

# 6. Muestreo

La forma en que se realiza la selección puede hacer la diferencia. Es más fácil obtener muestras malas que buenas.

No todo es tan simple como tomar sopa.

En la Sección 5.3 consideramos un estudio, realizado por un profesor universitario entre sus alumnos, para evaluar si los jóvenes que practican deportes consumen menos alcohol. Este es un ejemplo de participación voluntaria en un estudio, la muestra no es representativa de la población de interés.

Recordemos un ejemplo de la Sección 3.1. Interesaba conocer las opiniones respecto a si el correo electrónico debe responderse lo más rápido posible o no. Si la encuesta fue realizada vía el correo electrónico, las opiniones representan únicamente a los que tienen correo electrónico y les interesó responder la encuesta.

La próxima vez que se encuentre con un resultado de un estudio, averigüe qué composición tenía la muestra y pregúntese si la muestra representa a la población que interesa o a un subgrupo más restringido.

## □ 6.1 Muestreo aleatorio simple

Es bueno que la **muestra** se seleccione en forma **aleatoria**; esto significa que:

**Cada uno de los individuos de la población tiene la misma oportunidad de ser seleccionado.**

- Se utiliza algún mecanismo probabilístico para elegirlos.
- La gente no se selecciona a sí misma para participar.
- Nadie en la población es favorecido en el proceso de selección.

**Muestra aleatoria simple:** Una muestra aleatoria simple es la que se obtiene a partir de un mecanismo que le da a cada una de las unidades muestrales la misma probabilidad de ser elegida.

El muestreo aleatorio (el proceso por el cual se obtiene una muestra aleatoria) comienza con una lista de **unidades muestrales** de la que se extraerá la muestra. Esta lista se llama **marco muestral**. Idealmente, el marco muestral debería contener la lista de la totalidad de las unidades muestrales.

El **muestreo aleatorio simple** tiene dos propiedades que lo convierten en el procedimiento por excelencia de obtención de muestras.

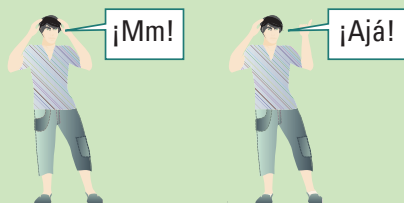
- Todas las unidades tienen la misma oportunidad de ser elegidas (es insesgado).
- La elección de una unidad no influye sobre la elección de otra (independencia).

**El Instituto Nacional de Estadísticas y Censos - INDEC** - realiza periódicamente **censos** para registrar las características básicas sobre población y vivienda, actividad económica y agropecuaria de nuestro país. Las unidades relevadas en los censos proveen el **marco muestral** para las **encuestas** que realiza durante los periodos intercensales.

Se espera que el muestreo aleatorio provea muestras representativas de la población.

Mediante un censo se intenta registrar todas las unidades muestrales de la población para proveer el marco muestral. Si se trata de un censo de población, deberán localizarse todas las personas. Si se trata de un censo económico, se registrarán todos los locales comerciales y productivos. Una vez que se dispone del marco muestral se abre la oportunidad de seleccionar la muestra.

Por otra parte, es necesario aclarar que una unidad muestral puede contener muchos individuos. Una escuela, con sus alumnos, puede ser una unidad muestral. El objetivo del estudio pueden ser las escuelas (por ej. interesa conocer la superficie cubierta por alumno) o ser los alumnos (por ej. interesa conocer el rendimiento en educación física).



¿Cómo? ¿Una unidad muestral puede estar constituida por muchos individuos?

Volvamos al ejemplo de la encuesta sobre la preferencia del candidato a presidente del Club Grande de Fútbol. Utilicemos la lista actualizada de todos los socios como marco muestral con los números de socio para identificarlos. Si se decide que 1 de cada 6 socios entrarán en la muestra podemos arrojar un dado tantas veces como socios tenemos en la lista y si sale 1 el socio es seleccionado.

