

# Encuestas y margen de error: una guía práctica

JOSU MEZO

En mis primeras colaboraciones en esta sección<sup>1</sup> hablé de algunos conocimientos básicos sobre las encuestas, que, en mi opinión, todo periodista debería tener. Por ejemplo, que solo debemos tomarnos en serio encuestas basadas en muestras aleatorias (es decir, con selección al azar de los entrevistados); que sí se puede hacer una encuesta solamente a unos cientos de personas y estimar lo que piensan o hacen millones, pero que esa estimación es aproximada; por lo que, al leer los resultados de las encuestas, hay que tener en cuenta el margen de error. Es decir, por poner un caso, si el 52 % de los entrevistados en una encuesta da una cierta respuesta, no podemos decir alegremente que “más de la mitad de los españoles” tienen tal opinión, porque ese 52 % es solo una aproximación al porcentaje que realmente existe en la población. Sabemos que no está lejos del 52 %, pero podría ser mayor o menor del 50 %.

En el presente artículo pretendo ahondar en este asunto del carácter aproximado de las encuestas, ayudando a los lectores sin formación en estadística a interpretar

mejor sus resultados. Para ello, he creado un archivo Excel, disponible *online* para su descarga, con el que será posible, en diferentes situaciones con encuestas, calcular el margen de error y los valores más probables en la población.

Antes de hacerlo, hay que recordar el significado del margen de error. Imaginemos que en una población con un 25 % de personas jóvenes (menores de 25) extrajéramos muchas muestras aleatorias de 600 personas y contáramos el porcentaje de jóvenes. En la mayoría de esas muestras, el porcentaje sería cercano al 25 %, pero no idéntico. Las matemáticas nos permiten calcular cuántas muestras se apartarían hasta un 1 %, un 2 % o un 3 % del porcentaje en la población. O al revés, podemos calcular cuánto se apartarían de ese porcentaje el 50 %, el 75 %, el 95 % o cualquier otro porcentaje de las muestras. En concreto, en este caso, el 95 % de las muestras no se apartarían más del 3,5 %.

Verdaderamente, solo elaboramos una encuesta. Cuando lo hacemos, podemos decir que hay un 95 % de probabilidades

**Josu Mezo** es profesor de la Universidad de Castilla-La Mancha y editor del blog *Malaprensa.com*.

---

1.- “Encuestas y muestras: tres dudas y tres reglas” (*Cuadernos de Periodistas*, n.º 6, pp. 174-185) y “Encuestas: el margen de error, la paja y el grano” (*Cuadernos de Periodistas*, n.º 7, pp. 90-102).

de que sea una de esas encuestas que no se aparta más del 3,5 % del porcentaje en la población. Decimos entonces que 3,5 % es el “margen de error” de esta encuesta, para un “nivel de confianza” del 95 %. Es decir, si al dato que obtenemos en la muestra le sumamos y restamos un 3,5 %, obtendremos un rango de datos al que llamamos “intervalo de confianza”, dentro del que estimamos que se encontrará el dato real en la población, con un 95 % de probabilidades de acertar.

Entendido esto, podemos empezar ya a practicar con la hoja de cálculo disponible *online*<sup>2</sup>. Tras descargarla y abrirla en un programa como Excel o similar, vamos a empezar con la primera hoja (“1. Un valor, una encuesta”), que sirve para situaciones como la inicial, en la que tenemos una encuesta, un porcentaje de personas que han escogido una determinada respuesta a una pregunta y queremos saber cuál es el porcentaje en la población. Las únicas celdas que es necesario rellenar son las que tienen un marco rojo (celdas B3 y B4). Opcionalmente, se pueden rellenar también las celdas con marco grueso negro (B6 y B7).

