

## Cuotas en investigación online: ¿qué son y cómo usarlas?

Obtener una muestra representativa del universo que queremos estudiar (lo que se conoce como “muestreo”), es una condición necesaria para obtener resultados fiables en una encuesta.

La forma más directa de obtener muestras representativas es el muestreo aleatorio simple (M.A.S.). En este tipo de muestreo, todos los individuos que constituyen el universo estudiado tienen igual probabilidad de ser seleccionados para formar parte de la muestra. Sería algo así como poner a todos los individuos que forman el universo en una urna e ir extrayendo por azar individuos hasta completar el tamaño de muestra deseado. El M.A.S. garantiza la ausencia de sesgos en el proceso de selección de la muestra.

Pero en la práctica es extremadamente difícil usar este método de muestreo. Para poder usar M.A.S. se requieren varias condiciones: la primera es disponer de un marco muestral de referencia fiable, es decir, un listado completo de los individuos que constituyen el universo. También necesitamos acceso a dichos individuos y la predisposición de los mismos a participar. Un individuo que rechace participar en una encuesta sería suficiente para invalidar un proceso de muestreo aleatorio perfecto.

### Muestreo por cuotas

Las condiciones anteriores son muy exigentes, haciendo que en la práctica el M.A.S. sea imposible o extremadamente costoso. Por ello, suelen usarse otras técnicas de muestreo, como el muestreo por cuotas.

El muestreo por cuotas pretende obtener una muestra representativa a partir de una selección de individuos no necesariamente aleatoria, forzando una distribución en relación a unas variables concretas idéntica a la distribución del universo estudiado.

Para usar muestreo por cuotas es necesario tener en cuenta dos cosas:

- (1) Necesitamos seleccionar las variables relevantes para el objeto de nuestro estudio. Por ejemplo, si estamos diseñando un estudio electoral, variables como la “edad”, “ubicación geográfica” o “clase social”, son variables relevantes porque pueden condicionar la orientación del voto. Dicho de otro modo, “edad” y “orientación del voto” son variables dependientes. Por el contrario, la “estatura” o el “nombre” del individuo son variables irrelevantes en un estudio electoral.
- (2) El segundo factor a considerar es que necesitamos conocer la distribución de dichas variables relevantes en el universo estudiado. De lo contrario, no podremos reproducir dicha distribución en la muestra. Por ejemplo, si queremos hacer un estudio sobre población general y deseamos fijar una cuota sobre la variable “edad”, podemos recurrir a estudios censales y otras fuentes de información proporcionados por los organismos oficiales de cada país. Estos estudios se realizan, o bien empleando acceso directo a todos los individuos del universo estudiado, o bien mediante un método de muestreo muy cercano al M.A.S.. Ambas cosas nos garantizan la fiabilidad de la información que manejamos para el universo.

### ¿Cuántas y qué variables debo emplear en mis cuotas?

Fijar cuotas en un diseño muestral tiene coste. Con independencia de la metodología empleada (cara a cara, telefónico u online), agregar cuotas encarece el trabajo de campo ya que obliga a descartar posibles participantes cuando exceden un objetivo que nos hemos fijado. Es por ello, que para seleccionar variables relevantes debemos tener en cuenta las siguientes consideraciones:

- (1) Algunas variables relevantes pueden ser redundantes (es decir, altamente correlacionadas), por lo que controlar sólo una de ellas puede ser suficiente. Por ejemplo, si fijamos una cuota sobre la variable “clase social” posiblemente pueda prescindir de otras variables como “nivel de ingresos”, “metros cuadrados de la vivienda” o “nivel de estudios”.
- (2) Si la forma en que selecciono individuos para la muestra, pese a no ser puramente aleatoria (M.A.S.), sí que garantiza aleatoriedad respecto a la variable relevante, puedo prescindir de incorporarla a las cuotas. Por ejemplo, podría estar obteniendo una muestra entre los lectores de una página web de noticias deportivas, y pese a que claramente no tendría una muestra aleatoria (todos los individuos son usuarios de Internet, aficionados al deporte y lectores del *site* en cuestión), podría darse el caso de que geográficamente los lectores de esta página web se distribuyesen de forma idéntica al universo estudiado. En este caso no necesitaría fijar cuotas por la variable “región”.

Considerando los dos criterios anteriores, deberíamos seleccionar las variables realmente esenciales que permitan garantizar un buen nivel de representatividad de la muestra sin encarecer excesivamente el estudio.

### ¿Cuotas cruzadas o cuotas no cruzadas?

Cuando realizamos un control de cuotas sobre dos o más variables, podemos definir dicho control de forma cruzada o no cruzada. Veámoslo con un ejemplo: supongamos que queremos obtener una muestra de 1.000 personas para un estudio electoral y hemos identificado dos variables relevantes: sexo (50% hombres y 50% mujeres) y edad (50% menores de 40 años y 50% mayores de 40 años).

Si imponemos un control de cuotas no cruzadas, exigiremos que del total de individuos de la muestra (1.000), 500 sean hombres y 500 mujeres, y que 500 sean menores de 40 años y 500 mayores. Dicho de otra forma, una muestra de 500 hombres menores de 40 años y 500 mujeres mayores de 40 años sería válida.

