

# Muestreo: Diseño y Análisis.

Por: Sharon L. Lohr

## Introducción

*"Cuando las estadísticas no se basan en cálculos estrictamente precisos, inducen a error en lugar de servir de guía.*

*La mente fácilmente se deja engañar por la falsa apariencia de exactitud que las estadísticas llevan a mantener en sus errores, y con confianza adopta errores vestidos de forma de verdad matemática."*

*-Alexis de Tocqueville, La democracia en América*

### La polémica muestra

Shere Hite presenta en su libro: *Mujeres y Amor: Una revolución cultural en curso* (1987) un número de resultados ampliamente citados:

- 84% de las mujeres "no están satisfechas emocionalmente con sus relaciones" (Pág. 804).
- 70% de todas las mujeres "que se casaron hace cinco o más años tienen relaciones sexuales fuera de sus matrimonios" (Pág. 856).
- 95% de las mujeres "llenen formularios de hostigamiento emocional y psicológico a causa de los hombres con los que están en relaciones amorosas" (Pág. 810).
- 84% de las mujeres "utilizan formas de condescendencia con los hombres en sus relaciones afectivas" (P. 809).

El libro fue muy criticado por los diarios y revistas de todo Estados Unidos. La revista Times señaló en su portada *"Back Off, Buddy" -retírate Buddy-* (12 de octubre de 1987). Esta revista señalaba las conclusiones del estudio de Hite como "dudosas" y "de valor limitado". ¿Por qué el estudio Hite es tan criticado? ¿Fue un error de Hite el informe que cita a mujeres que sienten que los hombres en sus vidas se niegan a tratarlas como iguales, que tal vez nunca han tenido la oportunidad de hablar antes? ¿Fue un error el informe de los porcentajes de estas mujeres que son infelices en sus relaciones con los hombres? Por supuesto que no.

La investigación de Hite le permitió a las mujeres discutir sus experiencias, y reflejar la riqueza de las mismas de manera que un cuestionario de selección múltiple no hubiera podido hacerlo. El error de Hite fue la generalización de estos resultados a todas las mujeres, ya sea que participaran de la encuesta o no. Los siguientes rasgos hacen a la encuesta inadecuada para generalizar sus resultados:

- La muestra es de **auto-seleccionados**, es decir, los destinatarios de los cuestionarios decidieron si estarían en la muestra o no. Hite envió 100.000 cuestionarios por correo; de estos, sólo el 4,5% fueron devueltos.

Los cuestionarios también fueron enviados a organizaciones como "Professional women's group", centros de asesoramiento, sociedades de iglesia, y centros de la tercera edad. Los miembros pueden diferir en las opiniones políticas, pero una respuesta de su grupo representaría a "todas las mujeres" del mismo. La encuesta tiene 127 preguntas de desarrollo, y la mayoría de las preguntas tienen varias partes. ¿Quiénes tenderán a devolver una encuesta? Muchas de las preguntas son vagas. Un ejemplo es el uso de palabras tales como "amor". El concepto de "amor" probablemente tiene tantas interpretaciones como personas, lo que hace imposible adjuntar una sola interpretación en cualquier estadística que tenga por objeto indicar cuantas mujeres están "enamoradas". Esta redacción de la pregunta funciona bien para provocar y obtener datos llamativos que componen la mayor parte del libro, pero hace que la interpretación matemática de los porcentajes sea difícil.

