

Encuestas controladas por cuotas: una aproximación empírica a su error total

Quota-controlled Surveys: an Empirical Approach to their Total Error

Alberto Penadés*
Araceli Mateos**
Ester Bejarano***

Recibido el 12 de mayo de 2024
Aceptado el 12 de junio de 2024

Resumen

La mayor parte de los datos de encuesta disponibles para la investigación de la opinión pública y la prospectiva electoral mediante muestras representativas, en España y América Latina, se han obtenido con selección final de las personas entrevistadas mediante cuotas, un método no probabilístico. La fiabilidad de las estimaciones hechas a partir de estos datos es mal conocida y tiene poco respaldo teórico: se sabe bastante sobre sus sesgos, pero casi nada sobre su varianza y “margen de error”. Esta investigación estudia con métodos empíricos la variabilidad de las muestras obtenidas en los barómetros periódicos del Centro de Investigaciones Sociológicas de España (CIS), aprovechando la longitud y regularidad de algunas de sus series de estimación de parámetros estables. Los resultados muestran que el crecimiento de la varianza es moderado y está dentro de lo esperable en un diseño complejo.

Palabras clave

Sondeos de opinión pública, encuestas electorales, métodos de cuota, fiabilidad.

* Doctor en Ciencia Política por la Universidad Autónoma de Madrid y Doctor Miembro del Instituto Juan March. Profesor Titular de Sociología en la Universidad de Salamanca (España). penades@usal.es. ORCID: 0000-0003-1599-2105.

** Doctora y Profesora Titular de Ciencia Política en la Universidad de Salamanca (España). amd@usal.es. ORCID: 0000-0001-6208-5917.

*** Doctora en Sociología y Profesora Asociada en el Departamento de Sociología y Comunicación de la Universidad de Salamanca (España). esbejarano@usal.es. ORCID: 0000-0002-8373-5701.

Abstract

The majority of survey data available for public opinion research and electoral forecasting through representative samples, in Spain and Latin America, has been obtained with the final selection of interviewees through quotas, a non-probabilistic method. The reliability of the estimates made from this data is poorly understood and has little theoretical support: we know quite a bit about its biases, but almost nothing about its variance and “margin of error.” This research empirically studies the variability of the samples obtained in the periodic barometers of the Center for Sociological Research of Spain (CIS), taking advantage of the length and regularity of some of its series estimating stable parameters. We find that the growth of variance is moderate and is within what is expected in a complex design.

Keywords

Public opinion polls, electoral polls, quota sampling, reliability.

¿Cuál es el “margen de error” de las estimaciones puntuales en las encuestas que usan métodos de selección no aleatoria, tales como el muestreo por cuotas? Es sabido que la estimación de error típico para construir intervalos de confianza que aparece normalmente en las fichas técnicas de las encuestas, basada en la teoría del muestreo aleatorio simple, es inválido cuando las muestras no son ni aleatorias ni simples, como suele ser el caso en las encuestas electorales y de opinión pública en muchos países de Europa y América Latina, entre ellos España y México. Es difícil señalar si el cálculo del error típico de las estimaciones es una buena o mala aproximación. En este trabajo se propone una respuesta empírica a las siguientes preguntas de investigación: ¿Cuál es la variabilidad realmente observada en estimaciones repetidas de parámetros estables a partir de muestras no aleatorias? ¿En qué medida esa variabilidad es distinta que la esperada si se tratase de un muestreo aleatorio simple con igual número de observaciones? ¿Cuánta precisión se pierde con respecto a los intervalos de confianza teóricos para cada tamaño de muestra?

Los barómetros del Centro de Investigaciones Sociológicas (CIS) ofrecen una oportunidad casi única de responder a estas preguntas, ya que efectúan mediciones muy reiteradas, con una misma metodología, de algunos parámetros sociales y de opinión que se mueven lentamente a lo largo del tiempo en la sociedad española. Además, estos barómetros son utilizados para llevar a cabo estimaciones electorales con gran proyección mediática mensualmente. Empleando este método observacional se comprueba que el muestreo de las encuestas del CIS da lugar a un incremento bastante moderado de la varianza con respecto a los supuestos teóricos del muestreo aleatorio simple.

La principal objeción al empleo de cuotas es que no existe una teoría para estimar el error de las inferencias basadas en el diseño muestral. Los intervalos de confianza que se establecen rutinariamente a partir del número de casos para las estimaciones apenas tienen sustento científico, pues la probabilidad de selección final de cada individuo es desconocida. Kish (1965, p. 562 y ss.) realiza una formulación clásica, aún muy influyente, de esta crítica. El consenso teórico casi total en la disciplina del análisis estadístico de encuestas es que “quienquiera que busque obtener inferencias a partir de una muestra por cuotas debe necesariamente adoptar un enfoque basado en un modelo” (Lohr, 2010, p. 97). En otras palabras, con este tipo de encuestas solo se deberían obtener expectativas condicionadas. La validez de los intervalos de confianza dependería en ese caso de que el modelo fuera correcto, no así del procedimiento de muestreo.

Sin embargo, esa no es la práctica general en la disciplina de análisis de encuestas, mucho menos entre sus usuarios menos orientados a la ciencia. Resulta inevitable que los datos de instituciones como el CIS se empleen para hacer estimaciones puntuales y de tendencias basadas en el diseño de la muestra, confiando en que esta sea “representativa”. Gran parte de los usuarios esperan

que los descriptivos se puedan generalizar con un riesgo de error moderado y previsible, sin necesidad de modelizar la respuesta. Es una demanda natural para los sondeos de opinión pública, comenzando por las estimaciones de intención de voto.¹ Por esta razón la Asociación Americana de Investigación en Opinión Pública pide expresamente que se reporten medidas de incertidumbre en todas las encuestas cualquiera que sea su metodología (Baker et al., 2013).

