

**1. En la encuesta los costos son altos y la administración es más compleja.
¿Por qué?**

Para responder esto se debe considerar contra qué se está comparando. Si la comparación es con un censo la encuesta es la más barata y presenta menos problemas de administración. Pero si es comparada con un estudio de participación de observadores como los grupos focales, efectivamente la encuesta sería más cara y más compleja. Es por eso que hicimos referencia a que es necesario tener claro cuando es más recomendable aplicar uno u otro, considerando lógicamente los objetivos de la investigación.

2. ¿Qué es una pregunta bidimensional?

Se refiere a la pregunta que, debido a su redacción, la persona al responder puede dar dos opciones de respuesta, lo cual no es correcto, ya que debe ser sólo una, por eso se habla de evitarlas.

3. Pasos para el diseño de la encuesta (...) Verificar la coherencia de los datos obtenidos, así como las omisiones en las encuestas muestrales realizadas. Un cuestionario demasiado largo produce una baja general de la calidad de las respuestas, tanto a las preguntas importantes como a las otras. ¿Cuál es el criterio para decidir si un cuestionario es demasiado largo o no?

Eso depende de una serie de factores, siendo el inicial el método de entrevista a utilizar. Si se habla de una entrevista cara a cara el tiempo puede ser mayor a lo que sería con una entrevista telefónica, y por lo tanto el cuestionario puede ser más largo. Por eso es que es de mucha importancia tener claro que es lo que fundamentalmente se debe incluir en el cuestionario para valorar lo que sería posible de quitar en caso de ser necesario.

Otro elemento a valorar para saber si es largo o no es cuando se hace la prueba piloto del cuestionario, puesto que ahí es posible verificar si existen problemas relativos al tiempo tiempo, o si lo que estimamos inicialmente en tiempos se está cumpliendo o no.

4. El marco. Antes de seleccionarse la muestra, la población debe ser dividida en partes llamadas unidades de muestreo. Estas deben cubrir la totalidad de la población y no traslaparse en el sentido de que todo elemento de la población se clasifique en una y solamente a una unidad de muestreo. Frecuentemente la construcción de esta lista de unidades de muestreo llamada marco es uno de los principales problemas prácticos (...) aunque se asegure lo contrario, estas listas preelaboradas suelen ser

incompletas, parcialmente ilegibles o contienen duplicaciones de magnitud desconocida. ¿A qué se refiere con "duplicaciones"?

La duplicación se refiere a que por ejemplo en un marco de muestreo, que básicamente es la lista de la población a trabajar, una persona pueda estar dos veces, lo cual le daría mayor probabilidad de ser seleccionado en comparación con los otros.

- 5. Resumen y análisis de los datos. El primer paso tras encuestar debe ser editar los cuestionarios obtenidos para corregir errores o cuando menos desechar los datos que obviamente están equivocados. Habrá la necesidad de tomar ciertas decisiones respecto al procedimiento de cálculo en los casos de omisión de respuestas de quienes responden o de eliminación de datos en el proceso de edición. Después, se realizarán los cálculos que conduzcan a las estimaciones. Puede haber diferentes métodos de estimación para los mismos datos. Una práctica aconsejable en la presentación de los datos es informar la magnitud esperada de error en las estimaciones más importantes. Una de las ventajas del muestreo de probabilidad es que se pueden hacer tales enunciados (de error esperado) aunque deberán ser muy clasificados si la cantidad de no-respuestas es importante. Como lo obvio a veces no es tan obvio, ¿qué se debe entender aquí por "cuando obviamente están equivocados"? Además, respecto a "deberán ser muy clasificados si la cantidad de no-respuestas es importante", ¿qué significa aquí "clasificados"? así como también ¿cómo se hace uno la idea (al menos cualitativa) de qué implica "muy" y qué implica "importante" (que en el fondo está haciendo referencia a "significativo", pero ¿bajo qué parámetros cualitativos se define tal significancia para este caso en concreto?)?**

"Cuando obviamente están equivocados". Esto es parte del proceso de validación de los datos, por ejemplo, que una sección solo debió ser respondida por mujeres, y al revisar vemos que tenemos respuestas de hombres, en este caso hablamos que hay un error obvio, dado que no fue planeado tener respuestas de hombres, y por lo tanto se debe corregir (revisar si fue un error al digitar y si era información de una mujer, o eliminarlo).