TABLA 6.1

Socio Número	Número aleatorio	Socio Número	Número aleatorio	Socio Número	Número aleatorio	Socio Número	Número aleatorio
1495	4	1.501	1	1.507	4	1.513	4
1496	8	1.502	6	1.508	4	1.514	7
1497	8	1.503	3	1.509	3	1.515	8
1498	7	1.504	7	1.510	8	1.516	8
1499	9	1.505	1	1.511	1	1.517	1
1500	5	1.506	7	1.512	7	1.518	3

Con este procedimiento, seleccionamos los socios no: 1.501, 1.505, 1.511 y 1.517 mediante un **muestreo aleatorio simple**.

También podríamos utilizar un programa de computadora para generar números entre 1 y 6 en forma aleatoria, sin necesidad de arrojar un dado.

#### Muestra aleatoria simple en dos pasos :

Paso 1. Se asigna una etiqueta numérica a cada individuo de la población.

Paso 2: Se utilizan números aleatorios para seleccionar las etiquetas al azar.

En la práctica, el primer paso del procedimiento es el más difícil. Esta dificultad da lugar a **muestreos alternativos** que **no** son **válidos** desde el punto de vista del análisis estadístico. Veremos algunos en la próxima sección.

## □ 6.2 Muestras malas

Todos los días encontramos ejemplos de **muestras malas**:

- Cuando se pide a los oyentes de un programa de radio que voten por tal o cual cantante, llamando por teléfono o enviando un mensaje de correo electrónico, se trata de **muestras de respuesta voluntaria**. Las encuestas de opinión en las que se llama, o se escribe, por propia iniciativa son ejemplos de muestras de respuesta voluntaria, poco satisfactorias desde un punto de vista estadístico.
- Otro tipo de muestra mala es la **muestra de conveniencia**. Si una pedagoga elige a sus propios alumnos, del último año de la escuela secundaria en la que trabaja, para evaluar un cambio en el método de enseñanza, los resultados no se podrán extender más allá de ese grupo.

Cada vez que mire los resultados de un estudio, busque la frase "muestra aleatoria". Si la encuentra, hile más fino para averiguar cómo fue obtenida y si en realidad fue elegida en forma aleatoria.

---

## □ 6.3 Sesgo

---

Alguna vez escuchamos **el sesgo es malo**. Pero, ¿qué es el sesgo? Es un favoritismo de alguna etapa del proceso de recolección de datos **beneficiando algunos resultados, perjudicando otros y desviando las conclusiones en direcciones equivocadas**.

Cuando alguna etapa del proceso de recolección de datos está sesgada, **utilizar una muestra grande no corrige el error**, simplemente lo repite.

Los datos en un estudio pueden estar sesgados por muchos motivos. A continuación, veremos algunos de ellos.

---

### 6.3.1 Sesgo por elección de la muestra

---

#### 6.3.1.1 Muestras por conveniencia

Exprimir las naranjas que se encuentran a la vista, en la parte de arriba del cajón, es un ejemplo de muestra de conveniencia. Las entrevistas en los centros comerciales (shopping) son otro ejemplo, porque los fabricantes y las agencias de publicidad suelen recolectar información respecto a los hábitos de compras de la población y el efecto de sus publicidades en grandes centros de compras. Obtener una muestra de esta manera es rápido y económico, pero la gente que contactan no es representativa de la mayoría de la población.

#### 6.3.1.2 Muestras con sesgo personal

Por simpatía, gusto o interés, quien está realizando la encuesta puede preferir encuestar a cierto tipo de personas y no a otras. Por ejemplo, es posible que un encuestador joven tienda a buscar chicas bonitas para preguntarles.

