunque transmitidos de forma diferente, dependiendo del tratamiento. Se controló, asimismo, la duración de la experiencia: en ambos grupos se dedicaron 45 minutos a la exposición. La animación con imágenes estuvo siempre controlada por el profesor, quien, a través del teclado del ordenador principal, controló el ritmo de la exposición, siendo éste

similar a su propio ritmo de enseñanza en cualquier clase normal. Al finalizar las sesiones (Pizarra y Ordenador) se repartió un cuestionario anónimo, igual para todos los alumnos, disponiendo éstos de 30 minutos para su cumplimentación.

Variables

Nuestras principales variables han sido:

1ª) El aprendizaje de los conceptos de la decisión estadística, medido a través de las respuestas de los alumnos al cuestionario. Estas respuestas fueron codificadas como aciertos o errores. En las cinco primeras preguntas del cuestionario, de respuestas cerradas, se contabilizó directamente el acierto o el error de la respuesta. En el resto de las preguntas, al ser de respuestas abiertas, fue preciso llevar a cabo una recodificación: en la pregunta 6A los alumnos debían desarrollar todas las fases necesarias para la toma de una decisión estadística: enunciado de la hipótesis nula, identificación de los resultados como poco o muy probables a la luz de la hipótesis nula, decisión (rechazo o mantenimiento de la hipótesis nula) y conclusión. Por su parte, en la pregunta 6B se trataba de identificar dos únicas fases: decisión y conclusión. La pregunta 6C sólo contaba con la fase de decisión.

2ª) El procedimiento de enseñanza con dos niveles: pizarra u ordenador.

Se ejerció un control riguroso en variables que podrían contaminar la experiencia como la duración de la misma, tanto en la exposición (45 minutos) como en la contestación al cuestionario (30 minutos). El lugar de la realización de la experiencia fue la misma (aula informatizada), siendo el mismo profesor para cada grupo de clase y para cada par de grupos de tratamiento. Se utilizó el mismo instrumento de medida, el mismo guión en las exposiciones y los mismos objetivos de aprendizaje.

La selección de los conceptos a desarrollar (decisión estadística) obedece a la razón de que su desconocimiento por parte de los alumnos fuera generalizado, que no hubieran sido impartidos ni en COU ni en LOGSE y que su enseñanza tuviera un nivel de dificultad suficiente (Vallecillos, 1996; Hubbard, 1997).

Por último, una variable importante, que en las evaluaciones de preguntas con respuestas abiertas es necesario controlar, es el criterio de evaluación de las mismas. Para garantizar la constancia de esta variable, decidimos que la corrección y codificación de todas las hojas de evaluación se llevaran a cabo por una única persona.

RESULTADOS

Las respuestas a las doce preguntas resultantes de la recodificación del cuestionario, fueron valoradas, inicialmente, como correctas o incorrectas (0 ó 1 punto, respectivamente).

En la Tabla 1 se reflejan las frecuencias (f) y porcentajes (%) de los aciertos para cada una de las preguntas, tanto para la situación de Pizarra como de ordenador, de entre todos los sujetos que respondieron a cada pregunta (n).

Asimismo, en la figura 1 mostramos gráficamente los porcentajes de aciertos según el tipo de tratamiento y la pregunta.

