

MANUAL DE MUESTREO POBLACIONAL

Aplicaciones en Salud Ambiental

Dra. Ma. de la Luz Kageyama
Dra. Luz Helena Sanín Aguirre
Dra. Isabelle Romieu



CENTRO PANAMERICANO DE ECOLOGÍA HUMANA Y SALUD
DIVISIÓN DE SALUD Y AMBIENTE
ORGANIZACIÓN PANAMERICANA DE LA SALUD
ORGANIZACIÓN MUNDIAL DE LA SALUD

ISBN 92 75 32200 7

Un agradecimiento especial por su colaboración en la revisión de este documento a la Dra. Silvia Ruiz Velasco del IIMAS/UNAM y a la Mtra. Martha María Téllez Rojo del INSP.

Diseño y formato: Ma. del Rosario Camacho
Revisión de estilo: Rocío Ledesma Saucedo

CONTENIDO

	Pág.
Prólogo	v
I. Introducción	1
II. Revisión de conceptos básicos de Estadística	7
III. Distribución Muestral de Medias	17
IV. Tipos de Muestreo	31
V. Aplicaciones con <i>MAS</i>	55
VI. Tamaño de la Muestra	67
VII. Marcos Muestrales	81
VIII. El Cuestionario	97
Apéndice 1. Estudio de Caso No. 1	111
Apéndice 2. Estudio de Caso No. 2	115
Apéndice 3. Guía de ejercicios	119
Bibliografía	121

PRÓLOGO

La Organización Panamericana de la Salud (OPS), a través de su Centro Panamericano de Ecología Humana y Salud (ECO) y con el apoyo de la Sociedad para la Cooperación Técnica de Alemania (GTZ), pone a su disposición el "Manual de Muestreo Poblacional. Aplicaciones en Salud Ambiental".

Este manual proporciona herramientas de bioestadística a los profesionales del área de salud ambiental para trabajos que requieren un muestreo poblacional con un mayor rigor metodológico.

Este documento forma parte de una serie de metodologías necesarias para apoyar a una red de centros colaborantes en el área de evaluación de riesgos para la salud por exposición a residuos peligrosos y se acompañará además de un manual para muestreo ambiental. Es parte de la labor editorial del Centro para desarrollar y poner al alcance de los profesionales interesados, material didáctico aplicable y de fácil manejo.

Complementa la serie de materiales desarrollados por la Red de Epidemiología Ambiental Global de la OMS (GEENET), como el texto de Epidemiología básica publicado en 1993 y el conjunto de publicaciones de la serie "Environmental Health Criteria", así como los documentos productos de los programas HEAL y HEADLAMP.

Dado que es un instrumento didáctico, es útil para el entrenamiento de personal que se dedica a trabajar en el campo. Mas que un resumen de libros de texto, el manual recoge la amplia experiencia de las autoras en preparar personal a nivel de pregrado y postgrado en bioestadística. Por ello, el manual muestra ejemplos prácticos, basados en experiencias reales, los cuales servirán al lector para comprender y aplicar mejor los conceptos.

Rob McConnell
Director ECO

Dependiendo de la información disponible, de los objetivos del estudio y del financiamiento para llevar a efecto esta encuesta, se pueden aplicar diversos esquemas de muestreo, por ejemplo, si sólo se cuenta con un listado de escuelas de la provincia, se podría comenzar seleccionando sólo algunas escuelas; de ellas, algunos grupos y finalmente algunos estudiantes, quienes integrarían la muestra definitiva. Este esquema involucra tres etapas escalonadas de selección que determinan las llamadas unidades de muestreo que corresponden a la primera, segunda y tercera etapa y quienes constituyen en este caso, el marco muestral.

2. ¿PARA QUÉ REALIZAR ESTUDIOS MEDIANTE MUESTREO?

- Evaluar riesgos de salud;
- determinar sitios específicos de monitoreo;
- determinar nuevos criterios de exposición;
- encontrar exposiciones no esperadas;
- evaluar efectos de nuevas opciones de control;
- cuantificar los efectos de diversas actividades, ocupaciones, estilos de vida, costumbres y creencias;
- validar diversos modelos de exposición y de riesgos;
- proyectar distribuciones de exposiciones futuras;
- disminuir costos e incrementar la profundidad de los estudios.

3. DISEÑO DE ENCUESTAS POR MUESTREO

3.1 Diseño de la investigación

En esta fase de planeación se deberán contemplar principalmente los siguientes aspectos:

3.1.1 Establecimiento de los objetivos del estudio

Para alcanzar resultados de calidad es necesario que desde la planeación los objetivos de la investigación queden claramente definidos, pues de ellos dependen las siguientes fases y aseguran el éxito de los proyectos.

Ejemplo 1.2 *Podríamos estar interesados en realizar un estudio en el interior de las viviendas para:*

- *Cuantificar la magnitud y la distribución geográfica de síntomas de salud y aspectos relativos al confort (ruidos, malos olores y aspectos visuales).*

▫ *Caracterizar los aspectos físicos, químicos y biológicos específicos en diferentes localizaciones.*

▫ *Estudiar asociaciones entre salud y los efectos de confort y factores ambientales, tomando en consideración factores que pudieran confundir tales asociaciones.*

▫ *Identificar áreas que no cuentan con estándares o guías.*

Ejemplo 1.3 *Se desea realizar un estudio de exposición humana a contaminantes, el cual mide exposiciones y niveles de concentración, bajo "tiempo efectivo" de vivir en ciertas condiciones, dirigido a realizar inferencias sobre una población blanco, para:*

▫ *Estimar la exposición individual de la población blanco en diferentes áreas geográficas a contaminantes específicos del aire, durante periodos de tiempo de elevada contaminación.*

▫ *Determinar la contribución de diferentes combustibles para cocinar (leña, petróleo) y/o exposición pasiva de no fumadores en concentraciones humanas, tanto residenciales como públicas.*

▫ *Comparar concentraciones de interiores y exteriores, estimar la intensidad de fuentes interiores y coeficientes de penetración de partículas.*

▫ *Proporcionar datos para la vigilancia ambiental que sirvan de base para estudios de problemas de salud humana o de exposición a algún contaminante ambiental.*

3.1.2 Metodología del estudio

En este apartado se definirán:

▫ El tipo de datos a ser recolectados.

▫ La metodología bajo la cual se realizará la recolección, ya sea mediante entrevistas, diarios u observación directa.

▫ El monitoreo ambiental que será efectuado, para qué sustancias, instrumentos a utilizarse, en qué sitios y en qué momentos.

3.1.3 Definición de la población blanco

Corresponde a aquella población acerca de la cual se quiere la información y que será representada por una muestra.

Los miembros de la población pueden ser personas, viviendas, edificios, vecindarios, ciudades, volúmenes de aire, piezas de material de construcción, etc.

La población blanco puede estar caracterizada por edad, sexo, grupos étnicos, hábitos de fumar, estado de salud, áreas geográficas, estilos de vida, características de la vivienda (tamaño y nivel socioeconómico), mascotas que la habitan y características culturales.

3.2 Diseño muestral

3.2.1 Plan de muestreo que determina la metodología que va a ser utilizada para seleccionar la muestra en una población.

3.2.2 Procedimientos de estimación y análisis de información, es decir, el tipo (técnicas) de análisis que serán aplicadas para obtener las estimaciones de los parámetros poblacionales a partir de una muestra probabilística.

3.3 Diseño de las formas para recolectar la información

Dependiendo de los datos a recolectarse se podrían diseñar cuestionarios para el hogar, trabajo y sobre hábitos. O diseñar cédulas para observación directa de las características de la vivienda o registros del monitoreo del aire, suelo o agua, en interiores o exteriores y en diferentes momentos.

3.4 Técnicas de medición

Se deberán explicitar todas aquellas técnicas para obtener las mediciones en una forma estandarizada y la mejor forma de realizar el control a fin de conseguir medidas precisas.

3.5 Operación de la encuesta

Incluye todas las operaciones de campo:

3.5.1 Prueba piloto

Para corregir posibles errores del cuestionario.

3.5.2 Supervisión y control de calidad

Estas actividades deberán ser programadas y efectuadas a fin de evitar posibles omisiones y evaluar la calidad de los datos obtenidos durante la fase de campo.

3.5.3 Codificación, captura y limpieza de los datos

Para facilitar el manejo y disminuir la posibilidad de errores en la base de datos.

3.5.4 Aspectos administrativos

En este rubro se contempla el flujo que se tiene que seguir de los formatos que han sido llenados, se determinarán las responsabilidades de los participantes y se fijarán los tiempos y movimientos del personal.

3.6 Plan del análisis estadístico

Determina el tipo de procesamiento y análisis estadístico para proveer las inferencias estadísticas acerca de los parámetros y de las distribuciones de exposición o niveles de concentración para una población blanco a través de una muestra.

Los atributos de esas inferencias debieran ser de precisión conocida, de preferencia insesgados, válidos y confiables.

En los estudios ambientales existen múltiples limitaciones para conseguir estos atributos, por ejemplo, el muestreo de varios días durante un período de estudio, no es aleatorio. Los participantes pueden modificar su conducta durante el transcurso del estudio, la tasa de participación entre los sujetos muestrales puede ser baja, la descompostura de los instrumentos de medición puede afectar adversamente la recolección de datos útiles, por lo que todos estos aspectos debieran ser revisados para disminuir los errores durante la investigación.

El plan deberá contemplar la presentación preliminar de las tablas y gráficas, las medidas de resumen esperadas, incluyendo las frecuencias, promedios, medidas de asociación, pruebas

de significancia e intervalos de confianza, así como los modelos para la estimación, validación y predicción.

3.7 Reporte

Señala los capítulos preliminares que se consideren importantes para incluir dentro del reporte.

CAPÍTULO II

REVISIÓN DE CONCEPTOS BÁSICOS DE ESTADÍSTICA

El manejo de los conceptos básicos de la estadística es indispensable para el diseño de encuestas por muestreo, por lo que en este capítulo se revisarán los más usuales y necesarios para desarrollar la teoría del muestreo.

El propósito principal de la estadística consiste en realizar **inferencias** acerca de una población basándose en las mediciones o información obtenida de una muestra. Para poder lograrlo, primero se tiene que describir esta información en una forma clara y concisa por lo que iniciaremos con algunas definiciones y medidas descriptivas de los datos más frecuentemente utilizados en el área de la salud ambiental.

1. TIPOS DE VARIABLES

1.1 Variables Nominales

En toda ciencia incluyendo el área ambiental, podemos encontrar muchos tipos de datos que tienen diferente estructura atendiendo a los valores posibles que pueden tomar. Uno de los más simples son los **datos nominales** en el que las observaciones caen en los grupos o categorías sin ningún orden determinado.

Cuando tratamos con datos nominales que sólo tienen dos distintos resultados se dice que son **dicotómicos** o **binarios** y si poseen más de dos categorías corresponde a **datos politómicos**. Sin embargo debemos tener en cuenta que sólo podremos realizar razones o porcentajes para describir esta serie de datos.