En esta investigación se ha adoptado un enfoque inspirado en el propuesto por primera vez por Stephan y McCarthy (1958, cap. 10): observar la muestra repetida con un mismo diseño a lo largo del tiempo, estudiando la evolución de la estimación de distintos parámetros estables y la distribución su error empírico en torno a una tendencia central. Es la solución empírica más sencilla y la única practicable con los datos disponibles.

Las investigaciones sobre las muestras de cuota suelen centrarse más en el sesgo que en la fiabilidad (Stephenson, 1979; Curtice y Sparrow, 1997; Lynn y Jowell, 1996; Díaz de Rada y Martínez Martín, 2014). Para la determinación de la variabilidad de las cuotas existen en la literatura tres aproximaciones empíricas y una teórica. Inicialmente, para las primeras muestras de cuota se propuso una estrategia experimental (Moser y Stuart, 1953). Pese a que los experimentos para la estimación de la varianza han sido defendidos por los críticos más exigentes con las cuotas (Kish, 1998), estos no se han prodigado, debido posiblemente a las dificultades técnicas y a su coste. La vía observacional, inaugurada por Stephan y McCarthy (1958, cap. 10), se ha retomado recientemente para obtener estimaciones empíricas de varianza a partir de sondeos repetidos por múltiples institutos durante el periodo electoral en EEUU (Shirani-Mehr et al., 2018) y en Alemania (Selb et al., 2023), aunque mezclando metodologías de campo muy diferentes. La tercera estrategia empírica posible consiste en estimar la varianza mediante la simulación de remuestreos (Sturgis et al., 2018). Los resultados en Reino Unido son prometedores, pero no pueden replicarse con las encuestas del CIS porque por razones de protección de datos no se facilita la información sobre qué entrevistas se han realizado en una misma sección censal.

Por último, algunos autores han defendido una aproximación teórica: bajo ciertos supuestos, se podría estimar la varianza de los métodos de cuota con las mismas fórmulas empleadas para estimar la varianza de una muestra compleja con selección probabilística, tratando a las cuotas como si fueran estratos en el diseño de la muestra (Deville, 1991; Sudman, 1966). De acuerdo con estos autores los métodos de cuota tal y como se aplican desde los años 60, con cuotas cruzadas combinadas con la aleatorización —o al menos dispersión— de los puntos últimos de muestreo, podrían satisfacer los supuestos teóricos necesarios. No cabe aquí juzgar la validez del argumento, pero sí señalar que ese cálculo podría ser más razonable para las fichas técnicas que el derivado de los supuestos del muestreo aleatorio simple. La presente investigación tiene relevancia práctica puesto que el

¹ La estimación basada en el diseño utiliza ponderaciones para corregir sesgos, no se deben confundir estas con la modelización de los datos.

muestreo con cuotas es el método más utilizado tanto en España como en buena parte de Europa y de América Latina para los sondeos electorales y de opinión pública de los institutos privados (Durand et al., 2018; Abundis et al., 2017; Sturgis et al., 2016; Moreno et al., 2014; Mateos y Penadés, 2013; Gschwend, 2005, Taylor, 1997; Taylor, 1995). Pero no es solo una cuestión de investigación comercial, pues en algunos países como España y Francia son dominantes también en las encuestas de orientación científica y pública y la cuestión de la validez de los datos de cuota se ha suscitado como problema para publicar resultados basados en ellos en revistas de investigación académica (Gschwend, 2005). La contribución de este análisis recupera el debate científico sobre la validez de los métodos de cuota y ofrece una respuesta optimista, al menos, por lo que se refiere a la confianza en los resultados de las encuestas hechas con el diseño muestral del CIS. En este artículo se valida de forma cuantitativa la observación cualitativa que hacen los investigadores de las series históricas de opinión pública: cuando los datos se recogen con muestras de cuota las estimaciones resultan coherentes y no muestran en su evolución ninguna volatilidad que parezca excesiva o inexplicable (Berinsky, 2006; Page y Shapiro, 1992; Cantril, 1948).

El muestreo del CIS en el panorama de las muestras por cuota

Lo que distingue a los métodos de cuota es que las personas seleccionadas en una unidad de muestreo se distribuyen en ciertas proporciones según características prefijadas. El sexo y los grupos de edad son las cuotas más habituales, pero estas pueden incluir nivel educativo, origen u otros rasgos del sujeto. Se trata de cualidades que 1) se desea que estén representadas en la muestra en proporciones conocidas porque son covariables importantes para la investigación; 2) no es posible usarlas como criterios de estratificación en el diseño de la muestra porque no se sabe si un individuo tiene la característica deseada hasta que no se produce el contacto; y 3) son parámetros muy relacionados con la tasa de respuesta, de modo que una muestra aleatoria los representaría de forma muy sesgada sin una considerable inversión de recursos.

Para una comparación pragmática entre la selección aleatoria y las cuotas conviene fijarse en dos cuestiones: 1) el tamaño de la última unidad de muestreo, es decir, cuánto se acerca el azar al individuo en la selección de las unidades (municipio, barrio, manzana...); y 2) la libertad de elección que tiene el entrevistador dentro de esa unidad.

Recuérdese que en el diseño aleatorio 1) el método dirige al entrevistador hasta una persona particular, con una probabilidad de selección conocida; y 2) la capacidad del entrevistador para intervenir en la selección se limita mucho (idealmente se anula) mediante reglas de reintento y sustitución.

