



UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS  
CENTRO DE ESTUDIOS DE OPINIÓN

## LA NO RESPUESTA EN LAS ENCUESTAS

Jean- Jacques DROESBEKE, Université Libre de Bruxelles

Pierre LAVALLÉE, Statistique Canada, Ottawa (Ontario)

Las preguntas no son jamás indiscretas, pero a veces las respuestas lo son. Oscar Wilde. Frases y Filosofía

**Abstract.** In the surveys it happens that it is impossible to obtain answer from the interviewed people for one or more variable than the survey tries to measure. One says then that one is in the presence of “nonanswers”. When the survey uses questionnaires, the lack of values can come from multiple factors, for example, bad ignorance or, will fear survey its of the required subject, the negligence of the poll that does not strive so that each question obtains an answer, the type of question, the length of the questionnaire, etc. In other words, all absence of answer is considered here like nonanswer, anyone is its origin, is this refusal to respond or the loss of the data previously obtained.

**Resumen.** En las encuestas ocurre que es imposible obtener respuesta de parte de los entrevistados para una o más variables que el survey procura medir. Se dice entonces que se está en presencia de “no respuestas”.

Cuando la encuesta utiliza cuestionarios, la falta de valores puede provenir de múltiples factores, por ejemplo, la mala voluntad, el temor del encuestado o su ignorancia del tema requerido, la negligencia del encuestador que no se esfuerza por que cada pregunta obtenga una respuesta, el tipo de pregunta, la longitud del cuestionario, etc. En otras palabras, se considera aquí como no respuesta toda

ausencia de respuesta, cualquiera sea su origen, sea este la negativa de responder o la pérdida de la información anteriormente obtenida.

## **1. Introducción**

En las encuestas, ya se trate de censos o de muestreos, ocurre inevitablemente que es imposible obtener, de parte de algunas de las unidades interrogadas, “valores” para una o más variables que la encuesta procura medir. Se dice entonces que se está en presencia de no respuestas. Cuando los valores se obtienen de sistemas automatizados, la no respuesta puede provenir de un defecto o de una avería. Por otra parte, cuando la encuesta utiliza cuestionarios, la falta de valores puede provenir de múltiples factores. Citemos, por ejemplo, la mala voluntad del encuestado o su ignorancia del valor requerido, la negligencia del encuestador que no se esfuerza por que cada pregunta obtenga una respuesta, la pérdida de un cuestionario, etc. En otras palabras, se considera aquí como no respuesta toda ausencia de respuesta, cualquiera sea su origen, sea este la negativa de responder o la pérdida de la información anteriormente obtenida. En cambio si la pregunta prevé la opción “sin opinión”, la elección de esta opción no es una no respuesta.

Según el tipo de encuesta, a veces es posible buscar los valores ausentes, sea repitiendo las mediciones, sea interrogando nuevamente a las personas encuestadas en cuyos cuestionarios faltan una o más respuestas. Sin embargo, esta manera de proceder acarrea a menudo costos, en recursos y en tiempo, que no siempre pueden ser soportados por la encuesta. En este caso se opta por limitar la repetición a sólo una fracción de las unidades concernidas por las no respuestas. Pero se da también el caso de encuestas en las cuales resulta imposible rehacer las mediciones o entrevistar una segunda vez a la persona. Por ejemplo, esta es excluida la posibilidad de interrogar nuevamente a las personas que, la primera vez, se negaron terminantemente a responder tal o cual pregunta. Puede ocurrir también que la persona concernida se haya mudado o fallecido. En resumen, a causa de las no respuestas, la muestra finalmente obtenida

corresponde raramente a la muestra inicialmente planificada por lo que concibieron la encuesta.

Es importante recalcar que aunque sea físicamente posible ponerse en contacto una y otra vez con una persona que no respondió, hasta obtener una respuesta, los valores así recabados no son necesariamente utilizables por el encuestador. En efecto la calidad de la información obtenida a cualquier precio puede ser tan mala que solo sirva para introducir un sesgo de las estimaciones. Por ejemplo, si una persona no responde porque no conoce el valor requerido, no es útil forzarla a responder, acosarla ciegamente conducirá sin duda alguna a un conjunto de valores erróneos. Cuando se repite una entrevista, es necesario recordar que no se trata únicamente de lograr una respuesta, sino más bien de obtener una “buena” respuesta. En ciertos casos, será más adecuado procesar la no respuesta mediante correcciones estadísticas, que tratar de reemplazar los valores ausentes con datos plagados de errores.