***Ejemplo 2.1** Si la variable fuera sexo, asignamos el valor 1 al masculino y 0 al femenino o viceversa y la secuencia no denota algún significado, por consiguiente el orden y la magnitud de estos números no son importantes. Del mismo modo, se puede asignar 1 al sexo femenino y 2 al masculino y la interpretación no cambiaría. En este caso, los números son utilizados para facilitar el manejo de los datos en análisis complejos de los datos.*

Ejemplo 2.2 *Si quisiéramos estudiar el tipo de agua de consumo humano, se podrían proponer las siguientes categorías:*

1 - agua hervida

2 - agua de pozo

3 - agua entubada

4 - agua purificada

Se puede identificar que corresponde a una variable nominal de tipo politómico, donde el orden de las categorías es arbitrario y los códigos asignados son para facilitar la captura de los datos.

1.2 Variables Ordinales

Si el orden en los grupos o clases llega a ser de importancia, los datos son referidos como ordinales.

Ejemplo 2.3 *Cuando se clasifica a un paciente por el grado de severidad de alguna enfermedad, se asigna el 1 si es ligera, 2 si es moderada, 3 como grave y 4 si es fatal. Se puede observar que existe un orden natural entre los grados y que los valores más pequeños corresponde a una menor severidad; sin embargo no existe relación entre la magnitud de esos números, es decir, si tiene valor 4 no es doblemente severa la enfermedad que para el que tiene valor 2.*

Ejemplo 2.4 *Frecuentemente se manejan preguntas de opinión con respecto a algún programa de control, como sería el de "hoy no circula" y podría preguntarse la opinión en una escala del 1 al 5, correspondiendo al 1 como totalmente en desacuerdo con ese programa, al 2 como en desacuerdo, 3 indeciso, 4 de acuerdo y 5 como totalmente de acuerdo con el programa.*

1.3 Variables discretas

Aquellos datos en que la magnitud y el orden son importantes, los números representan mediciones reales, más que meras etiquetas, es decir son aquellas variables que toman un número finito o numerable de valores sin importar cuáles sean estos. Para estos datos, puesto que adquiere significancia su magnitud, es posible realizar operaciones matemáticas.

Ejemplo 2.5 *Si quisiéramos conocer el número de abortos espontáneos en una población, estudiaríamos a todas las mujeres de 15 a 45 años de edad de un área específica y preguntaríamos por el número de abortos espontáneos sufridos por cada una de ellas, cuyas respuestas variarían del 0, 1, 2, 3, etc.; sin embargo, esta variable no acepta ninguna respuesta intermedia entre esos valores, por decir 0.5, mostrando su carácter discontinuo o discreto.*

1.4 Variables continuas

Los datos que representan mediciones y que pueden tomar cualquier valor de un intervalo se conocen como **datos continuos**. Puesto que en este caso podemos medir la distancia entre dos mediciones, las operaciones matemáticas pueden aplicarse y la única limitante es el grado de precisión con que se efectuó la medición.

Ejemplo 2.6 *La concentración de plomo en cabello en trabajadores de gasolineras, como $20 \mu\text{g g}^{-1}$ en el primer sujeto, 4.1 en el segundo, 3.6 en el tercero, etc.*

Ejemplo 2.7 *El tiempo de exposición por consumo de alimentos en vajillas de loza vidriada, es decir, 8 años en la primera persona, 3.2 años en la segunda, 0 en la tercera, etc.*

2. DEFINICIONES

Ahora bien, para poder caracterizar un conjunto de observaciones es necesario conocer algunas definiciones para obtener algunas medidas descriptivas.

2.1 Población

Es un conjunto de individuos, cuyas características son desconocidas y que a partir de una muestra, que es un subconjunto de individuos extraídos de esta población, permitirá inferir o estimar las características de tal conjunto.

Ejemplo 2.8 *Suponga usted que un investigador está interesado en conocer el nivel medio de cadmio en cabello en una población de 254 358 niños menores de 18 años de un país, con el fin de relacionarlos con algunos factores de riesgo. Si se estudiara a todos y cada uno de ellos se*

podría calcular el parámetro de esta población (es decir, el nivel de cadmio medio en la población menor de 18 años). Si él seleccionara una muestra de tamaño n de estos niños bajo algún criterio probabilístico, se reducirían los costos, el tiempo, etc. y se podría obtener un estimador para inferir el parámetro deseado.

El número de elementos que constituyen a la población es denotada por N y la medición de cada elemento está identificada por una literal, acompañada de un subíndice que va del 1 a N (X_i , Y_i o Z_i).

Ejemplo 2.9 *En un estudio efectuado en 1350 trabajadores de una empresa se obtuvieron los siguientes resultados de plomo en sangre ($\mu\text{g/dl}$):*

<i>Trabajador (x_i)</i>	<i>Plomo en sangre ($\mu\text{g/dl}$)</i>
x_1	3.4
x_2	6.9
x_3	12.8
x_4	10.5
.	.
.	.
.	.
x_{1247}	8.4
x_{1248}	16.3
x_{1249}	19.8
x_{1360}	7.5

donde $N = 1350$

Debido a que contar con el listado de mediciones resulta poco eficiente para darse cuenta de las características esenciales de un grupo de mediciones recurrimos a las medidas numéricas para resumir las características de esta serie de datos.

2.2 Total de la Población

Es la suma de las mediciones de una característica de todos los elementos que constituyen a la población.

$$X = \sum_{i=1}^N X_i \quad (2.1)$$

Ejemplo 2.10 Si el país tiene 254,358 niños menores de 18 años y a través de un censo, es decir, el estudio de todos los niños que constituyen esta población, se mide el nivel de cadmio en cabello ($\mu\text{g/g}$) en cada uno de ellos se podría calcular el total de cadmio que existe en el cabello de esta población, sumando todas las mediciones obtenidas.

$$X = \sum X_i = (0.351 + 2.118 + \dots + 0.932) = 103\,778.064$$

2.3 Media Poblacional

De la variable X es definida como aquel valor promedio o medio de la población y está denotada por:

$$\mu = \sum_{i=1}^N \frac{X_i}{N} \quad (2.2)$$

Ejemplo 2.11 Continuando con el ejemplo anterior, la media poblacional de cadmio en cabello en niños menores de este país sería:

$$\mu = \sum_{i=1}^N \frac{X_i}{N} = \frac{103,778.064}{254,358} = 0.408 \frac{\mu\text{g}}{\text{g}}$$

2.4 Varianza y desviación estándar poblacionales

La varianza y la desviación estándar de la distribución de una característica en una población son de interés porque miden la dispersión de la distribución, es decir, mide la desviación que hay de cada medición con respecto a su media (μ) a mayor dispersión mayores serán estas medidas. La varianza poblacional (σ^2) se expresa en términos cuadrados y está denotada por:

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N} \quad (2.3)$$

La desviación estándar poblacional denotada por (σ) es simplemente la raíz cuadrada de la varianza poblacional y está dada por:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}} \quad (2.4)$$

Ejemplo 2.12 La varianza poblacional del nivel de cadmio en cabello en los menores de 18 años resulta ser:

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N} = \frac{1}{254,358} [(0.351 - 0.408)^2 + \dots + (0.932 - 0.408)^2] = 0.012544$$

Ejemplo 2.13 La desviación estándar poblacional sería:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}} = \sqrt{0.012544} = 0.112$$

2.5 Proporción poblacional

Quando la característica que se está midiendo consiste en la presencia o ausencia de atributo con dos categorías, es decir, dicotómica o binaria, frecuentemente se desea estimar la proporción de elementos en la población que tienen ese atributo. Si el atributo es denotado por X y si $\sum X_i$ es el número total de elementos en la población que tiene el atributo, entonces P_x denota la proporción poblacional de elementos que tienen el atributo y está dada por:

$$P_x = \frac{\sum X_i}{N} \quad (2.5)$$

Del mismo modo, tenemos que (Q_x), la proporción poblacional de elementos que no tienen el atributo, está dada por:

$$Q_x = \frac{N - \sum X_i}{N} = (1 - P_x) \quad (2.6)$$

Por consiguiente:

$$(P_x + Q_x) = 1 \text{ y}$$

$$Q_x = (1 - P_x)$$

Debe notarse que una proporción poblacional es una media poblacional para la situación especial en la que la variable X está dada por:

$$X_i = \begin{cases} 1 \\ 0 \end{cases}$$

donde si el atributo está presente en el elemento i -ésimo le asignamos el número 1 y si está ausente se le asigna el 0.

Cuando la característica que se está considerando es una variable dicotómica, se puede mostrar que σ_p^2 , la varianza de la población definida anteriormente se reduce a la expresión:

$$\sigma_p^2 = P_x(1 - P_x) \quad (2.7)$$

Ejemplo 2.14 Si se deseara obtener la proporción de niños menores de 18 años que conviven con familiares que fuman dentro de su vivienda y los clasificamos como 1 si se cumple esta condición y 0 si no hay familiares que lo realicen, se estaría en posibilidad de calcular tal proporción y su varianza correspondiente, esto es:

$$P_x = \frac{\sum X_i}{N} = \frac{(1 + 1 + 0 + 0 + 0 + \dots + 0 + 1)}{254358} = 0.2832$$

$$\sigma_p^2 = P_x(1 - P_x) = (0.2832)(0.7168) = 0.2030$$

2.6 Parámetros y estimadores

Todas las medidas de resumen hasta aquí descritas se basan en las mediciones que se hacen en la población y se conocen como **parámetros de la población**, asimismo, todas ellas pueden ser obtenidas a partir de una muestra y se denominan **estimadores** ya que con ellas se realizará la estimación de esos parámetros. Estos estimadores serán identificados con letras minúsculas para poderlos diferenciar de las medidas poblacionales, como se muestra en la tabla 2.1.

TABLA NO. 2.1
PARÁMETROS Y ESTIMADORES MÁS USUALES EN
ENCUESTAS DEL ÁREA DE LA SALUD AMBIENTAL

PARÁMETROS	ESTIMADORES
<p>Media Poblacional:</p> $\mu = \frac{\sum_{i=1}^N X_i}{N}$ <p>Varianza Poblacional:</p> $\sigma^2 = \frac{\sum (X_i - \mu)^2}{N}$ <p>Desviación estándar poblacional:</p> $\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$ <p>Donde: X_i es el valor de cada medición desde que $i=1$ hasta N N es el número de elementos que constituyen la población bajo estudio</p>	<p>Media Muestral:</p> $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ <p>Varianza Muestral:</p> $S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$ <p>Desviación estándar muestral:</p> $S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$ <p>Donde: x_i es el valor de cada medición desde que $i=1$ hasta n n es el número de elementos que constituyen la muestra estudiada</p>
<p>Proporción poblacional de elementos con el atributo:</p> $P_x = \frac{\sum X_i}{N}$	<p>Proporción muestral de elementos con el atributo:</p> $p_x = \frac{\sum x_i}{n}$
<p>Varianza de la proporción poblacional:</p> $\sigma_p^2 = P_x(1 - P_x)$ <p>Donde: $X_i = \begin{cases} 1 & \text{tiene la característica} \\ 0 & \text{no la tiene} \end{cases}$ N es el número de elementos en la población</p>	<p>Varianza de la proporción muestral:</p> $\sigma_p^2 = p_x(1 - p_x)$ <p>Donde: $x_j = \begin{cases} 1 & \text{tiene la característica} \\ 0 & \text{no la tiene} \end{cases}$ n es el número de elementos en la muestra.</p>

2.7 Elemento

Es un sujeto u objeto al cual se le toman las mediciones.