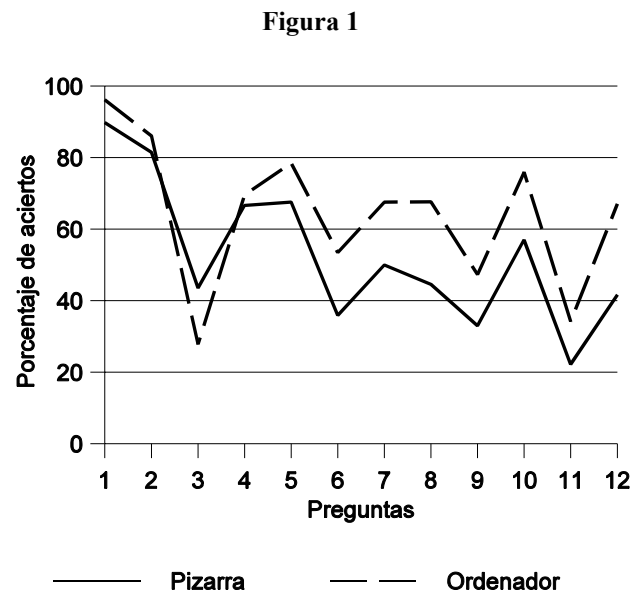


Tabla 1

| Preguntas | Pizarra | | | Ordenador | | |
|-----------|---------|----|-------|-----------|----|-------|
| | n | f | % | n | f | % |
| 1 | 108 | 97 | 89,81 | 79 | 76 | 96,20 |
| 2 | 108 | 88 | 81,48 | 79 | 68 | 86,08 |
| 3 | 108 | 47 | 43,52 | 79 | 22 | 27,85 |
| 4 | 108 | 72 | 66,67 | 79 | 55 | 69,62 |
| 5 | 108 | 73 | 67,59 | 79 | 62 | 78,48 |
| 6 | 78 | 28 | 35,90 | 58 | 31 | 53,45 |
| 7 | 100 | 50 | 50,00 | 74 | 50 | 67,57 |
| 8 | 83 | 37 | 44,58 | 65 | 44 | 67,69 |
| 9 | 100 | 33 | 33,00 | 76 | 36 | 47,37 |
| 10 | 107 | 61 | 57,01 | 79 | 60 | 75,95 |
| 11 | 108 | 24 | 22,22 | 79 | 27 | 34,18 |
| 12 | 108 | 45 | 41,67 | 79 | 53 | 67,09 |

A través de la tabla y la gráfica mencionadas, se puede observar que, salvo en el caso de la tercera pregunta, el grupo “ordenador” presenta un porcentaje de aciertos sensiblemente superior al grupo “pizarra”.

Se utilizó la U de Mann-Whitney para realizar cuatro pruebas de significación, según las puntuaciones en cuatro bloques de preguntas.

La primera prueba consistió en comprobar si existían diferencias estadísticamente significativas entre las respuestas globales del primer bloque de preguntas del cuestionario bajo los dos tratamientos. Para ello y dado que unas preguntas eran dicotómicas y otras de opción múltiple, se decidió ponderar las cinco preguntas en virtud de su probabilidad de acierto o error (0.5 ó 0.25), obteniendo la siguiente

variable (T_1):

$$T_1 = \frac{v_1 + v_2 + v_3}{2} + \frac{3}{4}(v_4 + v_5)$$

donde v_i es la valoración (0 ó 1) de la respuesta a la pregunta i .

De esta manera, T_1 ofrece puntuaciones comprendidas entre 0 y 3.

Al segundo bloque lo denominamos P6A (o apartado A de la pregunta 6) y permitía evaluar totalmente el único problema del cuestionario en el que el alumno debía enfrentarse a una tarea completa de decisión. La variable se expresó en los siguientes términos:

$$P6A = v_6 + v_7 + v_8 + v_9$$

Luego, P6A proporciona puntuaciones entre 0 y 4.

El tercer bloque, al que llamamos T_2 representa la suma directa de las variables en las preguntas abiertas, es decir:

$$T_2 = \sum_{i=0}^{12} v_i$$

Con T_2 las puntuaciones oscilan entre 0 y 7.

Por último, la puntuación total del cuestionario (Total), que es la suma directa de T_1 y T_2 y que, por lo tanto, supone una suma ponderada cuyos resultados van de 0 a 10.