A la derecha de la hoja (columnas D y E), hay un ejemplo explicativo de cómo

se rellenarían las celdas, y cuál sería el resultado (véase el Gráfico 1). En una encuesta a 1.100 personas, el 18 % dice que va a votar a un determinado partido. En la celda “valor de referencia” se ha escrito el número 20, que es el porcentaje de voto que tuvo el partido en la elección anterior. En las filas siguientes vemos que el margen de error, con un 95 % de confianza, es del 2,27 % y el intervalo de confianza va del 15,73 % al 20,27 %. Es decir, tenemos un 95 % de probabilidad de acertar si afirmamos que el valor real en la población está entre el 15,73 % y el 20,27 %. Ese intervalo de confianza incluye el valor 20. Por eso, en la celda final aparece un “No” bien grande que nos dice que no podemos estar seguros, al 95 % de confianza, de que, en realidad, el porcentaje de personas con intención de votar a ese partido no siga siendo el 20 %. [El Gráfico 1 se puede ver en la siguiente página].

Aquí se ve la utilidad de tener en cuenta el margen de error: una lectura ingenua de la encuesta diría que “el partido A pierde un 2 % de apoyo respecto a la elección anterior”. Entendiendo correctamente lo que significa una encuesta, no podemos decir eso realmente, puesto que no estamos seguros al 95 % de que sea así. Si copiamos los datos del ejemplo en la

---

2.- La hoja de cálculo se puede descargar de la dirección URL <http://tiny.cc/margenerror>. En ella se ve la hoja (fragmentada en páginas, como si se fuera a imprimir), aunque no se puede manipular. Para usarla, hay que descargarla usando el botón con el icono de descarga que aparece en la parte superior. En la dirección <http://tiny.cc/margenerror2> hay otra versión en formato de Excel compatible con versiones antiguas. Hasta donde he podido comprobar, en esa versión, las fórmulas funcionan correctamente, pero es posible que haya algunos errores de formato. En el archivo Excel se incluyen algunas aclaraciones y notas sobre su uso adicionales a las que se dan en este artículo.

GRÁFICO 1

	D	E	F	G	H
	Ejemplo: encuesta a 1.100 personas, 18% dice que votará al partido A. En la elección anterior tuvo un 20% de los votos				
Datos necesarios					
Tamaño muestra		1100			
Porcentaje en la muestra		18			
Datos opcionales:					
Nivel de confianza (%)		95			
Valor de referencia		20			
<b>RESULTADOS</b>					
<b>Margen de error</b>		2,27	Interpretación: podemos decir, con un 95% de probabilidades de acertar, que el porcentaje de personas que piensa votar al partido A en la población está entre el 15,73% y el 20,27%		
<b>Intervalo de confianza</b>					
Mínimo		15,73			
Máximo		20,27			
	<b>¿Podemos decir, con el nivel de confianza elegido, que la proporción en la población es diferente del valor de referencia?</b>				
	<b>No</b>		Interpretación: no estamos al 95% seguros de que el porcentaje de votantes del partido sea distinto del 20% que tuvo en la elección anterior		

columna B, pero cambiamos el nivel de confianza del 95 % al 90 %, veremos que el margen de error baja hasta el 1,91 % y el intervalo de confianza ya no incluye el valor 20. Por eso, en la celda inferior aparecerá entonces, con fondo verde, un gran “Sí”. Es decir, si nos conformamos con un 90 % de probabilidades de acertar, sí podemos decir que la intención de voto al partido A es menor que el 20 % que obtuvo en la elección anterior.

El valor de referencia es opcional porque no siempre tendremos un valor claro para la comparación. En las encuestas electorales tiene sentido utilizar el porcentaje de voto en una elección anterior. En las encuestas de otro tipo, por ejemplo, de apoyo o rechazo a una medida política, se puede utilizar el valor 50 para ver si el apoyo queda por debajo

o por encima de la mitad de la población. No obstante, en otros casos no habrá un valor de referencia claro, y podemos dejar esa celda en blanco, observando simplemente cuál es el margen de error y el intervalo de confianza (e ignorando la última celda con el “No” o el “Sí”).

