

PROBLEMAS ORIGINADOS POR LA NO RESPUESTA EN INVESTIGACIÓN SOCIAL: DEFINICIÓN, CONTROL Y TRATAMIENTO.

Presentamos un trabajo sobre la calidad en la recogida de datos mediante encuestas, en un momento en el que los estudios sociales han trascendido los ámbitos especializados y cada vez ocupan un lugar más destacado en los medios de comunicación. Hace tan sólo unos años los artículos sobre el tema tan sólo aparecían en revistas especializadas, mientras que en la actualidad es muy frecuente leer en la prensa trabajos sobre la "última encuesta del consumidor español", "los jóvenes son más tolerantes según una encuesta...", e incluso explicando las razones por las fallan los sondeos electorales" (Wert 1996, El País 1997).

La trascendencia de esta información es la que da origen al presente trabajo: el hecho de que este tipo de investigación social esté plenamente consolidado en nuestro país, unido a la elevada presencia de "resultados de encuestas" en los medio de comunicación -y de hecho en la vida de cada ciudadano- nos lleva a considerar que el desarrollo de esta forma de recogida de datos está plenamente consolidada, y que nuestra labor es trabajar con el fin de mejorar el uso de este instrumento.

Pese a esta realidad, un gran número de expertos consideran aún que "diversos aspectos centrales de la investigación mediante encuestas.. han sido escasamente investigados, o por lo menos no lo suficientemente como el amplio uso de encuestas merece" (García Ferrando, 1989). Es por ello por lo que presentamos un trabajo que tiene como objetivo exponer estrategias que permitan reducir la "ho respuesta" en las investigaciones que utilizan encuestas, independientemente del procedimiento de recogida de datos utilizado: personal (cara a cara), telefónico, postal, CAPI, TDE, VR, etc. En este trabajo no expondremos las ventajas de cada método de recogida de datos -aspecto que ha sido desarrollando en numerosos trabajos- sino que nuestro objetivo es analizar los problemas que genera la presencia de la no respuesta, y cómo utilizar determinadas estrategias para aumentar la tasa de respuesta. Lynn (1996) considera que la no respuesta genera dos tipos de problemas:

1. Se reduce el tamaño de la muestra, aumentando el error muestral y con ello el error típico de las estimaciones. Este aspecto tiene poca importancia puesto que en numerosas ocasiones se conoce el nivel de no respuesta, de modo que un aumento del tamaño muestral -en la fase del diseño- permitirá solucionar este problema: supongamos que deseamos realizar una investigación con una muestra de 1200 personas mediante una encuesta telefónica. Si la tasa media de respuesta en la encuesta telefónica es del 75%, un aumento de la muestra permitirá solucionar este problema.
2. Otro problema producido por la no respuesta es la introducción de sesgos: si las personas que no responden fueran similares a las que responden para reducir la no respuesta bastaría con utilizar el procedimiento realizado anteriormente. El problema es que la no respuesta no es aleatoria, no se produce por igual en todos los estratos de la población, de modo que la muestra obtenida genera sesgos a la hora de estimar la población total. Así una baja tasa de respuesta trae como consecuencia que los resultados sólo sean representativos de una parte de la población objeto de estudio, no siendo posible la generalización de los mismos, aspecto fundamental en la investigación social mediante encuesta.

Este segundo razonamiento está indicando el problema fundamental que aparece cuando se obtiene una baja tasa de respuesta, por el sesgo que cometemos al eliminar representatividad muestral al seleccionar siempre la misma tipología de personas. En el proceso de realización de una muestra de 1200 personas, por ejemplo, las ausencias en el domicilio suelen ser rápidamente sustituidas por los *reservas*; de modo que aunque siempre se terminará realizando las 1200 entrevistas, las personas seleccionadas pueden no tener las mismas características que aquellas que se planteaban elegir en un primer momento. El problema fundamental es que en la elección de los *reservas* se ha podido cometer un sesgo a la hora de seleccionar a los entrevistados, dando una mayor probabilidad de elección a las personas que se encontraban en casa en el momento de realizar la recogida de datos. No nos olvidemos que, a la hora de diseñar un muestreo, es mucho más importante la representatividad que el tamaño muestral (Alderete 1996); aunque parece que estamos más preocupados por el tamaño muestral (y por el error muestral que ello genera) que por el número de visitas que es necesario realizar para seleccionar un determinado entrevistado.