"Deberán ser muy clasificados si la cantidad de no-respuestas es importante". Esto se refiere a que por ejemplo en una respuesta clave tenemos un porcentaje alto de personas sin responderla, en este caso básicamente no vamos a poder utilizar la pregunta.

En el caso de los otros términos eso puede depender de la investigación, lógicamente si se tiene una alta cantidad de no respuestas el problema se complica.

Por eso es importante nunca dejar incompleta una encuesta, así como también usar el sondeo en caso de ser necesario para tener las respuestas.

6. **El objetivo de la teoría de muestro es que este sea más eficiente. Su objetivo es desarrollar métodos de selección de muestras y de estimación, que proporcionen, al menor costo posible, estimaciones con la exactitud suficiente para nuestros propósitos. Para cumplir el objetivo anterior es necesario poder predecir en cualquier metodología de muestreo considerada la precisión y el costo esperado de la misma. Respecto a la precisión, no es posible predecir cuál será el error de una estimación en una situación específica, porque eso implicaría el conocimiento del verdadero valor de la población. *En lugar de ello, la precisión de un procedimiento de muestreo se juzga al examinar la distribución de frecuencia generada para las estimaciones, suponiendo que el proceso de muestreo se aplica varias veces a la misma población. ¿Qué significa la afirmación en cursivas?, ¿cuáles son sus consecuencias operativas?***

Cuando se calcula una muestra con, por ejemplo, un 95% de confianza, lo que se está diciendo es que, si se selecciona 100 muestras diferentes, en 95% de ellas vamos a obtener valores similares a los que obtuvimos con nuestra muestra. Por ello, para efectos de calcular una muestra siempre es necesario tener información clave sobre las variables que deseamos estudiar, a efectos de realizar cálculos que se ajusten a la realidad. En este caso, lo mínimo es poder ver cuál fue la varianza y promedio en estudios anteriores a fin de incluirlos dentro del cálculo de la muestra.

7. **¿Qué son las muestras de tamaño común?**

Se refiere a tamaños de muestra tradicionales, por ejemplo, para una muestra pequeña un tamaño común es de cerca de 400 entrevistas, y para el caso de tener datos a nivel nacional en modalidad telefónica se habla de 800 a 1000, lógicamente puede variar según el deseo de la investigación.

8. **Hay dos diferencias entre la teoría estándar de encuestas por muestreo y la teoría clásica del muestreo como aparece en los libros de texto.**
 - 8.1. **En la teoría clásica, las mediciones hechas sobre las unidades de muestreo de la población suelen suponerse que siguen una distribución de frecuencia de forma matemática conocida, como sería la distribución normal, cuyos parámetros, media y varianza, por ejemplo, se estimarían a partir de los datos de las muestras. Por otro lado, en la teoría de las encuestas por muestreo, se supone que sólo se dispone de una información muy limitada sobre dicha distribución y, sobre todo, no se supone conocida su forma matemática, así que el**

enfoque se puede describir como independiente de un modelo o de una distribución de frecuencia. Esta es una actitud natural para encuestas muy grandes en las que se efectúan numerosas mediciones diferentes de las unidades que siguen diversas distribuciones de frecuencia. Para las encuestas en las que sólo se realizan pocas mediciones en cada unidad, el estudio de sus distribuciones de frecuencia puede justificar la hipótesis de que son de forma matemática conocida, lo que permite la aplicación de la teoría clásica. ¿Cuánto es, más o menos (al menos cualitativamente) "pocas mediciones"?

Esto es algo que ha variado mucho en el tiempo, dada las limitaciones de cómputo que había. Antes se decía que, al utilizar las distribuciones, se hablaba de muestras pequeñas para menos de 30-50 casos. Ahora ya eso no se considera tanto, ya que existe mayor capacidad. Pero es algo variante.