#### 6.3.1.3 Muestras de respuesta voluntaria

Surgen a partir de los individuos que se ofrecen voluntariamente a participar. Se trata, por ejemplo, de las que alimentan las votaciones organizadas por programas de radio, televisión o de algún sitio de Internet. No producen resultados que tengan algún significado en relación a la opinión de la población en general. Los participantes voluntarios, que por algún motivo decidieron participar, suelen tener opiniones más polarizadas.

### 6.3.2 Sesgo de respuesta



### 6.3.2.1 Debido a la presentación de las preguntas

Las diferentes palabras con las que se puede presentar una misma pregunta suele ser una fuente importante de sesgo en las respuestas.

En un curso de manejo organizado por un automóvil club se proyectó una película sobre un accidente de tránsito a dos grupos de alumnos. Ambos grupos eran similares respecto de la edad y el género. Al finalizar la proyección se preguntó:

- Al primer grupo: ¿a qué velocidad piensa que los dos autos chocaron? El promedio de las respuestas fue de 50,9 km/h.
- Al segundo grupo: ¿a qué velocidad piensa que los dos autos se colisionaron? El promedio de las respuestas fue de 65,9 km/h.

Ambos grupos vieron la misma película. El uso de la palabra **colisionaron** aumentó las estimaciones de la velocidad del accidente en **15 km/h**, esto es un aumento del 29,5 %

El sesgo debido a la forma en que se presenta una pregunta puede ser **intencional** o **no intencional**.

Las preguntas “¿No está usted harto de pagar impuestos para que todo siga igual de mal?” y “¿Le parece importante que se paguen impuestos para mejorar la educación, los servicios de salud y la seguridad?”, que apuntan al pago de impuestos, seguramente tendrán resultados muy diferentes. Ambas preguntas conllevan un **sesgo intencional**.

Una encuesta dirigida a alumnos de 7mo. grado que pregunte: “¿Cuáles son las 5 personas grandes que le gustaría conocer personalmente?” tendrá diferentes lecturas. Algunos de los alumnos podrán interpretar que se trata de personas mayores de edad, otros que son altos, otros que se refiere a gordos o tal vez a grandes estrellas de cine, de rock, políticos o deportistas, generando un sesgo **no intencional**.

### 6.3.2.2 Para tratar de agradar

A la gente no le gusta mostrarse con ideas que no están bien vistas socialmente. Por ejemplo, cuando esté cara a cara con un encuestador o llenando un formulario no anónimo, un varón evitará una respuesta que parezca machista, o una mujer responderá tratando de ocultar algún prejuicio.

### 6.3.2.3 Por recuerdo

Si la pregunta está referida a un acontecimiento ocurrido algún tiempo atrás, la respuesta tendrá un sesgo por recuerdo. Por ejemplo, si se le pregunta a una madre a qué edad comenzaron a caminar sus hijos, la veracidad y precisión de la respuesta dependerá de las características personales de la madre.

#### 6.3.2.4 Por no respuesta

Algunas veces las personas que han sido seleccionadas para una encuesta son muy difíciles de localizar o simplemente se niegan a responder. **Los individuos que no responden pueden ser muy diferentes de los que sí lo hacen.** Este tipo de sesgo se puede reducir sustituyendo a los que se niegan a responder por otros individuos con las mismas características de los que no responden, pero suele ser difícil.

Cuando mire los resultados de una encuesta que le interesa especialmente, antes de sacar sus propias conclusiones averigüe qué se preguntó, cómo fueron redactadas las preguntas, si las respuestas fueron dadas en forma anónima o no y cuántos se negaron a responder.

Es más fácil obtener muestras malas que buenas.

#### 6.3.2.5 Por subcubrimiento

Una encuesta telefónica ignora a todos los sujetos que no tienen teléfono. Una encuesta que realiza las entrevistas en hogares ignora a los que viven en la calle.