Tras esta exposición del problema el lector se preguntará si en las investigaciones actuales es tan elevada la tasa de no respuesta como para darle la importancia que le atribuimos aquí. La elección de un buen criterio que permita responder fielmente a la pregunta anterior no es tarea fácil, aunque consideramos que un indicador adecuado sobre la importancia de determinados problemas sociales puede ser "el número de trabajos científicos publicados sobre el tema". Así el análisis de las principales revistas científicas en nuestro país desvela una falta de interés sobre el tema, puesto que tan sólo han sido localizados cuatro artículos, y tres de éstos tienen más de 10 años. Además, no hay un solo libro publicado sobre el tema en España. Tras la exposición de estos datos la pregunta es clara: la no respuesta, ¿es un problema en la investigación social que se realiza en nuestro país?. ¿Merece la no respuesta, -un aspecto "olvidado" en la investigación social en nuestro país- la publicación de este libro?. Pese a que el autor de este trabajo no ha encontrado ninguna referencia escrita sobre la tasa de respuesta en nuestro país, conversaciones con profesionales de la investigación de opinión y mercado han puesto de manifiesto una tasa de respuesta de la encuesta personal cercana al 60%, que en el caso de la encuesta telefónica llega hasta el 76%. La falta de datos publicados sobre el tema en España nos lleva a consultar fuentes de otros países: un estudio realizado mediante un *meta-análisis* de todas las publicaciones recogidas en Psychological Abstracts, Sociological Abstracts, Dialog/SSCI, y la base de datos SRM proporciona una tasa de respuesta en la encuesta personal del 73.5%, que se reduce al 70.3% en la telefónica y al 68.2% en la postal (Hox y DeLeeuw, 1994).

Estos resultados muestran una alta tasa de respuesta, y nos sugieren que quizás estamos dando demasiada importancia al problema de la no respuesta en la metodología mediante encuesta. Sin duda estaríamos de acuerdo con esta conclusión si no fuera por el hecho de que la colaboración está bajando en los últimos años, como demostraremos en las páginas siguientes. Por esta razón, una vez elaborado el *meta-análisis* sobre tasa de respuesta expuesto en el párrafo anterior, Hox y DeLeeuw realizaron una encuesta en los Países Bajos para conocer la tasa de respuesta en la actualidad, encontrándose que los resultados diferían totalmente de los hallados en su *meta-análisis*: tasa de respuesta de la encuesta personal 46%, postal 64%, telefónica 64% (DeLeeuw, 1996). Investigaciones realizadas en otros países no difieren significativamente de los resultados de este estudio y colocan la tasa de respuesta en

las encuestas postales en el 58%, alcanzado el 65.3% en las entrevistas personales (cara a cara) y el 60.2% en las telefónicas (Goyder, 1987).

Como queda de manifiesto estos datos son muy similares a la información sobre cooperación en las encuestas en España que obteníamos en conversaciones con profesionales del sector. Esta concordancia nos plantea la siguiente reflexión: ¿de quién estamos hablando cuando, tras exponer los resultados de una investigación, decimos que el X% de los españoles piensa que..., o votará a cierto partido, o tiene unos hábitos....?. Si necesitamos visitar dos hogares para conseguir un entrevistado mediante la encuesta personal (tasa de respuesta del 46% en los Países Bajos, 60% en España), ¿de que parte de la población estamos hablando?. Si los que responden tuvieran el mismo perfil, mismas opiniones, iguales características,... que los que no lo hacen podríamos decir que unos representan a los otros; el problema es que numerosas investigaciones han señalado importantes diferencias entre los que responden y los que no lo hacen. Como veremos en detalle en las páginas siguientes los que responden suelen tener un mayor nivel educativo, un mayor estatus socioeconómico, son mayoritariamente mujeres, solteros, jóvenes, residen en hábitat rural,... (McDaniel y otros, 1987). Es decir, estamos exponiendo resultados de una parte de la población, extrapolando los resultados de esa parte de la población al conjunto del país.