Ejemplo 2.15 Elementos pueden ser una persona, una vivienda, un animal o una cosa.

2.8 Unidades de muestreo

Son colecciones no traslapadas de elementos de la población que la cubren completamente.

Ejemplo 2.16 Una unidad de muestreo podría ser un niño o si se quiere ser más eficiente podría ser el hogar, el cual incluye una colección de elementos. Si los hogares son las unidades de muestreo, deberá considerarse que ningún niño sea muestreado dos veces y que todos los niños tengan la misma probabilidad de ser elegidos.

Si la unidad de muestreo es el niño, es decir si sólo contiene a un elemento de la población, se dice que son idénticos.

2.9 Marco Muestral

Es una lista de unidades de muestreo. Algunos esquemas de muestreo pueden requerir múltiples marcos, dependiendo de las fases en que se realice este esquema. Por consiguiente, la muestra es un subconjunto de unidades muestrales seleccionadas bajo algún criterio probabilístico de uno o varios marcos muestrales, de modo que la teoría del muestreo sea aplicable.

2.10 Distribución normal

Es otro concepto importante y se refiere a una curva teórica con una forma semejante a una campana simétrica y que en la ecuación que la determina matemáticamente, la media y la desviación estándar constituyen los dos valores que la especifican. Esta densidad tiene una área bajo la curva igual a 1, por lo que en términos de ellos se pueden hacer varias afirmaciones con respecto al área que queda incluida entre dos valores cualesquiera.

Se puede demostrar que entre $\mu - \sigma$ y $\mu + \sigma$ encontramos al 68.26% del área, que entre $\mu - 1.96\sigma$ y $\mu + 1.96\sigma$ encontramos al 95% del área y que entre $\mu - 3\sigma$ y $\mu + 3\sigma$

encontramos al 99.73% del área bajo la curva normal. En general conociendo el valor de la media y la desviación estándar se puede determinar la proporción del área que se encuentra entre dos valores cualesquiera.

Ejemplo 2.17 Si la distribución de la concentración de cadmio en cabello en los niños menores de 18 años de este país siguiera una distribución normal, podríamos decir que entre la media mas y menos 1.96 desviaciones estándar encontramos al 95% de las mediciones de cadmio de esta población,

$$\mu - 1.96\sigma \leq X \leq \mu + 1.96\sigma \quad (2.8)$$

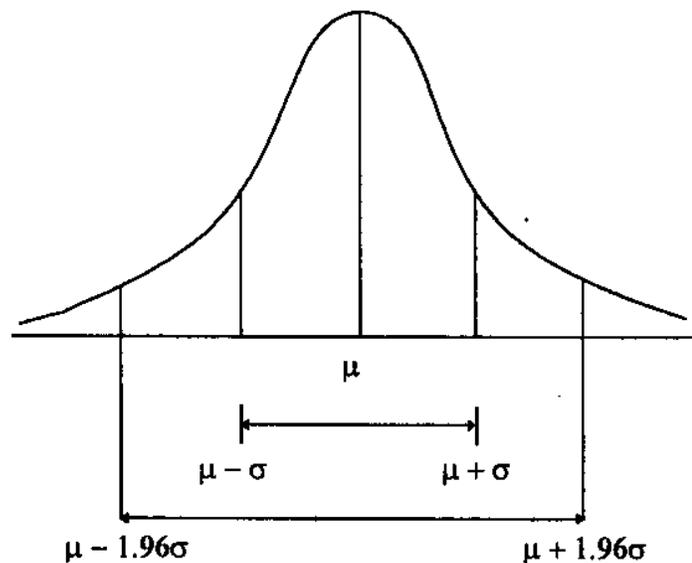
y sustituyendo,

$$0.408 - 1.96(0.112) \leq X \leq 0.408 + 1.96(0.112)$$

$$0.1884 \leq X \leq 0.6275$$

es decir, el 95% de las concentraciones de cadmio se encontrarían entre 0.1884 y 0.6275 $\mu\text{g/g}$.

FIG. 2.1
ÁREA BAJO LA CURVA NORMAL



CAPÍTULO III

DISTRIBUCIÓN MUESTRAL DE MEDIAS

1. INTRODUCCIÓN

Suponga que una comunidad hipotética (es decir, la población o universo) consiste de 6 elementos ($N = 6$). Esta población presenta los valores de un contaminante específico detectado en microgramos por 100 mililitros de sangre en la tabla 3.1.

TABLA NO. 3.1
MEDICIONES DE UN CONTAMINANTE ESPECÍFICO ($\mu\text{g}/100\text{ml}$ DE SANGRE)
DE UNA POBLACIÓN HIPOTÉTICA EN ESTUDIO DE TAMAÑO $N=6$

Elemento	Contaminante $\mu\text{g}/100\text{ml}$ de sangre
X_1	162
X_2	166
X_3	168
X_4	170
X_5	170
X_6	172

La **media poblacional**, es decir el parámetro que se desea estimar, es de 168 microgramos por 100 mililitros de sangre

$$\mu = \sum_{i=1}^N \frac{X_i}{N} = 168 \text{ microgramos por } 100 \text{ ml de sangre } (3.1)$$

Ahora, suponga que deseamos tomar una muestra de tamaño 2, con el propósito de estimar la media poblacional de este contaminante a partir de una muestra de dos elementos ($n = 2$).

Cualquiera que sea el procedimiento de selección aleatorio que podamos usar, sólo hay 15 posibles muestras o combinaciones de seis elementos tomados de dos en dos y se presentan en la tabla 3.2.

TABLA NO. 3.2
MUESTRAS POSIBLES DE TAMAÑO n=2 DE UNA POBLACIÓN TAMAÑO N=6,
CON SUS CORRESPONDIENTES MEDICIONES Y MEDIA MUESTRAL

Muestra No.	Miembros de la Muestra		Elementos de la Muestra (n= 2)		Media (\bar{x}) Muestral
1	X_1	X_2	162	166	164
2	X_1	X_3	162	168	165
3	X_1	X_4	162	170	166
4	X_1	X_5	162	170	166
5	X_1	X_6	162	172	167
6	X_2	X_3	166	168	167
7	X_2	X_4	166	170	168
8	X_2	X_5	166	170	168
9	X_2	X_6	166	172	169
10	X_3	X_4	168	170	169
11	X_3	X_5	168	170	169
12	X_3	X_6	168	172	170
13	X_4	X_5	170	170	170
14	X_4	X_6	170	172	171
15	X_5	X_6	170	172	171

Si observamos las 15 posibles muestras, algunas de ellas muestran valores bastante extremos. Si en nuestro proceso de selección se eligiera la primera muestra, la cual incluye los valores 162 y 166, daría como resultado una media muestral de 164:

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} = \frac{162 + 166}{2} = 164$$

Si por otro lado, nuestro proceso elige la muestra número 15, la media muestral sería 171:

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} = \frac{170 + 172}{2} = 171$$

Todas las demás muestras de tamaño 2 tendrían una media comprendida entre estas dos medias muestrales lo cual señala el hecho de que siempre en un muestreo aleatorio, se podría obtener una muestra extrema, es decir, el muestreo aleatorio no es garantía de representatividad de la población.

También podemos observar que únicamente dos medias muestrales, la 7 y la 8, coinciden con la media poblacional, por lo que no es suficiente obtener un valor muestral para extrapolar el parámetro buscado.

2. LA DISTRIBUCIÓN MUESTRAL DE UN ESTIMADOR

Si las medias muestrales obtenidas son ordenadas en cuanto a su magnitud podremos construir una distribución de frecuencias de medias muestrales, como se presenta en la tabla 3.3.

TABLA NO. 3.3
DISTRIBUCIÓN DE LAS MEDIAS MUESTRALES DE TODAS LAS POSIBLES
MUESTRAS DE TAMAÑO 2 DE LA POBLACIÓN EN ESTUDIO.

Valor de la media muestral	Frecuencia (f _i)
164	1
165	1
166	2
167	2
168	2
169	3
170	2
171	2
Total de medias (Σf_i)	15

Esta es conocida como la **distribución de medias muestrales**. Podemos decir que la media muestral es ahora una variable aleatoria, pues tiene una distribución muestral y por consiguiente varía de muestra en muestra como resultado de un proceso de selección muestral aleatorio.

3. PROPIEDADES PRINCIPALES DE LOS ESTIMADORES

Primero, se revisarán las características que posee un estimador, mediante su distribución muestral para finalmente estimar la media poblacional, lo cual se hará usando la media muestral como estimador.

3.1 Estimador insesgado

La primera propiedad que cualquier estimador debería poseer es que proporcionara sistemáticamente información correcta acerca del valor verdadero del parámetro. En otras palabras, es que la media de todas las medias muestrales de tamaño constante, debiera ser igual al valor del parámetro. Un estimador que cumple esta condición es denominado **estimador insesgado**.

En nuestro ejemplo, podemos confirmar esta condición, obteniendo la media de la distribución muestral de medias:

$$\mu_{\bar{x}} = \frac{\sum \bar{x}}{\sum f_i} = \frac{2520}{15} = 168 = \mu \quad (3.2)$$

Donde $\sum f_i$ es el número de muestras posibles de tamaño $n = 2$ de esta población de tamaño $N = 6$.

Si en promedio, la desviación de la media de medias muestrales del parámetro de la población que está siendo estimado es igual a cero, nos da una connotación de ausencia de sesgo en la estimación.

3.2 Error estándar

Sin embargo, el hecho de que el estimador sea insesgado no es suficiente, pues además se debería asegurar que el estimador seleccionado tenga una distribución muestral concentrada tan cerca como sea posible a su media (la cual, como vimos es igual al valor del parámetro).

La medida de variabilidad o dispersión más usado es la **desviación estándar de la distribución de las medias muestrales** que es denominada como **error estándar** o si se quiere en términos de sus cuadrados como la **varianza de su distribución** y cuyos cálculos se presentan en la tabla 3.4.

TABLA 3.4
CÁLCULO DEL ERROR ESTÁNDAR A PARTIR DE LA DISTRIBUCIÓN MUESTRAL DE MEDIAS.

Medias muestrales	Frecuencia f_i	$(\bar{x} - \mu_{\bar{x}})$	$(\bar{x} - \mu_{\bar{x}})^2$	$f_i(\bar{x} - \mu_{\bar{x}})^2$
164	1	-4	16	16
165	1	-3	9	9
166	2	-2	4	8
167	2	-1	1	2
168	2	0	0	0
169	3	1	1	3
170	2	2	4	8
171	2	3	9	18
Total	$\Sigma f_i = 15$			$64 = \Sigma f_i(\bar{x} - \mu_{\bar{x}})^2$

Por consiguiente, la **varianza del error estándar** es:

$$\sigma_{\bar{x}}^2 = \frac{\Sigma f_i(\bar{x} - \mu_{\bar{x}})^2}{\Sigma f_i} = \frac{64}{15} = 4.27 \quad (3.3)$$

y el **valor del error estándar** es:

$$\sigma_{\bar{x}} = \sqrt{\frac{\Sigma f_i(\bar{x} - \mu_{\bar{x}})^2}{\Sigma f_i}} = \sqrt{4.27} = 2.07 \quad (3.4)$$

Antes de buscar una fórmula que nos aproxime al valor del error estándar sin necesidad de recurrir directamente a la distribución de medias muestrales, tenemos que encontrar un estimador igualmente **insesgado** de la **varianza poblacional** y que tenga la **propiedad de tener la menor variabilidad posible**.