En un extremo se encuentran los primitivos métodos de cuota, en los que la unidad de muestreo podía ser toda una ciudad y el entrevistador no tenía otras instrucciones que cumplir las cuotas. Son estos los que se tienen en mente en los enunciados clásicos de la crítica a las cuotas. Por ejemplo, en la encuesta electoral de Gallup de 1946 las instrucciones para un entrevistador en la ciudad de San Louis (Misuri) requerían que entrevistara a 15 personas, 7 hombres y 8 mujeres. De las mujeres, 4 menores de 40 años y 4 mayores, 1 negra y 8 blancas, y todas repartidas por tramos de precio de alquiler. Se utilizaban así cuotas cruzadas por sexo y tres características (edad, raza y renta) tomadas de una en una, sin restricciones para encontrar a quien las cumpliera. (Stephan y McCarthy, 1958, p.86).

Casi en el otro extremo, en cuanto al tamaño de los conglomerados, está el diseño muestral de la General Social Survey (GSS, por sus siglas en inglés) entre 1972 y 1976 y que llamaron, con cierto eufemismo, “probabilístico modificado” o, internamente, “manzana y cuota”. La selección aleatoria llegaba a una unidad tan pequeña como la manzana, pero a partir de allí ya no se seguía al azar, sino que se empleaban cuotas (sexo, actividad y edad) para incluir segmentos difíciles de muestrear, abaratando costes en años de pocos recursos. Los entrevistadores solo se sujetaban a una pauta horaria. (Davis y Smith, 1992)

Si el clima intelectual del estudio de la opinión pública en Europa en los años 70 hubiera sido el de EEUU, posiblemente, el método adoptado por el CIS se habría llamado “probabilístico modificado”. En las encuestas presenciales del CIS se llega hasta los hogares como unidad seleccionada al azar mediante rutas aleatorias dentro de unidades de muestreo, a su vez, pequeñas, las secciones censales, que han sido elegidas en un diseño aleatorio estratificado por conglomerados. En cada sección se determina una cuota cruzada de sexo y edad (Martínez, 1999). Puede verse como un método en el que la selección puramente aleatoria llega hasta la sección censal, y en el que el cumplimiento de cuotas se acompaña de una restricción muy severa para el entrevistador a la hora de sustituir a potenciales entrevistados, aleatorizando la elección de viviendas donde intentarlo. También puede verse como un método que es aleatorio hasta la selección de hogares, pero con reemplazo inmediato de las personas entrevistadas si no hay respuesta o si la respuesta no cumple la cuota de la sección.

Las muestras representativas controladas mediante cuotas constituyeron un notable progreso metodológico para los sondeos electorales cuando fueron introducidas por Gallup en 1936. Sin embargo, la controversia ocasionada por el fracaso general de los pronósticos de las elecciones de 1946 hizo que en los años 50 el muestreo probabilístico se convirtiese en norma ideal de los institutos privados en EE UU., una noción disponible desde que el American National Election Survey (ANES, por sus siglas en inglés) recogiese su primera muestra probabilística en 1948 (Frankel y Frankel, 1987).

Los métodos de cuota siguen siendo usados en Europa y Latinoamérica, pero se han acercado a los probabilísticos en la selección de unidades de muestreo y en el control del efecto de entrevistador (Sudman, 1966). Al mismo tiempo, aun

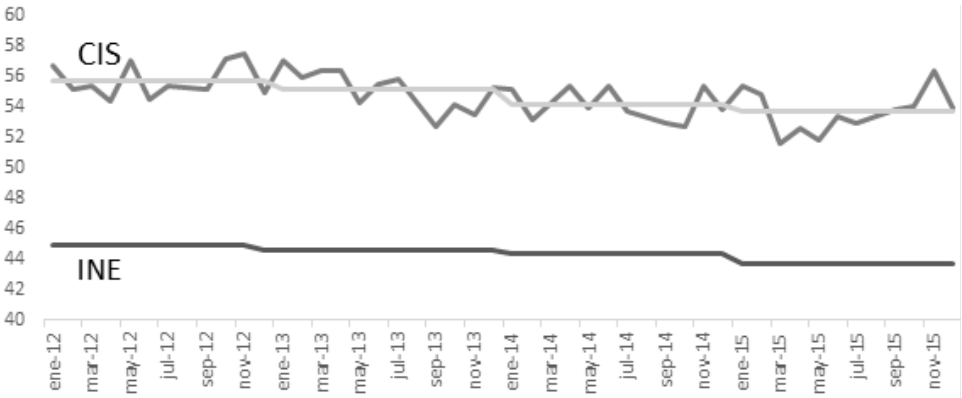
cuando se denomine muestra aleatoria, la variedad de técnicas que se emplean en la práctica profesional de los sondeos electorales para la selección de individuos dentro del hogar puede alejarse bastante de la norma teórica; y parecerse en algunos casos a los métodos de cuota (Gaziano, 2005). La diferencia real es menor que en los libros.

Método de estimación empírica

En el error de estimación de un parámetro hay dos componentes, la varianza y el sesgo. Las encuestas controladas por cuotas tienen unos sesgos propios como los tiene cualquier método de recogida de datos. Como ilustración, el gráfico 1 muestra la proporción de personas casadas encontradas en las muestras del CIS durante cuatro años. Por tratarse de un método de selección de entrevistados sin reintento las personas que viven solas tienen menor probabilidad de contacto. Como se aprecia en su comparación con los datos de registro anuales que ofrece el Instituto Nacional de Estadística (INE), las personas solteras aparecen infrarrepresentadas en todas las muestras. La inferencia puede corregir un sesgo como este de forma simple mediante ponderaciones.

