

## Métodos de recolección de datos

## CAPÍTULO 2

### 2.1. INTRODUCCIÓN

El capítulo 1 nos proveyó una visión general de los principios y procedimientos estadísticos para obtener y resumir información para la toma de decisiones.

Tenemos una **Teoría** que deseamos ensayar. **Juntamos datos** para poner a prueba la teoría. **Describimos** y **resumimos**, luego interpretamos los resultados y usamos los datos para confirmar o rechazar la teoría. Según la decisión tomada, puede que necesitemos volver a la teoría y juntar más datos o bien, una decisión puede ser tomada ahora, y luego en algún momento, estar sujeta a nuevos ensayos.

El método científico comprende los siguientes pasos:

- 1º. Formulación de la teoría.
- 2º. Recolección de datos para ensayar la teoría.
- 3º. Análisis de resultados.
- 4º. Interpretación de los resultados - toma de decisión.

En este capítulo, nos ubicaremos en el paso 2º, o sea en la forma de recoger los datos. Necesitamos juntar datos, sobre algún tema que nos interesa, para evaluar alguna teoría.

Los datos pueden ser recogidos por otros, o bien, puede que necesitemos producir nuestros propios datos. En ambos casos, la calidad de la toma de decisión dependerá de la calidad de los datos que obtuvimos. No todos los datos son producidos y presentados como datos "buenos" –fidedignos–. En este capítulo estudiaremos algunos métodos de muestreo (recordar que en tercer año, los alumnos tienen una materia: Teoría del muestreo).

La idea de muestrear es ganar información sobre el total, examinando sólo una parte.

¿Por qué sacar una muestra? Y ¿por qué no analizar a todos los elementos que componen la población objeto del estudio?

## 2.2. ¿POR QUÉ UNA MUESTRA?

Supongamos que deseamos estudiar una población. ¿Por qué no muestrear a toda la población? ¿Por qué no realizar un censo?

### Definición

Censo es una muestra formada por todos los elementos que componen la población.

Usando cualquier método, lea la siguiente oración durante unos segundos y determine el número de letras "F".

## FINISHED FILES ARE THE RESULT OF YEARS OF SCIENTIFIC STUDY COMBINED WITH THE EXPERIENCE OF MANY YEARS.

Números de "F": .....  
¿Cuántas "F" encontraron? ..... ¿tres? ¿cinco? ¿seis? .....

El censo de la población de USA (idem en República Argentina, ver página web INDEC: [www.indec.mecion.gob.ar](http://www.indec.mecion.gob.ar)), se realiza cada diez años, según lo establece la Constitución. El "Census Bureau", al igual que el INDEC, son más cuidadosos en la definición de Censo: lo definen como "un intento" de encuestar a toda la población. Un censo no es una tontería, como lo que acabamos de hacer con esta oración. Sin embargo a pesar de tratarse de una población de pocos elementos, casi ninguno arribó a la verdad. Un censo no ofrece garantía absoluta de calidad.

Si la población es grande, en lo que a número de elementos respecta, un censo puede ser muy caro. Más aún, supudiéramos económicamente hacer un censo, el mismo nos llevaría tanto tiempo, que la información no resultaría de interés. A veces es imposible realizar un censo: supongamos que estamos interesados en la duración de cierto tipo de pilas. No es lógico ensayar **todas** las baterías para ver cuánto duran. Cuando la medición destruye a los items, el censo no es factible.

En la mayoría de las situaciones, sacamos muestras aleatorias. En un

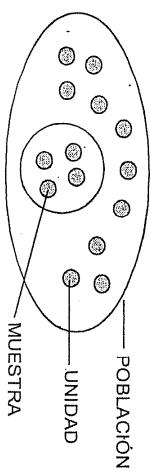
mundo tan cambiante, el muestreo nos permite conseguir una instantánea de lo que ocurre, en un momento determinado.

## 2.3. EL LENGUAJE DEL MUESTREO

El muestreo aleatorio consiste en seleccionar una parte de la población para observarla y analizarla y así poder estimar alguna característica sobre toda la población. Por ejemplo, en un estudio nacional para obtener información sobre el ingreso promedio de todas las familias, se podrá estimar el verdadero ingreso promedio de la nación tomando una muestra aleatoria de dicha población y considerando el ingreso promedio de aquellas familias que cayeron en la muestra.

En los departamentos de Control de Calidad es frecuente aceptar un lote de productos sólo si la proporción de defectuosos no supera ciertos límites. Examinando solamente una parte (muestra) de una gran parte (población) y considerando la proporción de defectuosos en la muestra, se puede evaluar si el lote es aceptable o no.

Recordemos las definiciones de: **población, muestra, unidad, variable**.



### Tamaño de la población $N = 15$ Tamaño de la muestra $n = 4$

#### Definiciones

➤ **POBLACIÓN:** conjunto total de objetos o individuos que presentan características comunes observables, definidas en un cierto tiempo y lugar.

➤ **UNIDAD:** objeto o persona individual, a menudo se la denomina sujeto si la población consiste en personas.

➤ **MUESTRA:** parte de la población que es usada para lograr información.

➤ **VARIABLE:** característica de interés a ser medida en cada unidad de la muestra.  
El tamaño de la muestra es denotado con la letra  $n$ .

## Ejemplo 2.1 Cruzando la calle

¿Puedo cruzar la calle en esta intersección? ¿Qué datos necesitaremos para tomar esta decisión? ¿Cuántos autos se están aproximando y a qué velocidad lo están haciendo?

La población estará formada por .....  
Las variables de interés .....  
Note que el número de autos no es tan importante como la "distancia" y la "velocidad".

A veces, resulta difícil compilar el listado de las unidades de la población. ¿Cómo podemos construir un listado de todos los peces de una laguna? ¿Cómo incorporamos a gente que vive en áreas inaccesibles o en áreas que, resultan muy costosas de visitar? Otras veces, la variable de interés es polémica. La población puede estar formada por los estudiantes de un determinado colegio, y la pregunta de interés ¿cuántos alumnos hacen "trampas" en sus exámenes? Analogamente, las decisiones no siempre son fáciles de tomar, porque no conocemos la situación completa. Supongamos que el presidente de un país, desea implementar un programa cuando el nivel de desempleo resulte muy alto. ¿Qué nivel es "muy alto"? ¿Podemos conocer la tasa real de desempleo?

Establecemos la diferencia entre una **población** y una **muestra**. Cuando deseamos conocer "algo" de la población, por ejemplo, el ingreso promedio o la proporción de votantes, el valor que es computado desde toda la población es llamado parámetro. El valor de un **parámetro** es un número fijo y puede ser computado sólo si examinamos la población entera. Sin embargo, en general, sólo se tiene una muestra de esa población. Cuando calculamos el ingreso promedio, o la proporción de votantes, con los datos provenientes de una muestra, esas cantidades se llaman **estadísticas**. Y, ¿qué hacemos si deseamos conocer el valor de un parámetro de nuestra población? si no podemos examinar la población entera, tomaremos una muestra aleatoria, y usaremos la información que ella nos brinda, la **estadística**, para estimar la correspondiente información de la población, el **parámetro**.

## Definiciones

Un **parámetro** es una medida resumen que se calcula a partir de todas las unidades de la población.  
Una **estadística** es una medida resumen que se calcula usando todas las unidades de la muestra de la población.

**Observación:** una manera de recordar esta distinción: la **letal** tra **p** es para la **población** y **parametro**, por lo tanto la **estadística** se la asocia a la **muestra**.

## Ejemplo 2.2. Partes defectuosas

Partes de un producto son embarcadas para ser enviadas a una compañía, que las comprará en "lotes", es decir, grandes cargamentos de 1000 partes. Luego de recibir un lote, el comprador examina una muestra de 20 partes para chequear el número de partes defectuosas, si las hay. El contrato entre el comprador y el abastecedor, generalmente, establece las condiciones de "aceptación del lote."

El lote de las partes es nuestra población 'bajo estudio'. Las unidades elementales serán las partes individuales. Las 20 partes seleccionadas al azar, de ese lote, constituyen la muestra. La variable a ser medida, será "la parte es defectuosa o no".

Si las 1.000 partes del lote son examinadas, podríamos conocer la verdadera proporción de partes defectuosas en el lote. Este valor será el parámetro. Si sólo 20 de las 1.000 partes son examinadas, podremos calcular la proporción de partes defectuosas muestral, pero no la del lote entero. Dicha proporción de partes defectuosas en la muestra es la estadística.