8.2. Las poblaciones en una encuesta tienen un número finito de unidades. Los resultados son ligeramente más complicados cuando el muestreo es de una población finita y no de una infinita. Por razones prácticas, a menudo se ignoran estas diferencias en los resultados para poblaciones finitas e infinitas. Los casos en que no se ignoren, los indicaremos. ¿A qué se refiere con que "Los resultados son ligeramente más complicados cuando el muestreo es de una población finita y no de una infinita"? ¿Por qué en algunos casos se ignoran estas diferencias (en los resultados para poblaciones finitas e infinitas) y en otros casos no?

Población finita es cuando se conoce la cantidad total de personas/unidades a las cuales se puede entrevistar, infinita lo opuesto. Lógicamente conocer o no esta información pueden hacer más complicado los cálculos de muestreo.

9. El encuestador no solo debe evitar sus propias expectativas, sino que debe ser sensible a las expectativas del entrevistado. ¿Qué significa ser sensible a las expectativas del entrevistado?

Significa poder empatizar con el entrevistado. Ello es importante por dos cuestiones. En primer lugar, por motivos éticos y, en segundo lugar, porque ello permite una mejor disposición (sea consciente o no el entrevistado) al brindar las respuestas que usualmente deriva en un incremento en la calidad de los datos obtenidos.

10. La cortesía, el tono, la aspiración social, el consentimiento y la posición pueden conducir a sesgos que sean fuente de error en las entrevistas. ¿Por qué?

Porque todos estos motivos, asociados también con las expectativas de los entrevistados, pueden conducir al entrevistado a sesgarse de tal o cual forma y, derivado de ello, que los resultados obtenidos tras la aplicación del instrumento sean erróneos de tal o cual forma.

11. El diseño de la encuesta es mucho más importante que el tamaño absoluto de la población. ¿Por qué?

Porque un adecuado diseño de la encuesta significa un adecuado diseño del instrumento de captura de datos, con todas las implicaciones de ello. La afirmación de Lohr es relativamente cierta, en el sentido de que es cierta únicamente si se da por hecho la existencia de un tamaño de muestra que no invalide la investigación, es decir, si con antelación en ambos escenarios (enfocarse en el diseño de la encuesta versus enfocarse en el tamaño absoluto de la población) se cuenta con un tamaño de la población suficiente para no invalidar (a causa de una muestra muy pequeña) la investigación en alguno de los escenarios.

12. Se deben especificar cuatro elementos para definir a la población: “contenido”, “unidades”, “extensión” y “tiempo”. “Extensión” es el territorio geográfico en que se realiza la encuesta y el “tiempo” es el período temporal en el que se hace, pero ¿qué son en general (más allá de ejemplos concretos) “contenido” y “unidades”?

Contenido: Implica la localización espacio-temporal de la población, así como sus características fundamentales.

Unidades: Hace referencia tanto a unidades de muestreo, unidades de observación, unidades de listado y unidades informantes. Los tres tipos de unidades son importantes al momento de planificar una investigación que implique el diseño de encuestas por muestreo.

Unidad de Muestreo: Localización espacio-temporal de las unidades de estudio. Cada uno de los elementos que forman un universo o conjunto sometido a estudio estadístico. Por ejemplo, si estudio niños, un lugar de muestreo podrían ser los kínder.

Unidad de Estudio: Es la unidad (*i.e.*, la propiedad de todo ser, en virtud de la cual no puede dividirse sin que su esencia se destruya o altere) de la cual se necesita información, es el individuo o conjunto de individuos de donde se obtiene el dato; la unidad de estudio corresponde a la entidad que va a ser objeto de medición y se refiere al qué o quién es sujeto de interés en una investigación. Un elemento (o unidad) es un objeto sobre el que se realiza una medición de una variable bajo estudio.

Unidad de Listado: Son las listas de las unidades de muestreo individualmente consideradas. Se usan para identificar y seleccionar unidades de muestreo.

Unidad Informante: Es la unidad que provee al investigador directamente la información requerida por este.

Unidad de Observación: La unidad a la cual observa directamente el investigador.

