
Opiniones sobre la profesión

Survey sampling: science or divination?

María del Mar Rueda García

Departamento de Estadística e Investigación Operativa
Universidad de Granada

✉ mrueda@ugr.es

Abstract

Sample surveys are one of the most important fields of application of statistics. Sampling is a day-to-day life experience for all of us: we are continually bombarded by opinion polls, which often do not have the right quality. This paper shows some important aspects to consider in conducting surveys and tries to claim the role of the statisticians in surveys.

1. Introducción

La celebración del Año Internacional de la Estadística ha servido para que instituciones, universidades, institutos de investigación, sociedades profesionales y agencias gubernamentales hayan aunado esfuerzos para celebrar la iniciativa y divulgar la importancia de la Estadística en la amplia comunidad científica, usuarios de datos en negocios y en el gobierno, en los medios de comunicación, estudiantes y el público en general. Aprovechando esta coyuntura, la invitación de la editora a escribir este artículo es una oportunidad que se me brindaba para unirme a esta actividad de divulgación, de forma que sirva no solo para difundir, sino principalmente para *reivindicar* el papel tan importante que tienen los estadísticos en la investigación en encuestas. En él quiero poner de manifiesto la necesidad de que éstos sean los que participen en el diseño, ejecución y análisis de una encuesta de calidad, cosa que en los últimos años se está obviando, y no se deje en las manos de sociólogos, psicólogos, politólogos, etc. la realización de encuestas sin un asesoramiento de un especialista en muestreo, cosa que ha llevado a la proliferación de encuestas sin ninguna base científica, de muy poca calidad y utilidad, pero con costes muy elevados para los organismos públicos y privados. Independientemente de esta reivindicación, espero que este artículo le sea de interés al público no especialista en este área concreta de la estadística como es el muestreo en poblaciones finitas.

Comenzaré justificando el título del artículo. Hace casi cuatro décadas la oficina nacional de estadística de Israel hizo una pequeña fiesta para la primera ministra, la señora Golda Meir. Ésta, en su charla de agradecimiento, les dijo a los estadísticos: *ustedes son realmente magos, preguntan a 1000 personas lo que opinan y conocen lo que el país completo piensa* (Pfefferman y Rao, 2009). Hoy día no es la única que considera que los que realizan encuestas son adivinos, y que las encuestas por muestreo no llevan detrás un fundamento científico sino más bien el arte de adivinar.

Adivinos o no, éste es el objetivo principal de las encuestas por muestreo: *conocer acerca de una población a partir de una muestra* (con frecuencia pequeña), tratando temas tales como: *¿cómo seleccionar la muestra?, ¿cómo procesar y analizar los datos?, ¿cómo calcular las estimaciones?, ¿qué hacer con los datos faltantes?, y en función de esto, puesto que no somos magos, también ¿cómo determinar el margen de error en las estimaciones?, dado que siempre hay error.*

Cada día nos vemos bombardeados por información obtenida de encuestas, fundamentalmente encuestas de opinión. Los periódicos y televisiones nos muestran resultados de encuestas sobre la opinión de los españoles acerca del último fichaje del Barcelona, sobre si Rajoy debe dimitir, sobre la tolerancia con los inmigrantes, sobre la intervención militar en Siria o sobre el nuevo peinado de la princesa de Asturias. Sería lamentable creer que el muestreo se limita a esto. Por espectaculares que sean, las encuestas de opinión sólo representan una parte pequeña de las encuestas por muestreo y desde luego suelen ser las peores elaboradas, más controvertidas y menos fiables.

La realización de encuestas por muestreo es una de las áreas más aplicadas en las que trabajan los estadísticos. El estadístico es solicitado, junto con otras personas, para evaluar el tamaño de las poblaciones, los tipos de bienes de los que disponen, y sus comportamientos frente a los problemas económicos y sociales. La actividad conocida como investigación de mercados depende en gran parte del uso del muestreo. Las estimaciones de la magnitud del auditorio de diferentes programas de radio y de televisión y de los lectores de periódicos y revistas se obtienen continuamente. Los fabricantes y detallistas quieren conocer las reacciones de la gente hacia un nuevo producto o nuevos métodos de presentación: sus quejas en relación a los productos antiguos y sus razones para preferir un producto frente a otro. El comercio y la industria utilizan el muestreo en un intento de aumentar la eficiencia de sus operaciones internas. Las actividades de control de calidad y muestreo de aceptación se fundamentan en decisiones que presuponen que los datos de la muestra son válidos para la producción completa. Además de las estadísticas nacionales y de empresas, los muestreos se utilizan en casi todos los campos científicos: ciencias de la educación, la salud, biología, meteorología, ecología, y por supuesto, en los estudios de opinión pública y los sondeos electorales, que han hecho tanto para generalizar la técnica de muestreo frente a los ojos del público.

En los últimos años más y más investigadores importantes en otras áreas de la estadística están moviendo su ámbito al de las encuestas por muestro investigando tópicos como estimación en áreas pequeñas, metodología de marcos imperfectos, datos incompletos, tratamiento de outliers, o métodos de remuestreo. Pero empezamos desde el principio.

2. Los inicios de las encuestas

Las encuestas por muestreo son un fenómeno que se desarrolla rápidamente en el siglo veinte, sobre todo se debió su rápido conocimiento a las encuestas de Gallup y Roper a mediados de los años treinta. En esos momentos la proporción de hogares con teléfono en casa era muy baja y estaba claramente marcada por pertenecer a las familias de altos ingresos. El mismo sesgo debido a altos ingresos se podía observar en la encuesta por correo llevada a cabo por la revista *Literary Digest*, que dependía de los dueños de teléfono y de los registros de automóviles para su elaboración.

Cuando la encuesta *Literary Digest* (con un número mayor de entrevistas) anunció que Landon ganaría a Roosevelt en las elecciones presidenciales, mientras que Gallup y Roper predijeron de forma correcta (con una muestra mucho menor) que Roosevelt sería el ganador, en ese momento se evidenció que el tipo de personas seleccionadas era más importante que el número de personas seleccionadas. El método que aplicaron Gallup y Roper se conoce como *muestreo por cuotas*. Las grandes ciudades se subdividen en subáreas y se les asigna un entrevistador y además existían otros entrevistadores que podían realizar entrevistas en cualquier zona que desearan pero debían seguir un control de hombres y mujeres encuestados en cada área, además del número de individuos con altos ingresos y bajos ingresos en cada área.