Por el contrario, si definimos una cuota cruzada sexo-edad, exigiremos que la muestra se componga de 250 hombres menores de 40, 250 hombres mayores de 40, 250 mujeres menores de 40 y 250 mujeres mayores de 40.

Definir cuotas cruzadas es más costoso que definir las no cruzadas. Para saber si es necesario cruzar cuotas, debemos valorar nuevamente relevancia y método de selección de los individuos:

- (1) ¿Influyen las variables en el objeto del estudio de idéntica forma al ser consideradas por separado o conjuntamente? Si dos variables relevantes son independientes entre ellas, podremos considerarlas como “no cruzadas”. En el ejemplo anterior, si la “edad” influye en la “orientación del voto” del mismo modo para hombres que para mujeres, podríamos emplear cuotas “no cruzadas”.
- (2) ¿La forma en que selecciono individuos para mi muestra puede favorecer la sobre-representación de alguno de los grupos definidos por la combinación de variables relevantes? Si la forma en que selecciono individuos garantiza *per se* que las proporciones de ambas variables cruzadas es la correcta, podría renunciar a definir este control.

### Cuotas offline y cuotas online

Las variables más empleadas para definir cuotas en cualquier metodología son las sociodemográficas: “sexo”, “edad”, “región”... Suelen ser variables con influencia en la mayor parte de temáticas objeto de estudio, y son fácilmente controlables, al existir fuentes de datos fiables con las cuáles comparar.

Sin embargo, cada metodología tiene sus particularidades que aconsejan adaptar el uso de las cuotas.

Cuando hacemos estudios mediante entrevista personal, es frecuente simplificar las cuotas geográficas por el enorme coste que tendría desplazar entrevistadores a todos los pueblos y ciudades de un país. Es por ello que suelen fijarse unas cuotas para las principales ciudades, asumiendo que dichas ciudades representan perfectamente el comportamiento de los individuos de una región amplia. Por ejemplo, en una muestra para Brasil, es frecuente solicitar un número determinado de individuos para São Paulo, Río de Janeiro y Recife. Para México, usaríamos Distrito Federal y Monterrey. O para España, Madrid y Barcelona.

Al hacer estudios telefónicos, es especialmente relevante considerar la variable “ocupación”, ya que al contactar con personas que se encuentran en su hogar, podríamos estar sobre-representando individuos desocupados y retirados.

El uso de cuotas en estudios online está todavía poco estudiado. Una práctica mayoritaria – y posiblemente errónea – es emplear las mismas cuotas que se definieron cuando el estudio se hacía offline.

A continuación, se listan una serie de consideraciones a tener en cuenta al usar Internet en la obtención de muestras:

- (1) La mayoría de investigaciones online emplean paneles: Es importante recordar cómo funcionan los paneles y con qué criterios reclutan a sus miembros. Por ejemplo: los paneles tratan de encuestar a todos sus miembros por igual. Por ello, si algunos perfiles concretos son más demandados en los estudios (mamás, personas con altos ingresos, etc.) pueden estar sobre-representados. Un panel online no es

representativo de la población por sí mismo, es un lugar en el que poder definir muestras representativas usando las cuotas adecuadas.

- (2) La dispersión geográfica no es un problema en online, al contrario, es una oportunidad. Si queremos obtener muestras representativas de una región, no necesitamos limitarnos a una ciudad representante de la misma como hacíamos en offline. Hacerlo reduce innecesariamente la representatividad y encarece el estudio, ya que a los paneles les resulta más natural reclutar miembros de regiones amplias que de una ciudad específica.
- (3) Todos los individuos usan Internet: Es una obviedad, pero es necesario tenerlo en cuenta. No podemos fijar cuotas sobre “no usuarios de Internet” o sobre “analfabetos”: en Internet todo el mundo sabe leer. Asimismo, el factor Internet puede correlacionar fuertemente con otras variables como el uso de tecnologías, lo cual aconsejaría fijar cuotas sobre este tipo de variables. Por ejemplo, si estamos haciendo un estudio sobre telefonía móvil, y el tipo de contrato (contrato/prepago) puede influir en los resultados que esperamos, en un panel online podemos encontrar sobre-representadas las modalidades de contrato propias de usuarios más avanzados (acceso a Internet móvil).
- (4) La adopción de Internet en cada país es clave para definir buenas cuotas: Si un grupo de población ha adoptado Internet en mayor medida que otro, su presencia en muestras online estará sobre-representada. Variables como “edad” y “clase social” son fundamentales, especialmente en América Latina, una región en la que existen diferencias sociales especialmente acusadas, que hacen que las clases más bajas accedan con mucha mayor dificultad a Internet.

En definitiva, cuando hacemos una encuesta online, las cuotas son una herramienta imprescindible para obtener representatividad, y la elección de las variables que empleamos en ellas debe realizarse atendiendo al problema objeto de estudio así como a la forma en que obtenemos la muestra.