Muchas de las preguntas guían la respuesta. Por ejemplo: *"¿Su esposo o amante la ven como una igual? ¿O hay momentos en los que parece que la tratan como un ser inferior? ¿La dejan fuera de las decisiones? Su palabra, ¿es la Ley superior?"* (Pág. 795). Hite escribe: *"¿Una investigación que no se base en la probabilidad o en una muestra al azar nos habilita a generalizar a partir de los resultados? Si un estudio es lo suficientemente grande y lo suficientemente amplia muestra, y si se generaliza con cuidado, sí"* (Pág. 778). La mayoría de los estadísticos respondería a la pregunta de Hite con un "no" rotundo. En la encuesta Hite, al elegir deliberadamente responder o no el cuestionario y al ser un porcentaje extremadamente pequeño las respuestas recibidas, las estadísticas calculadas a partir de estos datos no sirven para indicar actitudes generales. Porque objetivamente la muestra no es **representativa** de los Estados Unidos, y las estadísticas sólo describirían a las mujeres que han respondido. Hite afirma que los resultados de la muestra si pueden ser generalizado porque aborda características como la edad, la educación, y la ocupación coincidentes con las de la población total de mujeres norteamericanas. Sin embargo, hay un aspecto importante que Hite no toma en cuenta: Las mujeres que respondieron a su cuestionario estaban dispuestas a ocupar su tiempo en llenar un largo cuestionario sobre el acoso de los hombres y proporcionar información personal a un investigador. En cada grupo de edad y de clase socio-económica, las mujeres que deciden denunciar estas cuestiones han tenido experiencias diferentes que las mujeres que deciden no hacerlo; por lo que la muestra sólo representaría a este primer grupo.

## Requisitos de una buena muestra

En la película “*Magic Town*”, el encuestador interpretado por James Stewart descubre una ciudad que tiene exactamente las mismas características que todos los Estados Unidos: Grandview tiene la misma proporción de personas que votaron a un republicano, la misma proporción de personas bajo el umbral de la pobreza, la misma proporción de mecánicos automotrices, y así sucesivamente. Lo único que el personaje de Stewart tenía que hacer era entrevistar a la gente de Grandview, y conocería la “opinión pública” de todo Estados Unidos. Grandview sería el ejemplo perfecto; una versión a escala de la población, un reflejo de todas las características de la misma. Por supuesto, no existe una muestra tan perfecta en la realidad. Pero, una muestra como Grandview sería lo más cercano posible a una exactitud en representatividad. Representaría cada unidad de muestreo, las características de un número limitado de unidades de toda la población.

Algunas definiciones necesarias para la noción de una “buena muestra”: a) Observación de la “**unidad del objeto**” del que se toma la medida. Esta es la unidad de observación, a veces llamado “**elemento**”. En el estudio de las poblaciones humanas, “unidades de observación” son a menudo las personas. b) “**Población objetivo**”: La colección completa de las observaciones que queremos estudiar (la población destinataria). Por ejemplo, en una encuesta política, la “población objetivo”. ¿Son todos los adultos con derecho a voto? O ¿Todos los votantes inscritos? O ¿Todas las personas que votaron en las últimas elecciones? La definición de nuestra “población objetivo” afectara profundamente a las estadísticas que resulten. Ya que se selecciona un subconjunto de la población. c) “**Población de la muestra**”: La colección de todas las unidades de observación. La población sobre la que fue tomada la muestra. Se encuentra compuesta por “**unidades de muestreo**”. Los hogares pueden servir como unidades de muestreo, mientras que las unidades de observación serán los individuos que viven en esos hogares. d) El “**Marco de muestreo**” es la lista de unidades de muestreo. Para las encuestas telefónicas, “el marco” puede ser una lista de todos los números telefónicos residenciales en la ciudad (guía). Otro ejemplo, en el caso de entrevistas personales, sería la lista de todas las direcciones de los entrevistados, o en una encuesta agropecuaria, la lista de todas las explotaciones o un mapa de zonas en las que las explotaciones agrícolas.

En un estudio ideal, la población de la muestra sería idéntica a la población objetivo, pero este ideal rara vez se cumple. En las encuestas personales, la población de muestra es menor que la población objetivo. No todas las personas en la población objetivo se incluyen en el marco de la muestra, y siempre se estipula un número de personas que no responderán la encuesta. En el estudio de Hite, un dato de interés es el porcentaje de mujeres que son acosadas en su relación. Cada mujer en particular es un elemento. La población objetivo era todas las mujeres adultas en Estados Unidos. Mientras que la población de muestra con la que cuenta Hite es sólo las mujeres pertenecientes a organizaciones feministas que devolvieron el cuestionario. En consecuencia, las deducciones sólo se pueden hacer en base a la población analizada y no en base a la población de todas las mujeres adultas en Estados Unidos.