Al mismo tiempo, justo después del inicio de la segunda guerra mundial, estadísticos federales dirigidos por Hansen y cuyas funciones eran llevar a cabo una elaboración de encuestas más controladas para las encuestas del gobierno (Hansen, Dalenius y Tepping, 1985) dieron un gran vuelco en la forma de elaborar las encuestas. Estos autores desarrollaron muestreos probabilísticos, donde después de una estratificación se seleccionaban aleatoriamente áreas con probabilidades proporcionales a la población y después se realizaban selecciones con probabilidades iguales de hogares. Este muestreo probabilístico elimina las posibles influencias de los entrevistadores sobre la muestra y hace posible especificar la probabilidad de selección para cualquier hogar en la población y en consecuencia poder estimar el error de muestreo, cosa hasta entonces imposible. Sin embargo Gallup, Roper, y los investigadores de mercado eran muy reticentes a adoptar el muestreo probabilístico de áreas por la sencilla razón del elevado coste que supone si se compara con el muestreo por cuotas, cosa que sigue ocurriendo hoy día.

El principal empuje de los muestreos probabilísticos tuvo lugar después de las elecciones americanas de 1948 cuando la mayoría de las encuestas predijeron de forma errónea una victoria de los republicanos capitaneados por Dewey sobre el otro candidato Truman. Las encuestas fueron duramente criticadas por sus procedimientos en lo referente a la cuota de muestreo empleada, aunque más tarde los análisis demostraron que los mayores errores fueron debidos al paso de las encuestas demasiado pronto lo cual supuso grandes cambios con respecto a los últimos minutos del día de la votación, de tal forma que la preferencia que lideraba Truman no se observó en las encuestas de forma apreciable. Éste fue el hecho que motivó el cambio de tendencia y a partir de este momento los muestreos probabilísticos se convierten en los tipos de muestreo estándar en todos los organismos oficiales de estadística, aunque hoy día muchas empresas de opinión siguen utilizando el muestreo por cuotas y voluntarios, pese a su poca fiabilidad, por su menor coste y sobre todo porque no conlleva el manejo de una teoría estadística compleja.

Hoy día las encuestas sobre la política presidencial americana siguen siendo un buen medio de difusión del trabajo de los estadísticos frente a otros especialistas de encuestas. Véase el ejemplo de Nate Silver, que desde junio predecía una clara, aunque ajustada, victoria de Obama desde su blog. Los periodistas usaban sus artículos para tratar de desacreditar el cálculo estadístico utilizado, pero Silver acertó el resultado en todos los estados (frente a las predicciones de otras macroencuestas muy costosas) y éste éxito sirvió para dar difusión a los modelos estadísticos en la sociedad y poner en evidencia otra vez a las grandes empresas de encuestas de opinión.

3. Pero ¿cómo debe realizarse una encuesta?

Lo primero que comentamos brevemente es la complejidad que conlleva la realización de una buena encuesta por muestreo. El diseño de una encuesta por muestreo comprende varios aspectos íntimamente ligados, ya que el fallo de cualquiera de ellos puede invalidar la encuesta en su totalidad.

Básicamente, para realizar una encuesta, han de seguirse las siguientes etapas:

1. la determinación del objetivo de la encuesta,
2. el estudio de las condiciones generales en las cuales ésta se desarrollará, notando las restricciones a las cuales estará sometido el estadístico,
3. la elección de un sistema de trabajo para la recogida de los datos,
4. la observación de los datos en el terreno o trabajo de campo,
5. el proceso de datos y el control de su calidad,
6. el análisis estadístico y la interpretación de los resultados y

7. la publicación de los resultados obtenidos a partir de la encuesta y la calidad de éstos.

La importancia relativa de cada una de estas etapas depende mucho del tipo de encuesta efectuada, de los datos que hay que recoger, etcétera. La presentación de las diversas operaciones se hace siguiendo su orden natural, según el cual se desarrollarán generalmente. Sin embargo, es frecuente que se produzcan algunos pasos hacia atrás: el sistema de trabajo elegido o los objetivos fijados, pueden no ser compatibles con las restricciones existentes. Entonces hace falta modificar el programa, y cambiar alguna de las etapas anteriores.

El último apartado es esencial, y sin embargo se obvia en la mayoría de las encuestas. La mera publicación de los resultados de una encuesta, no arroja ninguna luz sobre la complejidad de las diversas operaciones que han debido realizarse. La descripción de éstas es la única forma de conocer algo sobre la calidad de la encuesta y, en consecuencia, de la confianza que podemos depositar en las estimaciones.

Dada la variedad de trabajo que conlleva la realización de las encuestas por muestreo es necesaria la colaboración de diversos profesionales desde el principio de la encuesta. Todos tienen su labor y sin la cooperación de cada uno de ellos no se puede obtener una encuesta útil y de calidad. Uno de estos profesionales desde luego es el estadístico, quien debe intervenir en varios apartados, pero es imprescindible en dos: la elección del sistema de trabajo para la obtención de la información y el análisis estadístico.

3.1. La elección de la muestra: el diseño muestral

La elección de un sistema de trabajo adecuado para la recolección de la información es fundamental. Existen muchas técnicas estadísticas más o menos complejas para aplicar a los datos, pero el método más elaborado no tiene validez si se aplica a datos que están mal tomados.

Así una de las líneas de trabajo más importantes para los estadísticos especialistas en encuestas es el estudio de diseños muestrales óptimos. El mecanismo aleatorio para la selección de una muestra de entre todas las posibles muestras que pueden obtenerse de la población se formaliza con el concepto de diseño muestral. Evidentemente todos los muestreos que no sean probabilísticos (muestreos por voluntarios, de cuotas, en el lugar,...) no pueden ser formalizados mediante un diseño muestral y por tanto toda la teoría de muestreo no puede aplicarse a dichos métodos, es decir no hay ninguna teoría que permita determinar para estos métodos la distribución de los estimadores y por tanto sus errores de muestreo. En consecuencia los estadísticos sólo tratamos con muestreos probabilísticos y los otros tipos de muestreo no son considerados. Esto no quiere decir que los muestreos de voluntarios o de cuotas no tengan ninguna utilidad: sí pueden usarse para estudios enfocados a detectar problemas o anomalías, influencias de algunas variables en otras en la muestra considerada,..., pero no

sirven para hacer inferencia, es decir sus resultados no pueden extrapolarse a toda la población.