## Ejemplo 2.3. ¿Parámetro o Estadística?

Título extractado de "Health Update": "Aerobics pieren los oídos"

Médicos del hospital Henry Ford en Detroit, estudiaron 125 clases de aeróbics en 5 clubes de salud, y encontraron que los niveles de sonido de la música en el 60% de las clases excedía el límite de seguridad.

Los resultados establecidos en el párrafo de arriba están basados

sobre un estudio de 125 clases de aerobics en 5 clubes de salud (no todas las clases de aerobics en todos los clubes de salud).  
Este 60% es una muestra es  $n = \dots$  y el tamaño de la muestra

PARA RESOLVER !!!

E) 9% de la población de los Estados Unidos tiene sangre tipo B. En una muestra aleatoria de 400 individuos de esa población, se encontró que el 12.5% tiene sangre tipo B.

a) En esta situación particular, el valor de 9% es un/a . . . . . (parámetro, estadística).

b) En esta situación particular, el valor de 12.5% es un/a . . . . . (parámetro, estadística)..

#### EJEMPLO 2.4. Una estadística varía, el parámetro es fijo

Suponga una población consistente en 5 estudiantes. La variable de interés, es el número de textos contenido en sus mochilas. El parámetro es el número promedio de textos en sus mochilas. Un estudiante tuvo 5 libros, otro 3, el tercero estudiante 1 libro, otro 2 y el último 4.

### ¿Cuál es el valor del parámetro?

Dado que conocemos todos los valores de la población, podemos calcularlo fácilmente ( $15/5 = 3$ ). El número promedio de libros en la población es 3. Si otra persona calcula el valor de este parámetro, también será 3. Este valor es **fijo (3)**.

Ahora supongamos que seleccionamos 2 estudiantes de esta población. Queremos conocer el número promedio de libros. ¿Qué haría Ud?

La vez que se calculan los promedios varían con el tipo de muestra. En el caso de los promedios de los dos alumnos seleccionados, este promedio es una **estadística**, porque está calculado en base a los valores de la muestra. El promedio dependerá de qué alumnos se seleccionaron.

Si seleccionamos los alumnos poseedores de uno y dos libros respectivamente, nuestra **estadística** es 1.5.

Si seleccionamos los estudiantes que llevan 3 y 5 libros, la estadística será **4**.

Por lo tanto, **es** el valor de la estadística necesariamente igual al parámetro 3?

El valor del parámetro es generalmente desconocido —a no ser que examinemos a toda la población—.

Luego en capítulos posteriores lo discutiremos.

## 2.4. ¿Buenos datos?

Producir "buenos" datos requiere algún esfuerzo. Deseamos juntar datos efectivos para ayudarnos a responder ciertas cuestiones, para producir buenas estimaciones, etc. Existen distintos métodos para extraer una muestra.

Consideremos un verdulero y el camionero que le trae las papas. Puede ser conveniente sacar una muestra de algunas bolsas de papas tomadas de la parte de arriba. Pero éstas no serán representativas del cargamento, dado que aquéllas que están abajo pueden contener papas dañadas por causa del embarque.

... las papas que aparezcan en la muestra deben ser tan "buenas" como las de la población?

El **vicio** (*bias*) o **sesgo** es definido como un perjuicio en una dirección. Un método de muestreo que produce resultados que sistemáticamente difieren de la verdadera población se dice que es **sesgado**.

Aunque muestrear papas de la parte de arriba del camión puede ser conveniente y fácil de hacer, esta muestra puede resultar sesgada.

Tales muestras "convenientes" son generalmente sesgadas, no representativas de la población y pueden conducir a inapropiadas conclusiones sobre la población.

Un diario impreso un cuestionario dirigido a los lectores: "Queremos saber qué piensa"  
*¿Responderán todos los lectores a ese relevamiento?*

Probablemente la mayoría **no**. Los resultados de tal estudio reflejarán sólo las opiniones de los que respondieron.

7 muestreo  $\rightarrow$  1000 hogares.

¿Qué lectores probablemente responderán? Aquellos que tienen una fuerte opinión sobre los ítems tratados. Tampoco sirven las respuestas voluntarias porque es posible que no sean representativas de todos los miembros de la población.

### PARA RESOLVER !!!: 2.3. Tamaño de la familia

Un **método de muestreo es sesgado** si produce resultados que difieren sistemáticamente de los verdaderos de la población.

Una **muestra conveniente** es una muestra que de aquellas unidades que se consiguen fácilmente.

**Muestras voluntarias** (ver ejemplo de Encuesta Periódico) o **convencionales** son generalmente sesgadas.

Veremos distintos tipos de sesgos en los siguientes ejemplos:

• **Sesgo de Selección:** es la tendencia sistemática sobre el procedimiento de muestreo para excluir o incluir cierto tipo de unidades.

• **Sesgo de No respuesta:** es la distorsión que se logra cuando un gran número de unidades seleccionadas para la muestra no responden o se niegan a responder y esas no respuestas tienen una tendencia a ser distintas de los que responden.

• **Sesgo de respuesta:** es la distorsión que se logra por la forma de preguntar o el comportamiento del entrevistador puede afectar la respuesta.

### PARA RESOLVER!!!: 2.2. ¿Es sesgado?

Un programa de TV llevó a cabo la siguiente encuesta de opinión:

¿Es resistido el control de armas? ¡Permitámos conocer su respuesta!! Esta noche!

- Si contesta Sí llame al 1-900-446-6444
- Si contesta No llame al 1-900-446-6445

La tarifa es 50 centavos el primer minuto.

¿Considera los resultados de esta encuesta de opiniones dignos? Explique.

Se llevó a cabo un estudio para estimar el tamaño promedio de las casas de familia (hogares) en USA.  
Un total de 1.000 personas fueron aleatoriamente seleccionadas e interrogadas respecto del número de personas que viven en sus casas.

El promedio de estas 1.000 respuestas fue 4.61.

④ ¿Cuál es la población de interés? Familias USA  
⑤ ¿Cuál es la variable en estudio? Nº de miembros de la familia  
⑥ ¿Cuál es el parámetro en estudio? Tamaño de la familia  
⑦ Un promedio calculado como el que se describió arriba tenderá a ser más grande que el verdadero promedio del tamaño de los hogares en USA.

Explique por qué sería éste el caso.

Conocer la manera en que los datos son obtenidos es muy importante. Para evaluar si una muestra es buena, debemos preguntnos ¿Cuál fue la población de interés? ¿Qué unidades se tomaron? ¿Cómo fue seleccionada la muestra? ¿Fue sesgada la selección? ¿Hubo sesgo de no respuesta? ¿Cuáles fueron las fuentes de datos? ¿Cuáles fueron exactamente las preguntas realizadas? ¿En qué orden se realizaron las preguntas?

Sabemos que las muestras convenientes y voluntarias darán generalmente resultados que son sesgados. La razón por lo que esto ocurre es que cuando se le pide a una persona que sea imparcial, difícilmente lo consiga. Un remedio para evitar la parcialidad humana es usar un sorteo para seleccionar la muestra.

La idea básica es que cada unidad de la población, tenga una chance conocida de ser seleccionada para la muestra. Los métodos de muestreo que tienen esta propiedad se denominan métodos de muestreo probabilísticos. Si se usa un método de muestreo aleatorio apropiado, aun utilizando una muestra relativamente pequeña, podrá reflejar correctamente lo que ocurre en una población. Con tales métodos, podemos cuantificar la precisión de usar una estadística muestral para estimar un parámetro poblacional correspondiente.

Un método de muestreo que asigna a cada unidad de la población una chance no negativa de ser seleccionada, se denomina **método de muestreo probabilístico**.

El tipo de muestreo aleatorio más sencillo es el **muestreo simple al azar**.

Con el método de muestreo aleatorio simple, todo grupo posible de unidades de un tamaño muestral requerido, tiene la misma chance de ser la muestra seleccionada.

En la próxima sección, veremos cómo seleccionar una muestra simple al azar. Sin embargo en la práctica, un muestreo simple al azar puede no resultar simple.

Existen otros métodos de muestreo probabilístico. En este apunte se tratarán los siguientes:

- Simple al azar
- Estratificado
- Sistemático
- De agrupación o clasificación o conglomeración
- Multietápico.

Su utilización dependerá de cada situación. Hay métodos que en determinada circunstancia son más fáciles de llevar a campo, y proveen mejor información que otros.

#### Observación

El uso de la palabra aleatorio no significa "sin plan o método". El sinónimo de seleccionar aleatoriamente es seleccionar indiscriminadamente.