Las estimaciones también tienen una variación en torno a la tendencia, que en este caso son las medias anuales. Se trata de la incertidumbre inevitable en una muestra. Esa variación es el objeto del análisis aquí desarrollado.

Gráfico 1. Sesgo y varianza. Ejemplo: proporción de casados/as en los barómetros (2012-15) y medias anuales



Fuente: elaboración propia.

El concepto de distribución en el muestreo de un estimador con un método de cuota es lógicamente coherente. La variabilidad la originan los entrevistadores. “Puesto que la variabilidad muestral es por definición la cantidad de variación que surge por la aplicación repetida de un determinado procedimiento de muestreo, la forma más directa de estimarla es utilizar situaciones en las que una misma variable ha sido medida en un cierto número de encuestas sucesivas” (Stephan y McCarthy, 1958, p. 214). Este es el enfoque que aquí se utiliza: se supone que la distribución en el muestreo es normal, y se utiliza una aproximación binomial, con un multiplicador adecuado para la varianza. Ese multiplicador puede compararse con el “efecto de diseño” de la encuesta. El término de comparación es el muestreo aleatorio simple.

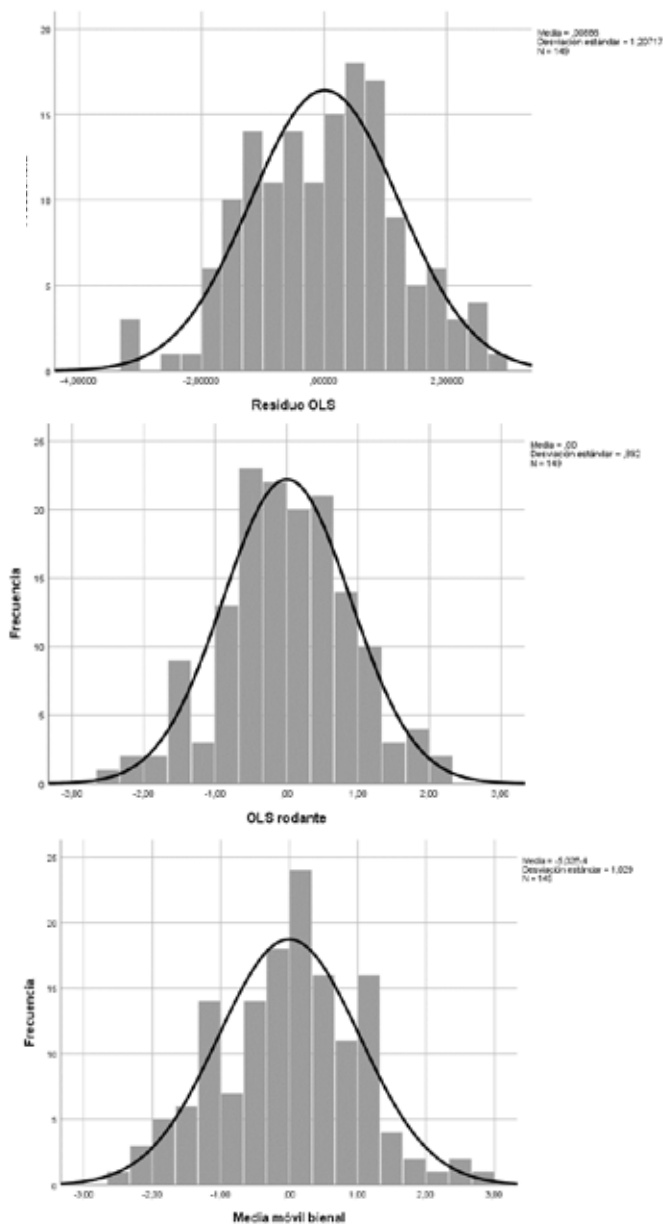
Uno de los asuntos más delicados es cómo escoger una representación de la “verdadera” tendencia que se supone que reflejan las repetidas encuestas. Para cuestiones que apenas varíen una simple media aritmética sería lo adecuado -así lo hicieron Stephan y McCarthy (1958)- o una media anual, como se sugiere en el gráfico 1. Si la cuestión varía de forma constante la tendencia podría estimarse con una recta de regresión, pero están considerándose series muy largas y es improbable que no se produzcan cambios en la verdadera pendiente a lo largo de los años. Por ejemplo, la proporción de personas que se identifican como católicas desciende gradualmente pero no sigue una tendencia estrictamente lineal, hay periodos donde se puede acelerar levemente. La variabilidad en torno a la tendencia lineal estricta posiblemente exagera el error de estimación del parámetro.

Por ello, se han incluido dos alternativas al puro ajuste lineal para el periodo de datos disponibles: el ajuste de un coeficiente de regresión lineal de mínimos cuadrados rodante (para el año anterior a cada observación, manteniendo la tendencia lineal para todo el primer año), y una media móvil, en este caso centrada en periodos de dos años. Los histogramas del gráfico 2 muestran las distribuciones de la variabilidad con respecto a estas tres aproximaciones al parámetro.

Como se ha comentado previamente, se supone que la distribución del error de muestreo es normal. Al menos, las pruebas de normalidad no permiten rechazar la hipótesis nula de que la distribución es normal en ninguno de los tres casos (tabla 1).

Por último, la desviación típica de estas distribuciones representa la aproximación al error típico estimado del estadístico en cuestión, es decir la proporción de católicos en España. Comparando las tres aproximaciones (tabla 2) puede observarse cómo la recta de regresión con un ajuste para todo el periodo es la estimación más conservadora sobre la precisión de las estimaciones, pues conlleva una estimación del error típico estimado mayor que las alternativas. La representación del parámetro como un coeficiente de regresión rodante conlleva unos residuos más concentrados y, por tanto, un error estimado menor. La media móvil representa una opción intermedia entre el ajuste excesivo y el ajuste más pobre de las otras dos alternativas.