Concretamente un diseño muestral d en una población finita \mathcal{U} , es un par $d = (S_d, P_d)$, donde S_d es un subconjunto del espacio muestral universal S y P_d es una función de probabilidad definida en S_d tal que:

1. $P_d(s) > 0 \forall s \in S_d$, y
2. $\forall u \in \mathcal{U} \quad \exists s \in S_d \mid s \ni u$.

Existen infinitos diseños muestrales sobre una población, tantos como funciones de probabilidad distintas pueden definirse sobre el conjunto de partes de \mathcal{U} .

La elección de un diseño muestral adecuado es fundamental pues las propiedades de los estimadores que luego se usen derivan del diseño muestral elegido, pues el diseño es el que determina la distribución de los estimadores y estadísticos que se utilicen posteriormente. El diseño muestral más básico es el muestreo aleatorio simple en el que todos los individuos tienen la misma probabilidad de ser elegidos. Este diseño es el más fácil de implementar y además los estimadores y errores de muestreo a usar para este tipo de diseño tienen un cálculo sencillo pero los errores suelen ser grandes en relación a otros tipos de diseños más complejos y sólo se utilizan en encuestas muy básicas. Los diseños muestrales estratificados y por conglomerados multietápico fueron desarrollados con el objeto de minimizar el coste de la encuesta a la vez que se controlaba la incertidumbre asociada a las estimaciones de las variables principales. Estos diseños fueron motivados por los procedimientos que se usaban en los años 1930 y 40 para encuestas de población y sus combinaciones siguen usándose en algunas encuestas oficiales, por ejemplo en la EPA o en la Encuesta nacional de salud, donde se utiliza una estratificación por provincias y conglomerados de hogares obtenidos en 3 o más etapas. Sin embargo, la existencia de nuevos métodos para recolección de datos, la disponibilidad de más información de las unidades en la población así como los nuevos usos de las encuestas conllevan una demanda de nuevos diseños muestrales.

- La existencia de información auxiliar de alta calidad permite construir diseños muestrales más complejos y eficientes. Por ejemplo, en una encuesta de salud interesa tener una representación alta de personas con riesgo alto de enfermedades coronarias, pero se desconoce cuáles de estas personas lo tienen antes de que la muestra sea seleccionada. Una muestra se obtiene en una primera fase en donde se les pregunta a los entrevistados por teléfono acerca de factores de riesgo y posteriormente se estratifican en función del riesgo. A la segunda submuestra obtenida entre los entrevistados originalmente se les realizan un examen médico con una afijación mayor entre el

estrato con riesgo alto. La ganancia usada mediante el muestreo en dos fases depende del coste relativo del muestreo en cada fase y de la eficiencia en la estratificación de la primera fase. Para los diseños muestrales en varias fases los estimadores sencillos no sirven, hay que formular estimadores específicos que combinen la información en cada fase y cuya distribución es muy compleja de estimar ya que la distribución de probabilidad en cada fase influye en las siguientes. Esperanzas y varianzas condicionadas tienen aquí un uso frecuente.

- Con la introducción de nuevos métodos de recogida de datos es difícil en algunas situaciones poner en práctica los diseños muestrales clásicos pues el marco no incluye la población completa. Por ejemplo los marcos obtenidos a partir de llamadas a dígitos aleatorios (Random Digit Dialing) no incluye hogares sin teléfono. Las encuestas telefónicas se pueden hacer a partir de teléfonos fijos y móviles, pero hay personas que disponen de los dos, por lo que los marcos no son incompatibles. Así se ha desarrollado la teoría de encuestas con marcos múltiples a partir de la unión de dos o más muestras obtenidas de distintos marcos que en su conjunto cubren a toda la población. El diseño de encuestas con múltiples marcos necesita incluir diseños adecuados para cada marco así como modificar los estimadores pues hay una sobreestimación importante de los parámetros.

Las encuestas en las cuales es importante ver la evolución de ciertas características en el tiempo (p.e. el cambio en la tasa de empleo de un mes a otro) tienen que diseñarse de forma especial. Las encuestas repetidas de sección cruzada sirven así para estimar el cambio en el desempleo de 2012 a 2013 pero no pueden usarse para responder preguntas relativas a la persistencia en el desempleo entre los individuos. En este caso se utilizan encuestas longitudinales para estimar la tendencia anual y la persistencia, pero necesitan consideraciones especiales en cuanto a la precisión y medidas de error que pueden variar con el tiempo.

- Las diferentes tipos de poblaciones investigadas también necesitan diferentes tipos de diseños muestrales.

El muestreo de unidades poco frecuentes exige procedimientos especiales. Se tiene una población de unidades poco frecuentes (Kalton y Anderson, 1986), que se supone inmersa o incluida en otra población mucho más numerosa por ejemplo, personas con una enfermedad muy rara o un tipo especial de especie en extinción en una región. Obtener estimaciones con precisión satisfactoria suele ser en general muy costoso, y puede llegar a ser prohibitivo.

Veamos el problema con un ejemplo. Supongamos que queremos estimar una proporción P cuyo valor es 0.02 con una muestra aleatoria simple de

tamaño 800, el error máximo admisible relativo en este caso sería 50 %, lo cual es muy alto. Si fijamos el error relativo en un 5 % el tamaño necesario sería $n=80000$. La necesidad de tomar una muestra tan grande indicaría la conveniencia de buscar procedimientos alternativos que permitieran obtener los resultados requeridos con un coste menor. Así aparecen métodos como:

- el muestreo inverso (Cochran 1977) en el que se van tomando unidades hasta que haya un mínimo de unidades con la característica deseada,

- el muestreo en bola de nieve (Goodman, 1961) que consiste en seleccionar una muestra inicial o básica de individuos y establecer en cada entrevista qué nuevos individuos de la población en estudio han de entrevistarse para constituir la muestra completa y que se utiliza para estudios de inmigrantes, minorías étnicas, grupos religiosos, personas con dificultades o marginados.