La población objetivo y población de la muestra en una encuesta telefónica de posibles votantes también presenta problemas. No todos los hogares tienen teléfono, así que un número de personas en la población objetivo no se asociara a un número de teléfono en el marco de muestreo. En algunos hogares con teléfono, los residentes no están registrados para votar y por lo tanto no son elegibles para la encuesta. Algunas personas elegibles en el marco de muestreo de la población no responden porque no pueden ser contactadas, algunos se niegan a responder a la encuesta, y algunos pueden ser reacios o incapaces de responder.

## Estadísticas.

Si la característica de interés fuera el número total de hogares en el Estados Unidos en los que algún habitante fue víctima de la delincuencia el último año, los elementos serían los hogares, la población objetivo estaría formada por todos los hogares en los Estados Unidos, y población de muestra estaría compuesta por los hogares en el “marco de muestra”, construido a partir del censo y los permisos de construcción.

El objetivo de la Encuesta Nacional de Plaguicidas, realizado por la Agencia de Protección del Medio Ambiente, fue el estudio de los plaguicidas y de nitratos en pozos de agua potable en todo el país. La población objetivo fueron todos los sistemas de agua comunitarios y los pozos rurales-domésticos de los Estados Unidos. La población analizada fue todos los sistemas de agua de la comunidad (todos los que se enumeran en la Ley Federal de informes de datos del sistema) y todos los pozos domésticos de identificación fuera de las reservas del gobierno que pertenecieran a familias dispuestas a cooperar con la encuesta. Estas encuestas de opinión pública a menudo se toman para predecir que candidato ganara las próximas elecciones. La población objetivo son las personas que votaran en las próximas elecciones, la población de la muestra son las personas que llegan al teléfono y dicen a quienes son propensos a votar en las próximas elecciones. Pocas encuestas nacionales en los Estados Unidos incluyen Alaska o Hawai, o personas en hospitales, residencias, o en las cárceles, porque no son parte de la marco de muestreo o de la población analizada y por lo tanto su información no es completa; se encuentra **sesgada**.

## Sesgo de selección

Una buena muestra estará libre de tanto “**sesgo de selección**” como sea posible. El “sesgo de selección” se produce cuando alguna parte de la población objetivo no esta en la población analizada. Si una encuesta diseñada para estudiar los ingresos del hogar omite personas transitorias, las estimaciones de la encuesta de los ingresos del hogar promedio o la mediana es probable que sean demasiado grandes. Una “**muestra de conveniencia**” es a menudo sesgada, ya que los elementos más fáciles de seleccionar y que tienen más tendencia a responder no suelen ser representativos de la sección de la población objetivo que es más difícil de seleccionar o de los elementos que no responden. Los siguientes ejemplos indican algunas formas de selección en el que sesgo puede ocurrir:

-Mediante un procedimiento de selección, desconocido para los investigadores, se depende de algunas características asociadas a las propiedades de interés. Por ejemplo, los investigadores tomaron muestras de conveniencia en los adolescentes para estudiar con que frecuencia los adolescentes hablan con sus padres y maestros sobre el SIDA. Pero los adolescentes dispuestos hablar con los investigadores sobre el SIDA son probablemente también los más propensos a hablar con otras figuras de autoridad sobre el SIDA. Los investigadores, que simplemente tomen un promedio de la cantidad de tiempo que los adolescentes de esta muestra indiquen, probablemente sobrestimen la cantidad de comunicación que ocurre entre padres y adolescentes en la población objetiva. En este sesgo deliberada o intencionalmente se manipula la selección a partir de un “**representante**”. Si queremos estimar la cantidad media del tiempo que un comprador pasa en un shopping de América y los compradores de la muestra seleccionada han pasado, de por sí, ya un tiempo en el shopping; hemos deliberadamente seleccionado una muestra para confirmar una opinión previa. Este tipo de muestra se llama “**muestra de juicio**” porque el investigador utiliza su juicio para seleccionar los elementos específicos que deben incluirse en la muestra. Manipulando finalmente la población objetivo. Por ejemplo, todas las encuestas de 1994 de las elecciones de gobernador de Arizona señalaron al candidato Eddie Basha como favorito en las encuestas por 9 puntos como mínimo. En la elección, Basha gano con un 37% de los votos y los otros dos candidatos obtuvieron el 35% y 28%, respectivamente. El problema fue que muchos votantes estaban indecisos al momento de las encuestas. En la elección primaria, Basha tuvo un fuerte apoyo de las zonas rurales y de grupos demográficos que no lo habían votado antes. Por lo tanto no se tomaron en cuenta para las encuestas.