Gráfico 2. Distribución de los errores de estimación con tres alternativas de medición del error: recta de regresión lineal, recta de regresión rodante anual y media móvil bienal.



Fuente: elaboración propia.

Tabla 1. Pruebas de normalidad

	Kolmogorov-Smirnova		Shapiro-Wilk	
	Estadístico	p	Estadístico	p
Residuo de la recta MC	0,047	>0,200	0,991	0,464
Residuo ajuste MC rodante	0,037	>0,200	0,994	0,81
Media móvil bienal	0,052	>0,200	0,991	0,524

Fuente: elaboración propia.

Tabla 2. Estadísticos de las estimaciones

	Error típico estimado (s)	Error medio	N
Residuo de la recta MC	1,207	0,009	149
Residuo ajuste MC rodante	0,892	0,004	149
Media móvil bienal	1,029	-0,001	145

Fuente: elaboración propia.

La media móvil es, además, la opción más flexible para poder comparar variables entre sí, las cuales pueden tener tendencias distintas y ser menos lineales que las del ejemplo. Por estas razones, y para completar el ejercicio, se revisa la variabilidad de una batería de indicadores utilizados en los cuestionarios aplicaciones por el CIS.

Datos

Para un análisis de este tipo se necesitan encuestas regulares realizadas con un mismo diseño muestral, un mismo método de administración y un mismo tamaño de muestra. Desde mayo de 1990 hasta julio de 2018 los barómetros del CIS se han llevado a cabo con regularidad, once meses al año, a partir de una muestra diseñada de 2500 entrevistas personales realizadas en el hogar. Este es el intervalo dentro del que se identifican los datos a utilizar en este trabajo. En este periodo ha habido modificaciones relativamente menores en el diseño de la muestra y en la administración (de papel y lápiz a medios electrónicos), pero dentro de un procedimiento muy estable. Con anterioridad a 1990 los tamaños de muestra eran más irregulares. Con posterioridad se han producido dos cambios que afectan a la comparabilidad temporal de los resultados: se ha aumentado el tamaño de muestra a 3,000 y más entrevistas desde septiembre de 2018 y se ha introducido la administración telefónica de los cuestionarios desde abril de 2020.

Dentro de este marco temporal se ha seleccionado un conjunto de variables que hayan sido medidas reiteradamente y con continuidad. Esta selección cubre tanto mediciones de actitudes, como de comportamientos y atributos de las personas entrevistadas. Lo ideal es encontrar indicadores que varíen gradualmente o permanezcan muy estables, para hacer una aproximación lo más posible a la noción teórica de muestreo repetido que mide un parámetro invariante. En casi todos los casos se cuenta con once muestras por año (en tres de las variables las mediciones son trimestrales), con una misma regularidad, un mismo tamaño de muestra (2,500 como muestra diseñada) y llevadas a cabo por una misma organización.

El conjunto de indicadores introducidos lo forman seis ítems de actitudes básicas, dos de comportamiento y cinco mediciones de atributos. Se dispone de un mínimo de 37 repeticiones y un máximo de 268, correspondientes a la serie de posición en la escala de ideología entre enero de 1996 hasta enero de 2016.

Tabla 3. Variables analizadas dentro del periodo 1996-2016

	muestras	regularidad
Católico	145	mensual menos agosto
No creyente	145	mensual menos agosto
Casi nunca va a misa	145	mensual menos agosto
Va a misa todos o casi todos los domingos	145	mensual menos agosto
Posición en el eje izda-dcha (media)	236	mensual menos agosto
Conservador	37	Trimestral
Liberal	37	Trimestral
Socialista	37	Trimestral
Jubilado	268	mensual menos agosto
Soltero	81	mensual menos agosto
Casado	81	mensual menos agosto
Sin estudios	150	mensual menos agosto
Estudios secundarios	150	mensual menos agosto

Fuente: elaboración propia.

Resultados

A continuación, se muestra cómo los errores estimados a partir de la variación empírica de las estimaciones en torno a la tendencia del parámetro (estimada mediante media móvil) implican una moderada inflación de los intervalos de confianza coherente con lo hallado en otras estimaciones empíricas y en la aproximación teórica del error en muestras de diseño complejo. En el anexo pueden consultarse las pruebas de normalidad en cuanto a la distribución empírica de los errores.

La tabla 4 presenta el resultado. En la primera columna se encuentra el error típico estimado a partir de la proporción media en el periodo de estudio y con una muestra de 2,500 observaciones. En la segunda columna se presenta el error típico estimado empíricamente a partir de la distribución de los errores con respecto a la tendencia medida por el indicador. En la tercera columna, el multiplicador que debería usarse con respecto a los intervalos clásicos a la hora de establecer un intervalo de confianza.

Tabla 4. Estimación empírica del error típico y factor de inflación de la desviación típica o “efecto de diseño” de las encuestas del CIS.

	Error típico para intervalo clásico	Error estimado como variabilidad empírica	Factor de inflación del intervalo de confianza	Media
Católico	0,00887	0,01029	1,161	
No creyente	0,00702	0,00796	1,134	
Posición en el eje izda-dcha	0,00044	0,00065	1,480	
Conservadores	0,00661	0,67157	1,015	
Liberal	0,00650	0,99275	1,527	1,294
Socialista	0,00698	0,88298	1,265	
Casi nunca va a misa	0,00990	0,01593	1,609	
Va a misa todos o casi todos los domingos	0,00702	0,00815	1,161	
Jubilados	0,00632	0,00777	0,814	
Solteros	0,00929	0,00922	0,993	
Casados	0,99625	108,416	1,088	0,952
Sin estudios	0,0053	0,0052	0,989	
Estudios secundarios	0,0067	0,0059	0,874	

Fuente: elaboración propia.