Si obtenemos las varianzas de cada una de las muestras posibles de tamaño dos encontramos las varianzas muestrales que se muestran en la tabla 3.5.

TABLA NO. 3.5
MUESTRAS POSIBLES DE TAMAÑO n=2 DE LA POBLACIÓN EN ESTUDIO,
CON SUS CORRESPONDIENTES MEDICIONES Y VARIANZA MUESTRAL.

Muestra No.	Miembros de la Muestra		Elementos de la Muestra (n=2)		Varianza Muestral ¹
1	X ₁	X ₂	162	168	4
2	X ₁	X ₃	162	168	9
3	X ₁	X ₄	162	170	16
4	X ₁	X ₅	162	170	16
5	X ₁	X ₆	162	172	25
6	X ₂	X ₃	166	168	1
7	X ₂	X ₄	166	170	4
8	X ₂	X ₅	166	170	4
9	X ₂	X ₆	166	172	9
10	X ₃	X ₄	168	170	1
11	X ₃	X ₅	168	170	1
12	X ₃	X ₆	168	172	4
13	X ₄	X ₅	170	170	0
14	X ₄	X ₆	170	172	1
15	X ₅	X ₆	170	172	1

¹ Ver ecuación de la tabla 2.1

Se presenta el cálculo de la primera varianza muestral:

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n} = \frac{(162 - 164)^2 + (166 - 164)^2}{2} = 4$$

Si ahora ordenamos en cuanto a su magnitud todas las varianzas muestrales para todas las posibles muestras de tamaño 2, encontramos la siguiente distribución de varianzas muestrales (tabla 3.6)

TABLA NO. 3.6
DISTRIBUCIÓN DE LAS VARIANZAS MUESTRALES DE TODAS LAS
POSIBLES MUESTRAS DE TAMAÑO $n=2$ DE LA POBLACIÓN EN ESTUDIO.

Valor de la varianza muestral	Frecuencia (f)
0	1
1	5
4	4
9	2
16	2
25	1
Total de varianzas	$\Sigma f_i = 15$

Podemos notar que 12 de las 15 varianzas son menores que la varianza poblacional:

$$\sigma^2 = \frac{\Sigma(x_i - \mu)^2}{N} = \frac{64}{8} = \frac{32}{3} \quad (3.6)$$

y la media de las varianzas muestrales (μ_s^2) resulta ser:

$$\mu_s^2 = \frac{\Sigma \sigma_i}{\Sigma f_i} = \frac{0 + 1 + 1 + \dots + 16 + 16 + 25}{15} = \frac{96}{15} = \frac{32}{5} \quad (3.6)$$

Por consiguiente, la varianza muestral es un **estimador sesgado** de la varianza poblacional, ya que en promedio la varianza muestral resulta diferente de la varianza poblacional.

No obstante existen métodos para corregir esta falla, alterando las varianzas, para restaurar y conseguir algún estimador insesgado que nos hable de la variabilidad poblacional. Esta modificación es muy simple y se consigue utilizando las definiciones de las varianzas tanto la poblacional como las muestrales, excepto que dividimos las desviaciones cuadráticas entre el valor correspondiente menos 1, esto es, la **varianza modificada de la población** ($\hat{\sigma}^2$) resulta ser:

$$\hat{\sigma}^2 = \frac{\sum(X_i - \mu)^2}{N - 1} = \frac{64}{5} \quad (3.7)$$

La varianza modificada de la muestra (σ^2)

$$\hat{s}^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1} \quad (3.8)$$

De este modo, la distribución de varianzas modificadas de las muestras se presenta en la tabla 3.7.

TABLA NO. 3.7
DISTRIBUCIÓN DE VARIANZAS MUESTRALES MODIFICADAS
DE MUESTRAS DE TAMAÑO $n=2$ DE LA POBLACIÓN EN ESTUDIO.

Valor de la varianza modificada muestral ²	Frecuencia (f_i)
0	1
2	5
8	4
18	2
32	2
50	1
Total de varianzas modificadas	$\sum f_i = 15$

² Ver ecuación 3.8

Donde la media de estas varianzas modificadas es de $\mu_{\hat{s}^2} = 64/5$ y es igual a la varianza modificada de la población, por lo que la varianza modificada de una muestra también resulta ser un estimador insesgado de la varianza poblacional modificada.

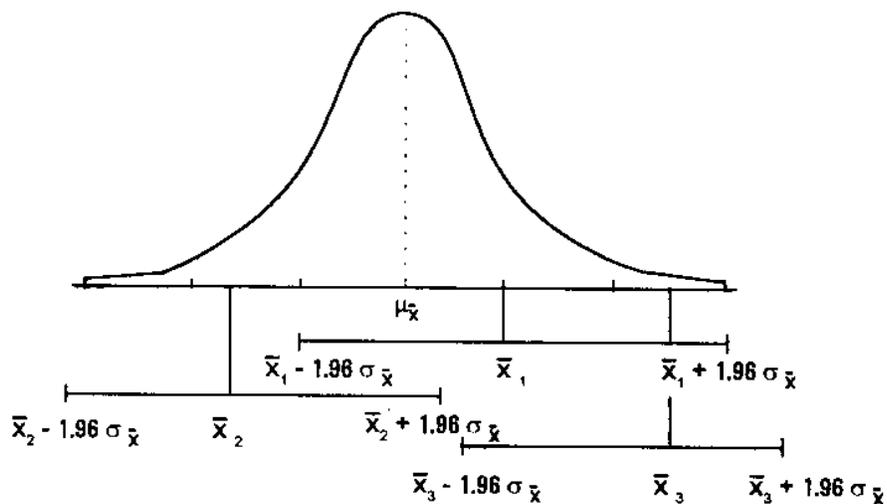
4. TEOREMA DE LÍMITE CENTRAL Y LA DISTRIBUCIÓN NORMAL

Este teorema nos dice que la distribución de medias muestrales se aproxima a la distribución normal, cuando el tamaño de la muestra es lo suficientemente grande. Mientras más irregular sea la distribución poblacional de donde procede la muestra el tamaño muestral necesario será mayor para que la distribución de medias muestrales se aproxime a la normal. Un tamaño muestral mínimo de 30 será requerido para asegurar que se adecúa a la aproximación normal.

Estamos ahora en capacidad para aplicar métodos elaborados para realizar inferencias sobre la población a partir de una muestra, es decir, ahora tenemos la posibilidad de utilizar las propiedades de la distribución normal que junto con la media y la varianza modificada de la muestra nos permite construir un intervalo que acota el valor del parámetro buscado con cierto nivel de confianza.

De este modo podemos expresar que un porcentaje predefinido de las muestras generarán un estimador tal, que el correspondiente intervalo de confianza construido contendrá el parámetro; así el valor del estimador (en este caso la media muestral) más menos un múltiplo de su error estándar tiene cierta probabilidad de "cubrir" al verdadero parámetro, por ejemplo podríamos decir que hay 95% de confianza de que la media muestral más y menos 1.96 veces su error estándar acotará a la media poblacional. Este es conocido como un intervalo del 95% de confianza para la media de la población.

FIGURA 3.1
INTERVALO DEL 95% DE CONFIANZA PARA LA MEDIA POBLACIONAL



Para calcular el error estándar usando un muestreo aleatorio simple fue necesario obtener la distribución muestral de medias de todas las posibles muestras de tamaño n , y obtener su desviación estándar lo que resulta contrario a lo que pretendemos al usar un muestreo. Sin embargo basados en estas construcciones se han derivado fórmulas que nos aproximan a la varianza de \bar{X} y que dependen de la varianza de la muestra así como de los tamaños de la muestra y de la población y los cuales son disponibles.

$$\hat{\sigma}_x^2 = \frac{\hat{S}^2}{n} \frac{N-n}{N} \quad (3.9)$$

Donde el coeficiente $(N-n)/N$ es conocido como el factor de corrección para poblaciones finitas, y es igual a $1-(n/N)$, que corresponde a 1, menos la fracción de muestreo, con respecto a la población. Este factor resulta importante cuando el tamaño de la muestra es grande con respecto a la población, sin embargo, cuando esto no sucede, la fracción de muestreo tiende a cero y el factor de corrección es aproximadamente 1, por lo que la varianza estimada del estimador \bar{x} se reduce prácticamente a:

$$\hat{\sigma}_x^2 = \frac{\hat{S}^2}{n} \quad (3.10)$$

Aunque el procedimiento del intervalo de confianza depende esencialmente de que el tamaño muestral sea lo suficientemente grande para que el efecto del teorema del límite central se sostenga, es interesante observar si nuestro ejemplo da resultados más o menos congruentes (recuerde que el ejemplo tiene una $n = 2$), los cuales se presentan en la tabla 3.8.

TABLA NO. 3.8
ESTIMADORES Y ERRORES ESTÁNDAR PARA LA CONSTRUCCIÓN DE
INTERVALOS DE CONFIANZA DE MUESTRAS DE TAMAÑO $n=2$ DE LA
POBLACIÓN EN ESTUDIO.

Muestra No.	Elementos de la Muestra (n=2)		Media Muestral \bar{x}	Varianza modificada de la muestra ³	Error estándar (95%)	EI IC ⁴ cubre el parámetro
1	162	166	164	8	1.63	no
2	162	168	165	18	2.45	si
3	162	170	166	32	3.27	si
4	162	170	166	32	3.27	si
5	162	172	167	50	4.08	si
6	166	168	167	2	0.82	si
7	166	170	168	8	1.63	si
8	166	170	168	8	1.63	si
9	166	172	169	18	2.45	si
10	168	170	169	2	0.82	si
11	168	170	169	2	0.82	si
12	168	172	170	8	1.63	si
13	170	170	170	0	0.00	no
14	170	172	171	2	0.82	no
15	170	172	171	2	0.82	no

³ Ver ecuación 3.8

⁴ IC= intervalo de confianza del 95%

La última columna presenta para cada muestra si la media muestral más menos 1.96 errores estándar si cubre o no al parámetro, el cual sabemos es de 168. Podemos ver que 11 de los 15 contienen (73.3%) al verdadero valor del parámetro. Considerando que el tamaño muestral es tan pequeño los resultados no son tan acertados, sin embargo si el tamaño muestral aumenta a 4 encontramos los valores que se muestran en la tabla 3.9.

TABLA NO. 3.9
ESTIMADORES Y ERRORES ESTÁNDAR PARA LA CONSTRUCCIÓN
DE INTERVALOS DE CONFIANZA A PARTIR DE MUESTRAS
DE TAMAÑO $n=4$ DE LA POBLACIÓN EN ESTUDIO.