- el muestreo de nexos , en el cual cada elemento puede estar asociado a más de una unidad de la lista (marco) de selección (por ejemplo un enfermo que ha sido tratado en varios hospitales, siendo las unidades de muestreo los hospitales)

- el muestreo adaptativo cuyo propósito es utilizar como información el patrón espacial de la población. Este método comienza seleccionando una muestra inicial de cuadrados y cuando en un cuadrado contiene el organismo de interés se añaden a la muestra los cuadrados en la vecindad del cuadrado originalmente seleccionado

- el muestreo de poblaciones evasivas (por ejemplo animales en libertad) también necesita diseños muestrales específicos (Thompson, 2012). Se han desarrollado muchos métodos de estimación del número de animales, pero casi todos ellos se pueden ver como variantes o combinaciones de las tres aproximaciones básicas siguientes que:

(1) usan la variación en el número de detecciones cuando alguna característica observable del estudio o de los individuos varía. Entre ellas se encuentran los muestreos de distancia y muestreos por parcelas.

(2) usan la variación en el número de detecciones cuando la población se reduce por eliminación de individuos. Entre ellas tenemos los métodos de eliminación, captura-esfuerzo y cambio de la razón.

(3) usan la proporción de individuos que son recapturados después de ser capturados y marcados. Un ejemplo son los métodos de marca-recaptura.

3.2. La estimación: diferentes tipos de inferencia

Una vez determinado el diseño muestral y seleccionada la muestra de acuerdo a ese diseño, el siguiente paso es la formulación de resultados acerca de la

población completa en base a la parte seleccionada. En este proceso el estadístico debe guiarse por el conjunto de principios y procedimientos de la Inferencia Estadística.

En el estudio de la inferencia en poblaciones finitas existen dos líneas fundamentales de desarrollo basadas en:

- el punto de vista prevalente en los estadísticos dedicados al diseño y ejecución de las encuestas por muestreo, de poblaciones finitas, reales y efectivas, en las que el elemento estocástico se debe al mecanismo de aleatorización que determina qué parte de la población se observa, y
- el punto de vista según el cual las muestras proceden de poblaciones que son realizaciones de superpoblaciones o poblaciones infinitas, previamente especificadas, tal y como se venía haciendo en la estadística matemática anterior al muestreo de poblaciones finitas, y en las que el elemento estocástico es introducido por el modelo probabilístico que fijamos a la población.

De acuerdo con estos puntos de vista, podemos considerar en relación con las encuestas por muestreo, dos tipos de inferencia: la inferencia en poblaciones fijas, y la inferencia en superpoblaciones o modelos.

La teoría clásica en muestreo de poblaciones finitas asume que la población finita que se investiga es una *población fija*. Más explícitamente, se trabaja en el contexto de una población que consta de N unidades bien identificadas y una variable de estudio y con valores fijados, a priori desconocidos, pero que el estadístico tiene acceso al verdadero valor y de la i -ésima unidad mediante una encuesta.

Cuando se trabaja con poblaciones reales algunas de estas consideraciones básicas no están totalmente justificadas y hay que proceder a una revisión de los problemas de inferencia y de los métodos. Además, desde un punto de vista teórico la aproximación por poblaciones fijas no permite establecer estrategias óptimas. La no existencia de estimadores uniformemente de mínima varianza de la media en la clase de estimadores insesgados, en la clase de estimadores lineales insesgados y en la clase de estimadores lineales homogéneos para diseños no unicluster (Hedayat y Sinha, 1991) motivó a los estadísticos para la construcción de otra modelización.

Dadas dos diseños muestrales y los estimadores asociados, generalmente no se puede establecer ningún resultado acerca de cuál de ellos es *mejor*. Entonces la formulación de un modelo que estructure los valores de la variable en estudio, puede resultar positivo para concluir algunos resultados. Esta es una de las razones para proponer modelos de superpoblación en el estudio de la comparación de estrategias muestrales para poblaciones finitas (Valliant et al. 2000). Ya Cochran (1946) propuso una estructuración de los valores poblacionales para poder comparar la eficacia del muestreo sistemático frente al muestreo aleatorio simple y al

muestreo estratificado, sin embargo hasta los trabajos de Royall en los años 70 no se introdujo este enfoque de forma matemática. Aunque el desarrollo de modelos de superpoblación es reciente, éstos fueron utilizados en los años cuarenta por diversos autores como Cochran (1946) o Deming y Stephan (1941).

Según esta modelización, los valores de la variable de interés se consideran como variables aleatorias Y_1, \dots, Y_N que tienen cierta distribución conjunta $\xi(Y_1, \dots, Y_N)$ y los valores actuales (desconocidos a menos que se haga una encuesta) y_1, \dots, y_N constituyen una realización de las variables Y_1, \dots, Y_N , respectivamente. La distribución $\xi(\cdot)$ va a determinar el modelo de superpoblación.

El modelo de superpoblación proporciona un resumen de nuestro conocimiento a priori de la naturaleza de la variable a estudiar en referencia a la población. En las encuestas es normal que se disponga de información auxiliar respecto a alguna variable relacionada con la variable objeto de estudio. Por ejemplo, características demográficas y socioeconómicas como sexo, edad, categoría profesional, o incluso observaciones de la misma variable de estudio pero en algún tiempo anterior. Esta información auxiliar puede incorporarse de forma natural mediante los modelos. Así en los últimos años se han considerado desde los modelos más básicos como el de razón ($Y_k = \beta x_k + e_k$, con la varianza del error e_k proporcional a una cierta potencia de la variable auxiliar, $v(e_k) = \sigma^2 x_k^\delta$) a modelos más complejos como los de regresión, regresión estratificada, modelos de conglomerados, etc. (Isaki y Fuller, 1982) que intentan incorporar la información auxiliar así como la estructura de la población, y dan lugar a estimadores óptimos bajo el modelo concreto de superpoblación considerado (estimadores de regresión generalizado, diferencia generalizado,...). Las técnicas de regresión no paramétrica también se aplican en esta modelización mediante modelos que asumen una relación entre las variables dada por una función de la que no se sabe la forma

$$Y_i = m(x_i) + e_i, \quad i = 1, \dots, N,$$

y esta función se aproxima usando técnicas de regresión polinómica local, spline, MARS,... (Breidt and Opsomer, 2000).