Este multiplicador de la desviación típica puede leerse como un efecto de diseño del CIS, incluye tanto el efecto de las cuotas como todos los demás aspectos de su diseño muestral. Al igual que sucede con las estimaciones del efecto de diseño para las muestras probabilísticas complejas, este depende de la pregunta concreta.

El efecto de diseño medio de los indicadores para actitudes y comportamientos es de 1,3. En promedio, habría que incrementar los intervalos de confianza para la estimación de variables actitudinales y de comportamiento de acuerdo con ese factor. La media para todas las variables estudiadas es de 1,13, aunque el factor de inflación puede llegar a ser de 1,6. Es interesante notar, no obstante, que los indicadores de atributos, donde el error de respuesta es menor, tienen un error muestral, en promedio, igual al error muestral teórico de una muestra aleatoria simple (algo más bajo: 0,95). Esto es un indicio de que la selección final mediante cuotas puede tener poca responsabilidad en la inflación de la varianza, aunque tal vez sí en su interacción con otras fuentes del error total.

La estimación del efecto de diseño de la Encuesta Social Europea, con una muestra probabilística también ejecutada por el CIS, arroja unos valores medianos para España de 1,45 en la Ronda II de la encuesta y un 1,95 en la Ronda I, mientras que las medianas en Europa son 1,53 y 1,61 (Ganninger, 2006). La comparación debe tomarse con extrema cautela, pero esta investigación ofrece una primera evidencia de que la selección por cuotas podría no añadir variabilidad al diseño muestral con respecto a la muestra aleatoria más comparable.

Estos resultados son coherentes con la investigación que existe, por escasa que sea, sobre determinación empírica de la varianza de las estimaciones realizadas con muestras controladas por cuotas. Los resultados revelan una fiabilidad mayor que los del tipo de encuestas de cuota que se realizaban hasta los años 50, como era de esperar, y se encuentran en línea con lo que se sabe de los métodos de cuota más semejantes al del CIS. Stephan y McCarthy (1958, p. 225) encontraron en 12 variables estudiadas que su varianza era entre 0,9 y 2,3 veces mayor (con una media total de 1,6) que en el muestreo aleatorio simple. Moser y Stuart (1953) encontraron experimentalmente que la varianza estimada para las muestras de cuota era entre una y tres veces mayor que la varianza estimada mediante una muestra aleatoria. Recientemente, Sturgis et al. (2018) han estimado que el efecto de diseño para un solo indicador, la intención de voto, de los distintos institutos de investigación británicos, estimado mediante simulaciones, se encontraba entre 1,04 y 1,68.

Conclusiones

Una de las objeciones al empleo del muestro por cuotas es la ausencia de una teoría robusta que permita estimar el error de las inferencias realizadas basadas en el diseño muestral. El análisis realizado en este trabajo muestra que el error típico de las estimaciones es normalmente mayor que el que reflejan las “fichas

técnicas” de las encuestas, aunque en el caso de los barómetros presenciales con muestra de cuota del CIS el incremento es moderado. En la estimación clásica, se hace necesario incrementar los intervalos de confianza: la probabilidad de que ciertas diferencias entre estimaciones se deban al azar es mayor de la que se podría suponer en base al número de observaciones. Para las variables de actitud, posiblemente un factor de 1,3 sea el adecuado como multiplicador del intervalo para reflejar el efecto de diseño.

No es evidente que la selección final por cuotas sea lo que contribuye principalmente a esta inflación de la varianza. Aunque puede llegar hasta el 60%, esto no es inusual en otras muestras de diseño complejo, sean o no de cuota, como muestra la comparación de nuestros resultados con los obtenidos a partir de la Encuesta Social Europea.

En general, la precisión será mayor para la medición de los atributos que para los comportamientos, y mayor en estos que en las actitudes. Pero también existen actitudes “claras” (no creyente) y comportamientos vagos (“casi nunca va a misa”). En el caso de la medición de atributos, la estimación de los barómetros analizados parece ser tan precisa o más de lo que se esperaría de un muestreo aleatorio simple.