### 2.5. MUESTREO SIMPLE AL AZAR

Una muestra simple al azar de tamaño  $n$  es una muestra de  $n$  unidades, seleccionadas de alguna manera tal que cada muestra posible de tamaño  $n$ , tenga la misma chance de ser seleccionada que cualquier otra muestra de tamaño  $n$ . Muestras de distinto tamaño tienen distintas chances de ser seleccionadas.

Puedo pensar en la siguiente muestra aleatoria simple: Colocar 1 bola (de igual tamaño, peso, color, etc.) con la variable de interés registrada en una canasta por cada unidad de la población. Mezclamos las bolas y una muestra aleatoria simple sería elegir, sin mirar,

algunas bolas de las canasta. Si no deseó seleccionar los items dos veces, seleccionaremos las bolas sin reemplazo.

En cualquier caso, con o sin reemplazo, en cada selección, cada unidad de la bolsa, tiene la misma chance de ser seleccionada. El método es imparcial y es considerado justo o insesgado.

El método es imparcial: toda unidad tiene la misma chance de participar en la muestra.

El muestreo simple no es necesariamente el procedimiento más económico o eficiente; pero proporciona la base a partir de la cual han evolucionado los procedimientos más complejos.

La clave de selección de muestras apropiadas es obtener y tener una lista actualizada de todos los individuos o elementos de los cuales se extraerá la muestra. Tal lista se conoce como **marco de población**. Este listado servirá como **población objetivo** de tal manera que si se extrajeran muchas muestras probabilísticas diferentes de tal lista, en el mejor de los casos, cada muestra sería una representación en miniatura de la población y produciría estimaciones razonables de sus características. Si el listado es inadecuado porque ciertos grupos de individuos o elementos de la población no estuvieron incluidos adecuadamente, las muestras probabilísticas aleatorias sólo proporcionarían estimaciones de las características de la población objetivo y no de la población real y ocurrirían sesgos en los resultados.

La idea "física" de obtener una muestra simple al azar se obtiene comenzando con un marco de la población y una fuente de números aleatorios. Los números aleatorios están publicados en tablas de números aleatorios o bien pueden ser generados por una calculadora o PC.

#### Ejemplo 2.5. Muestreo aleatorio simple de situación forestal

Como una parte del estudio sobre los efectos de la fragmentación del bosque en la declinación de las aves canoras, un equipo de investigadores necesita juntar información sobre la depredación de sus nidos.

Un área muy grande del bosque fue dividida en  $N=80$  sitios, de aproximadamente igual tamaño. Debido al tiempo y costo, los investigadores decidieron seleccionar 5 de estos sitios al azar y



Supongamos que queremos conocer la **verdadera proporción** de mujeres en su población. La denotaremos con **p**.

- ✓ Cuente el número de mujeres en su población:
- ✓ Cuente el número de personas en su población:
- ✓ Compute la proporción de mujeres en su población:  $p = \text{nº de mujeres}/N = \dots$

Miremos ahora los resultados en la muestra simple al azar de tamaño 3: En muchos casos a esta proporción muestral se la simboliza con " $\hat{p}$ ".

- ✓ Cuente el número de mujeres en su muestra:
- ✓ Compute la proporción muestral de mujeres  $\hat{p} = \text{número de mujeres}/n = \dots$

Recuerda que un **parámetro** es un número obtenido en la **parametría** y una **estadística** es computada a partir de una **muestra**.

- ✓ En este caso,  $p$  es un/a ..... (parámetro o estadístico)
- ✓  $\hat{p}$  es un/a ..... (parámetro o estadístico)
- ✓ ¿Es  $\hat{p} = p$ ? .....  
✓ ¿Será siempre así? .....

### PARA PENSAR !!!

- ◆ ¿Cuándo resulta fácil seleccionar una muestra simple al azar (m.s.a.)? ¿Ello siempre será posible?
- ◆ ¿Cuándo resulta difícil? ¿Por qué?
- ◆ ¿Cómo rotulamos las unidades si la población es de tamaño 78?, ¿292?, ¿2000?
- ◆ ¿Será más fácil hacerlo con T.I. o con tabla de números al azar?

### PARA RESOLVER !!!

#### 2.4. Muestra simple al azar

Forme un grupo de 10 estudiantes. La población de interés es su grupo. Su tarea consiste en seleccionar 1 muestra simple al azar de tamaño 3 de su grupo.

Pasos:

- 1) En el espacio previsto abajo, escriba los nombres de las personas de su grupo
- 2) Asigne un rótulo ( $n^{\circ}$ ) distinto a cada uno de los nombres de la lista.
- 3) Seleccione una muestra utilizando rótulos al azar.

Si usa la calculadora T.I. use semilla 21 y su  $N=10$   
Si usa Tabla números aleatorios: comience por la fila 13

¿Cuál es el 1er rótulo seleccionado?

¿Cuál es la 1era persona seleccionada?

¿Cuál es el 2do rótulo seleccionado?

¿Cuál es la 2da persona seleccionada?

¿Cuál es el 3er rótulo seleccionado?

¿Cuál es la 3ra persona seleccionada?

Un diario económico publica datos sobre ventas, ganancias, impuestos, dividendos, acciones, ganancia por acciones de las 500 compañías más importantes de la Nación.  
Ud. está contratado para seleccionar una muestra simple al azar de 10 compañías de un listado de 500.  
Explique cómo rotularía a las compañías y luego use la calculadora.

enviar a un equipo de investigadores a cada sitio para su observación.  
Por lo tanto necesitamos seleccionar una muestra simple al azar de tamaño 5 de una población de  $N=80$ .

Usando **T.I.** para seleccionar 5 sitios al azar.

(

**Paso 1:** Rotule o enumere de 1 a 80 los sitios.

(

**Paso 2:** Use T.I. con semilla 29  
(

Resultado: los sitios muestreados son #50; #60, #43; #49, #74

( tabla

Si uso la fila 10: 85.475 36.857 53.342 etc.,  
(

85, 47, 53, 68, 57, etc.  $\Rightarrow mn = 5$  simple al azar: # 47; # 53; # 68; # 57; # 34  
↓  
Lo descarto.

### PARA RESOLVER !!! 2.4. Muestra simple al azar

dora (con semilla igual a 53) o Tabla de número al azar (fila 26 leyendo de izquierda a derecha) para identificar los rótulos que integraron la muestra.

El m.s.a. es la mejor elección para cierto tipo de ensayos, sin embargo hay situaciones en este método que resulta difícil de manejar y/o demanda mucho tiempo.

Supongamos que es importante conseguir información de distintas regiones, distintos subgrupos de la población. A veces las unidades de una población caen dentro de grupos naturales llamados estratos. Por ejemplo, si deseamos estimar el costo medio del cuidado de niños pequeños en una comunidad, podría ser conveniente mirar a la población, como el agregado de tres subgrupos:

- 1) hogares naturales,
  - 2) guarderías pequeñas,
  - 3) guarderías grandes.
- En vez de tomar una muestra de centros para el cuidado de niños, se podía tomar una muestra aleatoria simple de cada grupo, que nos daría información sobre cada grupo, y también sobre el total de centros.

## 2.6. MUESTREO ALEATORIO ESTRATIFICADO

Una muestra aleatoria estratificada se obtiene dividiendo o estratificando la población en subgrupos o estratos mutuamente excluyentes y sacando una m.s.a. de unidades de cada estrato.

Subgrupos mutuamente excluyentes: cada unidad de la población pertenece sólo a un estrato.

Una muestra aleatoria estratificada puede ser usada si se desea obtener información sobre las unidades dentro de cada estrato separadamente. Además, en algunos casos este método proporciona estimaciones más precisas sobre la población que la que se obtendría con el método de muestreo simple al azar.

Es más efectivo cuando las unidades dentro del estrato individual son muy semejantes con respecto a la característica que se mide y los estratos son muy diferentes entre sí.

Supongamos que se sabe que los hombres responden de una misma manera a una determinada pregunta, pero que las mujeres lo hacen de

forma más variada. Si sólo se pueden elegir 4 personas para interrogar, sería conveniente dividir la muestra en dos, y encuestar a tres mujeres y sólo a un hombre.

### Ejemplo 2.6. Dieta baja en grasas, alta en carbohidratos

Supongamos que un experto en alimentos desea determinar los efectos de una dieta baja en grasas, alta en carbohidratos en una Universidad. El experto cree que la dieta puede tener un efecto distinto en los estudiantes de distintas edades y género.