Lo que no hay que hacer es utilizar como valor de referencia el porcentaje obtenido en una encuesta anterior, ya que ese valor está sometido, a su vez, a un margen de error. De hecho, precisamente la segunda pestaña de la hoja de cálculo (“2. Dos encuestas”) está pensada para comparar dos encuestas. Las celdas que tenemos que rellenar en este caso (de nuevo, enmarcadas en rojo) son cuatro, con el tamaño de la muestra y el porcentaje de cada una de las encuestas (celdas B2, B3, B5 y B6). Como en la pestaña anterior, también

podemos, opcionalmente, cambiar el valor de la celda con el nivel de confianza (B8); por defecto, lleva el valor 95.

El funcionamiento de esta hoja se ve en el ejemplo de la derecha (columnas D y E), reproducido en el Gráfico 2, en el que se ha supuesto que estamos ante dos encuestas en una ciudad –la primera con una muestra de 600 personas y la segunda con una muestra de 400–, y se quiere comparar el porcentaje que cree que la ciudad está cada vez más limpia. Ese porcentaje ha pasado del 25 % al 20 %, con una diferencia, por tanto, de -5 %. La hoja de cálculo nos dice que el margen de error para la diferencia es 5,23 %, con un intervalo de confianza entre -10,23 % y 0,23 %. Es decir, que estimamos, con una probabilidad de

acertar del 95 %, que la variación real de la gente que piensa que la ciudad está cada vez más limpia en la población va de -10,23 % (ha bajado algo más de 10 puntos) hasta +0,23 % (ha subido muy ligeramente). Dicho de otra forma, como el intervalo de confianza de la diferencia incluye el valor cero, no estamos seguros de que la diferencia en realidad no sea cero. Pudiera ser que la opinión pública en ese asunto no hubiera cambiado en absoluto, pero por puro azar, por la variación normal de una encuesta a otra, nos sale ahora un resultado que es un 5 % más bajo que en la encuesta anterior. Por ello, la última celda del ejemplo responde con un “No” grande a la pregunta “¿Podemos decir, con el nivel de confianza elegido, que la proporción en las dos poblaciones es distinta?”.

GRÁFICO 2

D	E	F	G	H
Ejemplo: dos encuestas en una ciudad. Con seis meses de diferencia. Preguntamos porcentaje que cree que la ciudad está cada vez más limpia.				
Primera encuesta				
Tamaño muestra	600			
Porcentaje	25			
Segunda encuesta				
Tamaño muestra	400			
Porcentaje	20			
Nivel de confianza (95%)	95			
<b>RESULTADOS</b>				
Diferencia (2ª-1ª)	-5			
Margen de error de la diferencia	5,23			
<b>Intervalo de confianza de la diferencia</b>				
Mínimo	-10,23			
Máximo	0,23			
<b>¿Podemos decir, con el nivel de confianza elegido, que la proporción en las dos poblaciones es distinta?</b>				
<b>No</b>		Interpretación: no estamos al 95% seguros de que el porcentaje que piensa que la ciudad está cada vez más limpia haya cambiado		

Si copiamos los datos del ejemplo en la columna B, podemos ver que el margen de error de la diferencia se reduce, y el intervalo de confianza ya no incluye el valor cero, si bajamos el nivel de confianza al 90 % o si suponemos que la primera muestra no es de 600, sino de 800 personas. Con esos datos sí podríamos decir que han disminuido, en la población, los que piensan que la ciudad está cada vez más limpia.

Esta pestaña no solo sirve para dos encuestas realizadas en el mismo lugar en distinto momento. También permite comparar dos encuestas realizadas en dos lugares (dos países, dos ciudades), y se puede utilizar de la misma forma si lo que tenemos es una sola encuesta, pero queremos comparar las respuestas de subgrupos de la población: hombres y mujeres, jóvenes y viejos, o parados y ocupados. El único problema, en este último caso, es que necesitaremos averiguar cuál es el tamaño de cada una de las submuestras. Es decir, por ejemplo, si nos llegan los datos con una muestra de 1.000 personas, y nos dan el porcentaje de hombres y mujeres que tienen una opinión, en las celdas correspondientes al tamaño de la muestra no deberíamos introducir el valor 1.000, sino el número de hombres y mujeres entrevistados, dato que no siempre estará disponible. La misma dificultad se plantea para comparar grupos de edad, por territorios, por ideologías...