### □ 6.4 Otros tipos de muestreos

#### 6.4.1 Muestreo sistemático

Veamos un ejemplo de la utilidad de este método. Si nos interesa la opinión de las alumnas de una escuela respecto del aumento de las horas destinadas a la práctica de deportes, podríamos entrevistar a las alumnas a la salida y elegir una de cada diez (suponiendo que salgan de a una) hasta que hayan salido todas. De esta manera, si la escuela tiene 227 alumnas, la muestra tendrá 22 alumnas.

**Muestreo sistemático:** El muestreo comienza con una unidad elegida al azar y a partir de allí continúa cada  $k$  unidades. Si  $n$  es el tamaño muestral y  $N$  es el tamaño de la población entonces  $k$  es aproximadamente  $N/n$ .

Este tipo de muestreo permite evitar el sesgo personal y es más sencillo que el muestreo aleatorio. Es útil cuando la población está ordenada naturalmente (si no lo está, para utilizar este tipo de muestreo es necesario ordenarla, pero al ordenarla, se pierden las ventajas que tiene).

Por su simplicidad, se suele utilizar para **control de calidad** durante, o al finalizar, la fabricación de diversos productos.

En una producción continua de tubos de pasta dentífrica, se elige un tubo por hora y se lo analiza para verificar que cumple con las especificaciones.



**Advertencia:** Este muestreo no es adecuado cuando el período de la selección está relacionado con alguna característica que nos interesa evaluar.

Podría ocurrir que cada hora (una hora es el período de la selección) se produzca una leve caída de tensión que hace que los tubos de pasta dentífrica se llenen más o menos. No detectaríamos esa variación con el muestreo cada hora.

Al realizar un muestreo sistemático es importante estar alerta para identificar los factores que puedan estar invalidando los resultados.

## 6.4.2 Muestreo aleatorio estratificado

En un muestreo estratificado la población se divide en **grupos homogéneos** llamados estratos. Luego se realiza un muestreo aleatorio simple de unidades muestrales dentro de cada estrato.

Los estratos se eligen de acuerdo con los valores conocidos de algunas variables, de manera que haya **poca variabilidad dentro del estrato** (los valores de dichas variables para las unidades de un estrato particular difieren poco), pero que haya **mucha variabilidad entre estratos** (los valores de dichas variables para las unidades de distintos estratos difieren mucho).

### Ejemplo 1:

La población de una ciudad podría estratificarse por

- **grupo de edad:** menos de 6 años, entre 6 y 12 años, entre 13 y 18 años y mayores de 18 años.
- **género:** femenino, masculino.

Así obtenemos 8 estratos, dentro de los cuales los individuos tienen 2 características similares: grupo de edad y género. Podríamos realizar un muestreo proporcional a la cantidad de individuos que tiene cada estrato, de manera que el tamaño de la muestra dentro de cada estrato dependa de la proporción de la población total que dicho estrato representa.

### Ejemplo 2:

En una encuesta diseñada para conocer la situación de la industria en una provincia podrían utilizarse estratos por tamaño y actividad. Para cada actividad industrial podrían incluirse **todos** los locales industriales con 500 ó más obreros ocupados (**inclusión forzosa** - la muestra los contiene a todos), **la mitad** de los que tuvieran entre 499 y 200, **la cuarta parte** entre 199 a 50 y **1 de cada 20** para los de menos de 50. Tendríamos así 4 estratos:



- Estrato 1: Locales con 500 ó más obreros
- Estrato 2: Locales con 499-200 obreros
- Estrato 3: Locales con 199-50 obreros
- Estrato 4: Locales con 50-0 obreros

Si además se dividiera la actividad industrial en dos: 1) industria alimenticia, 2) industria no alimenticia, ¿cuántos estratos tendría la muestra? Tendría 8 estratos, dos por cada uno de los 4 estratos anteriores.