No incluir a toda la población objetivo en el marco de muestra, se llama: “**falta de cobertura**”. Muchos estudios al usar el censo de los EE.UU. (que se realiza cada diez años) para construir el marco de muestreo, pierden de vista que el censo no puede enumerar un gran número de viviendas, y que subestima la producción de una serie de grupos de población. Fay (1988) estiman que en el censo de 1980 se perdió un 8% de todos los hombres de color. Por lo que cualquier encuesta que utilice los datos del censo de 1980 como la única fuente para la construcción de un marco de muestreo automáticamente perderá el 8% de los hombres de color, y plantea un error anterior incluso a la toma de la encuesta. .

Al no obtener respuestas de toda la muestra elegida se distorsionan los resultados, incluso en aquellos datos en los que las encuestas han sido cuidadosamente diseñadas para reducir al mínimo otros peligros de sesgo de selección.

A menudo, “**las no respuestas**” difieren críticamente los resultados de las encuestas pero su alcance sólo se conocerá más adelante. Muchas encuestas recogidas por los periódicos o revistas de investigación tienen tasas de respuesta menores al 10% en muchos casos. Es difícil, entonces, la generalización de los datos cuando el 90% de la población seleccionada se niega a participar. Por ejemplo, la base de datos de la Encuesta de Salud de Adolescentes fue diseñada para obtener una muestra de Minnesota, en el caso de los estudiantes de la escuela secundaria y preparatoria en escuelas públicas (Remafedi 1992). En general, participó el 49% de los distritos escolares que fueron invitados. La tasa de respuesta varió con el tamaño del distrito escolar:

| Tipo de Tasa de Participación del Distrito Escolar (%) |       |
|--|-------|
| Urbana   | → 100 |
| Metropolitana suburbana                                | → 25  |
| No metropolitanas con más de 2000 estudiantes          | → 62  |
| No metropolitanas 1000-1999                            | → 27  |
| No metropolitanas con 500 a 999 estudiantes            | → 61  |
| No metropolitanas con menos de 500 estudiantes         | → 53  |

En cada uno de los distritos escolares que han participado, las encuestas fueron distribuidas a estudiantes, y la participación de los estudiantes fue voluntaria. De las 52.553 encuestas distribuidas, 36.741 fueron completados y devueltos, lo que resulta una tasa de respuesta del 70%. La encuesta contenía preguntas sobre los hábitos de salud, religiosos, la situación psicosocial y la orientación sexual. Es probable que las respuestas de los distritos escolares ausentes tengan diferentes niveles de salud y actividad. Pero parece más probable aún que los estudiantes que si respondieron a la encuesta en promedio, tengan un perfil de salud diferente a los estudiantes que no respondieron.

Muchos estudios compararon los encuestados que si y que no respondieron y han encontrado diferencias en los dos grupos. En "Salud de la Mujer" de Iowa el estudio, 41.836 mujeres respondieron a un cuestionario enviado por correo en 1986. Bisgard (1994) compararon estas respuestas con las respuestas de los encuestados que no respondieron a los 55.323 registros de control en el Estado de la Salud Registro. Encontraron que la mortalidad cambia de acuerdo a la edad y la tasa de cáncer fue significativamente mayor en el grupo de los que no respondieron que en el grupo de los que si lo hicieron. Permitir que la muestra conste enteramente de voluntarios afectaría la misma. Tales son los casos de las encuestas telefónicas y las estadísticas de la radio y la televisión por su composición no se puede confiar en ellas.