Numerosos artículos y libros se ocupan de la no respuesta. Nos ha parecido interesante proponer una visión de conjunto utilizando el esquema siguiente. Una encuesta puede ser realizada mediante un muestreo o un censo. Este último es menos frecuente y puede ser considerado como un muestreo exhaustivo. Por otra parte, la selección de una muestra puede ser realizada esencialmente de dos maneras.

La primera consiste en elaborar un plan de muestreo al azar, lo que permite estudiar el problema de estimación de parámetros de la población en condiciones a menudo favorables (búsqueda de estimadores insesgados, etc.). En este contexto, el tratamiento de las no respuesta presenta facetas tan diversas como interesantes. Una muestra puede también provenir de un procedimiento empírico, e incluso puede ser construida con ningún procedimiento específico. En éstas condiciones, ya no es posible hablar científicamente de estimación. El tratamiento de las no respuestas, entonces, debe formar parte del análisis de la estructura de

los datos colectados, análisis que, por supuesto, también puede aplicarse al caso del muestreo al azar.

Con el objeto de ser lo más completos posible, hemos optado por situarnos en el contexto más favorable : en el de las respuestas aleatorias. Por otra parte, concentraremos nuestra atención en los tratamientos que permiten cuantificar la no respuesta.

En una primera etapa, haremos un boceto de la encuesta por muestreo aleatorio con el objeto de precisar la notación utilizada y de suministrar, a los lectores poco habituados a trabajar en este contexto, algunos elementos generales de base. En segundo lugar, presentaremos los tipos de no respuesta clasificados según diferentes criterios. Como tercer paso abordaremos la definición de las diferentes tasas de respuesta posibles. Finalmente, estudiaremos el tratamiento de las no respuestas totales y de las no respuestas parciales. Además en cada etapa evocaremos la situación particular de las encuestas repetidas en el tiempo.

## **2. Tipos de no respuesta**

Es posible realizar una clasificación de las no respuesta según diferentes criterios.

### **No respuesta total o parcial :**

La no respuesta total ocurre cuando no se ha podido medir ninguna de las variables de interés. Por ejemplo, la persona encuestada se ha negado simplemente a responder. Cuando la no respuesta parcial ocurre cuando sólo se ha podido medir un subgrupo de variables de interés. La persona encuestada, por ejemplo no conocía la respuesta a una de las preguntas.

### **No respuesta ignorable o no ignorable :**

la no respuesta es ignorable cuando la probabilidad de responder a una pregunta dada no depende del valor de la variable medida. El hecho de que la persona responda o no a una pregunta, no está ligado a la respuesta de esta pregunta.

Un ejemplo de no respuesta ignorable es el caso de la persona que no reexpide el cuestionario sobre la satisfacción de empleados, simplemente por pereza. Un ejemplo de no respuesta no ignorable es el caso donde solo los empleados no satisfechos devuelven el cuestionario. Este último caso tiende evidentemente a sesgar los resultados de la encuesta si los datos son utilizados sin corrección alguna.

### **La respuesta crónica y de “oleada”:**

Esta última clasificación de las no respuesta es específica a las encuestas longitudinales y a las encuestas repetidas. La no respuesta crónica ocurre cuando una persona deja de responder definitivamente a partir de una determinada “oleada” de entrevistas o envío de cuestionarios. Por ejemplo, la persona puede haber dejado la empresa, o bien ya no se la puede encontrar. Notemos que las personas fallecidas son consideradas como realmente encuestadas, pero para quienes la variante  $y_i$  es igual a cero. Las no respuesta de “oleada” ocurre cuando una persona deja de responder a las encuestas de una o varias “oleadas”, pero sólo provisionalmente.

