

**COMENTARIOS A
OPTIMIZACIÓN DEL ERROR MUESTRAL EN ENCUESTAS A POBLACIÓN
GENERAL: CRITERIOS DE ESTRATIFICACIÓN Y CÁLCULO DEL ERROR
CON “WESVAR COMPLEX SAMPLES” DE M. TRUJILLO**

Rafael Serrano del Rosal

Carlos Girón

Gonzalo Sánchez-Crespo

Carlos Camacho

Rafael Serrano del Rosal

IESAA-CSIC

En raras ocasiones se tiene la oportunidad de comentar un artículo, tan interesante y didáctico como pegado a la práctica de la investigación social mediante encuestas, como el escrito por Manuel Trujillo. Éste aborda lo que, a mi juicio, son dos de los grandes problemas de la utilización de esta técnica de investigación. Por un lado, la necesidad de rebajar lo más posible el error al que está sometida toda investigación mediante encuesta. Por otro lado, el de hacerlo al menor coste posible, es decir, sin que la reducción del error pase necesariamente por el incremento del tamaño muestral. Como ejemplo, el autor se centra en las encuestas que habitualmente se denominan de población general, argumentando desde un punto de vista pragmático y económico la necesidad que hay de recurrir a muestreos complejos en lugar de a los aleatorios simples. No obstante, es en las apreciaciones que hace sobre los muestreos estratificados, por conglomerados o no, donde radica la importancia del artículo. De hecho, en un momento donde las críticas a los diseños muestrales complejos comienza a ser habitual, el autor conocedor de las dificultades prácticas y las limitaciones económicas con las que habitualmente contamos los investigadores sociales, trata de demostrar y ejemplificar la utilidad de los mismos siempre y cuando se diseñen adecuadamente. Para ello critica acertadamente el hecho, más habitual

de lo que debiera, al que gusto de llamar “Kit de estratificación”. A saber, el hecho de estratificar siempre por las mismas variables, olvidando, tanto cuál es el fundamento de este tipo de muestreo, como la forma en que se debería realizar el mismo. De hecho hoy en día, en muchas ocasiones, hay disponible información sobre otra u otras variables cuya relación con la variable que queremos estudiar es suficientemente estrecha, como para que los estimadores que se obtengan con ella/as discriminen mejor entre los diferentes estratos, con lo que realmente se alcanzaría el principal objetivo que se debe tener en mente a la hora de diseñar una muestra: reducir el error muestral. Todo ello, sin los perjuicios económicos y prácticos que conlleva, tanto aumentar el tamaño muestral, como utilizar un muestreo aleatorio simple (como se demuestra con claridad en los ejemplos que pone Manuel Trujillo).

El artículo aborda también otro de los problemas que, a mi entender, ha denotado entre algunos científicos sociales el uso de los muestreos estratificados, por otra parte tan habituales en la investigación mediante encuesta. Éste no es otro que la mala medición que en ocasiones se hace del error muestral, lo que se debe habitualmente, como señala el autor con elegancia, al uso de programas estadísticos que sólo obtienen errores para el supuesto de un muestreo aleatorio simple. El autor opta por solucionar este problema con un programa y procedimiento concreto (Wesvar, JK_n) aunque, a mi juicio, la aportación más importante del artículo en este sentido no es la excelencia o no de dicho programa y método concreto, sino el énfasis que pone en la necesidad de utilizar una herramienta válida para la medición del error, especialmente, cuando se diseñan muestras complejas, ya que de no poder o no saber hacerlo así lo mejor es no utilizar este tipo de diseños. La dureza de esta afirmación se debe, principalmente, a que considero que buena parte de las críticas que se hacen a la utilización de muestreos complejos, se basan, o bien al desconocimiento de las enormes dificultades económicas y técnicas que, en la mayoría de las ocasiones, se tienen para implementar otro tipo de muestreos; o bien, en la mala utilización que en muchas ocasiones se hace de los mismos y no tanto a los verdaderos problemas que éstos tienen.

Finalmente, deseo señalar la dificultad de encontrar trabajos que, de forma tan clara y rigurosa, aborden una cuestión tan importante como es la del uso que hacemos los científicos sociales de los diseños muestrales.