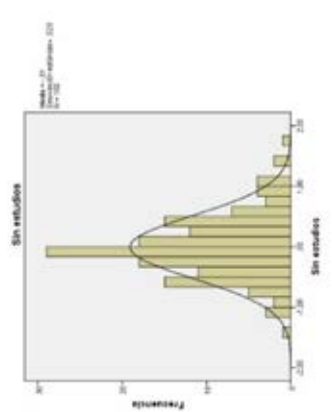
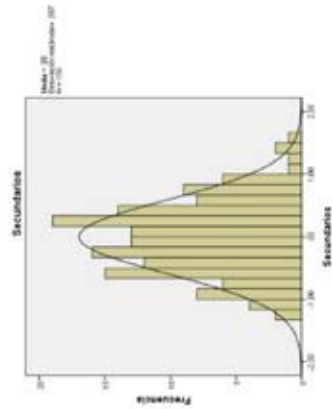
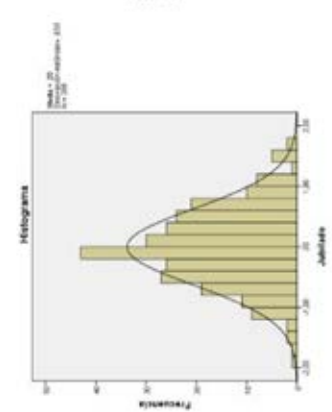
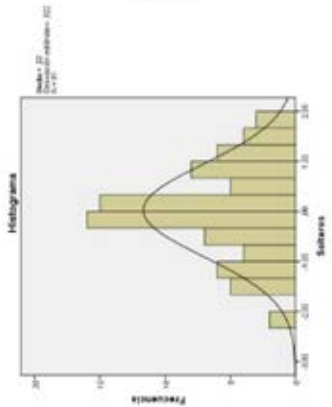
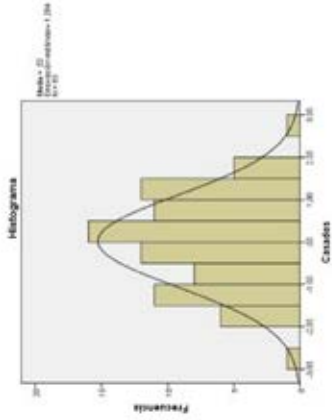
Hoy en día existe una importante erosión de la confianza en las encuestas electorales, y esto hace que sea aún más relevante incrementar los análisis e instrumentos que proporcionen una mayor fiabilidad en los procesos de recogida de la información. Por ello, sería aconsejable que las instituciones o empresas demoscópicas dedicadas al análisis de la opinión pública (o a la realización específica de encuestas electorales) que utilizan muestras de cuota realizaran una estimación *a priori* del error muestral utilizando fórmulas teóricas. Esta práctica ya es habitual en la Encuesta Social Europea, considerando dichas cuotas como si fueran parte del diseño muestral completo, lo que amplía el margen de confianza. Sería aconsejable que encuestas aplicadas en el contexto latinoamericano, que utilizan este tipo de diseño muestral, replicasen dicha práctica para incrementar dicho margen de confianza.

Como investigaciones pendientes, podrían utilizarse estrategias empíricas alternativas para la estimación del error, como el remuestreo simulado, siempre que se permita el acceso a los datos que lo hacen posible, como es tener la distribución censal de las personas entrevistadas. Esta dificultad ha supuesto una limitación al trabajo aquí realizado, ya que para garantizar la anonimización, no se dispuso de dicha información. También se podría analizar la variación repetida en encuestas sobre un tema estable donde poder comprobar si buenas formulaciones de preguntas, al tener menor variación de interpretación que aquellas con algún error de formulación, tendrían menor variación. Finalmente, sería interesante en futuros trabajos comprobar estas estrategias analíticas con encuestas que utilizan otros modos de administración (telefónicos, online, mixtos).

- Abundis, F., Becerra, L., Berrueto, F., Berumen, E.F., De La Rosa, M., Durand, C., Flores, C., Juárez, J., Moreno, A., y Penagos, D. (2017). *La precisión de las encuestas electorales: un paradigma en movimiento*. Instituto Nacional Electoral. https://www.ine.mx/wp-content/uploads/2019/11/la_precision_de_las_encuestas_Vol-II.pdf
- Baker, R. J., Brick, M. Bates, N.A., Battaglia, M., Couper, M.P., Dever, J.A., Gile, K.J. y Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, 1(2), 90–143. <https://doi.org/10.1093/jssam/smt008>
- Berinsky, A. J. (2006). American Public Opinion in the 1930s and 1940s. The Analysis of Quota-Controlled Sample Survey Data. *Public Opinion Quarterly*, 70(4), 499–529. <https://doi.org/10.1093/poq/nfl021>
- Cantril, H. (1948). Opinion Trends in World War II: Some Guides to Interpretation. *Public Opinion Quarterly*, 12(1), 30–44. <https://doi.org/10.1086/265918>
- Curtice, J. y Sparrow, N. (1997). How Accurate are Traditional Quota Opinion Polls. *International Journal of Market Research*, 39(3), 1-14. <https://doi.org/10.1177/147078539703900302>
- Davis, J. A., y Smith, T. W. (1992). *The NORC General Social Survey: A user's guide*. Sage Publications, Inc. <https://doi.org/10.4135/9781483345246>
- Deville, J.C. (1991). A Theory of Quota Surveys. *Survey Methodology*, 17(2), 163-181.
- Díaz De Rada, V. y Martínez Martín, V. (2014). Random Route and Quota Sampling: Do They Offer Any Advantage over Probably Sampling Methods?. *Open Journal of Statistics*, 4, 391-401. <http://doi.org/10.4236/ojs.2014.45038>
- Durand, C., Johnson, T., Moreno, A. y Traugott, M. (2018). *Report of the WAPOR Committee Reviewing the Pre-election Polls in the 2017 Presidential Election in Chile*. World Association for Public Opinion Research (WAPOR). <https://wapor.org/wp-content/uploads/Final-WAPOR-Report-onChile-2017-Election.pdf>
- Frankel, M.R. y Frankel, L.R. (1987). Fifty Years of Survey Sampling in the United States. *The Public Opinion Quarterly*, 51(2) Suppl., S127-S138. https://doi.org/10.1093/poq/51.4_PART_2.S127
- Ganninger, M. (2006). *Estimation of Design Effects for ESS Round II*. European Social Survey Documentation. https://www.europeansocialsurvey.org/docs/round2/methods/ESS2_estimation_of_design_effects.pdf
- Gaziano, C. (2005). Comparative Analysis of Within-Household Respondent Selection Techniques. *Public Opinion Quarterly*, 69(1), 124-157. <https://doi.org/10.1093/poq/nfi006>
- Gschwend, T. (2005). Analyzing Quota Sample Data and the Peer-Review Process. *French Politics*, 3, 88-91. <https://doi.org/10.1057/palgrave.fp.8200068>
- Kish, L. (1998). *Quota sampling: Old Plus New Thought*. Web Survey Methodology. http://www.websm.org/uploadi/editor/doc/1458891348Kish_1988_Quota_Sampling.pdf
- Kish, L. (1965). *Survey Sampling*. John Wiley.
- Lohr, S.L. (2010). *Sampling: Design and Analysis*. Brooks/Cole
- Lynn, P. y Jowell, R. (1996). How Might Opinion Polls be Improved? The Case for Probability Sampling. *Journal of the Royal Statistical Society*, 159(1), 21-28. <https://doi.org/10.2307/2983465>