Preocupados por este problema planteamos el presente trabajo cuyo fin es conocer las características de las personas que tienen unas mayores tasas de no respuesta, para seguidamente exponer una serie de estrategias que permitan reducir las tasas de no respuesta. No obstante, antes de plantear estas estrategias este trabajo comienza planteándose hasta que punto puede considerarse la "no respuesta" como un problema importante en la investigación social mediante encuesta que se realiza actualmente.

Para realizar un adecuado planteamiento del problema será necesario considerar no sólo los momentos en los cuales se produce la no respuesta, sino fundamentalmente las consecuencias producidas por la no consideración de todos los entrevistados (capítulo dos). Un adecuado planteamiento del problema obliga a reflexionar sobre los factores que intervienen en la no respuesta, teniendo en cuenta tanto las características sociodemográficas de los que no responden como las características "sociales" de la no respuesta (capítulo tres).

Expuesta la importancia de la no respuesta en la investigación social, y tras analizar los diferentes factores que intervienen en la misma llega el momento de presentar diversas estrategias que permitan solucionar este problema, aspectos que serán analizados en los capítulos cuatro y cinco.

En el capítulo cuatro se exponen diversas estrategias que permiten reducir la no respuesta: las primeras reciben el nombre de "estrategias de prevención" puesto que son empleadas antes de la recogida de datos (aumento del número de contactos y empleo de gratificaciones), mientras que otras son empleadas en la realización del trabajo de campo: 1) insistencia en el anonimato y confidencialidad de las respuestas, 2) sustitución de los que no responden, 3) re-entrevistar los que han rechazado responder, 4) realizar un segundo muestreo entre los que no responden, y 5) muestreo de preguntas: buscando información fundamental.

En el capítulo cinco se presentan otras estrategias para reducir el efecto de la no respuesta, pero en este caso no se trata de "prevenir" ni de "intervenir" durante la recogida de datos, sino más bien de "tratar de arreglar lo que tenemos". Estas estrategias serán utilizadas cuando ya se ha finalizado la recogida de datos, una vez que la información obtenida ha sido introducida en un formato magnético. En primer lugar se exponen diversas técnicas para analizar las características de la no respuesta, y a continuación se exponen diversos procedimientos para solucionar el problema, basados fundamentalmente en la ponderación de las respuestas obtenidas.

*PROBLEMAS ORIGINADOS POR LA NO RESPUESTA
EN INVESTIGACIÓN SOCIAL:
DEFINICIÓN, CONTROL Y TRATAMIENTO.*

1.- INTRODUCCIÓN.

2.- LA NO RESPUESTA COMO PROBLEMA EN LA INVESTIGACIÓN SOCIAL.

- A.- Definición de conceptos.
- B.- Categorías de entrevistados según el rol que ocupan en la investigación.
- C.- Momentos en los que se produce la no respuesta.
- D.- Cálculo e interpretación de medidas relativas de no respuesta.
- E.- Consecuencias producidas por la no consideración de todos los entrevistados.

3.- FACTORES QUE INTERVIENEN EN LA NO-RESPUESTA.

- A.- Características sociodemográficas de los que no responden.
- B.- La respuesta en relación con el entorno social en el que se produce.
- C.- Otros factores que inciden en la no respuesta (Meta-Análisis).
- D.- Un intento de teorización sobre los factores que intervienen en la no respuesta:

4.- ESTRATEGIAS PARA REDUCIR LA NO RESPUESTA.

- A.- Estrategias de prevención.
 - 1.- Aumento del número de contactos.
 - * Al comienzo del estudio (A priori).
 - * Insistencia en la cooperación de los entrevistados (Notificación a posteriori).
 - 2.- Utilización de gratificaciones.
 - 3.- Utilización conjunta de ambos elementos.
- B.- Actuaciones durante la realización del trabajo de campo.
 - 1.- Insistencia en el anonimato y confidencialidad de las respuestas.
 - 2.- Sustitución de los que no responden.
 - 3.- Volver a entrevistar los que han rechazado responder.
 - 4.- Realizar un segundo muestreo entre los que no responden.
 - 5.- Muestreo de preguntas: buscando información fundamental.

5.- ESTRATEGIAS PARA MITIGAR EL EFECTO DE LA NO RESPUESTA.