Muestra No.	Elementos de la (n=4)				Media Muestral \bar{x}	Varianza modificada de la muestra ⁵	Error estándar ⁶	El IC ⁷ cubre el parámetro
1	162	168	168	170	168.5	11.67	0.99	si
2	162	166	168	170	168.5	11.67	0.99	si
3	162	166	168	172	167.0	17.33	1.20	si
4	162	166	170	170	167.0	14.67	1.11	si
5	162	166	170	172	167.5	19.67	1.28	si
6	162	166	170	172	167.5	19.67	1.28	si
7	162	168	170	170	167.5	14.33	1.09	si
8	162	168	170	172	168.0	18.67	0.64	si
9	162	168	170	172	168.0	18.67	0.64	si
10	162	170	170	172	168.5	19.67	1.64	si
11	166	168	170	170	168.5	03.67	0.55	si
12	166	168	170	172	169.0	06.67	0.75	si
13	166	168	170	172	169.0	06.67	0.75	si
14	166	170	170	172	169.5	06.33	0.13	no
15	168	170	170	172	170.0	02.67	0.47	no

⁵ Ver ecuación 3.8

⁶ Ver ecuación 3.9

⁷ IC = intervalo de confianza del 95%

Estos resultados son considerablemente más cercanos a los esperados que los anteriores ya que 13 de los 15 intervalos (86.67%) incluyen al verdadero valor del parámetro. Así encontramos que el teorema del límite central resulta ser aproximado para muestras pequeñas y completamente adecuado para muestras de 30 o más elementos.

Si la población original no es normal, el Teorema Central del Límite siempre será una aproximación no importa el tamaño de la muestra, pero entre más grande sea, la aproximación será mejor.

Luego entonces, el fin principal del uso del muestreo probabilístico es que mediante un estimador y su correspondiente error estándar multiplicado por algún valor de z para controlar el nivel de confiabilidad nos permite construir un intervalo que acote al valor poblacional. Y que mientras mayor sea el tamaño muestral estaremos en capacidad de aumentar la precisión de este intervalo.

Así, si obtuviéramos una muestra al azar y correspondiera a la número 10 de la tabla, el intervalo al nivel del 95% de confianza sería el siguiente:

$$\bar{x} - Z \sigma_{\bar{x}} \leq \mu \leq \bar{x} + Z \sigma_{\bar{x}} \quad (3.11)$$

$$168.5 - 1.96 (1.64) \leq \mu \leq 168.5 + 1.96 (1.64)$$

$$165.29 \leq \mu \leq 171.71$$

Por tanto, entre 165.29 y 171.71 unidades encontraremos a la media poblacional con una confianza del 95%, es decir, de cada 100 intervalos que se construyesen de este modo, en 5 de ellos vamos a cometer un error, pues el intervalo de confianza no acotará el parámetro buscado, sin embargo, los otros 95 sí incluirán al verdadero parámetro poblacional.

Hasta aquí, nos hemos basado en datos continuos para presentar la distribución muestral de medias. De manera equivalente, la distribución muestral de proporciones, es donde en caso de variables dicotómicas, se utiliza la proporción muestral en lugar de la media y el valor de $p(1-p)$ en sustitución de la varianza muestral, para poder estimar, bajo un cierto nivel, la proporción poblacional.

Ejemplo 3.1 *Se desea estimar la proporción de personas que consumen cotidianamente sus alimentos en vajillas de loza vidriada, en cierto país que tiene una población de 2,825,345 personas, para lo cual se selecciona una muestra de 250 personas a las cuales se les pregunta sobre esta cuestión y se obtienen los siguientes resultados:*

$x = 72$ personas consumen sus alimentos habitualmente en vajilla de loza vidriada.

$(n-x) = (250-72) = 178$ personas no los consumen habitualmente en vajilla de loza vidriada.

La proporción de personas que usan loza (p_x) es:

$$p_x = \frac{\sum x_i}{n} = \frac{72}{250} = 0.288 \quad (3.12)$$

La proporción que no la usan (q_x) es:

$$q_x = \frac{n - \sum x_i}{n} = \frac{250 - 72}{250} = 0.712 = 1 - 0.288 \quad (3.13)$$

La varianza estimada del estimador resulta ser:

$$\hat{\sigma}_{p_x}^2 = \frac{p_x q_x}{n} \frac{N - n}{N} = \frac{(0.288)(0.712)}{250} \frac{2825345 - 250}{2825345} = 0.00082 \quad (3.14)$$

Y el error estándar es:

$$\hat{\sigma}_p = \sqrt{0.00082} = 0.028 \quad (3.15)$$

Si construimos un intervalo del 95% de confianza tenemos que:

$$p_x - 1.96 \sigma_p \leq P \leq p_x + 1.96 \sigma_p \quad (3.16)$$

$$0.288 - 1.96(0.0286) \leq P \leq 0.288 + 1.96(0.0286)$$

$$0.2319 \leq P \leq 0.3441$$

Expresando estos resultados en términos de porcentajes podemos decir que la proporción que consume sus alimentos cotidianamente en vajillas de loza vidriada de esta población se encuentra entre el 23.19% y 34.41%, con un nivel de confianza del 95%.

CAPÍTULO IV

TIPOS DE MUESTREO

En este capítulo revisaremos los siguientes tópicos:

- Definición de muestreo.
- Tipos de muestreo.
- Tipos de muestreo probabilístico.
- Tipos de muestreo no probabilístico.

Estrategia: Mediante un ejemplo hipotético, se hará la conceptualización general de los tipos de muestreo probabilístico.

Se usarán ejemplos simples a fin de que las relaciones necesarias se puedan describir e interpretar fácilmente. Si bien son ejemplos algo artificiales ya que en la práctica pocas veces tenemos distribuciones tan ordenadas, la extensión de los métodos a situaciones más prácticas, de mayor tamaño y complejidad, resultará clara.

Ejercicios: Tendientes a fijar y aclarar los conceptos. Se harán al final del capítulo. Se hará un taller de manejo de la tabla de números aleatorios.

1. DEFINICIÓN DE MUESTREO

Es la operación para tomar una muestra del Universo.

El objetivo del muestreo es contar con los datos necesarios para estimar parámetros en la población, es decir poder hacer una inferencia estadística con la mayor confiabilidad posible.

Una vez definida por el investigador el tamaño de la muestra (n), mediante técnicas que veremos más adelante (CAPÍTULO VI), él mismo se enfrenta al problema de cómo obtener esa muestra de la forma que sea más útil para los fines que persigue y por supuesto, que le rinda la mejor relación costo-beneficio. Esas diferentes formas o maneras de seleccionar la muestra son conocidas como tipos y en un sentido amplio son sólo dos.

2. TIPOS DE MUESTREO

- Muestreo probabilístico.
- Muestreo no probabilístico.

2.1 Muestreo probabilístico

2.1.1 Probabilidad. Es la medida del grado de ocurrencia de un suceso. Siempre es un número entre cero y uno y se define como el cociente entre el número de resultados favorables al suceso y el número total de resultados posibles. Por ejemplo, una moneda tiene dos lados, el número de posibles resultados al lanzarla al aire es 2. La probabilidad de que el resultado sea águila es $1/2 = 0.5$.

2.1.2 Definición. Es el tipo de muestreo en el cual cada unidad del universo tiene una probabilidad conocida y diferente de cero, de ser incluida en la muestra. Recordemos que la muestra, además de dar las estimaciones de las características de la población, debe proporcionar medidas de la precisión de tales estimaciones. Esas medidas de la precisión se podrán usar para determinar el error máximo que razonablemente puede esperarse en esas estimaciones si el procedimiento se cumple en la forma especificada y si la muestra tiene el tamaño suficiente. No se puede estimar la precisión a menos que la selección (tipo de muestreo) se efectúe de tal modo que se conozca la probabilidad de selección de cada unidad y se utilice algún tipo de muestra probabilística.

Hay varios tipos de muestreo probabilístico y a ellos nos referiremos con más detalle en este capítulo. El muestreo probabilístico permite estimar parámetros para el universo, es decir inferir.

2.2 Muestreo no probabilístico

La probabilidad de las unidades del universo para ser incluidas en la muestra se desconoce o, existen de antemano unidades con probabilidad cero. Este tipo de muestreo, por supuesto, no permite la aplicación de las técnicas de la inferencia estadística para sacar conclusiones acerca del Universo. Es necesario, sin embargo, destacar la existencia de situaciones, muy especialmente en el área ambiental, en las cuales lo único razonable es optar por este tipo de elección en lugar de un método probabilístico, por ello dedicaremos una parte de este capítulo al análisis de este apartado.

3. TIPOS DE MUESTREO PROBABILÍSTICO

- Muestreo Aleatorio Simple
- Muestreo Sistemático
- Muestreo Estratificado
- Muestreo de Conglomerados
- Diseños Complejos
(Mixtos, Polietápicos, etc.).

3.1 Muestreo aleatorio simple (MAS)

Se da este nombre al tipo más sencillo de muestreo probabilístico.

3.1.1 Definición. Procedimiento de selección de una muestra mediante el cual cada subconjunto de tamaño n de la población tiene la misma probabilidad de ser seleccionado, es decir ninguna combinación de n elementos tiene más probabilidad de ser seleccionada que otra, siendo esta probabilidad igual a:

$$\text{prob} = \frac{1}{\binom{N}{n}} = \frac{1}{\frac{n!(N-n)!}{N!}}$$

Ejemplo 4.1 *Vamos a suponer la existencia de una hipotética población agrícola a la cual llamaremos Pavarandocito. Es esta una población en la cual por sus características agrícolas y epidemiológicas, dado que es una región palúdica y con antecedentes de haber presentado casos de fiebre amarilla en el pasado, se han usado ampliamente diversos tipos de compuestos químicos. Recientes hallazgos de investigación hacen sospechar la relación de uno de estos compuestos con una patología específica en la población escolar, razón por la cual se desea estudiar una muestra de tamaño n , de los niños entre 6 y 12 años de edad con el fin de aplicarles una entrevista, realizarles un examen físico y tomarles una muestra de sangre para la determinación de un biomarcador, específicamente relacionado con el compuesto químico en cuestión (nótese que igual pudiera tratarse de una población con alta contaminación ambiental por partículas o un pueblo minero etc., es decir, el ejemplo es válido para muchas otras circunstancias de interés ambiental).*

La población de escolares en esta población es de 10,000 niños (N) y el tamaño de muestra a seleccionar se ha establecido en 400 niños (n) (Según el procedimiento explicado en el capítulo VI). El pueblo tiene cuatro grandes zonas determinadas por su ubicación geográfica, sus características socioeconómicas y, por supuesto, diferentes características en cuanto al uso de los compuestos químicos en cuestión; en cada una de estas zonas hay 50 escuelas primarias con un total de cincuenta niños cada una (ver los esquemas anexos de la población de Pavarandocito 4.1 y 4.1.1 páginas 45 y 52 respectivamente).