Desde hace años, vivimos un proceso lleno de connotaciones muy variadas, de popularización de la informática y las computadoras. Gracias a este fenómeno, herramientas que sólo estaban disponibles para unos pocos han llegado hoy a una mayoría de hogares y empresas. En términos generales, esta tendencia es positiva. Sin embargo, cuenta con algunos inconvenientes. Los programas de computadora están hechos por empresas, es decir, por instituciones con ánimo de lucro, no por organizaciones benéficas. El objetivo no es ayudar a la Humanidad, sino ingresar

dinero, mejor cuanto más sea y durante más tiempo. En este terreno, pues, lo que interesa es establecer estrategias tales que permitan cumplir con dos principios:

1. Vender más copias de una versión.
2. Lanzar versiones continuamente, con una asiduidad tal que no viole el primer principio.

Cada nueva versión tiene que ofrecer algo nuevo; por lo general, tal vez sea más potente y, con seguridad, más fácil de usar que las anteriores. Y éste es el panorama, para cualquier campo con aplicación en los programas para computadora, sin excluir los sistemas de análisis de datos.

Los programas de análisis de datos deberían ser conceptualizados como herramientas que utilizan los expertos en análisis de datos, para facilitar su tarea. Es impensable realizar cálculos a mano hoy en día. Es más, la mayoría de los tratamientos que hacemos con los datos son extraordinariamente complejos y requieren una cantidad inmanejable de operaciones para un mortal.

Sin embargo, si observamos qué es lo que ocurre realmente en la práctica, encontraremos que la mayoría de las personas que recurren a los programas de análisis de datos no sólo no son “expertas”, sino que muchas de ellas carecen de conceptos mínimos y básicos sobre la ciencia estadística. La popularización de estos programas ha generado situaciones tan alarmantes como que existan usuarios para quienes la estadística consiste en pulsar botones con el ratón, dejando que el sistema tome múltiples decisiones escondidas bajo las opciones por defecto.

El artículo de Manuel Trujillo habla, entre otros aspectos como algunos hábitos indeseables en la estratificación, de *efecto del diseño*. Es ésta una expresión que apunta a un concepto tan antiguo como los diseños de muestreo. Sería deseable que el trabajo de Trujillo no se hubiera escrito, al igual como otros de corte similar nacidos desde hace años hasta la fecha y que denuncian el mismo hecho: “por favor, utilicen ustedes errores tipo adecuados al muestreo que han realizado”. Es, pues, un tópico. Pero es importante que se insista en ello (y por las vías apropiadas, como esta revista) porque el problema no sólo no se reduce, sino que aumenta con la misma rapidez con que se extiende el uso de los programas de computadora para el análisis de los datos. Debería alarmarnos el hecho de que las herramientas de análisis más extendidas y conocidas (más populares) no contemplan otro modelo de muestreo que el aleatorio simple para poblaciones de tamaño infinito y que se distribuyen con consignas tales como “¿El análisis de datos? ¡Pero si es muy fácil!”.

Un perfil muy habitual de investigador que “quiere hacer *bien* las cosas” es aquél que se toma el trabajo de decidir un tamaño para la muestra, consultando el apéndice de un libro casi genérico que utiliza (si bien, este investigador tal vez no lo sabe) el modelo de un muestreo aleatorio simple con reposición. Acto seguido recoge los datos de su encuesta mediante una estrategia *libre* que quizá se asemeja a un muestreo bietápico y, por último, hace los análisis con el programa *que tiene a mano* porque es el que conoce o el que ha adquirido su institución, realizando estimaciones incorrectas, si bien orgulloso de su pulcritud *científica*.

Creo que es una obligación moral para los expertos en metodología de su ciencia concreta (sea sociología, estadística, psicología, economía...) velar por que este problema se reduzca hasta desaparecer. Es una batalla doble. Implica, por un lado, trabajar por una adecuada formación de los profesionales. Por otro, exige ejercer una labor real de selección de los sistemas informatizados y de presión sobre el diseño de éstos.