### **3. Tasa de respuesta**

Las tasas de respuesta revisten una importancia particular en las encuestas por muestreo. Éstas pueden servir, por una parte a medir el avance o el éxito de la colecta de datos de una encuesta, y por otra, pueden ayudar a corregir las estimaciones para tomar en cuenta la no respuestas. Así se puede distinguir dos categorías de tasas de respuesta. Así se puede distinguir dos categorías de tasa de respuesta: las tasas de respuesta operacionales y las que calificaremos de tasas de respuesta correctoras.

#### **Tasas de respuesta operacionales:**

Las tasas de respuesta operacionales sirven para medir la importancia de la no respuesta en una encuesta. La idea generalmente aceptada es que las tasas de respuesta corresponden al porcentaje de los que sí han respondido en un

conjunto dado. Pero según Groves [1989], hay varias maneras de definir una tasa de respuesta operacional. De hecho, ocurre que prácticamente cada encuesta tiene su propia definición de la tasa de respuesta, lo que puede engendrar problemas, principalmente de mala interpretación. La definición de la tasa operacional de respuesta depende de cuatro factores ligados al plan de muestreo de la encuesta.

El primer factor corresponde a la opción de incluir o no, en el cálculo de la tasa de respuesta, ciertas unidades de la base de muestreo que no son elegibles para la encuesta. En efecto, ocurren casos en los que solo un subconjunto de las unidades de la base de muestreo sean elegibles para la encuesta. Decidir incluir o no las unidades no elegibles de la base de muestreo, influirá inevitablemente sobre el valor de la tasa de respuesta. Así, en una encuesta sobre la educación nacional, se puede interrogar a los miembros de una universidad incluyendo o no a los encuestados potenciales de nacionalidad extranjera.

El segundo factor depende de que las unidades muestreadas contengan una sola o varias unidades de encuesta. Por ejemplo, encontramos este último caso en las encuestas sociales donde se seleccionan hogares con el fin de entrevistar a personas. La tasa de respuesta puede basarse sobre los hogares o bien sobre las respuestas.

El tercer factor está ligado a las probabilidades de inclusión de las unidades de encuesta, probabilidades que pueden ser desiguales o no. Si las probabilidades son desiguales, el hecho de incluir o no en los cálculos los factores de ponderación del muestreo (es decir, el inverso de las probabilidades de inclusión), influirá sobre el valor de la tasa de muestreo resultante.

Finalmente, el cuarto factor depende de la posibilidad de sustituir una unidad por otra después de la colecta de datos. El tomar en cuenta en la fracción del

muestreo las unidades que no respondieron y fueron sustituidas, antes de obtener una unidad que responda, influirá de nuevo sobre el valor de la tasa de respuesta.

#### **4. Tratamiento de la no respuesta total**

Mencionamos anteriormente que puede ser difícil, incluso imposible, entrar en contacto nuevamente con las personas encuestadas con el fin de obtener los valores que faltan. La presente sección presenta los métodos que pueden servir a una cierta corrección de los efectos nefastos de la no respuesta total.

##### **4.1. Nuevo contacto con los que no respondieron**

Este método propuesto por Hansen y Hurwitz [1946] se parece al muestreo en dos etapas. Con este método se abandona el tratamiento de la no respuesta mediante la utilización de las probabilidades de respuesta. La idea de base de este método es, en primer lugar, de separar la muestra en dos grupos distintos : por una parte los que respondieron (para los cuales se tienen valores medidos por las variables de interés) y por otra parte los que no respondieron. A continuación se extrae una submuestra del grupo de los que no respondieron, para los cuales se procura medir los valores que faltan de las variables de interés.

##### **4.2. Imputación**

La imputación consiste en asignar un valor pertinente a una variable donde falta el valor.

Imputación por la media : Un primer método que se encuentra frecuentemente es la imputación por la media. Consiste en reemplazar el valor que falta por una media calculada a partir de las unidades que respondieron.

El “**hot deck**”: Un segundo método de imputación que puede ser utilizado es el “hot deck”. Consiste en extraer al azar uno de los que sí respondió, llamado en este caso donador. Los valores de la variable del donador son posteriormente asignados a una unidad dada de los que no respondieron.

**El “cold deck”:** Un tercer método es la imputación por “cold deck”. Difiere del hot deck por el hecho de que elige el donador a partir de una fuente externa de datos. Frecuentemente esta fuente corresponde a los datos de encuestas precedentes.