En los últimos años, (Opsomer y Miller, 2005), una cuestión muy debatida entre los especialistas en muestreo en poblaciones finitas es el uso o no de modelos de superpoblación (Little, 2004). En el muestreo en poblaciones finitas la experiencia de muchos años y trabajos dice que la inferencia debe estar basada fundamentalmente en la distribución resultante de la aleatorización de la muestra más que en los modelos estadísticos. Hoy día sigue existiendo una gran controversia entre los partidarios de la inferencia basada en el diseño y los partidarios de la inferencia basada en el modelo.

La inferencia basada en el diseño toma en cuenta la información proporcio-

nada por el diseño muestral y proporciona buenas estimaciones para grandes muestras, sin necesidad de imponer fuertes restricciones de modelización de las variables. Por otro lado, es esencialmente asintótica y entonces su uso para muestras pequeñas puede no ser adecuado. También la inferencia basada en modelos tiene sus ventajas: proporciona una aproximación unificada para la inferencia más de acorde con las principales teorías estadísticas en otras áreas, permite utilizar la metodología bayesiana (Bolfarine and Zacks, 1992, Mukhopadhyay, 2001), que no tiene cabida en la aproximación clásica y, si el modelo está bien especificado, puede dar lugar a estimaciones más precisas que las basadas en el diseño. Esto suele ocurrir en los estudios en dominios o áreas pequeñas (Rao, 2005) donde hay más información auxiliar de la población y recursos para su estudio con lo que se puede definir y comprobar bien el modelo de superpoblación.

La mayor debilidad de la inferencia basada en el modelo es que si el modelo está mal definido, puede dar lugar a inferencias mucho peores que las basadas en el diseño. La no robustez de los estimadores basados en el modelo (Chambers y Dunstan, 1986) es una de las razones utilizadas para preferir la inferencia basada en el diseño.

En los últimos años hay intentos de reconciliar ambas teorías. Así ha aparecido un nuevo enfoque: la inferencia basada en el diseño, modelo asistida (Särndall, et al., 1992; Särndall, 2007)

En el enfoque modelo asistido la inferencia está basada en el diseño igual que en la aproximación clásica (el objetivo es minimizar el error cuadrático medio bajo el diseño) pero se utilizan modelos de superpoblación para motivar la elección del estimador; en particular los estimadores clásicos que incorporan información auxiliar (como los estimadores de razón y de regresión) pueden motivarse utilizando modelos lineales de superpoblación.

En los últimos años han aparecido dos metodologías importantes para el uso de información auxiliar: el método de calibración introducido por Deville y Särndall (1992) y el de verosimilitud empírica (Chen y Qin, 1993). Estos estimadores con buenas propiedades teóricas y prácticas son modelo asistidos pues utilizan modelos para explicar la relación entre la variable objeto de estudio y la variable auxiliar.

3.3. El error en la estimación

Independientemente de la modelización seguida, una vez elegido el estimador óptimo para el parámetro se calcula la estimación a partir de la muestra seleccionada. Esta estimación fluctúa según las diferentes muestras susceptibles de ser seleccionadas, y conllevan así un error al estimar el parámetro llamado error de muestreo. A primera vista puede parecer que los errores de muestreo imponen una seria limitación a la utilización de muestras. Esta primera impresión es por completo injustificada, ya que el error de muestreo es sólo una parte del error total, y además se dispone de métodos para controlarlo y medirlo.

Existe un procedimiento científico que permite establecer a priori el error de muestreo que estamos dispuestos a tolerar y elegir la muestra de tal forma que permita afirmar, prácticamente con certeza, que las conclusiones que hagamos sobre los valores de la población no vendrán afectadas por errores de muestreo superiores a los límites previamente establecidos. La determinación de una expresión explícita para el error de muestreo que se exprese en función de parámetros poblacionales sencillos de estimar a partir de la muestra, no siempre tiene solución. Sólo para los diseños muestrales más sencillos (mas, muestreo estratificado, por conglomerados) y parámetros sencillos (medias, totales o proporciones) suelen obtenerse estas expresiones. En el caso de parámetros complejos (funciones de distribución, quantiles, varianzas, coeficientes de variación, ...) o para diseños muestrales complejos hay que acudir a procedimientos de aproximación. Además de las técnicas de linealización, surgen nuevos métodos de remuestreo específicos de poblaciones finitas que se definen teniendo en cuenta el procedimiento por el que se ha seleccionado la muestra obtenida. (p.e. grupos aleatorios, semimuestras, jackknife, bootstrap paramétrico y no paramétrico,...)

La teoría de muestras asume que la variable de interés es medida en cada unidad de la población sin error, entonces los errores en las estimaciones se deben exclusivamente al muestreo. Además de los errores de muestreo, existen otros errores ajenos al muestreo que se presentan tanto en las encuestas como en los censos, aunque con mayor intensidad y menos posibilidades de corrección en estos últimos. Entre ellos están los errores de observación, de cobertura, de medida, de procesamiento y sobre todo los debidos la falta de respuesta.

La falta de respuesta ocurre en casi todas, sino a todas, las encuestas. El efecto de la no-respuesta es doble: sesgo en la estimación y crecimiento en el error de muestreo al disminuir el tamaño muestral. El principal objetivo de tratar la no-respuesta es la reducción del sesgo de no-respuesta, que ocurre si los que responden y lo no respondientes son diferentes respecto a las variables en estudio.

Existen muchos procedimientos para tratar la falta de respuesta, tanto en la fase de recogida de los datos (Encuestas repetidas, Encuesta delegada, Sustitución de unidades, uso de incentivos, Método de Hansen y Hurwitz, Método de Politz y Simmons, respuesta aleatorizada,...) así como en las etapas de su procesamiento y análisis (estimación basada en datos faltantes, imputación, ponderación,...).