De la teoría de probabilidades nosotros sabemos que el número de subconjuntos diferentes de n elementos, de un conjunto N es

$$C_n^N = \frac{N!}{(N-n)! n!} \quad (4.1)$$

Lo cual, en el caso de Pavarandocito, nos arroja un número que no cabría en esta página y el cual no nos interesa, ya que lo que nos importa es que cualquiera de esos subconjuntos (muestras) de elementos tenga la misma probabilidad de ser seleccionado.

3.1.2 Procedimiento

Es necesario contar con un marco muestral muy completo el cual tenga un listado de 1 a 10,000, de cada uno de los niños que asisten a las escuelas primarias de dicha población. Hay varias opciones:

3.1.2.1 Primero hacer una ficha con cada nombre, meter las fichas en una bolsa y, utilizando el método de la lotería, sacar las 400 fichas necesarias para completar nuestra n , en un muestreo sin reemplazo, en este caso concreto.

3.1.2.2 Utilizar una tabla de números aleatorios y mediante un procedimiento sin reemplazo, completar los 400 elementos necesarios a nuestros fines. Al final del capítulo se presenta el instructivo No. 4.1 "Uso de la tabla de dígitos aleatorios" con un ejemplo basado en la tablas 4.1 "Tabla de dígitos aleatorios" y tabla 4.2 "Valores de plomo en sangre en 325 mujeres puérperas de la Ciudad de México".

3.1.2.3 Tener este marco muestral como una base de datos en una computadora y solicitarle a la misma una muestra del tamaño deseado, utilizando el paquete apropiado (STATA 3.1, 1992 por ejemplo). Esto es igual al punto número dos ya que la computadora hace este trabajo utilizando un procedimiento equivalente al uso de la tabla de números aleatorios.

3.1.3 Ventajas

- Todos los elementos tienen igual probabilidad.
- Los cálculos matemáticos son sencillos.

3.1.4 Desventajas

- Se requiere un marco muestral completo y detallado.
- La muestra puede quedar muy dispersa.
- Puede ser necesario visitar una localidad por un solo elemento.

3.1.5 Consideraciones especiales

El MAS no requiere ningún esfuerzo especial para obligar a la muestra a ser representativa de la población. La tendencia a ese carácter es inherente al proceso mismo, aunque no lo garantiza.

En el MAS, el único camino para reducir el error de muestreo es aumentar el tamaño de la muestra. Sin embargo, antes de elegir la muestra, si se tiene algún conocimiento acerca de la población, se puede utilizar esa información en la estratificación y así reducir el error de muestreo. El Muestreo Aleatorio Estratificado (MAE) que veremos más adelante, es una alternativa para el MAS cuando tenemos la posibilidad de dividir la población en grupos.

3.2 Muestreo Sistemático (MS)

3.2.1 Definición

Es aquel tipo de muestreo que nos permite obtener la muestra deseada al seleccionar aleatoriamente un elemento de los primeros k elementos en el marco muestral y después cada k -ésimo elemento hasta completar n .

Ejemplo 4.2 *En el caso que nos ocupa en la población de Pavarandocito, nuestra fracción de muestreo es*

$$n/N = 400/10,000 = 4/100 = 1/25 \quad (4.2)$$

Es decir para completar nuestro tamaño muestral en una forma sistemática, nosotros debemos tomar uno de cada 25 escolares en la ciudad. Este número 25, inverso de la fracción de muestreo, es llamado k (intervalo de muestreo).

3.2.2 Procedimiento

La muestra se extrae tomando un número al azar entre 1 y k (con la tabla de números aleatorios) el cual se conoce como **arranque aleatorio** y, que para efectos de este ejercicio, diremos que fue el número 11, mismo al que se le va sumando la k , hasta obtener la muestra deseada.

Ejemplo 4.3 *Veamos de nuestro hipotético marco muestral numerado del 1 al 10,000; ya sabemos que debemos tomar los niños que aparezcan en la lista con los siguientes números: 11 (arranque aleatorio), 36 (arranque aleatorio más k , es decir, $(11 + 25)$), 61 ($36 + 25$), 86, 111, 136 y así sucesivamente hasta completar la n de tamaño 400 al llegar al niño marcado en el marco muestral con el número 9,986, suponiendo que todo se hace idealmente.*

3.2.3 Ventajas

- Sólo es necesario usar la tabla de números aleatorios una sola vez.
- La probabilidad de cada elemento es conocida, diferente a la del MAS, pues aquí la probabilidad es $1/k$.
- Garantiza la distribución de la muestra.
- Los cálculos son iguales a los del MAS.

3.2.4 Desventajas

- Es necesario un marco muestral muy detallado.
- No sirve si hay algún ordenamiento en el marco muestral o que se presente algún comportamiento periódico. Por ejemplo que se trate de registros matrimoniales y todos los números impares sean mujeres.

3.3 Muestreo Aleatorio Estratificado (MAE)

3.3.1 Definición

Es un método de muestreo que consiste en clasificar primero los elementos de la población en grupos (estratos) y seleccionar luego, en cada grupo, una muestra aleatoria simple, tomando al menos un elemento de cada grupo (por lo general se necesitan mínimo dos para hacer estimaciones de precisión suficiente y si se quisieran hacer estimaciones por estrato).

Los estratos pueden reflejar distintos grupos de edad, regiones geográficas, sexo, diferente grado de exposición, etc.

Ejemplo 4.4 *Los estratos del ejemplo fueron definidos con base en tres criterios: ubicación geográfica, características socioeconómicas y diferente grado de exposición. Es claro ver que nuestro interés es estimar el parámetro (media del biomarcador) para cada uno de los estratos, puesto que suponemos diferente riesgo en cada uno de ellos y para la población total. (Esquema 4.1.2).*

En este caso, dado que en aras del ejemplo los estratos son de igual tamaño, podemos hacer una afijación proporcional de la muestra y tomar 100 niños de cada estrato, ya sea por MAS o por MS. Tendremos así la estimación para cada una de las zonas (estratos) de Pavarandocito y para la población total.

3.3.2 Procedimiento

El proceso que se sigue para establecer los grupos ya mencionados se conoce como estratificación y los distintos grupos se llaman estratos. Los estratos pueden reflejar regiones geográficas de un país, áreas densa o escasamente pobladas, diferentes niveles de contaminación o cualesquiera otros grupos.

En la estratificación se agrupan elementos similares a fin de que la varianza dentro de cada grupo sea pequeña; al mismo tiempo es deseable que las medias de los distintos estratos sean diferentes entre sí. Asimismo, todos los elementos deben de quedar incluidos en uno y sólo uno de los estratos.

No sólo las grandes encuestas tienen la estratificación incorporada en su diseño, es muy útil por ejemplo en un hospital, obtener datos para cada uno de los servicios o tener la información para hacer estimaciones por sexo. En el área ambiental la estratificación cobra gran relevancia por cuanto permite agrupar por zonas de riesgo similar, pero diferentes entre sí. Es muy importante el criterio experimentado en la definición y selección de los estratos ya que quizá sea este el paso más trascendental en este tipo de muestreo. Todos los estratos deben quedar representados.

En el MAE no siempre se hace una afijación proporcional de la muestra, ya que pueden tomarse en cuenta otros criterios que permiten llegar a la llamada afijación óptima, mediante la aplicación de una tasa de muestreo en cada estrato, proporcional a la desviación estándar en ese mismo estrato, es decir una especie de ponderación. En ocasiones los costos de obtención de los datos varían considerablemente de un estrato a otro (ej.: área rural y urbana) y se prefiere concentrar la muestra mayoritaria en el estrato más económico, considerando la afijación

óptima para un costo fijo, en lugar de un tamaño de muestra fijo. En otras palabras, se trata de afijar la muestra entre los estratos de tal forma, que se obtenga el error estándar más bajo, con un presupuesto fijo, para lo cual se necesita una "función del costo", la cual es una función matemática que expresa el costo de obtener la información en función del tamaño de la muestra (n). Este es un procedimiento más complejo, por ello el lector interesado puede consultar el texto de Scheaffer (Scheaffer, 1986).

Ejemplo 4.5 Afijación proporcional. (En el caso de estratos de diferentes tamaños).

Supongamos un universo de 10,000 y una muestra de tamaño 400; el universo tiene 3 estratos de tamaño 5,000, 3,000 y 2,000 respectivamente.

Universo	N	%	n
Estrato I	5,000	50	200
Estrato II	3,000	30	120
Estrato III	2,000	20	80
Total	10,000	100	400

El porcentaje con el que contribuye cada estrato al total del universo, se aplica al tamaño de muestra calculado n y ese es el número de elementos que debe tomarse en cada estrato.

De tal manera, que el Estrato I con 5,000 elementos, contribuye al 50% del Universo. La muestra calculada fue de 400. El 50% de 400 es igual a 200, ese es el número de elementos que debe tomarse del Estrato I y así sucesivamente.

3.3.3 Ventajas

- La estimación final del parámetro puede tener un error mucho menor que el obtenido para una muestra aleatoria simple de tamaño comparable.

- Por lo anterior puede tener una mejor relación costo-beneficio.
- Permite estimar parámetros para cada estrato y para la población total.
- Asegura la participación de todos los estratos.

3.3.4 Desventajas

- Mayor complejidad de los cálculos
- Necesario un marco muestral detallado por estrato.
- Criterio muy especializado en la selección y definición de los estratos.

3.4 Muestreo de Conglomerados (MC)

3.4.1 Definición

Se llama así a aquel tipo de muestreo, en el cual las unidades de muestreo son agrupaciones o colecciones de elementos llamadas conglomerados.

Los conglomerados pueden ser hogares, manzanas, fábricas, etc. (Ver Tabla No. 4.3. Ejemplos de Conglomerados y esquema 4.1.3 páginas 51 y 53 respectivamente).

3.4.2 Consideraciones especiales

Muchos problemas reales plantean el reto de muestrear ante la ausencia de un marco muestral muy detallado (por ejemplo, una lista de todos los escolares de Pavarandocito) o con un costo muy elevado en la construcción del mismo. Hay otras ocasiones en que este marco existe, pero el muestreo de conglomerados resulta mucho más eficiente dadas las economías que reporta en lo que se refiere a la etapa de campo.

Ejemplo 4.6 *En nuestro ejemplo los conglomerados son las escuelas, las cuales agrupan varias unidades de análisis (escolares). En términos generales la característica que define un conglomerado es su proximidad geográfica, al reunir varios elementos en una misma área. Un conglomerado no implica necesariamente la homogeneidad que se mencionaba para el estrato y representa más una facilidad operativa. De hecho, se busca que sean heterogéneos al interior de cada conglomerado y homogéneos entre sí.*

3.4.3 Procedimiento

En un sentido estricto el muestreo de conglomerados implica que se muestree a todos los elementos incluidos en el conglomerado muestreado. Si en nuestro ejemplo el conglomerado es la escuela, sólo necesitaríamos una lista de las doscientas escuelas de Pavarandocito, de la cual tomaríamos una muestra de ocho escuelas (ya sea por MAS o por MS) luego de lo cual estudiaríamos en cada escuela la totalidad de los niños de dicha escuela. Es decir, tenemos un universo de 10,000 elementos, repartidos en 200 conglomerados (escuelas) de 50 niños cada uno. Como la muestra es de 400, necesitamos un listado de las 200 escuelas y mediante la tabla de números aleatorios escogemos 8 de ellas, en estos 8 tenemos el total de la muestra ($8 \times 50 = 400$), por eso hay que estudiar a la totalidad de los niños de cada escuela.