Subdividió la población en subpoblaciones:

Mujeres > 21 años, Hombres > 21 años  
Mujeres ≤ 21 años, Hombres ≤ 21 años

Y luego extraigo una muestra simple al azar de cada uno de estos estratos mutuamente excluyentes, formando así la muestra estratificada.

## 2.6. Muestreo aleatorio estratificado

Forme un grupo de 4 estudiantes. Necesitará por lo menos contar con una mujer y un hombre en su grupo. La población de interés es su grupo entero.

Una cuestión de interés: ¿Cuántas clases de Inglés tomará este mes?

Deseamos conocer el número promedio de clases de Inglés para su población.

✓ Escriba los nombres de las personas de su grupo e interrogué cuántas clases tomará en el mes.

✓ Compute el promedio para su población (sume todas las respuestas y luego divida por el número de estudiantes en su población).

Promedio =  $\text{Sum}/N = \dots$   
Este valor es un/a ..... (parámetro o estadística)

✓ Tome una muestra simple al azar de tamaño  $n = 2$

1) Asigne un rótulo diferente a cada nombre de su lista.

2) Seleccione un lugar en su tabla de números al azar (fila 10, columna 22), o bien use su calculadora (semilla 104).  
 ✓ ¿Quiénes fueron seleccionados y cuáles fueron sus respuestas?

3) Calcule el promedio a partir de esa muestra simple al azar de  $n=2$

$$\text{Promedio} = \frac{\text{sum}/n}{\dots} = \dots$$

Este valor es un/a ..... (parámetro o estadística)

Ahora está capacitado para tomar una muestra de tamaño  $n=2$  pero deseas tomar una mujer y un varón en tu muestra **¿cómo estratificar?**

Pasos:

1) Escriba un listado de los Varones y Mujeres en su grupo, es decir forme estratos e incluya sus respuestas al lado de sus nombres.  
**Mujeres** (estrato 1)  
**Varones** (estrato 2)

2) Asigne un rótulo a cada unidad.

3) Seleccione una muestra aleatoria simple de tamaño  $n=1$  del estrato de mujeres (comience en fila 14 y col. 1 o semilla 24) y extraiga una muestra de tamaño  $n=1$  de varones (comience en fila 23, col. 20 o semilla 35). Registre las respuestas.

4) compute la respuesta promedio para cada estrato separadamente:  
 Estrato 1      promedio = suma/n = .....  
 Estrato 2      promedio = suma/n = .....  
 Por lo tanto tenemos una estimación para cada estrato ¿cómo usaremos estas dos estimaciones para obtener el promedio de toda la población?

5) Calcule el promedio general combinando el promedio de los dos estratos.  
 Dado que el tamaño de los estratos puede diferir, tomamos un promedio ponderado por la proporción de unidades en cada estrato respecto de la población total.

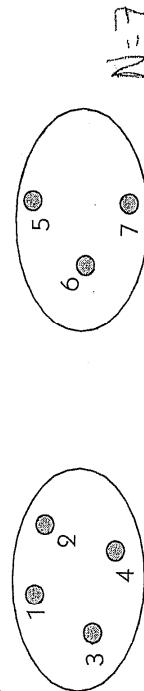
Promedio General:

$$\frac{\# \text{unidades en el estrato 1}}{N} \left[ \frac{\text{estimación promedio}}{\text{Estrato 1}} \right] + \frac{\# \text{unidades en el estrato 2}}{N} \left[ \frac{\text{estimación promedio}}{\text{Estrato 2}} \right]$$

- ✓ Este número es un/a ..... (parámetro, estadística)
- ✓ ¿Cómo resulta este valor comparado con el promedio poblacional?

### Ejemplo 2.7. Promedio ponderado

Supongamos una población que consiste en 4 mujeres con valores de respuesta 1, 2, 3 y 4; y 3 hombres cuyos valores de respuesta son 5, 6 y 7.

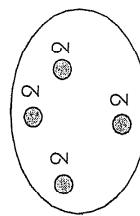


### Tamaño de la población N=7

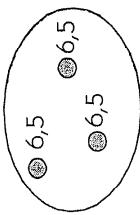
- ✓ Una muestra simple al azar de mujeres de tamaño  $n=3$  resultó en las respuestas  $\{1, 2, 3\}$  y así la estimación promedio del estrato M es 2.
- ✓ Una muestra simple al azar de hombres de tamaño  $n=2$  resultó en las respuestas  $\{6, 7\}$  y así la estimación promedio del estrato H es 6,5. El estrato M tiene 4 unidades poblacionales. El estrato H tiene 3 unidades poblacionales.

Para computar el promedio ponderado general, reemplazamos cada uno de los 4 valores correspondiente al estrato M por el promedio estimado que es 2 y cada uno de los 3 valores del estrato H por el promedio estimado que es 6,5.

### Estrato mujeres (M)



### Estrato hombres (H)



### Tamaño de la población N=7

$$\begin{aligned} \text{Promedio General} &= [(2+2+2)+(6,5+6,5+6,5)]/7 = [4(2)+3(6,5)]/7 = \\ &= (4/7)2 + (3/7)6,5 = 3,93 \end{aligned}$$

- ✓ ¿Cómo hicimos?
- ✓ ¿Cómo es esta estimación basada en la muestra aleatoria estratificada comparada con el promedio poblacional?

El promedio poblacional es: .....

### **PARA PENSAR !!!**

- ✓ ¿Cuándo se toma un tamaño de muestra más grande que en un estrato que en otro?
- ✓ Cuando forma los estratos, ¿cómo debe ser la variabilidad de las unidades dentro de cada estrato comparada con la variabilidad entre los estratos?
- ✓ Una muestra estratificada ... ¿es una muestra simple al azar? Explique.

### **Ejemplo 2.8. ¿Cuál es el diseño?**

En un estudio de mercado, una compañía de servicios desea estimar el monto promedio de las facturas mensuales de sus clientes habituales.

Unos 9,050 clientes pertenecen a un área geográfica particular bajo estudio. Se conoce también que 2,210 de tales clientes, recibieron montos de facturas mensual mínimo de \$86.5. Se propusieron dos diseños distintos de muestreo:

- I. Seleccionar una m.s.a. de 100 clientes del total de la población.
  - II. Seleccionar una m.s.a. de 100 clientes de los 6,840 clientes que recibieron montos de facturas superiores a \$86.5.
- ¿Cuál diseño recomendaría? .....

La población puede clasificarse en dos estratos según el monto mensual de facturas.

#### **Estrato I:** mínimo de \$86.5

#### **Estrato II:** mayores de \$86.5

Conociendo que las cuentas mensuales reales para aquellos del Estrato I es \$8.65, hemos muestreado efectivamente el 100% de ese estrato.

Necesitamos información de las cuentas mensuales para aquellos del Estrato II. Por lo tanto muestreamos tantas como podamos, por ejemplo,  $n=100$  para este estrato. El promedio obtenido de cada estrato será combinado usando el promedio ponderado.

### **2.7. Prácticas contables**

#### **PARA RESOLVER !!!**

Los contadores a menudo usan muestras aleatorias estratificadas durante las auditorías para verificar los registros de valores recibidos. La estratificación está basada en el monto de los ítems y a menudo incluye el 100% de los ítems más grandes.

Una compañía informa haber recibido 5,000 valores. De estos, 200 son de montos superiores a \$100,000, otros 1,000 están entre \$10,000 y \$100,000 y el remanente –3,800– corresponde a montos menores a \$10,000.

Usando estos grupos como estratos, Ud. decide verificar todos los valores de montos más grandes ( $> \$100,000$ ) y tomar una m.s.a. del 5% de los valores del estrato (\$10,000 a \$100,000) y el 1% de los valores menores ( $< \$10,000$ ).

- a) Basado sobre este diseño muestral ¿cuántos valores serán muestreados?

# de valores de gran monto a ser muestreados = .....

# de valores de menor monto a ser muestreados = .....

- b) Describa cómo rotularía los valores en el "Estrato pequeño" para seleccionar algunos valores de monto pequeño. Use calculadora (semilla=25) o la Tabla de números al azar (fila 18, col.1) para seleccionar sólo los 5 primeros valores de montos pequeños que luego serán verificados.
- Piense cuidadosamente sobre cuántos valores de montos pequeños hay para rotular.

#### **2.8. Asistentes de día**

Una ciudad tiene otorgadas 40 licencias a centros diurnos de asistencia a niños.

Un inspector de dicha ciudad necesita verificar la aceptación de esos centros con respecto a una nueva ordenanza de seguridad. Debido a la escasez de tiempo, el inspector sólo inspeccionará 12 centros mensualmente. Estratificará los centros según el número de asistidos a su cargo.