13. El dominio es el conjunto de las subclases, ¿cuál es la definición de subclase y, por consiguiente, de clase?

Dominio: Subclases planeadas específicamente en el diseño de la muestra.

Estratos: Subdivisiones de la población en segmentos de diferente tipo.

Clase: Una clase o intervalo es un rango en el cual se distribuyen los datos que tienen características en común. Una subclase es entonces un subconjunto de una clase.

Marco Muestral: Listado de todas las unidades de muestreo.

14. ¿Qué son las variables Y_i ?

Son las características de la unidad de estudio.

15. ¿Cuál es el significado de Selección MESIP?

MESIP: Muestreo con igual probabilidad (en referencia a la selección de los elementos de una muestra).

16. ¿Qué significa “muestreo irrestricto aleatorio”?

Muestreo aleatorio tal y como se conoce.

17. El error cuadrático medio se relaciona con la varianza. ¿Cómo se relacionan exactamente?

Se relacionan por el hecho de que el ECM es la suma de la varianza del estimador y del cuadrado de su sesgo, *i.e.*, del cuadrado de la diferencia entre su valor observado y su esperanza.

18. ¿Qué son los conglomerados en el contexto de las encuestas por muestreo?

Diseño de muestreo	Unidad de muestreo			Probabilidades de selección	
	Elementos	Conglomerados		Iguales	Desiguales
		Iguales	Desiguales		
Muestreo aleatorio simple	1	4	8	12 (Mesip)	16
Muestreo sistemático	2	5	9	13 (Mesip)	17
Muestreo estratificado	3	6	10	14 (Mesip)	18
Muestreo politápico		7	11	15 PPT (dos, tres o + etapas) es Mesip	19 PPT (dos, tres o + etapas)

Al igual que en el contexto del Aprendizaje Automático, los *conglomerados* en el contexto de las encuestas por muestreo son agrupamientos de observaciones con determinadas características comunes. Estos conglomerados pueden ser iguales o diferentes entre sí, es decir, tener las mismas características o divergir en una o más de ellas. En el contexto antes descrito, en el de las encuestas por muestreo, los conglomerados son también conocidos como *Unidades Primarias de Muestreo (UPM)*. Las UPM son lo primero que el investigador debe seleccionar al iniciar el proceso de muestreo. Son los segmentos censales fundamentales. De hecho, eran conocidos bajo este nombre (segmento censal o SC) en la metodología utilizada anteriormente.

Se define por unidad de muestreo a los conjuntos no solapados de la población que cubren la población completamente. Si cada unidad de muestreo contiene uno y solamente un elemento de la población se le suele denominar unidad primaria. Si por el contrario contiene varios se le denomina compuesta.

19. ¿Qué significa el “margen de error” en el contexto de una encuesta?

Es la cantidad de error de muestreo aleatorio resultado de la elaboración de una encuesta.

20. ¿Por qué razones se prefiere el muestreo aleatorio? ¿Sólo una muestra obtenida vía muestreo aleatorio es entonces representativa? ¿Por qué?

El muestreo no aleatorio, conocido en algunas ocasiones como muestreo a juicio, es un muestreo válido para condiciones particulares. Por ejemplo, en casos en que el investigador realice un estudio en donde se sabe con antelación que al buscar personas que cumplan con tales o cuales características deseadas de forma aleatoria, no será posible encontrarlas, pero que también el investigador conoce con antelación que un grupo de personas localizado espacio-temporalmente en algún lugar, tiene esas características deseadas. Por ello, es significativamente más sencillo hacer el estudio con ese grupo y derivar de ello conclusiones.

Sin embargo, cuando se realice un muestreo no aleatorio, no necesariamente es posible realizar inferencias sobre la totalidad de la población con la que se está trabajando. Por ejemplo, al trabajar con un grupo de personas de una ciudad específica que tienen una característica de elevado interés para el investigador (y que no se encuentra presente en el resto de la población del país), no se podría decir por ello que el país al que pertenece esa ciudad presenta la misma característica del grupo de personas entrevistado en una de sus ciudades.