### Tres pasos de un muestreo aleatorio estratificado:

- **Paso 1:** las unidades se agrupan en estratos. Los estratos se eligen teniendo en cuenta que estos grupos tienen un interés especial dentro de la población, o porque los individuos en el estrato se parecen mucho.
- **Paso 2:** se establece la proporción de unidades, o **fracción de muestreo**, que se incluirá para cada estrato
- **Paso 3:** dentro de cada estrato se realiza un muestreo aleatorio simple y la **proporción de individuos** que se incluye en la muestra es la establecida en el paso 2. La unión de las muestras de cada estrato constituye la muestra completa.

## 6.4.3 Muestreo por conglomerados

En un muestro por conglomerados la población se divide en **grupos heterogéneos** llamados **conglomerados**. Luego se realiza un muestreo aleatorio simple en el que las unidades muestrales son los conglomerados.

La idea del agrupamiento para un **muestreo aleatorio por conglomerados** (también llamados aglomerados) es opuesta a la del muestreo estratificado. Interesa que los individuos que componen cada grupo sean lo más heterogéneos posibles y se espera que cada conglomerado sea representativo de la población. Los **conglomerados** son las **unidades del muestreo**, pero las unidades de interés son los individuos dentro de los conglomerados. Se selecciona una muestra aleatoria de conglomerados, y **se observan todos los individuos dentro de cada conglomerado** ó se selecciona una muestra aleatoria simple dentro del conglomerado. Este tipo de muestreo puede tener mejor rendimiento costo-efectividad que un muestreo aleatorio simple, en especial si los costos de traslado son altos.

### Ejemplo 1:

Una encuesta de viviendas. Se divide la ciudad en manzanas, se seleccionan las manzanas mediante un muestreo aleatorio simple y se visitan todas las casas de cada manzana seleccionada.

### Ejemplo 2:

En un estudio interesa evaluar la capacidad de lectoescritura de alumnos de 7mo. grado. Se seleccionarán al azar las escuelas y luego se realizará la prueba en todos los alumnos de 7mo. grado de las escuelas seleccionadas.

### Tres pasos de un muestreo aleatorio por conglomerados:

- **Paso 1:** Los individuos se agrupan en conglomerados. Los conglomerados generalmente tienen una proximidad física, pero dentro de cada conglomerado las unidades son heterogéneas.
- **Paso 2:** Los conglomerados son las unidades muestrales. Se establece la proporción de unidades que se incluirá.
- **Paso 3:** Se realiza un muestreo aleatorio simple de conglomerados y se estudian todos los individuos de cada conglomerado seleccionado. El tamaño final de la muestra es la cantidad de individuos que componen todos los conglomerados seleccionados.

---

## 6.4.4 Muestreo multietápico

---

Un muestreo multietápico tiene dos o más pasos y, en cada uno de ellos se aplica cualquiera de los procedimientos de selección anteriores.

### Ejemplo 1:

Una encuesta de viviendas. En **la primera etapa** se divide la ciudad en barrios, se toma una muestra aleatoria simple de barrios. En **la segunda etapa**, cada barrio seleccionado en la primera etapa se divide en manzanas, se seleccionan las manzanas mediante un muestreo aleatorio simple, y se visitan todas las casas de cada manzana seleccionada.

### Ejemplo 2:

Estudio para evaluar la capacidad de lectoescritura de alumnos de 7mo. grado. En **la primera etapa** se seleccionan al azar las escuelas, y en **la segunda etapa** se selecciona dentro de cada escuela un cierto número de cursos de 7mo. grado. La prueba se realiza en todos los alumnos de 7mo. grado de los cursos seleccionados en la segunda etapa.

## □ 6.5 Actividades y ejercicios

### 1. ¿Cuál es la Población? ¿Cuál es la muestra?

Para cada uno de los siguientes estudios indicar la población lo más detalladamente posible, es decir describir a los individuos que la componen. Si la información es insuficiente, completarla de la forma que se considere más adecuada. También indicar cuál es la muestra.