Un centro con 15 asistidos o menos es considerado un pequeño centro.

- El estrato I consta de 10 grandes centros ( $n^o$  de asistidos > 15)  
 El estrato II consta de 30 pequeñas centros.  
 El inspector planea tomar una m.s.a. del Estrato I (grandes asistencias) y otra del Estrato II (pequeños asistencias).  
 La variable "número de asistidos" está considerada como la variable de estratificación: variable cuyos valores determinan cuáles unidades están en cada estrato.
- ✓ ¿Cuántos "centros grandes" le recomendaría al inspector muestrear del Estrato I?
  - ✓ ¿Cuántos "centros pequeños" del Estrato II?

 **!No existe una única solución!**

Si no hay información previa sobre la variabilidad de los estratos, podríamos muestrear en proporción al tamaño del estrato y el tamaño de la población.

Tamaño Muestral  
 del Estrato I: .....  
 Tamaño Muestral  
 del Estrato II: .....

En la parte inferior se muestra la población y su respuesta:  
 < una lista con 40 centros de día; con **sí** se indica que el centro aceptaría la nueva ordenanza y con **no** que la ordenanza no es aceptada. Por supuesto que esta lista no es conocida por el inspector.

Estrato I (grandes centros)		Estrato II (pequeños centros)									
1	Sí		1	Sí	11	No	21	No			
2	Sí		2	No	12	No	22	Sí			
3	No		3	No	13	Sí	23	Sí			
4	Sí		4	Sí	14	No	24	No			
5	No		5	No	15	Sí	25	Sí			
6	Sí		6	No	16	Sí	26	No			
7	Sí		7	No	17	No	27	No			
8	No		8	Sí	18	Sí	28	Sí			
9	Sí		9	Sí	19	No	29	No			
10	Sí		10	No	20	No	30	Sí			

Use una calculadora o Tabla de números al azar para elegir una muestra estratificada de tamaño 12 según lo especificado arriba.

Para el Estrato I use semilla 21 y N=10 o fila 20 y col 11.  
 Para el Estrato II use semilla 121 y N=30 o fila 30 y col 6.

Liste los centros de su muestra y sus respuestas.

Basado en sus resultados muestrales, estime la proporción de centros diurnos que están de acuerdo con la nueva ordenanza.  
 Halle la estimación ponderada.

$$T.I. 83 \hat{\rho} = \text{estadístico} = \dots \dots \dots$$

$$T.N.A. \hat{\rho} = \text{estadístico} = \dots \dots \dots$$

- ✓ ¿Cuál es la **Verdadera proporción** poblacional de las asistencias de día que fueron cumplidas?  
 $\rho$  = verdadero parámetro = .....  
 ....
- ✓ ¿Cómo resultó su estimador o estadístico comparado con el parámetro poblacional?

El muestreo aleatorio estratificado trabaja bien en situaciones donde la variabilidad entre los valores dentro del estrato es menor comparado con la variabilidad entre estratos.

La información es obtenida para cada estrato separadamente. Esta información es combinada a través del promedio ponderado para obtener una estimación general de toda la población.

En el promedio ponderado, los pesos son las fracciones de la población en cada estrato, sin importar el tamaño de la muestra dentro de cada estrato. El método de muestreo estratificado es **insensgado** cuando se usa el promedio ponderado.

A veces, implementar el método de muestreo simple aleatorio o

el estratificado aleatorio puede resultar costoso. En tales casos se usan otros métodos como el **SISTEMÁTICO** y la **CONGLOMERACIÓN** (muestreo en grupos).

## 2.7. MUESTREO SISTEMÁTICO

Supongamos tener un listado de 2.000 nombres de estudiantes de los cuales deseamos seleccionar una muestra de tamaño  $n=100$ . Podríamos seleccionar un nombre cada 20 de la lista. Esta es la idea que encierra el método de muestreo sistemático.

### Definición

Para una muestra sistemática "1-en- $K$ ", ordenamos las unidades poblacionales de alguna manera y seleccionamos aleatoriamente 1 de las  $k$  primeras unidades de dicha lista ordenada. Esta unidad seleccionada es la primera unidad a ser incluida en la muestra. Continuaremos seleccionando cada  $k$ -ésima unidad de dicho listado hasta completar el tamaño de la muestra deseado.

Para nuestro listado de 2.000 nombres elegiríamos un punto de comienzo entre los primeros 20 nombres. Comenzando con ese nombre seleccionado, muestreáramos cada 20 nombres después de él:

La muestra sistemática 1-20 consistirá de una persona de cada grupo de 20, igualmente espaciados a través del listado.

Una vez que se determina el punto de comienzo, el muestreo continúa sistemáticamente.

Antes de seleccionar el punto de partida, hay 20 muestras sistemáticas 1-en-20 posibles, y cada una es igualmente probable de ser tomada como la muestra actual.

Una ventaja es que la muestra es tomada de la población ordenada. Este método es muy rápido y conveniente cuando se cuenta con el listado de las unidades que integran la población. Por ejemplo, podríamos tener en un archivo un listado de pacientes con un cierto tipo de dolor que visitan al clínico y deseamos administrar un tratamiento equivalente a cada 7<sup>o</sup> paciente de la lista, comenzando con el paciente seleccionado aleatoriamente de entre los 7 primeros pacientes del listado.

Una desventaja es que puede conducir a una muestra sesgada. Una lista de pasajeros de un mismo vuelo ubicados por número de asiento. Hay 200 pasajeros sentados en 40 filas con 5 asientos cada uno. Si

deseamos tomar una muestra sistemática de cada 5 pasajeros, tendríamos pasajeros sentados en la misma ubicación de la fila, por ejemplo todos sobre un extremo o todos en el asiento del medio. Esto podría conducirnos a una respuesta sesgada sobre el confort de los vuelos.

## Ejemplo 2.9. Una muestra sistemática 1-en-4

Suponga que cuenta con 19 unidades de una población, rotuladas desde la "A" a la "S" y deseamos tomar una **muestra sistemática** de 1 en 4. Dividamos la población en grupo de 4 elementos

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

Las primeras 4 unidades son rotuladas con 1, 2, 3 y 4 seleccionamos aleatoriamente un número al azar de estos valores usando la Tabla de números al azar o bien la calculadora. Si el rótulo seleccionado es el "2", la muestra sistemática consistirá del 2º, 6º, 10º y 18º elemento. Las unidades serán: B, F, J, N, R.

Si el rótulo seleccionado es un "4", la muestra sistemática será: D, H, L, P. El tamaño muestral dependerá del rótulo que se seleccionó en el primer grupo.

**Observación:** Ud. no necesita dividir a la población en grupos de tamaño 4, sólo necesita un listado de las unidades de la población y aleatoriamente elige una de las 4 primeras unidades de la población y luego selecciona cada 4 unidades para incluir en la muestra.

### PARA PENSAR !!!

Consideremos nuevamente las 19 unidades que componen una población.

Las mismas están rotuladas de la A a la S. Se plantea tomar una muestra sistemática de 1 en 4

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

a) ¿Cuál es la chance de que la letra A sea seleccionada?

- b) ¿Cuál es la chance de que la letra J sea seleccionada? .. .

c) ¿Cuál es la chance de que la letra P sea seleccionada? .. .

d) ¿Cada unidad tiene la misma chance de ser seleccionada?  Sí  No

e) ¿Una muestra sistemática es una m.s.a.?  Sí  No

f) ¿Un muestreo sistemático es un muestreo estratificado?  Sí  No

g) ¿El tamaño de la muestra es fijo? Explique . . . . .

**PARA RESOLVER !!!**

**2.9. Muestreo sistemático en el censo de manzanas de una misma región**

La figura subsiguiente es el mapa del empadronamiento de una región, #5395 en Detroit, Michigan. Las regiones empadronadas son áreas homogéneas con una población promedio de alrededor de 4000. Cada manzana en la región está marcada con un número identificatorio de la oficina de Censo.

Los números comienzan con 101 y finalizan con 514. Note que algunos números están salteados, por ejemplo de 414 pasa a 501. A los encuestadores se les asignará un coníunto de manzanas

✓ Su tarea: tome una muestra sistemática (1 en 10) de manzanas. ✓ Siguiendo los números de identificación de manzanas en la población, use un lápiz para seguir la huella en el diagrama; es un cuadro que incluye un conjunto de manzanas.

- dicir, conecte las manzanas con una línea en orden ascendente de los números identificatorios.