- Martínez, V. (1999). Diseño de encuestas de opinión: Barómetro CIS. *QÜESTIÓ*, 23(2), 343-362. <http://hdl.handle.net/2099/4111>
- Mateos, A. y Penadés, A. (2013). Las encuestas electorales en la prensa escrita (2008-2011). Errores, sesgos y transparencia. *Metodología de Encuestas*, 15, 99-119. <http://casus.usal.es/pkp/index.php/MdE/article/view/1039>
- Moreno, A., Aguilar, R., y Romero, V. (2014). Estimaciones de encuestas preelectorales en México: en busca de las principales fuentes de error. *Revista Latinoamericana de Opinión Pública*, (4), 44-93. <http://hdl.handle.net/10366/142669>
- Moser, C. y Stuart, A. (1953). An experimental study of quota sampling. *Journal of the Royal Statistical Society, series A* 116, (4), 349-405. <https://doi.org/10.2307/2343021>
- Page, B.I. y Shapiro, R.Y. (1992). *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences*. University of Chicago Press.
- Selb, P. Chen, S., Körtner, J. y Bosch, P. (2023). Bias and Variance in Multiparty Election Polls. *Public Opinion Quarterly*, 87(4), 1025-1037, <https://doi.org/10.1093/poq/nfad046>
- Shirani-Mehr, H., Rothschild, D., Goel, S. y Gelman, A. (2018). Disentangling Bias and Variance in Election Polls. *Journal of the American Statistical Association*, 113(522), 607-614. <https://doi.org/10.1080/01621459.2018.1448823>
- Stephan, F. F. y McCarthy, P.J. (1958). *Sampling Opinions. An Analysis of Survey Procedure*. John Wiley.
- Stephenson, C. B. (1979). Probability Sampling With Quotas: An Experiment. *Public Opinion Quarterly*, 43(4), 477-496. <http://www.jstor.org/stable/2748547>
- Sturgis, P., Baker, N., Callegaro, M., Fisher, S., Green, J., Jennings, W., Kuha, W., Lauderdale, B. y Smith, P. (2016). *Report of the Inquiry into the 2015 British general election opinion polls*. Market Research Society and British Polling Council. <https://eprints.ncrm.ac.uk/id/eprint/3789/>
- Sturgis, P., Kuha, J., Baker, N., Callegaro, M., Fisher, S., Green, J., Jennings, W., Lauderdale, B.E. y Smith, P. (2018). An assessment of the causes of the errors in the 2015 UK general election opinion polls. *Journal of the Royal Statistical Society, series A* 181, (3), 757-781. <https://doi.org/10.1111/rssa.12329>
- Sudman, S. (1966). Probability Sampling with Quotas. *Journal of the American Statistical Association*, 61(315), 749-771. <https://doi.org/10.1080/01621459.1966.10480903>
- Taylor, H. (1995). Horses for Courses: How Survey Firms in Different Countries Measure Public Opinion with Very Different Methods. *International Journal of Market Research*, 37(3), 1-9. <https://doi.org/10.1177/14707853950370030>
- Taylor, H. (1997). The Very Different Methods Used to Conduct Telephone Surveys of the Public. *International Journal of Market Research*, 39(3), 1-13. <https://doi.org/10.1177/14707853970390030>

En el gráfico A1 se muestran los trece histogramas con la distribución de los errores de las trece variables escogidas, tomando la media móvil como representación de la verdadera tendencia. Como puede apreciarse visualmente, no todos los casos se ajustan igualmente al supuesto de distribución normal de los errores. La prueba de normalidad de Kolmogorov-Smirnova (tabla A1) afirma que se puede rechazar que sean normales dos casos (asistencia a misa y estar casado), pero esto no se confirma por la prueba de Shapiro-Wilk; esta segunda prueba recomienda rechazar la hipótesis de normalidad para el caso de los “no creyentes”, aunque eso no se confirma por la primera prueba. A simple vista, un problema que se deriva del uso de la media móvil como parámetro es el apuntamiento de la distribución, con una concentración no normal de valores cercanos al centro.



Fuente: elaboración propia.

Tabla A1. Pruebas de normalidad

	Kolmogorov-Smirnova			Shapiro-Wilk		
		gl	p		gl	p
Ideología media (1-10)	0,029	236	>0,200	0,995	236	0,591
Conservador	0,101	37	<,200	0,977	37	0,637
Liberal	0,081	37	<,200	0,988	37	0,959
Socialista	0,087	37	<,200	0,977	37	0,631
Católico	0,052	145	<,200	0,991	145	0,524
No creyente	0,057	145	<,200	0,969	145	0,003
Misa casi nunca	0,054	145	<,200	0,993	145	0,74
Misa casi todos los domingos	0,076	145	0,039	0,989	145	0,287
Jubilado	0,029	268	<,200	0,997	268	0,924
Casados	0,056	81	0,039	0,991	81	0,827
Solteros	0,067	81	<,200	0,979	81	0,219

Fuente: elaboración propia.