Recuerdo ahora una conversación con una colega, metodóloga de las ciencias sociales, hace ya cuatro años. Ella me indicaba "Si yo fuera muy escrupulosa con este tema, no habría dado mi voto en ninguno de los tribunales de tesis a los que he pertenecido, ni hubiera dado el visto bueno a casi ninguno de los artículos de investigación empírica que he revisado". El problema existe y nosotros tenemos algo de responsabilidad en ello. Un pequeño grano de arena: dar a leer a nuestros alumnos el trabajo de Trujillo u otro texto que denuncie este hecho y que contenga simulaciones como las tablas 1 a 4, acompañando esta valiosa información con una alternativa práctica: la formación en cualquiera de los sistemas informatizados de análisis que sí contemplan el diseño original de muestreo.

Seamos realistas con respecto al software que utilizamos. Y esto en dos sentidos: ni podemos pedirle lo que no puede darnos, ni tenemos por qué utilizar lo que no necesitamos.

Gonzalo Sánchez-Crespo

INE

La estratificación es el instrumento más importante entre los que cuenta el diseñador de encuestas. Una estratificación competente precisa información del marco, dispuesta de forma adecuada para la investigación que se quiere llevar a cabo. Cada investigación por muestreo debe plantearse aisladamente con un diseño específico.

El caso óptimo, donde se consigue el efecto de reducción del error mayor, es para muestras de tamaño uno, dentro de cada estrato. Este caso convierte el muestreo estratificado en un tipo de muestreo sistemático. Se trata de aquel diseño donde únicamente se obtiene una unidad dentro de cada grupo, estrato, de forma aleatoria y no por ocupar una determinada posición relativa. El método estratificado es especialmente adecuado sobre el muestreo sistemático al evitar posibles sesgos producidos por la existencia de ciclos o tendencias en la población. Estas tendencias y ciclos son frecuentes al llevar a cabo procesos de estratificación. El principal inconveniente de diseños óptimos como el señalado u otros es que se complica la obtención de estimaciones del error.

La forma práctica de agrupar la población en grupos homogéneos, es ordenar las unidades del marco por cada una de las variables que se consideren relevantes en la estratificación. En este punto es preciso ser cuidadoso con el método de selección aplicado puesto que pueden haberse introducido en el diseño factores que pueden dar lugar al aumento del coeficiente de correlación intraclásica. Es preciso, en esta

fase del diseño, al plantear el método de selección, realizar estudios y simulaciones que permitan estimar la ganancia en precisión obtenida en la estratificación.

La ventaja de utilizar las secciones censales es que se dispone de abundante información procedente de los censos generales realizados por el Instituto Nacional de Estadística cada diez años. Exactamente 2098 variables por sección permiten plantear una caracterización orientada a diferentes objetivos. La ventaja añadida por estos datos es que se encuentran en un soporte (*CD-ROM – CERCA*) fácilmente utilizable por ordenador. El número de libros necesarios para contener estas tablas y el trabajo asociado a introducirlas en el diseño de la encuesta podría ser muy desmotivador, por el coste e, incluso, por imposibilidad física. Desde los censos del 90 ésta es una magnífica realidad, tal vez no siempre bien aprovechada.

El inconveniente de utilizar la sección censal viene dado cuando se produce un fenómeno de homogeneización. Hace relativamente poco, un pueblo de pescadores en la costa era algo más o menos bien definido. Hoy podemos encontrar estudiantes universitarios, comerciantes, agricultores y trabajadores de la industria o de los servicios en prácticamente cualquier área urbanizada. El aumento y mejora de los medios de transporte, la proliferación de centros de estudios superiores, las tendencias de la población a núcleos urbanos de tamaños medios situados en los alrededores de las ciudades, hace cada vez más difícil caracterizar zonas como exclusivamente agrarias, industriales o de otro tipo. A medida que sea más intenso el proceso de homogeneización del seccionado, más difícil resultará utilizarlo para la estratificación, y más necesarios serán estudios como el que desarrolla Trujillo. Con la información disponible, hoy en día en *áreas pequeñas* inferiores a la sección habría que plantear la estratificación, incluso, por edificios. Ésta es una posibilidad impensable hace pocos años pero que, hoy, con la información censal completa, es una realidad. El *CD-ROM CERCA+100* contiene 1.386 datos individuales de cada una de las 89.514 unidades territoriales de España.