✓ Para una muestra sistemática (1 en 10) las primeras 10 manzanas forman su primer grupo. Rotule esas 10 manzanas y use su calculadora (semilla = 33 y N=10) o la T.N.A. (fila 16, col. 1) para seleccionar aleatoriamente su manzana inicial que será la primera manzana en su muestra.  
¿Cuál es su primer manzana seleccionada? : .....  
.....

✓ Ahora comenzando con la primera manzana seleccionada, elegimos cada 10º manzana, en orden, para ser incluida en su muestra. Liste las manzanas que integran la muestra con sus correspondientes números identificatorios.

## **2.8. MUESTREO DE CONGLOMERADOS**

Supongamos que se desea conocer acerca de los hábitos de mirar TV que tienen las familias, en particular los niños, adultos femeninos y adultos masculinos.

Esperamos que las respuestas de los niños sean muy similares entre sí. Lo mismo ocurre con los adultos masculinos, pero sus hábitos no son necesariamente los mismos que los de los niños y las respuestas de los adultos femeninos son muy similares entre sí, pero no necesariamente las mismas que la de los adultos masculinos y niños.

Supongamos que queremos obtener una muestra que contenga respuesta para los tres grupos

Su primer pensamiento podría haber sido una muestra estratificada aleatoria con tres estratos: **Chicos, Mujeres Adultas, Hombres Adultos**. Sin embargo, tal plan de muestreo podría resultar muy costoso, especialmente si a los individuos seleccionados se los entrevista personalmente. Por ejemplo un chico seleccionado podría vivir en el norte de la región, un adulto femenino en el oeste y un adulto masculino en el sur. En este caso en vez de usar muestreo estratificado, se usará el método de **muestreo de conglomerados** (clustering).

Cada familia que tiene un adulto masculino, un adulto femenino y un chico formará una nueva unidad llamada **conglomerado**, **raci-mo o cluster**.

Entonces seleccionaremos una o más familias, es decir uno o más conglomerados y se entrevistará a cada uno del conglomerado seleccionado.

#### Definición

En el muestreo de conglomerados las unidades de la población se agrupan en conglomerados o clusters.

Uno o más conglomerados (clusters) son seleccionados aleatoriamente.

Si un cluster es seleccionado, todas las unidades serán incluidas en la muestra.

A menudo se confunde el muestreo por conglomerados con el muestreo estratificado. En ambos casos, las unidades de la población son divididas en grupos.

En el **muestreo estratificado**, los estratos están formados de tal manera que las unidades dentro del estrato son **homogéneas** con cada unidad de la población perteneciendo a uno y sólo a uno de los estratos. Aleatoriamente seleccionamos algunas unidades dentro de cada estrato para formar la muestra.

En el caso de **muestreo de conglomerados**, los conglomerados están formados como si fuesen una "**mini población**", una pequeña versión de la gran población, con cada unidad de la población perteneciente a uno y sólo uno de los clusters. Aleatoriamente seleccionamos uno o más conglomerados (clusters), así las unidades de la población entran en la muestra por conglomerados; no individualmente.

El beneficio del muestreo por conglomerados está en el ahorro del tiempo y en el costo. Menos energías y monedas se necesitarán si el entrevistador está en un área geográfica determinada y no viajando para atravesar distintas áreas.

Algunos conglomerados o clusters están naturalmente definidos tales como: familias, hospitales, etc.

Puede ser que la definición natural de clusters no refleje la variación que está presente en la población. Podría no parecer una "mini población". En este caso aunque ahorraremos \$, perdemos precisión.

#### Ejemplo 2.10. Una muestra de conglomerados de estudiantes

El presidente de una universidad privada religiosa desea conocer la opinión de los 280 alumnos ingresantes respecto del entrenamiento o práctica religiosa que recibieron.

En esa universidad todos los alumnos ingresantes son invitados a tomar una clase introductoria de religión. Un diseño de muestreo por conglomerados será el siguiente:

- Escribir un listado de todas las clases introductorias de religión
  - Enumerar o rotular cada clase en la lista.
- Recuerda que si usa T.N.A., todos los rótulos deben tener el mismo número de dígitos, de allí que cada clase tiene la misma chance de ser seleccionada.

• Tomar una muestra simple al azar de 5 clases usando la calculadora o T.N.A. para seleccionar los rótulos de clases.

• La muestra consistirá de todos los estudiantes de esas 5 clases seleccionadas.

#### PARA RESOLVER !!!

#### 2.10. Muestreo de conglomerados de las manzanas

La figura siguiente es el mapa de empadronamiento de la región #5375 en Detroit, Michigan. Las manzanas en la región han sido agrupadas para formar conglomerados con sus correspondientes números de identificación dado por la oficina de Censos.

404	403	402	401	308	301	211	204	203	108	101
405	406	407	408	308	301	210	205	202	110	109
504	503	502	501	306	303	209	206	201	111	106
505	506	507	508	305	304	208	207			105
										104

Se observan 5 conglomerados o "clusters" basadas en los "cientos" según su número de identificación.

- Tarea:** Tome una muestra aleatoria de conglomerados de manzanas de tamaño 2.
- Rotule los 5 conglomerados.
  - Tome una muestra aleatoria de dos conglomerados. Use la calculadora ( $\text{semilla}=10$ ;  $N=5$ ) o la T.N.A. (fila 24; col. 31).
  - ¿Cuál es el rótulo de su primer conglomerado seleccionado?
  - La muestra consistirá de todas las manzanas de estos 2 conglomerados seleccionados. Liste los números identificatorios de las manzanas en su muestra.
  - Cálculadora:  
T.N.A.:  
¿Cuál es la chance de que una manzana sea seleccionada por su muestra?

**PARA RESOLVER !!!**  
**2.11. Muestreo de conglomerados de alumnos**

La población de interés está formada por los estudiantes de la clase de hoy. Cada una de las filas de los alumnos en esta clase forma un conglomerado.

Fila #10  
Fila #9  
Fila #8  
Fila #7  
•  
•  
•  
Fila #1

**Frente de la clase**

- Asigne un rótulo a cada cluster (cada fila) y seleccione un cluster aleatoriamente [use su calculadora; semilla=279 o T.N.A.; fila 9; col. 21]
- ¿Cuál clusters o conglomerado seleccionó?

- ¿Cuál es la chance de que cualquier estudiante de la población sea seleccionado?
  - Pregunte algo de interés a todos los alumnos sentados en la fila seleccionada y registre su respuesta.
  - Compute la respuesta promedio para el conglomerado muestral.

$$\text{Promedio} = (\bar{x}) = \text{Sum}/n = \dots ; \text{este número es un parámetro, estadística)$$

**PARA PENSAR !!!**

- ¿Una muestra por conglomerados es una muestra simple al azar? Explique
- ¿Una muestra por conglomerados es una muestra aleatoria estratificada? Explique
- Cuando formamos los conglomerados ¿Cómo debería ser la variabilidad de las unidades dentro de cada conglomerado comparado con la variabilidad entre conglomerados?
- ¿Este criterio es el mismo que para el muestreo estratificado?

**PARA RESOLVER !!!**

**2.12. Menciona el método de muestreo**

Lea cada situación e identifique el método de muestreo descrito [simple al azar; estratificado; sistemático, por conglomerados; convencional].

Discuta sus respuestas con sus compañeros.

- #1. Un cargamento de 1.000 botellas de colonia de 80 ml han llegado a un comerciante. Estas botellas fueron embarcadas en 50 cajas con 20 botellas cada una. De las 50 cajas 5 fueron seleccionadas aleatoriamente y se obtuvo el contenido promedio de estas 100 botellas seleccionadas.
- Método: .....  
.....  
.....  
.....  
.....

#2. Un miembro de una facultad desea tomar una muestra de 1.600 alumnos de una carrera.

Cada alumno tiene un código o número identificatorio. Se dispone del listado de todos los números identificatorios.

El profesor selecciona un número aleatorio entre los 16 primeros números identificatorios de dicho listado y luego elige los otros integrantes de la muestra a intervalos de 16 números identificatorios de la lista.

Método: ..... .

#3. El elecano desea tomar una muestra de los 1600 alumnos de una carrera. El mismo decide entrevistar a los 100 primeros alumnos que ingresen a un mismo curso el próximo lunes a la mañana.

Método: ..... .

### PARA RESOLVER !!!

#### 2.13. ¿Qué método de muestreo podría haber usado?

Una población consta de 12 personas. Estas 12 personas están listadas según la tabla adjunta:

Población	Estrato I	Estrato II
Congl. I	NORA (1)	GABRIEL (2)
Congl. II	AUGUSTO (8)	EUGENIO (7)
Congl. III	LUCÍA (9)	PABLITO (10)
		GERMÁN (11)
		MEMI (12)

Observación: Los números debajo de cada nombre son sus números identificatorios.