Aunque la imputación múltiple está ganando popularidad en los institutos nacionales de estadística, la inmensa mayoría de encuestas utilizan alguna forma de imputación simple, tanto determinística (imputación por razón, regresión, de la media, de valores auxiliares, vecino más próximo) que tienden a distorsionar la distribución de variables imputadas, como estocásticos (regresión aleatoria, hot-deck...) que tienden a preservar la distribución de la variable imputada pero añaden una componente adicional a la varianza debida al mecanismo aleatorio

de imputación. Ahora bien, la imputación, siendo el principal método usado para tratar la falta de respuesta, tiene también muchos inconvenientes y en algunas agencias, como en la oficina de estadística de Suecia está prohibida por motivos legales. Algunas importantes agencias oficiales de estadística utilizan como alternativa para el tratamiento de datos faltantes la reponderación por calibración (Lundström y Särndal, 2001).

4. Las encuestas de opinión en nuestro país

Continuamente vemos publicados en los periódicos encuestas que se presentan al público general como una verdad absoluta pues una gran parte de este público no está preparado para evaluar la calidad de la encuesta y poder medir así la confianza que puede depositar en los resultados. Las encuestas por muestreo sólo muestran con exactitud la opinión de los individuos preguntados, y la extrapolación de los resultados a toda la población conlleva siempre errores que son imposibles de cuantificar en su totalidad. Además la inmensa mayoría de encuestas de opinión ni calculan ni publican los errores muestrales (y mucho menos los no muestrales) por ello cada individuo puede dar una credibilidad distinta a los resultados obtenidos en esta encuesta de opinión, siempre subjetiva.

A continuación se presentan algunos ejemplos de encuestas de opinión que se hacen en nuestro país.

Los barómetros del Centro de Investigaciones Sociológicas (CIS) son las encuestas de opinión que, con periodicidad mensual, se realizan para hacer el seguimiento de las opiniones y actitudes de la sociedad española ante distintas situaciones o acontecimientos de la actualidad. En la página web del CIS, aparece un apartado de metodología en el que se sólo se describen los indicadores utilizados. Sin embargo no aparece ninguna ficha técnica, ni información acerca del diseño muestral usado, el estimador empleado, el error de muestreo, los procedimientos usados para tratar la falta de respuesta, los posibles sesgos,...

En contraposición tenemos las encuestas realizadas por el Instituto Nacional de Estadística (INE). Por ejemplo, en la Encuesta de Población Activa aparece un documento con una descripción bastante amplia del marco, diseño muestral, estimadores, errores de muestreo, de no muestreo,.... También el Instituto de Estudios Sociales Avanzados (centro público de investigación científica especializado en Ciencias Sociales) elabora muchas encuestas de opinión en cuyas fichas técnicas aparece información relativa a los muestreos probabilísticos usados, estimadores, errores, etc., aunque no tan detallada como en el caso de las encuestas del INE.

Aparte de los organismos oficiales, cualquier medio de comunicación ofrece continuamente encuestas de opinión. Por ejemplo la página del diario el País, en la que dicen expresamente que no es una encuesta científica y sólo representa a los individuos que voluntariamente quieren votar.

Efectivamente las encuestas de voluntarios no tienen ninguna base científica. No es un muestreo probabilístico y además de no poder calcular los errores de estimación pueden dar lugar a estimaciones muy sesgadas. Un ejemplo significativo es la encuesta realizada sobre acoso laboral en la Universidad de Granada. Dicha encuesta se realizó mediante entrevistas a voluntarios por internet. En la breve ficha técnica aparece un error muestral imposible de determinar: lo que han hecho es aplicar la fórmula correspondiente al estimador de la proporción muestral si se hubiera obtenido una muestra aleatoria simple (lo cual no es así). El sesgo cuando los resultados de una encuesta de voluntarios se extrapola a toda la población es muy grande, debido a que las características de las personas que intervienen voluntarias en este estudio difieren mucho de las que no participan. Como resultado de dicha encuesta se afirma que el 60,44 % de los trabajadores de dicha universidad confirman el acoso laboral. Este resultado refleja la opinión de los individuos que voluntariamente han participado en el estudio (que evidentemente están muy motivados por el tema del acoso laboral) pero no refleja el sentimiento general que tenemos los trabajadores de esta universidad.

Ejemplos de éstos tenemos miles desgraciadamente. Y es que hoy día cualquiera se encuentra capacitado para elaborar una encuesta y el estudio estadístico correspondiente. Una parte de la culpa la tiene la facilidad del manejo de programas para la elaboración y tratamiento de datos estadísticos que son de amplia difusión. Son muchos los que saben hoy meter datos y manejar el SPSS, el S-Plus, el R-Commander,... pero muy pocos saben que la mayoría de los procedimientos estadísticos que tienen implementados sólo sirven si las unidades han sido seleccionadas aleatoriamente y mediante diseños muestrales sencillos.

Por otra parte, hoy está muy de moda la realización de encuestas puesto que se utilizan como instrumento de investigación en todas las áreas de las ciencias sociales. Así vemos en la mayoría de los planes de estudios contenidos de muestreo e inferencia estadística que se han incorporado dentro de asignaturas con nombres enmascarados y que se imparten por profesores que no tienen formación estadística ninguna, pero que no tienen problema en explicar los diseños muestrales (aunque desconozcan qué significa este término), tamaños de muestra (aplicando una fórmula que sólo sirve para muestreo aleatorio simple y para estimar proporciones) y a qué tecla darle para que salgan los distintos test con el SPSS (aunque no tenga sentido esa técnica en ese problema concreto). El ser usuario de las encuestas no significa ser un experto para su impartición. El desconocimiento de las bases científicas de la estadística por muchos de los profesores que actualmente enseñan encuestas en la universidad es una de las principales razones en la proliferación de encuestas sin ningún rigor científico con el que constantemente nos encontramos.