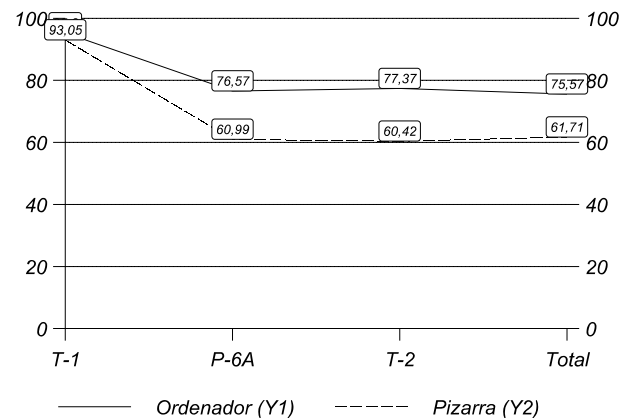
La Tabla 2 muestra los resultados del cruce entre las variables a través del estadístico U de Mann-Whitney, así como las medias en rangos de los dos tratamientos (pizarra y ordenador) y su significación:

Tabla 2

| | Rangos medios | | Estadístico | Significación |
|----------------|---------------|---------|-------------|---------------|
| | Ordenador | Pizarra | U | |
| T ₁ | 95,30 | 93,05 | 4163,5 | 0,7705 |
| P6A | 76,57 | 60,99 | 1676 | 0,0184 |
| T ₂ | 77,37 | 60,42 | 1631,5 | 0,0115 |
| Total | 75,57 | 61,71 | 1732 | 0,0410 |

Datos que, en la figura 2, reflejan una clara tendencia a mostrar al grupo de ordenador superior en éxitos al de pizarra, salvo la diferencia en T₁, que no ha resultado significativa.

Figura 2. Rangos medios



DISCUSIÓN

Volvemos a la pregunta inicial: ¿Animación por ordenador o desarrollo tradicional? A la luz de los resultados parecería clara la decisión. Y así concluimos que existe una clara ventaja en la utilización de la animación por ordenador. Sin embargo, esta conclusión debe ser matizada, pues no se trata en sentido estricto de un estudio experimental dado que los grupos son naturales, por mucho que los autores nos hayamos preocupado de realizar todo tipo de controles.

El segundo problema que nos preocupa es que los resultados de la animación no han sido constantes a lo largo de todo el proceso; en efecto, se invierten los resultados a favor de la enseñanza tradicional (Pizarra) en la pregunta 3. Por último, queda por comprobar el efecto debido a la “novedad” de la situación. Es lícito preguntar qué habría ocurrido con un grupo de clase acostumbrado a recibir la información con imágenes en movimiento, el día que el profesor aparece dispuesto a una clase a la antigua usanza.

Como conclusión, aunque es cierto que los resultados alcanzados permiten determinar un rendimiento superior con la utilización de la animación, resulta conveniente llevar a cabo nuevos estudios que determinen las situaciones favorables y desfavorables para la utilización de la técnica, proponiendo experimentos más ortodoxos.

ANEXO**PREGUNTAS CONTENIDAS EN EL CUESTIONARIO DE EVALUACION**

Las siguientes preguntas (1 a 3) se responden rodeando la opción que se considere correcta: V si el enunciado es verdadero o F si es falso.

1. En general, a medida que las consecuencias de mantener o rechazar la H_0 sean más trascendentes, es aconsejable escoger un α también más alto V F
2. En la decisión estadística, al aumentar el valor de α , aumenta la posibilidad de rechazar la hipótesis nula ~~V~~ F
3. En la decisión estadística el término α indica el riesgo de equivocación al mantener la hipótesis nula V F

En cada una de las 2 siguientes preguntas (4 a 7) hay cuatro opciones de respuesta, de entre las cuales sólo una es correcta. Redondea la letra que la encabeza.