Se obtuvo una muestra de tamaño 4 (personas) de esta población. Consideré los siguientes métodos de muestreo: i) simple al azar ii) aleatorio estratificado con igual tamaño muestral de cada estrato iii) de conglomerado de filas.

Para cada muestra indique cuál método lo generó, encerrando **Sí** o **No** en cada caso.

**Ayuda:**  
Más de un método es posible.

Muestra	Método de muestreo		
	i) simple al azar <input checked="" type="radio"/>	ii) estratificado <input checked="" type="radio"/>	iii) conglomerado <input checked="" type="radio"/>
a) Augusto, Eugenio, Gastón, Dante	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
b) Nora, Pablito, Augusto, Gabriel	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
c) Augusto, Eugenio, Marta, Memi	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>

d) Tome una muestra sistemática 1-en-3 de esta población. Use los números identificatorios como un listado ordenado de las unidades de esta población  
[Use calculadora científica con semilla=78; N=3 o T.N.A. con fila=3; col. 3. Liste los nombres de las personas en su muestra]

### 2.9. MUESTREO MULTIETÁPICO

Hemos analizado distintos métodos de muestreo, cada uno con sus ventajas y desventajas.

Actualmente muchos ensayos muestrales nacionales usan una combinación de tales métodos de muestreo.

Por ejemplo podrían dividir al país en regiones, dentro de cada región estratificar por comunidad (rural, urbana, suburbana). Dentro de las comunidades seleccionadas elegir pocas manzanas aleatoriamente.

Luego seleccionar sistemáticamente algunas direcciones dentro de las manzanas seleccionadas y entrevistar a un miembro del domicilio seleccionado.

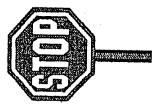
Este método de muestreo en etapas es llamado **Muestreo Multietápico**. En cada etapa se puede elegir algunos de los métodos de muestreo analizados. Los items que son muestreados en cada etapa se tornan gradualmente más pequeños en tamaño (o número).

Los items seleccionados en una etapa i), son seleccionados de cada "item" que fueron seleccionados en la etapa previa.

Una ventaja del muestreo multietápico es que no se necesita un listado completo con las unidades finales que fueron seleccionadas.

Las unidades seleccionadas en la última etapa fueron los domicicios, pero no necesitamos comenzar con un gran listado con todas las direcciones del país.

Sólo necesitamos un listado de unidades muestrales en cada etapa, es decir, un listado de las comunidades dentro de cada región,

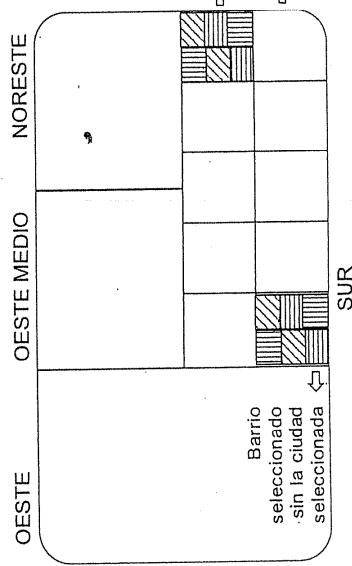


Luego un listado de las manzanas dentro de las comunidades seleccionadas. Un entrevistado podría ser enviado a la manzana seleccionada y listar los domicilios y luego muestrear los de esa lista. Obviamente otro de los beneficios es el tiempo y la reducción del costo.

## Ejemplo 2.11. Procedimiento para un ensayo pre-electoral nacional

Se llevó a cabo un estudio en 4 regiones geográficas de EEUU (Nor-este, Sur, Medio Oeste, Oeste) Dentro de cada región las ciudades fueron agrupadas por tamaño, según número de habitantes.

- Etapa 1: muestrear algunos de estos grupos de ciudades.
- Etapa 2: muestrear algunas ciudades dentro de cada grupo seleccionado.
- Etapa 3: muestrear algunos distritos dentro de cada ciudad seleccionada.
- Etapa 4: muestrear algunos barrios dentro de cada distrito seleccionado.
- Etapa 5: muestrear algunos hogares dentro de cada barrio seleccionado.
- Etapa 6: entrevistar a todos o a algunos miembros de cada hogar seleccionado.



PALABRAS CLAVES

- CENSO
- POBLACIÓN
- UNIDADES
- MUESTRA
- VARIABLES
- TAMAÑO DE LA POBLACIÓN ( $N$ )
- TAMAÑO DE LA MUESTRA ( $n$ )
- PARÁMETRO
- ESTADÍSTICA
- SESGO (VÍCIO O BIAS)
- MUESTREO CONVENCIONAL
- MUESTREO VOLUNTARIO
- SELECCIÓN
- SESGO DE NO RESPUESTA
- SESGO DE RESPUESTA
- SESGO DE LONGITUD DE MUESTREO
- MÉTODOS DE MUESTREO PROBABILÍSTICO
- MUESTRA SIMPLE ALEATORIA
- TINA. TABLA DE NÚMEROS AL AZAR
- SEMILLA
- ESTRATO
- MUESTRA ESTRATIFICADA ALEATORIA
- MUESTRA SISTEMÁTICA
- MUESTRA DE CONGLOMERADO (O CLUSTERS)
- MUESTRA MULTITÁPICA

**EJERCICIOS**

**2.1.** Los registros de una ciudad muestran que el 35% de todos los residentes son mayores de 60 años. Para ensayar Ud. llamará a 200 números telefónicos de los residentes de esa ciudad elegidos aleatoriamente. De los residentes contactados el 28% resultó mayor de 60 años. Observando el párrafo de abajo, encierre en un círculo la mejor respuesta:

- ... el 35% es un (parámetro, estadístico).
- ... el 28% es un (parámetro, estadístico).

**2.2.** De cuatro marcas de autos lujosos se determinaron aleatoriamente 5 mediciones y se registraron las millas recorridas por galón, o sea el consumo:

	Mpg				
Rolls Royce	10.1	12.3	13.1	11.8	12.6
Porsche	14.8	13.1	13.7	14.5	13.9
Lexus	17.0	16.1	16.8	16.5	17.5
BMW	18.1	19.0	18.9	19.9	20.5

Supongamos que se conoce el verdadero promedio de mpg de todos los autos lujosos (de estos 4 tipos) y es de 17.4 mpg.

**a)** Si la información sobre las mpg se obtiene de un comerciante de autos lujosos de San Diego. Se podría cometer un sesgo:

- i) sesgo de selección
- ii) sesgo de no respuesta
- iii) sesgo de respuesta

**b)** Si la información de las mpg está basada en un diseño muestral verdadero pero de propietarios "reticentes a contestar" sobre las características de dichos autos de los cuales sólo el 8% respondió al ensayo. El nombre del sesgo que podría ocurrir en este caso:

- i) sesgo de selección
- ii) sesgo de no respuesta
- iii) sesgo de respuesta

**c)** Si los datos muestrales de mpg están basados en un ensayo muestral de comerciantes elegidos aleatoriamente pero generalmente creen que

un bajo millaje por galón es atribuido a autos lujosos. El nombre del sesgo que podría ocurrir es:

- i) sesgo de selección
- ii) sesgo de no respuesta
- iii) sesgo de respuesta

**2.3.** Un departamento educativo realizó un relevamiento deseando conocer el ingreso obtenido por los alumnos. Una pregunta del cuestionario era: ¿Cuál es su ingreso corriente? Respondió aproximadamente el 30% de los alumnos.

Realice un comentario sobre algún tipo de sesgo, si corresponde, considerando también cómo resultaría el ingreso promedio entre los alumnos que respondieron y el ingreso promedio actual de todos los alumnos.

**2.4.** Seleccione una de las siguientes alternativas:

- En el muestreo simple aleatorio
- El sesgo de no respuesta será pequeño
  - El sesgo de respuesta será pequeño
  - Dos muestras de igual tamaño tendrán la misma chance de ser extraídas.
  - A cada sujeto se le hacen preguntas fáciles y se permiten respuestas aleatorias.

**2.5.** Una clase consiste de 100 alumnos. A cada estudiante se le asigna uno y sólo un número del 00 al 99. Supongamos que una caja tiene 100 tarjetas numeradas con dichos números. Revolvemos bien y con los ojos vendados extraemos 20 tarjetas simultáneamente. Los alumnos asignados con tales números integran la muestra:

- a)** ¿Es una muestra aleatoria simple de la clase? Explique.
- b)** Si su respuesta en (a) es sí, ¿es la muestra extraída con o sin reemplazo?