- A.- Identificación/evaluación de la no respuesta.
 - 1.- Comparación con la población total.
 - 2.- Comparación de distintas características de la muestra.
 - 3.- Comparación entre los rasgos de los que responden y los que no responden.
- B.- Estrategias para solucionar el problema: procesos de ajuste.
 - 1.- Imputación.
 - 2.- Ajustes basados en la ponderación.
 - a.- Ajuste de Politz Simmons.
 - b.- Ajustes Muestrales.
 - c.- Ajustes basados en la población (post-estratificación).

- d.- Otros métodos.
- f.- Implicaciones prácticas.

6.- CONCLUSIÓN.

REFERENCIAS BIBLIOGRÁFICAS.

TÉCNICAS DE ANÁLISIS DE DATOS PARA INVESTIGADORES SOCIALES

Presentamos una obra en la que se exponen algunas de las técnicas de análisis de datos más utilizadas en la investigación social mediante encuesta, contextualizadas dentro de un diseño de investigación concreto. El objetivo de este trabajo no se centra tanto en una exposición de distintas técnicas de análisis de datos, sino que se explica cómo solucionar problemas de una investigación utilizando determinadas técnicas de análisis de datos. Con este fin se plantea una investigación completa, desde la formulación de hipótesis hasta la exposición de resultados y generalización de los mismos. Este proceso de investigación requerirá utilizar técnicas de análisis de datos, oportunidad que aprovecharemos para exponer el uso y la interpretación de cada una de éstas.

El elevado número de técnicas de análisis de datos utilizadas en la investigación social aconseja realizar una selección a fin de explicar algunas de ellas en profundidad. Uno de los criterios más importantes para diferenciar entre las técnicas de análisis de datos hace referencia al tipo de métrica de las variables utilizadas; cualitativas y cuantitativas. El hecho que la mayor parte de la investigación que se realiza en ciencias sociales utilice variables cualitativas, unido a la escasa atención que han recibido éstas en los manuales de investigación social nos anima a centrarnos en las técnicas de análisis de datos para variables cualitativas.

Antes de continuar deseamos realizar algunas consideraciones sobre el contenido de esta obra. Señalar en primer lugar que es un libro de *Técnicas de Investigación Social* ya que nuestro interés es el planteamiento y posterior desarrollo de una investigación completa, investigación que requiere utilizar procedimientos de análisis de datos, como también precisa la elaboración de unas hipótesis de trabajo, un análisis de los datos a recoger, construcción de un cuestionario y prueba del mismo, etc. Dentro del análisis de datos nuestro interés se centrará en razonar porqué se utiliza una determinada técnica y como interpretar los resultados que proporciona. No nos preocupa la comprensión de fórmulas complejas, que además trataremos de evitar en la medida de lo posible, sino que deseamos profundizar en la utilización de determinadas técnicas y en la interpretación de los resultados que proporcionan.

El enorme desarrollo de los paquetes estadísticos, unida a la amplia difusión de ordenadores personales requiere prestar una mayor importancia a los criterios de utilización de las técnicas, y a la interpretación de los resultados proporcionados. En este trabajo se ha elegido uno de los programas estadísticos más utilizados por los investigadores sociales de todo el mundo, el SPSS (versión 8) para windows. Las razones que han motivado esta elección son, entre otras, la facilidad de su uso, la explicación de determinados procesos estadísticos en sus menús de ayuda, y la existencia de una versión para estudiantes que posibilita -a un precio muy asequible- que éstos puedan adquirirlo y familiarizarse con él. Deseamos volver a incidir que no es la explicación del programa estadístico la que guiará la exposición, sino que serán las necesidades de la investigación la que aconseje la utilización -y explicación- de determinados procesos de análisis de datos. Respecto a la utilización del programa estadístico, es preciso indicar que todas las menciones a cuadros de diálogo del SPSS, recuadros, etc. aparecerán entrecomillados y en letra cursiva.