4. En una toma de decisión estadística, utilizando un nivel de confianza del 80%, el estadístico se sitúa entre el parámetro y el límite inferior del intervalo de probabilidad construido. Si el nivel de confianza fuera del 85% ¿Cuál sería la toma de decisión correcta?
 - a) Rechazar la hipótesis nula.
 - b) Mantener la hipótesis nula.
 - c) Depende de las consecuencias en la toma de decisión.
 - d) Depende del planteamiento de la hipótesis nula.
5. En un estudio, si se ha rechazado la hipótesis nula, con un nivel de significación muy pequeño, esto implica que:
 - a) La diferencia entre el estadístico y el parámetro es insignificante.
 - b) La diferencia entre el estadístico y el parámetro es estadísticamente debida al azar.
 - c) La diferencia entre el estadístico y el parámetro es estadísticamente significativa.
 - d) Existe una probabilidad muy pequeña de que el estadístico sea diferente del parámetro.
6. Se supone que el nivel de agresividad de los escolares sevillanos, según una determinada prueba, es de $\mu = 80$ y $\sigma = 14$. Con objeto de comprobar el efecto de unos dibujos animados en la conducta agresiva, en una muestra de 49 escolares, y después de ver dichos dibujos, se observa un nivel medio de agresividad de 84.5. Con estos supuestos:
 - a) Con un nivel de confianza del 95%, ¿se puede afirmar que dichos dibujos animados influyen en el nivel de agresividad de los escolares sevillanos?
 - b) ¿Cuál sería la toma de decisión (mantener o rechazar H_0) y la conclusión, si el nivel de confianza fuera del 99%?
 - c) ¿Cuál sería la toma de decisión si el nivel de confianza fuera del 90%?

Utiliza la parte posterior a esta hoja para las soluciones a este problema.

GRACIAS POR TU COLABORACION

REFERENCIAS BIBLIOGRÁFICAS

- Aaker, D. A. y Day, G.S. (1989). *Investigación de mercados*. México: McGraw-Hill.
- Abramson, J. H. (1990): *Survey methods in community medicine*. London: Churhill Livingstone.
- AEDEMO (1986). *Código Internacional CCI-ESOMAR de prácticas legales en materia de investigación de mercado y opinión*. Barcelona: AEDEMO.
- Aldenderfer, M.S. y Blashfield, R.K. (1991). *Cluster Analysis*. Londres: Sage Publications.
- Anastasi, A. (1986) Evolving concepts on test validation. *Annual review of psychology*. 37. 1-5.
- Andrés Orizo, F. (1996): *Sistema de valores en la España de los 90*, Madrid: CIS.
- Anguera, M. T.; Arnau, J.; Ato, M.; Martínez, R.; Pascual, J. y Vallejo, G. (1995) *Métodos de investigación en psicología*. Madrid: Síntesis.
- Arce, C. (1993). *Escalamiento multidimensional. Una técnica multivariante para el análisis de datos de proximidad y preferencia*. Barcelona: PPU.
- Arias, A y Fernández, B. (1998) La encuesta como técnica de investigación social, en A. J. Rojas; J. S. Fernández, y C. Pérez (ed) *Investigar mediante encuestas*. Madrid: Síntesis.
- Asher, H.B. (1984). *Causal modeling*. Beverly Hills: Sage.
- Asok, C. y Sukhatme, B.V. (1976): On Sampford's procedure of unequal probability sampling without replacement. *Journal of the American Statistical Association*, 71, 912-918.
- Batista, J.M. y Coenders, G. (1997): Modelos de ecuaciones estructurales: el análisis factorial confirmatorio para la depuración del cuestionario. En D. Renom (Ed.) *Tratamiento informatizado de datos*. Barcelona: Masson.
- Bentler, P.M. (1995): *EQS structural equations program manual*. Encino: Multivariate Software.
- Berka, K. (1983): *Measurement*. London: Reidel