Hay una variante de este muestreo, que puede quedar también incluida en lo que hemos llamado **diseños complejos** y es el conocido como muestreo **multietápico o polietápico de conglomerados** (bietápico, trietápico, etc.). En esta variante se seleccionan los conglomerados y puede procederse a una segunda selección dentro de los conglomerados seleccionados y lo mismo en las etapas que sean necesarias. Así la necesidad de listados se va limitando a aquellas unidades de muestreo que vayan siendo seleccionadas en cada etapa.

En este caso la primera etapa podría ser escoger conglomerados y luego dentro de cada conglomerado escoger elementos. Un caso muy usual es definir los conglomerados como manzanas (de vivienda) y luego dentro de cada manzana elegir viviendas (esquema 4.1.4, página 53).

Casi todos los países construyen sus marcos muestrales definiendo conglomerados básicos que pueden agruparse o desagruparse en conjuntos y subconjuntos de los mismos. Ello se verá con mayor detalle en el Capítulo VII dedicado a los marcos muestrales, pero basta señalar el ejemplo de México que ha definido las llamadas **áreas geoestadísticas básicas (AGEB)**, las cuales constituyen un segmento geográfico definido que puede abarcar tanto viviendas como negocios, fincas o terrenos y situarse en una zona rural o urbana. Un conglomerado puede ser un conjunto de AGEBs o un subconjunto del mismo, según se opere para los fines del estudio que se esté realizando.

3.4.4 Ventajas

- Disminución en los costos de trabajo de campo.
- Poder trabajar en ausencia de listados muy complejos.

3.4.5 Desventajas

- Complicación de los cálculos matemáticos para la estimación de los parámetros, lo cual en los tiempos actuales, con el uso masivo de las computadoras ha sido

obviado, ya que existe un buen número de paquetes computacionales que permiten obtener estos datos con gran facilidad: SUDAAN, PSDUM (Silva, 1993).

Antes de terminar este apartado vale la pena enfatizar la diferencia entre estratos y conglomerados. Los estratos deben ser tan homogéneos dentro de ellos como sea posible, pero un estrato debe diferir de otro tanto como se pueda en relación a la característica que está siendo medida. Los conglomerados por su parte deben ser tan heterogéneos dentro de sí mismos como sea posible, pero similares a los otros para aprovechar las ventajas económicas de su diseño. Los estratos deben quedar representados todos, es decir contribuir a la muestra. Los conglomerados no tienen que quedar representados todos, pero si uno queda en la muestra debe ser muestreado totalmente (al menos en el monoetápico).

3.5 Diseños Complejos

En ocasiones la realidad presenta retos que superan los esquemas clásicos presentados aquí, es el caso por ejemplo de las encuestas nacionales, especialmente en países con una población numerosa.

Es usual que ante la ausencia de marcos muestrales tan complejos y detallados como los que se requerirían para llevar a cabo una estrategia simple, se opta por estratificar y luego dentro de cada estrato se definen conglomerados que pueden muestrearse en una o varias etapas, estos diseños reciben el nombre de mixtos o complejos (esquema 4.1.5, página 54).

Regresemos a Pavarandocito y reflexionemos en las ventajas y economías que representaría la elección de un muestreo mixto, para nuestra muestra de 400 escolares. Podríamos simplemente tomar 2 escuelas de cada uno de los estratos y examinar la totalidad de los escolares incluidos en ellas, estaríamos frente a un tipo de muestreo mixto (estratificado y de conglomerados monoetápico) que simplificaría grandemente las actividades de campo, sin necesidad de un marco muestral complejo, ya que sólo sería necesario contar con el listado de alumnos de las escuelas incluidas. La visita a una escuela tendría así una relación costo-beneficio muy satisfactoria puesto que se examina al total de alumnos.

Un ejemplo real de diseño complejo lo constituye la segunda encuesta nacional de salud llevada a cabo en México en 1991 y cuyo diseño puede consultarse en las publicaciones al respecto. El texto de Silva-Aycaguer (1993) se extiende ampliamente en las características de este tipo de diseños, por lo cual remitimos a él, al lector interesado en profundizar en este tema.

4. MUESTRAS NO PROBABILÍSTICAS

Pese a la superioridad reconocida de los métodos de muestreo probabilístico, a veces se emplean en su lugar métodos no probabilísticos, habitualmente para situaciones en las cuales el primero es supremamente costoso o en casos que lo ameritan por la circunscripción de un problema específico, como suele suceder en algunas contingencias ambientales, en las cuales sólo es posible muestrear a aquellos sujetos expuestos o en los cuales se sospecha la presencia de alguna situación pre o patológica concreta, en cuyo caso hablaríamos de un estudio de brote (cluster).

Hay otros casos en los cuales no es posible acceder a la población de origen y se trabaja con los datos que se han podido obtener en un proceso que pareciera el inverso del muestreo. Un ejemplo concreto de esta caso es el que se presenta cuando tenemos datos de un programa de vigilancia epidemiológica (ej. Vigilancia de los niveles de plomo en el binomio madre-hijo)¹ que sin ser una muestra probabilística en el sentido estricto, nos permite inferir acerca de la población de origen. Una situación similar se presenta en los estudios analíticos donde importa más la comparabilidad que la representatividad.

Los métodos básicos de muestreo no probabilístico son:

- El muestreo de cuota.
- El muestreo de juicio (criterio).
- El muestreo de sujetos disponibles.

4.1 El muestreo por cuota

Esta empieza con una matriz que describe las características generales de la población que se desea estudiar. Por ejemplo distribución por edad y sexo, por zona urbana y rural etc. Una vez generada dicha matriz y asignada una proporción relativa a cada una de sus células se asume que la muestra ha de estar dispersa por toda la población y ha de contener la misma proporción de individuos con cierta característica, que la población entera. Se establece pues que la muestra tenga sus elementos distribuidos por las subclases de tal manera que las proporciones muestrales coincidan con las de la población. Puede asignarse al o los entrevistadores una cuota fija de muestreo con las restricciones que implique la cuota (12 entrevistas, seis en niños y seis en niñas por ejemplo).

¹ En este caso se toman muestras de sangre de la madre y de cordón umbilical para medir los niveles de plomo en sangre. Se toman madres con embarazo y parto normales que acepten participar voluntariamente. Se hace en los hospitales principales de la ciudad, con una cuota fija diariamente, de lunes a viernes.

4.2 El muestreo de juicio (criterio)

En el cual se selecciona específicamente un grupo que a juicio del criterio de un experto, represente adecuadamente a la población en cuanto a la característica que queremos estudiar. Este es un tipo de muestreo muy importante dado que es el único razonable por ejemplo en aquellos casos, en que el tamaño planeado de la muestra es muy pequeño, sobre todo si el tamaño poblacional también lo es. En la aplicación de este método el experto debe tener suficiente información acerca de la población y sopesar muy concienzudamente todos los elementos, para elegir aquellos que conformen el modelo más parecido a la realidad. Puede ser muy útil este muestreo pero nunca se le podrá aplicar la teoría del muestreo probabilístico.

Ejemplo 4.7 Escoger trabajadores de las gasolineras para estudiar efectos de exposición a hidrocarburos. No pueden aplicarse los resultados a la población general.

4.3 El muestreo de sujetos disponibles (o de conveniencia)

Es el que hace con sujetos que aceptan participar voluntariamente en el estudio, por algún interés personal o por una gratificación económica. Se ha utilizado mucho en clínica, farmacología, toxicología, etc. Por ejemplo, el estudio de Rabinowitz sobre la cinética del plomo (Rabinowitz, 1979).

Algunos autores hablan de otro tipo de muestreo al que llaman **semiprobabilístico** y que consiste en la selección aleatoria de una unidad muestral, después de lo cual se escogen los elementos a juicio de un experto, por lo cual según nuestro criterio, es del tipo conocido como muestreo de juicio. Se usa en estudios de tipo ambiental y ocupacional, en los cuales puede seleccionarse aleatoriamente una o varias unidades de muestreo, por ejemplo: determinadas áreas de trabajo en una fábrica, de cada una de ellas escoger los elementos a incluir.

En general debe recomendarse al investigador que trate con mucha cautela el tipo de diseño no probabilístico y que cuando ello sea necesario, se haga cuidando el rigor científico.

CONCLUSIÓN

El tipo de diseño a elegir depende pues del problema que se desea resolver y de las facilidades operativas y económicas de que se disponga. Parece una paradoja ya que hacemos un muestreo para conocer algo acerca de una población, pero para poder hacer el muestreo bien hecho debemos conocer muy bien la población que queremos muestrear, pero este conocimiento es más bien a nivel global que particular.

EJERCICIOS

- 1. Enumere las ventajas y desventajas de cada uno de los tipos de muestreo explicados en este capítulo.**
- 2. Utilizando la tabla de números aleatorios seleccione una muestra de tamaño 50, del universo de valores que se anexa en la tabla No. 4.2**
- 3. Utilizando el mismo ejemplo de Pavarandocito, seleccione una muestra de tamaño 800. Para el MAS, solamente mencione los 10 primeros elementos.**
- 4. Mencione qué marco muestral disponible para población general, existe en su país o región.**
- 5. Usted necesita una muestra de escolares de 6-7 años y de 12-13, para participar en un estudio multicéntrico sobre prevalencia de asma. ¿Qué estrategia podría utilizar en su región específica? ¿De qué marco muestral podría disponer? ¿Qué fallas tendría dicho marco? Presente las respuestas con un esquema que permita entender la propuesta.**

ESQUEMA No. 4.1
POBLACIÓN DE PAVARANDOCITO

<p>ZONA I</p> <p>$N_1 - 2,500$</p> <p>Escuelas - 50</p> <p>UG: Montaña NSE: Medio GE: Nula</p>	<p>ZONA II</p> <p>$N_2 - 2,500$</p> <p>Escuelas - 50</p> <p>UG: Llano NSE: Muy bajo GE: Muy alta</p>
<p>ZONA III</p> <p>$N_3 - 2,500$</p> <p>Escuelas - 50</p> <p>UG: Costa NSE: Bajo GE: Alta</p>	<p>ZONA IV</p> <p>$N_4 - 2,500$</p> <p>Escuelas - 50</p> <p>UG: Desierto NSE: Alto GE: Media</p>

N - 10,000

n - 400

200 escuelas con 50 alumnos cada una

UG - Ubicación Geográfica

NSE - Nivel Socioeconómico

GE - Grado de exposición

INSTRUCTIVO 4.1

USO DE LA TABLA DE NÚMEROS ALEATORIOS

Una manera de seleccionar una muestra aleatoria simple es utilizar una tabla de números aleatorios como la que aquí se ilustra. Esta consta de 20 columnas de números, agrupadas de cuatro en cuatro para ayuda visual y 20 renglones agrupados de cinco en cinco por la misma razón (tabla 4.1).

Nuestra población de interés son los 325 valores de plomo en sangre en mujeres puérperas que se indican en la tabla No. 4.2. Se desea extraer una muestra de tamaño 10, utilizando la tabla de números aleatorios.