En definitiva creo que estamos ante un artículo que muestra de forma clara las posibilidades que actualmente tiene el diseñador de encuestas a su disposición. Estos procesos estaban hasta hace poco sólo a disposición de los grandes centros estadísticos, puesto que precisaban unos medios de cómputo no accesibles para compañías pequeñas o para investigadores aislados. Hoy, estos procedimientos de optimización se hacen necesarios, y lo que es más importante, están a disposición de la mejora de las investigaciones por muestreo, a un coste relativamente pequeño en comparación con los costes que ahorran y con el aumento en la precisión de las estimaciones que aportan. No sólo es necesario hacer las cosas bien desde un punto de vista científico o ético, es, además, en este caso, rentable y tecnológicamente posible.

Creo que estudios como el reseñado mejoran el nivel y la aceptación de la profesión estadística en el mercado de trabajo. Permitiendo una valoración positiva de las ganancias que aporta emplear métodos eficientes en el diseño de investigaciones por muestreo. Apartando y denostando las fichas técnicas de encuestas normalizadas, extendidas y, en muchas ocasiones, inadecuadas. Felicito y animo al autor para que continúe en esta línea de divulgación y de rigor con éste y otros temas y, en

concreto, me permito sugerirle abordar la reducción del sesgo introducido por la falta de respuesta en las investigaciones por muestreo.

Carlos Camacho

Universidad de Sevilla

Las investigaciones están llenas de paradojas que a casi todos nos gustaría pasar por alto y así no tener que modificar los esquemas preconcebidos que frecuentemente nos complican la vida, aunque en estas complicaciones y en su solución radica el sentido del investigador.

Hay, como digo, muchas paradojas: la distribución de los datos no sigue ciertas leyes de probabilidad deseadas, la linealidad en la relación entre las variables no siempre se cumple y los muestreos no son lo suficientemente aleatorios. Al final, solemos proceder como si nada de lo ocurrido fuera cierto para así poder actuar como queramos (o podamos). Las consecuencias son obvias, pero como no sabemos (o no queremos saber) en lo que nos equivocamos la referencia de lo real es precisamente nuestra medida, pues todo parece verdaderamente bien hecho. Además, la parafernalia matemática ayuda a la exaltación de nuestra obra. Todo a falta de alguna voz discordante.

En esta línea se mueve el artículo de Manuel Trujillo. No precisamente en ocultar la cabeza bajo el hoyo o mirar a otro lado, sino en denunciar lo que es un secreto a voces y poner coto a tantos desmanes cometidos en nombre de la Estadística. No se pone la necesaria atención en evitar algo perfectamente evitable teniendo, como tenemos, nuestras miradas en objetivos más altos. No se realizan correctamente los muestreos, o mejor dicho, no se *diseñan* adecuadamente y, si por casualidad se hace, pocas veces se refleja en el tratamiento estadístico posterior de los datos, por la sencilla razón de que los programas al uso (léase SPSS) no lo contemplan. Así pues, dos cometidos importantes se propone el autor: a) disminuir el error muestral sin recurrir a la fuerza bruta (aumentando el número de observaciones), esto es, mejorando el diseño del muestreo. Y una vez logrado tal propósito: b) actuar en consecuencia, lo que implica no tratar el error muestral como si fuera aleatorio simple sino al que corresponda, según el procedimiento de muestreo aplicado. Para ello no hace falta desarrollar rutinas que debamos implementar en los paquetes estadísticos habituales, sino tan sólo utilizar un programa específicamente pensado para ello: *Wesvar Complex Samples*.

Es además de agradecer su vocación didáctica tan poco frecuente en artículos metodológicos. Así, algunos ejemplos prácticos nos sirven de referencia para conocer el procedimiento utilizado y la cuantía de los errores asociados, que a su vez, se comparan, merced a la simulación, con los que habrían si se utilizara exclusivamente un muestreo aleatorio simple. Y, para más impacto, el ahorro logrado en términos de sujetos y económico para una misma precisión en el muestreo.

En definitiva, se trata de un trabajo de obligada lectura para todos aquellos que nos movemos más allá del estrecho cerco de lo aleatorio simple. Altamente recomendable para cualquier curso (texto) sobre muestreo. Después de esto, ya no quedan excusas convincentes.