La novedad de este texto, a nuestro juicio, está en la carencia en el mercado de una obra que explique a fondo el proceso completo de una investigación social. Pese a que en los últimos años se ha producido una proliferación de publicaciones sobre la utilización de programas estadísticos, así como otros textos que utilizan ejemplos donde se exponen los primeros pasos en el diseño de la investigación (entre otros Soler y Perdíguer 1992, Abascal y Grande 1994, etc.), el texto que aquí presentamos unifica ambos planteamientos al exponer problemas concretos del diseño de la investigación social y las dificultades del análisis de datos. Creemos que hay una importante carencia en el mercado español de un texto que considere conjuntamente estos planteamientos: por un lado los criterios metodológicos de la investigación social, por otro el análisis de datos reales de una investigación, y por último la capacidad de realizar conclusiones y análisis "especiales" utilizando las propiedades de los paquetes estadísticos.

En cuanto a su estructura este libro está dividido en dos partes; 1) *Del problema social al planteamiento metodológico* y 2) *De la metodología al análisis de datos*. La primera parte comienza con una breve exposición del proceso de investigación social, proceso que será utilizado a la hora de plantear los diferentes pasos de nuestra investigación. Una justificación metodológica y la contextualización teórica del fenómeno a investigar permitirá formular las hipótesis de trabajo que guiarán la investigación. Este primer capítulo termina con una exposición de la lógica del análisis de datos en la investigación social.

Planteada la investigación, será necesario elaborar un sistema de medición que permita *verificar* la adecuación de las hipótesis planteadas en la realidad objeto de estudio. La aplicación de un sistema de medición en una investigación social requiere tener en cuenta los problemas unidos a la medición de hechos sociales, así como una evaluación de estas mediciones analizando la fiabilidad y validez de las mismas. Al final del segundo capítulo se expondrán los criterios que han guiado la construcción del cuestionario utilizado en la presente investigación.

El capítulo tres introduce al lector en la tarea de seleccionar un programa estadístico de análisis de datos, al tiempo que se realizan unas pequeñas explicaciones del programa estadístico elegido. El último capítulo de la primera parte está dedicado a la preparación de los datos para ser procesados con un paquete estadístico. Desde que los datos son recogidos hasta que es posible trabajar con ellos es preciso realizar una serie de verificaciones y comprobaciones de la calidad de los mismos. Este capítulo termina explicando las características de los datos recogidos con el cuestionario utilizado.

Estos cuatro capítulos preparan al lector para la exposición de la segunda parte de esta obra, *de la metodología al análisis de datos*, en la que se exponen diversas técnicas de análisis de datos según el número de variables analizadas: técnicas univariantes, bivariantes y multivariantes. El primer capítulo está dedicado a la descripción univariante de cada una de las variables que forman las hipótesis del trabajo. Aprovecharemos determinadas características de las hipótesis para explicar diversas estrategias de transformación de datos, utilización de gráficos, etc.

En el capítulo seis se exponen diversas técnicas que se ocupan de la relación entre variables, aspecto crucial para solucionar determinadas hipótesis. La relación entre dos

variables es también el objetivo del capítulo siete, si bien aquí el interés se centra en la localización y cuantificación de la relación entre las variables.

El análisis de dos variables puede generar, en ocasiones, que se interpreten relaciones entre variables que no existen, al estar generadas por la influencia de terceras variables (influencias espúreas). La localización y eliminación de estas influencias permitirá definir con precisión los factores que inciden en cada una de las conductas planteadas. Más de dos variables utiliza también la técnica expuesta en el penúltimo capítulo, si bien el objetivo de ésta es conocer como determinadas variables permiten "predecir" los resultados de una variable dependiente. El capítulo diez pondrá fin a este libro, exponiendo las conclusiones del trabajo y animando a continuar la investigación utilizando otras técnicas de análisis de datos.

En cuanto al público para el que se ha escrito de este libro, y aunque en un primer momento fue realizado pensando en utilizarlo como manual en un curso de Técnicas de Investigación Social, las sucesivas modificaciones realizadas en los últimos años –tratando de hacer más fácil la comprensión del mismo- lo hacen muy apropiado para todo tipo de personas interesadas en la investigación mediante encuesta: sociólogos, trabajadores sociales, investigadores de mercado, psicólogos, economistas, etc. El libro está escrito en un lenguaje claro y sencillo, y considera que el lector tiene algunos conocimientos previos sobre investigación social; principalmente sobre muestreo y estadística básica. No obstante, dejar claro que este libro es para gente que se inicia, no para especialistas en la materia.