- Bickel, P. y Freedman, D. (1981): Some asymptotic theory for the bootstrap. *The Annals of Statistics*, Vol. 91 (6). pag. 1196-1217.
- Bollen, K.A. (1989): *Structural equations with latent variables*. New York: Wiley.
- Bollen, K.A. (1993): *Testing structural equation models*. Londres: Sage.
- Campbell, D. T. (1988): Evolutionary Epistemology. En Overman, E. S. (Ed.) *Methodology and epistemology for social science: selected papers. Donald T. Campbell*. Chicago: University of Chicago Press.
- CIRES (1994): *La realidad Social en España 1992-1993*, Madrid: Fundación BBV.
- Cochran, W. G. (1977) *Sampling Techniques*. New York: John Wiley.
- Cook, T. D. y Campbell, D. T. (1979) *Quasi-experimentation. Design & analysis issues for field settings*. Boston: Houghton Mifflin.
- Cook, T., Levinton, L. y Shadish, W. (1985): Program evaluation. En G. Lindzey y E. Arosen (eds.) *Handbook of Social Psychology*. pp. 699-777. Nueva York: Random House.
- Crocker, L y Algina, J. (1986) *Introduction to classical and modern test theory*. Nueva York: Holt, Rinehart and Wilston.
- Cronbach, L. J. (1988) Five perspectives on validity argument. En: H. Wainer y H. I. Brown (Eds). *Test validity*, pp(3- 17). Hillsdale. New Jersey: Lawrence Erlbaum.
- Chacón, S., y Martínez, R. y Moreno, R. (en prensa): Criterios para la enseñanza y la construcción de materiales docentes de fundamentos de metodología científica. En G. Mares y Y. Guevara. *Avances en Psicología Interconductual*. México: UNAM.
- Chambers, R.L. y Dunstan, R. (1986): Estimating distribution functions from survey data. *Biometrika*, 73, 597-604.
- De Miguel, A. (1994): *La Sociedad Española 1993-94*, Madrid: Editorial Complutense.
- De Miguel, A. (1997): *La Sociedad Española 1996-97*, Madrid: Editorial Complutense.
- Del Pino Artacho, J. y Bericat, E. (1998): *Valores sociales en la cultura andaluza. Encuesta Mundial de valores. Andalucía 1996*, Madrid: CIS.
- Diaconis, P. and Efron, B. (1983): Computer intensive methods in statistics. *Sci. Amer.* 248, 116-130.

- Efron, B. (1979): Bootstrap methods: another look at the jackknife. *The Annals of Statistics* Vol.7 (1), 1-26
- Efron, B. (1979a): Computers and the Theory of Statistics: Thinking the Unthinkable. *SIAM*, 21.
- Efron, B. y Tibshirani, R.J. (1993): *An introduction to the bootstrap*. New York: Chapman and Hall.
- Fariña, F. y Arce, R. (1993). *Ansiedad ante los ordenadores*. Colección Recursos Humanos. Serie Psicología. Madrid: Eudema.
- Fazeli, H.; Manzano, V y Pérez Santamaría, F.J. (1998): *Análisis de datos en Psicología*. Madrid: Pirámide.
- Fernández, F.R. y Mayor, J.A. (1995): Método de exploración secuencial cuasi equilibrado en muestreo en poblaciones finitas. [A sampling scheme with sequential quasi balanced exploration]. *Estadística Española*, 135, 143-162.
- Fernández, F.R. y Mayor, J.A. (1996): Diseños muestrales π -equivalentes y equivalentes de primer orden. [π -Equivalent and first order equivalent sampling designs]. *Questiío*, 20, 45-70.
- Ferrater Mora, J. (1984): *Diccionario de Filosofía*, Madrid: Alianza Editorial.
- Fink, A. (1995) *How to report on surveys*. Thousand Oaks, California: Sage.
- Fowler, F. J. (1993) *Survey research methods*. Newbury Park, California: Sage.
- García Ferrando, M.; Ibáñez, J. y Alvira, F. (1992): *El análisis de la realidad social. Métodos y técnicas de investigación*. Madrid: Alianza Editorial.
- Groves, R. M. (1989) *Survey errors and survey costs*. Nueva York: Wiley.
- Gómez Benito, J. (1996): Aplicaciones de los modelos de estructuras de covarianza al análisis psicométrico. En J. Muñiz (Coord.) *Psicometría*. Madrid: Universitas.
- Harvatopoulus, y. et al. (1992): *El arte de la encuesta*. Bilbao: Ediciones Deusto.
- Hill, B.M. (1968): Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, 63, 677-691.
- Hood, R.W. Jr. (1970): Religious orientation and the reported religious experience. *Journal for the Scientific Study of Religion*, 9(4). 285-291.
- Hubbard, R. (1997). Assesment and the Process of learning statistics. *Journal of Statistical Education*, 5, 1, 21-29.