**2.6.** Un periódico de Illinois informó: "Un ensayo reportó que la mayoría de los trabajadores consume droga en el trabajo". El artículo dio la siguiente información: ... "el relevamiento realizado entre las personas que llamaron a la línea nacional de ayuda (consumidores de cocaína) de las cuales 227 fueron elegidas aleatoriamente durante un periodo de 6 semanas en el mes de febrero y marzo..." El 92% de los que llamaron dijeron que a veces trabajaron bajo influencia de drogas. Marzo 1995.

- a) ¿Qué clase de muestreo se usó?  
 b) ¿De qué población provino la muestra?  
 c) ¿Piensa que el título de dicho artículo es correcto? Explique.

**2.7.** Juana tiene dos billeteras. Una contiene dos billetes de 1 peso y 2 billetes de \$5 mientras que la otra billetera tiene 4 billetes de \$1. Decide seleccionar aleatoriamente una billetera y sacar dos billetes. Usará los billetes seleccionados para decidir cuál de las siguientes hipótesis sustentar:

**H<sub>0</sub>:** la billetera contiene: \$1, \$1, \$1, \$5  
**H<sub>1</sub>:** la billetera contiene: \$1, \$1, \$1, \$1

La regla de decisión de Juana es:

Rechazar  $H_0$  y decidir que la billetera contiene 4 billetes de \$1, si los dos billetes seleccionados son de \$1.  
 (Suponga que ambas billeteras y el dinero contenido en ellas están lo suficientemente bien mezcladas como para pensar en una muestra simple al azar de tamaño 2)

Dato: Juana seleccionó dos billetes de \$1.

a) ¿Cuál es la decisión de Juana?  
 b) ¿Fueron los datos estadísticamente significativos?

c) ¿Puede Juana estar en lo cierto pensando que ella tiene la billetera con 4 billetes de \$1 y que no es la que tiene dos billetes de \$1 y dos de \$5?  
 d) ¿Qué clase de error pudo haber cometido Juana?

e) Suponga que cada billete individual es distingüible:  
 Billetera bajo  $H_0$ : \$11, \$12, \$51, \$52

Billetera bajo  $H_1$ : \$11, \$12, \$13, \$14

Liste las posibles muestras aleatorias de tamaño dos que podrá conseguir si las selecciona de la billetera bajo  $H_0$ .  
 f) Idem e) pero seleccionando de la billetera bajo  $H_1$ .

g) Calcule el p-value para este resultado observado de 2 billetes de \$1.  
 Se puede concluir que:

- a) El porcentaje de ítems defectuosos en la población es 2,5%.  
 b) Fueras de cada 200 ítems, habrá 5 ítems defectuosos.  
 c) El porcentaje de ítems defectuosos en la muestra es 2,5%.  
 d) El porcentaje de ítems defectuosos en la población debe ser mayor que 2,5% dado que la población es siempre más grande que la muestra.  
 e) Ninguno de los de arriba.

**2.9.** Una empresa emplea 1,000 hombres y 200 mujeres. La tarea de un gerente consiste en entrevistar a 100 hombres y 100 mujeres empleados, integrantes de una muestra estratificada:

- a) ¿Cuál es la chance de que un empleado particular femenino sea entrevistado?  
 b) ¿Cuál es la chance de que un empleado particular masculino sea entrevistado?  
 c) Si elegimos una muestra aleatoria estratificada de empleados de 100 varones y 20 mujeres de esa empresa, cada empleado tendrá la misma chance de ser seleccionado,

¿Cuál es la chance? .....  
 ¿Es esta una muestra simple al azar? .....  
 Explique.

**2.10.** Una organización estudiantil desea conocer las actitudes de los estudiantes sobre una propuesta de cambio en el horario de la biblioteca. Seleccionaron aleatoriamente 100 ingresantes, 100 del ciclo básico y 100 alumnos de los últimos años. ¿Qué tipo de diseño muestral se utilizó?

**2.11.** Una oficina de servicios desea investigar acerca de la discriminación del sexo de los pagadores de bancos de cierto país. El estudio estará concentrado en 85 bancos de dicho país.  
 Hay 300 mujeres pagadoras y 100 hombres pagadores a través de los 85 bancos.  
 De las 300 mujeres pagadoras se seleccionaron aleatoriamente 75 y de los 100 pagadores se seleccionaron aleatoriamente 25. ¿Qué tipo de diseño muestral se usó en este estudio?

**2.12.** La secretaría estudiantil de un Colegio, desea estimar el monto promedio de dinero que los estudiantes gastan cada semestre en libros de estudio.  
 Se propuso un método de muestreo estratificado. Se sugieren distintas variables de estratificación (para formar los estratos). Para cada opción indique si dicha variable es buena para estratificar y explique su respuesta.

- a) estratificar por carrera  
 b) estratificar según género (F o M)  
 c) estratificar según antigüedad del estudiante.

**2.13.** Supongamos que Ud. cuenta con 4 libros que planea leerlos en el

futuro. Dos son de Ciencia Ficción, los llamaremos A Y B, conteniendo 212 y 379 páginas respectivamente. Los otros dos libros son de Autoayuda, C Y D con 350 y 575 páginas respectivamente.

Supongamos que selecciona aleatoriamente uno de los libros de

Ciencia Ficción y también uno de los de Autoayuda.

**a)** ¿Cuál es el nombre del método de muestreo utilizado?

**b)** ¿Cuál es la chance de que ambos libros A y B sea seleccionados con el método?

**c)** Dé el nombre de dos libros que puedan ser seleccionados con este método.

**d)** Dé el nombre de otros dos libros que puedan ser seleccionados con este método.

**e)** ¿Cuál es la chance que el número total de páginas en los dos libros seleccionados exceda de 800 páginas?

**2.14.** Suponga un archivo que contiene 500 domicilios de los cuales 5 deben ser seleccionados. Un muestreo aleatorio sistemático se usará para seleccionar 5 domicilios entre 500. Los 500 domicilios ya han sido ordenados e identificados. Tome una muestra aleatoria sistemática -1 en 100- para seleccionar 5 domicilios [use semilla=48; N=100 o TNA: fila=55,col=1]

**a)** Describa su procedimiento de rotulación y liste los domicilios seleccionados.

**b)** ¿Cuál es la chance de que cualquier domicilio sea seleccionado? Explique su respuesta

**2.15.** Considere 975 docentes de una facultad privada. El decano desea conducir un ensayo sobre la propuesta de un nuevo método de enseñanza para obtener mayores ingresos para la institución.

El decano confecciona un listado de los 60 departamentos que allí funcionan. Planea tomar una m.s.a. de 6 departamentos. Cada docente pertenece a un solo departamento. Su muestra consistirá de todos los docentes de esos 6 departamentos.

**a)** Establezca el tipo de muestreo que se describe.

**b)** Los 60 departamentos están identificados como departamento #1, departamento #2, departamento #3 ... a departamento #60. Usted debe tomar una m.s.a. de 6 departamentos del listado de 60. Use su Tl. (semilla=79; y T.N.A.: fila=60; col=1). Sea cuidadoso al describir su procedimiento de rotulación.

**c)** ¿Puede determinar la chance de que algún profesor sea seleccionado?

Si su respuesta es sí, ¿Cuál es la chance?

Si su respuesta es no, explique su respuesta.

**2.16.** Suponga que el área final de un diseño de muestreo multietápico contiene 200 direcciones de las cuales 5 deben ser seleccionadas. La muestra de tamaño 5 es elegida como sigue:

**Paso 1:** Elegir aleatoriamente una de las primeras 40 direcciones de la lista.

**Paso 2:** La muestra consiste en la dirección elegida en el paso 1 y las direcciones que están en las posiciones: 40, 80, 120 y 160 luego de haber elegido el primer domicilio del listado. Por ejemplo si la 10<sup>a</sup> dirección fue elegida en el paso 1, luego nuestra muestra consiste de la 10<sup>a</sup>, 50<sup>a</sup>, 90<sup>a</sup>, 130<sup>a</sup>, y 170<sup>a</sup> dirección del listado.

**a)** ¿Qué método de muestreo se describió en los pasos 1 y 2, es decir, en la última etapa del diseño de muestreo?

**b)** ¿Cuál es la chance de que alguna dirección específica sea elegida entre las 200 direcciones elegidas de la etapa final? Explique su respuesta.