**Imputación múltiple:** Un último método es la imputación múltiple. La imputación múltiple consiste en imputar diferentes valores a cada persona del grupo de los que no respondieron.

### **4.3. Encuestas Longitudinales**

Las encuestas longitudinales plantean un problema particular en el tratamiento de la no respuesta total. A título de ejemplo, supongamos que la encuesta sobre la satisfacción de los empleados sea repetida con los mismos empleados de un año a otro. Se puede así medir la evolución de la satisfacción de los empleados en el transcurso del tiempo.

Se ha visto precedentemente que en las encuestas longitudinales se encuentran dos tipos específicos de no respuesta : la no respuesta crónica y la no respuesta de “oleada”. Estudiaremos los dos casos sucesivamente.

**La no respuesta crónica:** A nivel de la no respuesta crónica, la estimación de las probabilidades de respuesta desempeña un papel primordial, ya que, con el tiempo, las unidades que tienden a quedarse en la muestra tienen particularidades diferentes de las unidades que salen de ella. Por ejemplo, los empleados que cuesta localizar, son generalmente aquellos que cambian frecuentemente de sector de actividad. Por lo tanto, su probabilidad de respuesta es mucho menor que la de las personas con una función estable. Es importante tomar en cuenta este factor cuando se estiman las probabilidades de respuesta. Observemos que la información colectada a lo largo de varias "oleadas", ayuda a la estimación de estas probabilidades. Esto da una cierta ventaja a las encuestas longitudinales en relación a las encuestas puntuales, en las cuales se carece frecuentemente de información sobre las características de los que no responden.



**Las no respuestas de "oleada":** Frecuentemente, las encuestas longitudinales tienen como objetivo medir los cambios en dos "oleadas" de encuesta, de ahí el interés de tratar las no respuesta de "oleada". Podemos interesarnos en medir los cambios globales, comparando las estimaciones a nivel de la población entre dos "oleadas" diferentes. Podemos también interesarnos en medir los cambios (o transiciones) a nivel de las unidades encuestadas.

Imputación histórica: Un método de imputación empleado frecuentemente para el tratamiento de no respuestas de "oleada" es la imputación histórica.

Imputación por el vecino más próximo: Un segundo método es la imputación por el vecino más próximo. Este método se parece en una cierta manera al método "hot deck". Sin embargo, con la imputación por el vecino más próximo el donador no es elegido al azar sino que se toma más bien aquel que se encuentra más próximo, según una distancia calculada utilizando la información de las "oleadas" precedente y siguiente.

Imputación por modelización: Un tercer método de imputación útil para el tratamiento de la no respuesta en las encuestas longitudinales es la imputación por modelización (o imputación con la ayuda de un modelo).

## **5. Tratamiento de la no respuesta parcial**

En comparación con la no respuesta total, la no respuesta parcial requiere un tratamiento particular. El problema es sobre todo de orden práctico. En efecto, para cada variable y donde ocurren no respuestas, podría utilizarse el mismo tratamiento que para la no respuesta total, es decir, corregir las ponderaciones del muestreo de manera a tomar en cuenta la no respuesta.

### **5.1. Imputación**

La imputación en el caso de la no respuesta parcial consiste en asignar un valor plausible a una de las variables cuyo valor está ausente, sin modificar las variables que poseen ya valores.

Imputación por el vecino más próximo: Un método es la imputación por el vecino más próximo. Para la no respuesta parcial se elige el donador tomando aquel que encuentra más cerca según una distancia calculada utilizando la variable conocida.

### **5.2. Encuestas longitudinales**

Como para la no respuesta total, las encuestas longitudinales plantean un problema particular en el tratamiento de la no respuesta parcial.

## **6. CONCLUSIONES**

Hemos presentado algunos métodos destinados a facilitar el tratamiento de las no respuestas. Sin embargo, está claro que cada encuesta es un caso único que puede ser objeto de un tratamiento específico. Por otro lado, como lo hemos precisado al inicio de este artículo, la manera de proceder depende del método de muestro utilizado. Aquí hemos considerado el caso de un muestreo aleatorio. En el caso contrario, la utilización de los métodos de imputación que respetan la estructura de los datos, así como la exploración de la estabilidad de los planos factoriales, constituyen las formas habituales de proceder.