El Literary Digest (1932, 1936a, b, c) comenzó a tomar las encuestas para pronosticar el resultado de las elecciones presidenciales de EE.UU. en 1912, alcanzado una reputación por su precisión, ya que el pronóstico ganador fue correcto en todas las elecciones entre 1912 y 1932. En 1932, por ejemplo, la encuesta predijo que Roosevelt recibiría 56% de los votos populares y 474 votos en el colegio electoral, en la elección real, Roosevelt recibió el 58% de los populares votos y 472 votos en el colegio electoral. Con un sólido historial de precisión, no es de extrañar que los editores de la Literary Digest tuvieran una gran confianza en sus métodos de votación en 1936.

Lanzamiento de la encuesta de 1936, que dijo: La encuesta representa la evolución de treinta años de constante evolución y perfeccionamiento. Sobre la base de un "muestreo comercial", sus métodos utilizados durante más de un siglo por las editoriales para impulsar venta de libros, la lista de distribución actual se extrae de cada guía telefónica en los Estados Unidos, de las listas de los clubes y asociaciones profesionales de la ciudad, las listas de los electores inscritos, clasificados por correo y datos profesionales. (1936a, 3). Pero el 31 de octubre, la encuesta predijo que republicano Alf Landon recibiría 55% de los votos, frente al 41% para el presidente Roosevelt. El artículo se tituló "Landon, 1.293.669; Roosevelt, 972 897: Vuelta final en La Encuesta Digest de los Diez Millones de votantes"

En ella figura la siguiente afirmación: "Nosotros no hacemos ninguna pretensión de infalibilidad. No hicimos acunar la frase "exactitud asombrosa", pero ha sido tan libremente aplicado a nuestras encuestas " (I 936b). Lo positivo de esta defensa fue que en la elección, Roosevelt recibió el 61% de los votos; Landon, el 37%. ¿Qué salió mal? Uno de los problemas puede haber sido falta de cobertura en el muestreo marco, que se basaba en gran medida en las guías telefónicas y el registro de automóviles. Los hogares con un teléfono o un automóvil en 1936 eran, en general, más ricos que otros hogares, y la opinión de las políticas económicas de Roosevelt se relacionaba generalmente con la clase más humilde. Sin embargo, el sesgo de muestreo no explica todas las discrepancias.

Análisis pos-mortem de la encuesta realizada por Squire (1988) y Calahan (1989) indican que incluso las personas con dos coches y un teléfono tendían a favorecer a Roosevelt y a veces las personas sin coche ni teléfono no lo apoyaron. La baja tasa de respuesta fue probablemente la fuente de gran parte del error. Miles de millones de cuestionarios fueron enviados por correo, y sólo 2,3 millones fueron devueltos, una enorme muestra, pero con una tasa de respuesta inferior a 25%. En Allentown, Pennsylvania, por ejemplo, la encuesta fue enviada a todos los votantes registrados, pero los resultados de la encuesta de Allentown estaban siendo incorrectos, ya que sólo un tercio de las papeletas fueron devueltas. Para Squire (1988) las personas que apoyaban a Landon eran mucho más propensas a responder la encuesta. De hecho, muchos partidarios de Roosevelt ni siquiera recuerdan haber recibido una encuesta, a pesar de que estaban en la lista de correo.

Una de las lecciones que pueden extraerse de la encuesta Literary Digest es que el tamaño de una muestra no es garantía de su exactitud. Los editores se sentían complacidos porque se enviaron cuestionarios a más de una cuarta parte de todos los votantes registrados y se obtuvo una muestra enorme de 2,3 millones de personas. Pero grandes muestras no representativas puede realizar tanto daño como pequeñas muestras no representativas. Una gran muestra puede hacer más daño que una pequeña porque mucha gente piensa que las grandes muestras son siempre mejores que las pequeñas.

El diseño de la encuesta es mucho más importante que el tamaño absoluto de la muestra. ¿De qué sirven las muestras con Sesgo de selección? Preferimos tener muestras sin sesgo de selección, que sirvan como un microcosmos de la población. Cuando el interés principal esta en estimar el numero total de víctimas de delitos violentos en Estados Unidos o el porcentaje de de los posibles votantes en el Reino Unido que tengan la intención de voto para el Partido Laborista en las próximas elecciones; un sesgo de selección en la muestra puede causar una estimación no valida. El esquema de muestreo intencional funciona bien para justificar una conclusión en el estudio. Un plan de muestreo debe evitar el sesgo de selección.