- Una encuesta de opinión contacta a 1.243 adultos y les pregunta, ¿ha comprado un billete de lotería en los últimos 12 meses?
- Durante la reunión anual del colegio de abogados, todos los presentes (2.500), llenaron una encuesta referida al tipo de seguro que prefería para su automóvil.
- En 1968 se realizó en Holanda un test de inteligencia a todos los varones de 18 años que estaban realizando el Servicio Militar Obligatorio.
- El INDEC lleva a cabo la Encuesta Permanente de Hogares (EPH) en la que se encuestan 25.000 hogares para captar información sobre la realidad económico-social de la República Argentina.

### 2. Voto secreto y obligatorio.

- ¿Qué tipos de sesgos se pueden producir cuando una elección para presidente se realiza en forma voluntaria?
- ¿Qué tipos de sesgos se pueden producir si el voto en la Comisión Directiva de un club o en la Cámara de Diputados no es secreto?

### 3. Se quiere realizar una encuesta entre los alumnos de una escuela secundaria, de 2.500 alumnos (500 alumnos por cada año, de 1 ro. a 5 to.), utilizando una muestra de tamaño 100. El propósito de la encuesta es determinar si a los/as alumno/as les interesa discutir el siguiente tema: “Debe reducirse la edad de imputabilidad penal para los menores de edad, que establece la ley nacional 22.278, a dieciséis años de edad; como respuesta al incremento en la cantidad de delitos graves cometidos por jóvenes y adolescentes”.

### 4. Indicar cuál es el tipo de muestreo realizado en cada caso.

- Cada alumno escribe su nombre en un papel, lo pone en una bolsa y el director elige 100 papeles.
- A cada alumno se le asigna un número entre 1 y 2.500 y se seleccionan generando 100 números al azar de cuatro dígitos utilizando algún programa de computación.
- Para cada año se asigna a cada alumno un número entre 1 y 500, y se elige 1 de cada 25 alumnos.
- Se eligen al azar una división de cada uno de los años y se seleccionan 20 alumnos de cada división.
- Se eligen al azar 60 alumnos de los primeros 3 años y 40 alumnos de los últimos dos años

- Se eligen al azar 60 alumnos de los primeros 3 años y 40 alumnos de los últimos dos años. Se seleccionan en forma separada los varones y las mujeres de acuerdo con la proporción de mujeres y varones que tiene la escuela.
5. En un programa de radio se invitó a las/los oyentes a contestar la siguiente pregunta: “¿Si pudiera volver el tiempo atrás volvería a tener hijos?” De más de 10.000 respuestas el 70% dijo no. ¿Qué muestra esto?
- Elegir, entre las cinco siguientes, la respuesta que mejor responde a esta última pregunta.
- a. La encuesta no dice nada porque arrastra el sesgo por respuesta voluntaria.
  - b. No se puede decir nada sin saber las características de los oyentes.
  - c. Para sacar una conclusión, es necesario separar las respuestas entre hombres y mujeres.
  - d. Hubiese tenido más sentido tomar una muestra aleatoria de las 10.000 respuestas para sacar conclusiones.
  - e. Es una muestra legítima elegida al azar entre todos los que escuchan ese programa y tiene un tamaño suficiente como para concluir que la mayoría de los oyentes lo pensarían dos veces antes de tener más hijos.
6. Indicar cuál o cuáles de las siguientes afirmaciones son válidas.
- a. Las respuestas que se obtienen al utilizar un cuestionario expresado en términos no neutrales tendrán “sesgo por respuestas”.
  - b. Las encuestas de respuesta voluntaria subestiman a la gente con opiniones muy firmes.
  - c. Las encuestas de respuesta voluntaria generalmente sobre representan las respuestas negativas.
  - d. En general, es posible reducir el sesgo tomando muestras muy grandes, cuanto más grande es el tamaño de la muestra mejor.
  - e. El tamaño de la muestra no tiene nada que ver con el sesgo.
  - f. Los resultados que se obtienen de un censo son siempre más precisos que los que se obtienen de una muestra, sin que importe cuán cuidadoso haya sido el diseño y su aplicación.