- IESA (1997): *Normas para la aplicación del cuestionario*. Córdoba: Instituto de Estudios Sociales de Andalucía.
- Johnsey, A.; Morrison, G.R. y Ross, S.M. (1992). Using elaboration strategies training in computer based instruction to promote generative learning. *Contemporary Educational Psychology*, 17, 125-135.
- Jöreskog, K.G. y Sörbom, D. (1993): *LISREL8: Structural equation modelling with the Simplis command language*. Hillsdale: Lawrence Erlbaum.
- Kozma, R.B. (1991). Learning with Media. *Review of Educational Research*, 61, 179-211.
- Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 1-27.
- Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29, 115-129.
- Kuk, A.Y.C. (1988): Estimation of distribution functions and medians under sampling with unequal probabilities. *Biometrika*, 75, 97-103.
- Kuk, A.Y.C. y Mak, T.K. (1989): Median estimation in the presence of auxiliary information. *Journal of the Royal Statistical Society, Series B*, 51, 261-269.
- Lambert, J. (1989): *Psicología Social*. Madrid: Pirámide.
- León y Montero (1997): *Diseño de investigaciones*. Mc Craw Hill.
- Lininger, Ch. A. y Warwick, D.P. (1984). *La encuesta por muestreo: teoría y práctica*. México: Compañía Editorial Continental S.A.
- Manzano, V. (1993): *Análisis estadístico con SPSS/PC+. Fundamentos de análisis, preliminares, estudios descriptivos y utilidades*. Madrid: Rama.
- Manzano, V. (1995): *Inferencia estadística. Aplicaciones con SPSS/PC+*. Madrid: Rama.
- Manzano, V. y González, A. (1998): Selección del encuestado. En A.J. Rojas, J. S. Fernández y C. Pérez (Eds.) *Investigar mediante encuesta*. Madrid: Síntesis.
- Manzano, V.; Rojas, A.J. y Fernández, J. (1996): *Manual para encuestadores*. Barcelona: Ariel.
- Martínez, R. (1995) El método de encuestas por muestreo: Conceptos básicos. En: M. T. Anguera.; J. Arnau.; M. Ato; R. Martínez; J. Pascual y G. Vallejo. (1995) *Métodos de investigación en psicología*. Madrid: Síntesis

- Martínez, R. (1995): *Psicometría: Teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- Maxwell, S. y Delaney, H. (1990): *Designing experiments and analyzing data*. California: Brooks/Cole Pub.
- Mayor, J.A. (1997): Diseños Muestrales apropiados para la estimación de la función de distribución sobre poblaciones finitas. *XXIII Congreso Nacional de Estadística e Investigación Operativa*. Valencia.
- McGuire, W.J. (1994): The psychology of behavior and the behavior of psychologists. *Mexican Journal of Behavior Analysis*, 20. *Monografic Issue*, 75-98.
- McGuire, W.J. (1997): Creative hypothesis generating in Psychology: some useful heuristics. *Annual Review of Psychology*, 48, 1-30.
- Messick, S (1989) Meaning and values in test validation: the science and ethics of assessment. *Educational research*, 18. 5-11.
- Micceri, T. (1989): The unicorn, the normal curve and other improbable creatures. *Psychological Bulletin*, 105, 156-166.
- Mittchell, J. (1990) *An introduction to the logic of psychological measurement*. Hillsdale, N. J.: Lawrence Erlbaum,
- Morales, P. (1988): *Medición de actitudes en psicología y educación*. San Sebastián: Tattalo.
- Moreno, R., Martínez, R. y Chacón, S. (en preparación): *Fundamentos de metodología científica*.
- Muñoz Pérez, J. (1990): Dispersive ordering by the spread function. *Statistics & Probability Letters*, 10, 407-410.
- Muñoz Pérez, J. y Sánchez Gómez, A. (1990): Dispersive ordering by dilation. *J. Appl. Prob.*, 27, 440-444.
- Muñoz Reyes, A.M. (1994): *Simulaciones bootstrap: correcciones a su variabilidad muestral*. Tesis doctoral. Sevilla.: Universidad de Sevilla.
- Nunnally, J. C. y Bernstein, I. J. (1995): *Teoría psicométrica*. Madrid: McGraw-Hill
- Padilla, J. L. (1998): El cuestionario. En A.J. Rojas, J.S. Fernández y C. Pérez (Eds.) *Investigar mediante encuesta*. Madrid: Síntesis.