Finalmente, la tercera pestaña de la hoja de cálculo (“3. 2 respuestas a 1 pregunta”) es la que se debe utilizar cuando queremos comparar los porcentajes de dos respuestas a una misma pregunta en una misma encuesta. Por ejemplo, los porcentajes de apoyo a dos partidos, los favorables o contrarios a una idea o, en general, los que escogen dos respuestas en cualquier pregunta con varias respuestas posibles, y en la que cada persona solo puede escoger una opción. En este caso, las celdas a rellenar obligatoriamente (marco rojo) son tres: el tamaño de la muestra y los porcentajes que eligieron las dos respuestas que queremos comparar (celdas B2, B3 y B4). Opcionalmente, como las otras veces, se puede cambiar el nivel de confianza (B6), que, por defecto, está rellena con el valor 95.

En el ejemplo de las celdas D y E (Gráfico 3) se puede ver cuál sería el resultado si tuviéramos una encuesta con una muestra de 800 personas, que ante un referéndum nos dijera que el 51 % de los encuestados apoya el sí y el 46 % apoya el no (el resto escoge otras respuestas). El resultado es un margen de confianza de la diferencia (en este caso, 6,82 %) y un intervalo de confianza de la diferencia (de -1,82 % a +11,82 %). La interpretación es similar al caso anterior: con un nivel de confianza del 95 % solo podemos decir que la diferencia entre los partidarios del sí y del no en la población va de -1,82 % a +11,82 %. Por eso, de nuevo, en la última celda, un “No” nos

GRÁFICO 3

	D	E	F	G	H
Ejemplo: intención de voto en un referéndum. Encuesta a 800 personas, 51% responde "sí", 46% responde "no".					
Tamaño muestra		800			
Porcentaje respuesta A		51			
Porcentaje respuesta B		46			
Nivel de confianza		95			
Diferencia (A-B)		5			
Margen de error de la diferencia		6,82	Interpretación: pensamos, con 95% de probabilidades de acertar, que en la población, el porcentaje de personas que apoya el sí en el referéndum es entre un 1,82% menor y un 11,82% mayor que el que apoya el no		
Intervalo de confianza de la diferencia					
Mínimo		-1,82			
Máximo		11,82			
<b>¿Podemos decir, con el nivel de confianza elegido, que la proporción en la población de las dos respuestas es distinta?</b>					
<b>No</b>			Interpretación: no estamos al 95% seguros de que, en la población, el porcentaje que apoya el sí es mayor que el que apoya el no		

avisa de que no podemos decir, con el nivel de confianza escogido, que los dos porcentajes no sean iguales en la población.

En este caso, si copiamos los datos del ejemplo en la columna B, veremos que, aunque bajemos el nivel de confianza al 90 %, seguimos sin estar seguros de si son más los partidarios del sí o los del no. Solo bajando el nivel de confianza al 84 % (es decir, aceptando un riesgo de equivocarnos del 16 % o casi una de cada seis veces), podríamos decir que los partidarios del sí son más que los del no.

Con esta herramienta, en conclusión, es posible calcular el margen de error y los

valores más probables en la población en tres tipos de porcentajes derivados de encuestas: una encuesta con un solo porcentaje, dos encuestas con porcentajes de la misma pregunta y dos porcentajes de dos respuestas a la misma pregunta. Podemos evitar con ella las interpretaciones demasiado ingenuas de las encuestas, que toman sus valores por ciertos absolutamente e ignoran el margen de error. El resultado puede ser menos excitante y rotundo, porque muchas veces nos llevará a una conclusión del tipo “no estamos seguros” o “no sabemos”. La alternativa es dar por seguras cosas que no lo son, y creernos a pies juntillas afirmaciones muy dudosas. Puede dar para mejores titulares, aunque para peor periodismo. ■