La ética profesional demanda especificar siempre el tipo de muestreo realizado y no hacerlo pasar por otro tipo de muestreo al presentar los resultados de la investigación.

21. La aleatorización es una operación física. ¿A qué se refiere con que debe ser exacta con un modelo probabilístico?

Hace referencia a que es un proceso con etapas cuya existencia es de carácter objetivo (en cuanto no-arbitrario, para que sea lo más independiente posible de las subjetividades), lo cual se mal define como “operación física”. Cuando se dice que “debe ser exacta con un modelo probabilístico” se está haciendo referencia a que esta “operación física” (cuyo significado se definió antes en términos objetivos) debe poder ser descrita con algún modelo probabilístico o, lo que es lo mismo, que este proceso y sus etapas debe poder ser descrito con precisión estadística por alguna distribución de probabilidad, algún conjunto de estas o algún modelo derivado de alguna forma de ellas.

22. El marco es un concepto más general: incluye listas físicas y también los procedimientos que pueden tener en cuenta a todas las unidades de muestreo sin el esfuerzo físico de listarlas efectivamente. ¿A qué se refiere con “listas físicas”? Por otro lado, ¿cuáles son esos “procedimientos que pueden tener en cuenta a todas las unidades de muestreo sin el esfuerzo físico de listarlas efectivamente”? Y, finalmente, ¿a qué se refiere con “listarlas efectivamente”?

En este contexto, al hablar de “esfuerzo físico” se está empleando la palabra “físico” en su significado literal, en contraste con los métodos computacionales automatizados mediante la Informática.

23. Por lo general, ese tipo de listados no existe y de existir, sería un proceso caro y administrativamente complejo. ¿Qué características tiene entonces tal listado que hacen se incrementen tanto sus costos y su complejidad administrativa al aplicarlos en poblaciones humanas? ¿O las características de las poblaciones humanas son las que explican eso? ¿O ambas y en qué medida (cualitativamente hablando, por supuesto)?

Puesto que los marcos muestrales son el listado de a quienes voy a entrevistar. Se supone que nadie en ningún país, salvo instituciones como el Tribunal Supremo Electoral, posee el listado completo de los habitantes de un país. Y si algún particular lo quisiese construir por su cuenta, tendría que enfrentar costos altísimos. Además, poder establecer la ubicación de todas las personas de un país en determinada fecha y hora que se desee entrevistarlas no es tarea simple, por ejemplo, para un país de seis millones de habitantes, puesto que las personas pueden encontrarse trabajando, estudiando o en otro lugar que no sea su domicilio electoral. Finalmente, debe considerarse la viabilidad económica de los desplazamientos asociados a localizar unidades informantes fuera de su domicilio electoral, y ello en caso sea posible determinar tales localizaciones ajenas al domicilio electoral de tales informantes.

24. ¿A qué procedimiento concreto se hace referencia cuando se establece que “El interés, entonces, es seleccionar SC o UPM para luego seleccionar las unidades de estudio (cualquiera que ésta sea), proceso que puede aplicarse a nivel nacional o en zonas menores (regiones, provincias, cantones, ¿urbano, rural, etc.)”?

Significa ir realizando progresivamente la selección en sus diferentes etapas en el diseño de la muestra. Por ello se habla de *muestreo polietápico*.

25. Uno de los problemas usuales en los marcos muestrales es encontrar conglomerados juntos en un mismo listado. ¿Operativamente qué significa esto y por qué es un problema?

En el contexto de los marcos muestrales (listados de todas las unidades muestrales), lo anterior significa que dentro de un marco muestral puede existir un conglomerado presente dentro de otro marco muestral. Un ejemplo de ello sería, en el contexto de encuestas a poblaciones humanas, el caso de dos teléfonos asignados para una misma vivienda, puesto que la información obtenida de dos unidades de muestreo (los teléfonos asociados a los domicilios electorales) correspondería a una unidad de estudio (una persona o familia) y una unidad informante (una persona).

Esto es un problema porque se viola el principio clásico de que todos los elementos deben tener la misma probabilidad de selección para que el muestreo sea aleatorio.