Por último quisiera mostrar que esta situación no ocurre solo en nuestro país. La encuesta que voy a ilustrar a continuación es el eurobarómetro 151 titulada: Irak y la paz en el mundo. El conjunto de ítems que contempló versa

sobre diversos elementos de opinión acerca del conflicto en Irak, el papel de diferentes países y regiones en el asunto y qué se esperaba de la Unión Europea. Uno de los ítems planteaba a la persona entrevistada que señalara qué países consideraba que constituyen una mayor amenaza a la paz en el mundo. Entre otros aspectos, los resultados señalaron a EEUU y, principalmente, a Israel como los dos países que los habitantes de Europa consideraban más peligrosos. Tal circunstancia levantó una polvareda de opiniones encontradas en los más diversos foros, incluyendo disculpas políticas sin antecedentes. La polémica suscitada se basa en interpretaciones subjetivas de la encuesta y no en información resultante de la propia encuesta. Algunas de las afirmaciones que se han realizado: "Los resultados son sesgados y antisemitas", "brote de antisemitismo combinado con un incomprensible antiamericanismo."

La encuesta fue realizada por la empresa EOS Gallup Europa, con gran experiencia en el campo de las encuestas de opinión. Ahora bien, a la hora de evaluar nuestra confianza en los resultados presentados que están basados en la opinión de una muestra concreta (de aproximadamente 500 personas de cada país) hemos de releer y analizar seriamente el informe técnico. El informe técnico presentado en el eurobarómetro es escaso y no del todo clarificador y presenta grandes carencias y aspectos confusos, como por ejemplo:

- No se indicó cómo se construyó el marco. Puesto que la encuesta se realizó telefónicamente, parece lógico pensar que el marco se elaboró a partir de las guías de teléfonos de cada país, lo cual conllevaría importantes errores en la cobertura de la encuesta.
- En ninguna parte del informe se indica cómo se seleccionaron las unidades de la muestra. Ni siquiera dice si el muestreo fue probabilístico o no. Pueden haber utilizado algún tipo de muestreo no probabilístico como el muestreo de cuotas o bien el de unidades representativas.
- Suponiendo que fuera un muestreo probabilístico, parece lógico pensar que hubieran realizado un muestreo estratificado aleatorio. Ahora bien ¿por qué han repartido uniformemente la muestra entre los estratos (naciones)? Cualquier alumno de algún curso básico de muestreo sabe que es el peor reparto (afijación) que se puede hacer, pues en la muestra se le da la misma importancia a países como Alemania y Luxemburgo, cuando la población de la primera es 186 veces mayor que la segunda. En Luxemburgo una muestra de 500 personas pudiera ser representativa, pero en Alemania resulta a todas luces insuficiente, siendo el error de muestreo mayor.
- Un aspecto fundamental en la planificación de la encuesta es la determinación del tamaño muestral. Tampoco aparece en el informe con qué criterio se ha seleccionado el tamaño muestral. Si el criterio utilizado ha sido el

habitual de fijar un máximo para alguna medida del error de muestreo, ¿en cuánto se ha fijado dicho error máximo a cometer en las estimaciones?

- En cuanto al proceso de estimación, no se indica expresamente el estimador que se ha utilizado para las proporciones de cada nación ni para las proporciones en toda la Unión Europea.
- Otro aspecto muy importante a resaltar es que en el informe no aparece mención alguna a los errores ajenos al muestreo, que suelen ser mayores y más difíciles de controlar en las encuestas que los errores de muestreo. ¿Cuál ha sido la tasa de falta de respuesta? ¿Qué procedimiento se ha utilizado para la sustitución de las unidades? ¿Cómo se ha tratado la falta de respuesta a una cierta pregunta: se han imputado los datos o se han suprimido las unidades con algún campo faltante? ¿Cuál ha sido entonces el tamaño muestral efectivo en cada pregunta? ¿Qué acción se ha tomado en cada caso?

Así pues, después de la lectura del informe técnico (que se supone que es el completo dirigido a expertos y no tiene ni siquiera la información que se le exige al informe resumido), se sacan muy pocas cosas en claro acerca del diseño, realización y tratamiento de esta encuesta. Probablemente la encuesta esté bien planteada y ejecutada, pero la información que la empresa da acerca de ella es insuficiente y confusa, de forma que a nosotros (que se nos supone expertos) nos es difícil decidir en base al informe, qué confianza depositar en los resultados publicados. La credibilidad puede darse en función de otros criterios (por ejemplo, la experiencia en encuestas de opinión de la empresa ejecutora, la confianza en la decisión de la Comisión Europea en encargar la encuesta a la empresa más preparada,...) pero estos no son criterios científicos. Esta encuesta es, pues, un buen ejemplo de que no sólo hay que hacer bien las encuestas de opinión, sino que además hay que dar información amplia a los ciudadanos de cómo se han realizado.

5. ¿El futuro?

Actualmente, los avances metodológicos en muestreo en poblaciones finitas, la experiencia acumulada en redacción de cuestionarios, el mejor adiestramiento en los entrevistadores, el perfeccionamiento de los procesos de tabulación, de depuración y análisis de datos y los medios informáticos disponibles, hacen que las encuestas de opinión sean un instrumento de análisis útil e imprescindible para gobiernos democráticos que necesitan conocer las opiniones de sus ciudadanos. En el siglo XXI disponemos de conocimientos y medios técnicos que garanticen la realización de buenas encuestas que puedan reflejar algunas opiniones de la ciudadanía. Lo fundamental es utilizar a los expertos en cada una de estas áreas conformando un grupo de trabajo en el que participen estadísticos (que diseñen

el muestreo, y realicen el análisis de datos), informáticos (que supervisen los medios informáticos para adquisición, codificación y almacenamiento de datos), psicólogos (para la elaboración de cuestionarios adecuados, formación de entrevistadores,...) así como expertos en el ámbito específico de investigación de la encuesta (que determinarán los objetivos de la encuesta, los medios e información disponible de la población a investigar y cómo utilizar los resultados derivados de la encuesta).

Pero se puede avanzar todavía más en la realización de encuestas de calidad, y aquí el papel del estadístico es fundamental mediante la investigación. En el área de muestreo, en los últimos años han aumentado considerablemente las publicaciones científicas, fundamentalmente desde la aparición de la metodología de modelos de superpoblación, que ha permitido incluir metodologías bien desarrolladas en la inferencia clásica (como los modelos lineales, la inferencia bayesiana,...) al caso del muestreo en poblaciones finitas. Un ejemplo claro es el ámbito de las áreas pequeñas, que ha dado lugar a un gran número de publicaciones en los últimos años, muchas de ellas de grupos de investigadores de nuestro país.