- Pargament, K.I., Kennell, J., Hathaway, W., Grevengoed, N. ; Newman, J. y Jones, W. (1988): Religion and the problem-solving process: three styles of coping. *Journal for the Scientific Study of Religion*, 27(1), 90-104.
- Paz, M.D. (1996): Validez. En J. Muñiz (Coord.) *Psicometría*. Madrid: Universitas.
- Pedhazur, E.J. (1982). *Multiple regression in behavioral reserar. Explanatio and prediction (2nd ed.)*. New York: Halt Rinehart.
- Popper, K.R. (1963): *Conjectures and refutations. The growth of scientific knowledge*. London: Routledge and Kegan Paul.
- Rao, J.N.K. y Wu, C.F.L.(1988): Resampling Inference with complex Survey Data. *JASA* Vol 83, (401).
- Rao, J.N.K.; kovar, J.G. y Mantel, H.J. (1990): On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*, 77, 365-375.
- Rincón Igea, D. et al (1995): *Técnicas de investigación en ciencias sociales*. Madrid: Dykinson.
- Rojas, A. J.; Fernández, J. S. y Pérez, C. (ed)(1998) *Investigar mediante encuestas*. Madrid: Síntesis.
- Rojas, A.J. (1998): *Aplicación del modelo de Crédito Parcial y del modelo de Escalas de Clasificación a la medición de actitudes*. Tesis doctoral. Granada: Universidad de Granada.
- Rossi, P.; Wright, J. y Anderson, A.B: (1998). *Handbook of Survey Researchs*. Massachussets: Academic Press.
- Rubenstein, R. (1981): *Simulation and Monte Carlo Method*. New York: Wiley.
- Santestmases, M. (1997): *DYANE, Diseño y análisis de encuestas en investigación social y de mercados*. Madrid: Pirámide.
- Sampford, M.R. (1967): On sampling without replacement with unequal probabilities of selection. *Biometrika*, 54, 499-513.
- Saris, W.E. y Stronkhorst, L.H. (1984): *Causal modelling in Nonexperimental Research*. Amsterdam: Sociometric Research Foundation.
- Sedrask, J. y Meyer, J. (1978): Confidence intervals for the quantiles of a finite population: simple random and stratified simple random sampling. *Journal of the Royal Statistical Society, Series B*, 40, 239-252.

- Shepard, R.N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distances function (I y II). *Psychometrika*, 27, 125-139, 219-246.
- Singh, K. (1981): On asymptotics accuracy of Efron's Bootstrap. *The Annals of Statistics*. Vol. 9, 1187-1195.
- Smith, S.M. (1990). *PC multidimensional statistics package*. Provo, UT: Institute of Business Management, Brigham Young University.
- Steffens, K. (1996). Learning with multimedia computer programs problems of evaluating processes and outcomes. Comunicación presentada al Congreso Europeo de Investigación Educativa. Sevilla.
- Stevens, S. S. (1946): On the theory of scales of measurement. *Science*, 103, 667-680.
- Stevens, S. S. (1946) On the theory of scales of measurement. *Science*. 103. 667-680.
- Suppes, P., Krantz, D.H., Luce, R.D. y Tversky, A.A. (1989): *Foundations of measurement. Vol II*. London: Academic Press.
- Toulmin, S.E. (1979): *La comprensión humana*. Madrid: Alianza.
- Trochim, W. M. (1997) *Introduction to validity*. URL: <http://trochim.human.cornell.edu/kb/INTROVAL.HTM>
- Vallecillos, A. (1996). *Inferencia estadística y enseñanza: un análisis didáctico del contraste de hipótesis estadísticas*. Colección Mathema. Granada: Comares.
- Varela, J.; Rial, A. y Braña, T. (1998). ¿Qué valora el consumidor en una promoción de ventas?. *Esic Market* (en prensa).
- Woodruff, R.S. (1952): Confidence intervals for medians and other position measures. *Journal of the American Statistical Association*, 47, 635-646.