Sin embargo creo que el gran reto que tiene la investigación en muestreo viene de la mano de las nuevas tecnologías. Internet ha tenido también un efecto importante en el negocio de la realización de encuestas. El rápido desarrollo de encuestas a través de la WWW ha hecho que hoy día las encuestas web reemplacen en muchos casos a los tradicionales métodos de recogida de datos. Las encuestas web son una forma muy sencilla de tener acceso a un grupo grande de personas con un coste muy reducido, de forma muy rápida y que proporciona nuevas posibilidades (como el uso de material multimedia). Sin embargo este tipo de encuestas web de voluntarios tiene algunos problemas metodológicos importantes como la falta de cobertura de la población, los sesgos de autoselección, los problemas de construcción del marco, etc. que hacen muy difícil, sino imposible, hacer inferencias válidas acerca de la población investigada. Sin embargo, el gran potencial de este tipo de encuestas no se puede obviar. Se hace pues necesario dar nuevos procedimientos que resuelvan los problemas metodológicos que desde el punto de vista estadístico tienen estas encuestas. Ha habido alguna idea muy novedosa para dar una estructura probabilística a este tipo de encuestas, pero desgraciadamente esta investigación no ha obtenido financiación: cosas de la crisis.

Por último, comentar que hoy día el futuro puede venir de la realización de encuestas con diversas formas de recolección de los datos (teléfonos fijos, móviles, entrevista personal, encuestas web,...) que combinen las ventajas de cada procedimiento y se necesita por tanto investigar acerca de nuevas metodologías para integrar los datos obtenidos mediante los distintas formas de encuestación de forma que se aumente la calidad de los resultados.

Referencias

- [1] Bolfarine, H.I., Zacks, S. (1992). *Prediction Theory for Finite Populations*. Springer. New York.
- [2] Breidt, F.J. and Opsomer, J.D. (2000). Local polynomial regression estimators in survey sampling. *The Annals of Statistics*, **28**(4), 1026-1053.
- [3] Chambers, R.L. and Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika*, **73**, 597-604.
- [4] Chen, J. and Sitter, R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica*, **9**, 385-406.
- [5] Chen, J. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, **80**, 107-116.
- [6] Cochran, W. G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *Annals of Mathematical Statistics* **17**, 164-177.
- [7] Cochran, W.G. (1977). *Sampling Techniques*. Wiley: New York (USA).
- [8] Deming, W. E. and Stephan, F. F. (1941). On the interpretation of censuses as samples. *Journal of the American Statistical Association*, **36**, 45-49
- [9] Deville, J.C. and Särndal, C.E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, **87**, 376-382.
- [10] Goodman, L.A. (1961). Snowball sampling. *Annals of Mathematical Statistics*, **32**, 148-170.
- [11] Hansen, M.H., Dalenius, T. and Tepping, B.J. (1985). The development of sample surveys of finite populations. En Atkinson and Fienberg (eds), *A Celebration of Statistics: The ISI Centenary Volume*, 327-354. Springer-Verlag. New York.
- [12] Hartley, H. and Rao, J. (1968). A new estimation theory for sample surveys. *Biometrika* **55**(3), 547-557.
- [13] Hedayat, A.S. and Sinha, B.K. (1991). *Design and Inference in Finite Population Sampling*. Wiley: New York (USA)
- [14] Isaki, C. and Fuller, W. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, **77**(377), 89-96.

-
- [15] Justicia, F. El Acoso Laboral en la Universidad de Granada. En: www.ugr.es/ccoo/informeacosouniversidad.pdf
- [16] Kalton, G. and Anderson, D. (1986). Sampling rare populations. *Journal of the Royal Statistical Society. Series A*, **149(1)**, 65–82.
- [17] Kish, L. (1995). The hundred years wars of survey sampling. *Statistics in Transition* **2**, 813–830.
- [18] Little, R.J.(2004). To model or not to model? Competing modes of inference for finite population sampling. *Journal of the American Statistical Association* **99(446)**, 546–556.
- [19] Lundström, S. and Särndal, C.E. (2001). *Estimation in the presence of Non-response and Frame Imperfections* SCB. Statistics Sweden, Örebro.
- [20] Mukhopadhyay, P. (2001). *Topics in Survey Sampling* Springer. New York.
- [21] Opsomer, J.D. and Miller, C.P. (2005). Selecting the amount of smoothing in nonparametric regression estimation for complex. surveys. *Nonparametric Statistics*, **17,1**: 593–611.
- [22] Pfefferman, D., Rao, C.R. (2009). *Handbook of Statistics. Volume 29: Sample Surveys: Design, Methods and Applications*. North-Holland. Amsterdam
- [23] Rao, J.N.K. (2005). *Small area estimation*. John Wiley and Sons. New Jersey.
- [24] Särndal, C.E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, **33(2)**, 99–119.
- [25] Särndal, C.E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag. New York.
- [26] Singh, S. (2003). *Advanced sampling theory with applications: How Michael "selected" Amy*. Kluwer Academic Publisher. The Netherlands.
- [27] Thomson, S.K. (2012). *Sampling*. John Wiley & Sons. New Jersey.
- [28] Valliant, R., Dorfman, A.H. and Royall, R.M. (2000). Finite population sampling and inference: A prediction approach. Wiley Series in Probability and Statistics, Survey Methodology Section. John Wiley and Sons, Inc. New York.
- [29] Centro de Investigaciones Sociológicas. En: <http://www.cis.es/>
- [30] Instituto Nacional de Estadística. En: <http://www.ine.es/>
- [31] Instituto de Estudios Sociales Avanzados. En: <http://www.iesa.csic.es/>

Acerca del autor

María del Mar Rueda es Catedrática de Estadística e Investigación Operativa de la Universidad de Granada. Es Licenciada y Doctora en Matemáticas por la Universidad de Granada y su línea principal de investigación es el uso de información auxiliar en el contexto del muestreo en poblaciones finitas. Aparte de sus publicaciones en revistas internacionales de prestigio en este área, ha colaborado también con empresas privadas y públicas en la realización de encuestas por muestreo.