Como primer paso se debe localizar un punto de partida, lo cual puede hacerse en varias formas, simplemente dejando caer la punta del lápiz sin mirar la hoja, el punto de partida será el dígito más próximo a la punta del lápiz. Supóngase que en este caso ese punto corresponde a la intersección del renglón 5 con la columna 9, es decir el dígito 2 (ver tabla).

Dado que se tienen 325 valores para elegir, es necesario seleccionar números de tres dígitos y aceptar aquellos que estén comprendidos entre 001 y 325 que es nuestro rango de posibilidades.

El primer número de tres dígitos, iniciando en el punto mencionado y siguiendo un esquema de lectura como el que se usa en la mayoría de los países occidentales (de izquierda a derecha y de arriba hacia abajo) es el número 232, el cual corresponde entonces a nuestro primer elemento de la muestra. Se aclara que se van tomando dígitos de tres en tres pero se incluyen en la muestra sólo los que estén dentro del rango de interés ya mencionado. En ese entendido los siguientes elementos de nuestra muestra serían los marcados con los números: 033, 219, 026, 045, 094, 132, 058, 106 y 066. Si un número se repite y nosotros estamos haciendo un muestreo sin reemplazo, lo incluimos solamente una vez.

TABLA 4.1
TABLA DE NÚMEROS ALEATORIOS

67 28	96 25	68 36	24 72	03 85
85 86	94 78	32 59	51 82	86 43
40 10	60 09	05 88	78 44	63 13
94 55	89 48	90 80	77 80	26 89
11 63	77 77	23 20	33 62	62 19
64 00	28 04	54 55	38 57	94 82
50 94	13 23	78 41	60 58	10 80
86 98	37 96	44 13	45 06	34 59
66 91	42 83	60 77	90 91	60 90
33 58	12 18	02 07	19 40	21 29
52 49	40 16	72 40	73 05	50 90
76 98	93 99	78 30	79 47	96 82
50 26	54 30	01 88	69 57	54 45
49 46	61 89	33 79	96 84	28 34
19 65	13 44	78 39	73 88	63 03
64 17	47 67	87 59	81 40	72 81
18 43	97 37	68 97	58 58	57 95
65 58	60 87	51 09	98 81	15 53
79 90	31 00	91 14	85 65	31 75
07 23	00 15	59 05	16 09	94 42

TABLA 4.2
VALORES DE PLOMO EN SANGRE EN 325 MUJERES
PUERPERAS DE LA CIUDAD DE MÉXICO

PLOMO $\mu\text{g/dl}$		PLOMO $\mu\text{g/dl}$		PLOMO $\mu\text{g/dl}$		PLOMO $\mu\text{g/dl}$	
1.	8.1	35.	15.9	69.	11.9	103.	8.7
2.	8.9	36.	6.9	70.	6.6	104.	5.9
3.	9.1	37.	6.6	71.	16.8	105.	13.3
4.	12.9	38.	20.3	72.	4.9	106.	10.7
5.	12.4	39.	7.6	73.	9.6	107.	9.8
6.	6.1	40.	11.5	74.	14.9	108.	14.4
7.	15.9	41.	5.1	75.	11.6	109.	5.2
8.	8.2	42.	7.8	76.	5.6	110.	4.2
9.	7.1	43.	11.4	77.	24.5	111.	5.4
10.	7.8	44.	14.6	78.	9.1	112.	7.3
11.	6.5	45.	6.3	79.	9.2	113.	4.0
12.	8.7	46.	16.2	80.	8.2	114.	3.2
13.	7.0	47.	8.8	81.	9.9	115.	2.9
14.	7.2	48.	6.4	82.	8.1	116.	8.3
15.	5.9	49.	10.0	83.	7.7	117.	10.5
16.	16.9	50.	7.4	84.	16.3	118.	21.0
17.	10.7	51.	11.7	85.	13.1	119.	8.1
18.	16.6	52.	17	86.	17.0	120.	9.2
19.	8.0	53.	14.4	87.	12.0	121.	8.5
20.	5.6	54.	10.0	88.	9.1	122.	10.0
21.	11.5	55.	7.9	89.	9.3	123.	6.3
22.	9.7	56.	8.2	90.	4.2	124.	8.5
23.	12.9	57.	8.3	91.	4.5	125.	23.1
24.	3.2	58.	11.3	92.	4.8	126.	13.3
25.	6.3	59.	5.1	93.	10.3	127.	10.5
26.	7.8	60.	23.3	94.	5.2	128.	6.8
27.	4.8	61.	10.0	95.	5.4	129.	14.4
28.	8.2	62.	10.4	96.	4.0	130.	7.6
29.	9.1	63.	6.5	97.	7.8	131.	5.9
30.	8.4	64.	6.4	98.	6.3	132.	7.4
31.	3.2	65.	12.3	99.	3.6	133.	7.4
32.	4.7	66.	5.9	100.	14.0	134.	6.4
33.	12.4	67.	9.3	101.	9.1	135.	9.3
34.	5.3	68.	23.0	102.	6.6	136.	11.9

PLOMO $\mu\text{g/dl}$	PLOMO $\mu\text{g/dl}$	PLOMO $\mu\text{g/dl}$	PLOMO $\mu\text{g/dl}$
137. 6.3	174. 11.9	211. 11.9	248. 7.9
138. 5.0	175. 22.4	212. 9.9	249. 6.3
139. 11.2	176. 11.7	213. 11.8	250. 5.8
140. 6.8	177. 15.8	214. 12.5	251. 5.9
141. 6.9	178. 9.2	215. 6.7	252. 16.5
142. 5.7	179. 7.4	216. 9.1	253. 11.7
143. 4.8	180. 11.4	217. 5.1	254. 9.5
144. 11.2	181. 23.9	218. 8.4	255. 7.9
145. 10.0	182. 14.3	219. 12.9	256. 6.3
146. 5.9	183. 9.8	220. 14.9	257. 10.9
147. 11.4	184. 10.6	221. 4.9	258. 11.1
148. 6.5	185. 7.0	222. 8.9	259. 9.1
149. 7.8	186. 7.9	223. 14.2	260. 6.2
150. 10.3	187. 6.8	224. 12	261. 8.8
151. 14.5	188. 12.4	225. 11.7	262. 7.3
152. 14.4	189. 5.1	226. 8.2	263. 9.7
153. 7.4	190. 5.5	227. 7.3	264. 9.2
154. 5.2	191. 7.4	228. 7.0	265. 10.8
155. 18.9	192. 17.1	229. 13.5	266. 7.0
156. 16.1	193. 7.1	230. 8.1	267. 6.3
157. 29.1	194. 9.7	231. 7.5	268. 14.3
158. 4.0	195. 14.2	232. 8.5	269. 15.4
159. 19.5	196. 9.4	233. 8.5	270. 6.9
160. 12.5	197. 9.4	234. 10.0	271. 8.6
161. 23.1	198. 8.7	235. 6.2	272. 4.4
162. 11.8	199. 4.0	236. 16.9	273. 8.0
163. 6.0	200. 10.7	237. 9.0	274. 9.0
164. 8.4	201. 13.1	238. 21.0	275. 4.4
165. 11.0	202. 8.8	239. 9.2	276. 7.7
166. 7.3	203. 8.7	240. 5.7	277. 7.2
167. 11.6	204. 11.9	241. 9.4	278. 18.7
168. 13.4	205. 11.5	242. 6.2	279. 7.3
169. 6.4	206. 11.4	243. 6.6	280. 7.9
170. 9.2	207. 11.9	244. 12.3	281. 8.9
171. 12.7	208. 10.0	245. 8.7	282. 12.6
172. 8.5	209. 10.5	246. 8.1	283. 7.0
173. 15.0	210. 7.3	247. 5.6	284. 9.1

PLOMO $\mu\text{g/dl}$		PLOMO $\mu\text{g/dl}$		PLOMO $\mu\text{g/dl}$		PLOMO $\mu\text{g/dl}$	
285.	4.6	296.	5.0	307.	9.3	318.	10.5
286.	6.8	297.	11.1	308.	7.9	319.	13.0
287.	4.9	298.	9.7	309.	7.2	320.	12.0
288.	6.1	299.	12.1	310.	4.5	321.	10.7
289.	5.2	300.	5.8	311.	6.4	322.	9.9
290.	7.0	301.	4.4	312.	7.3	323.	8.8
291.	8.7	302.	7.3	313.	7.8	324.	8.0
292.	15.2	303.	7.8	314.	14.9	325.	8.0
293.	5.5	304.	4.3	315.	13.5		
294.	6.8	305.	5.2	316.	8.0		
295.	9.9	306.	8.9	317.	9.8		

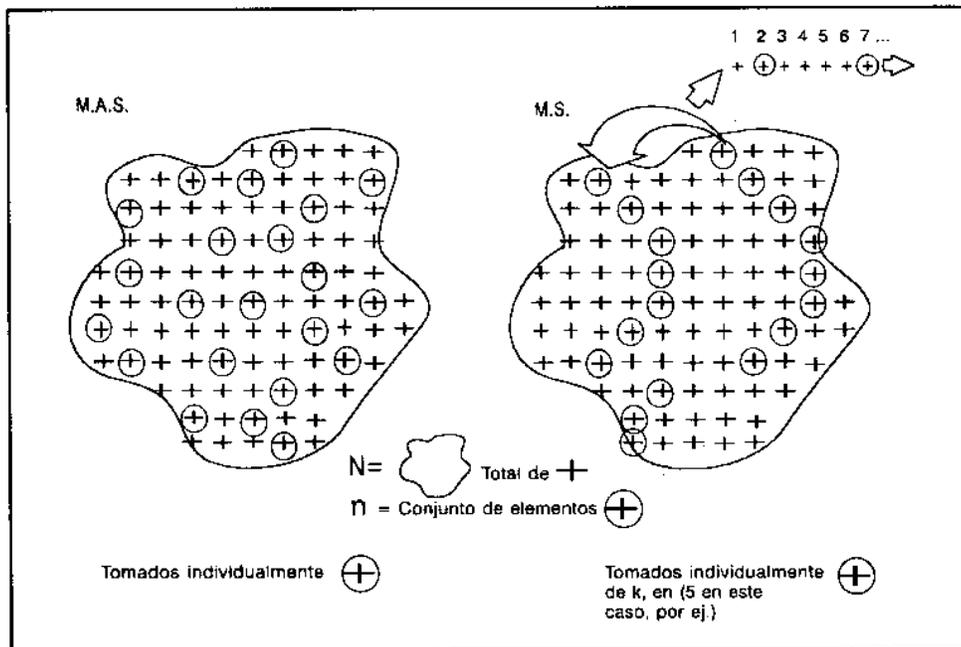
A continuación se presentan los diez elementos incluidos en la muestra con sus respectivos valores:

No.	Tabla	Pb	No.	Tabla	Valor Pb
1.	232	8.5	6.	094	5.2
2.	033	12.4	7.	132	7.4
3.	219	12.9	8.	058	11.3
4.	028	7.8	9.	106	10.7
5.	045	6.3	10.	066	5.9

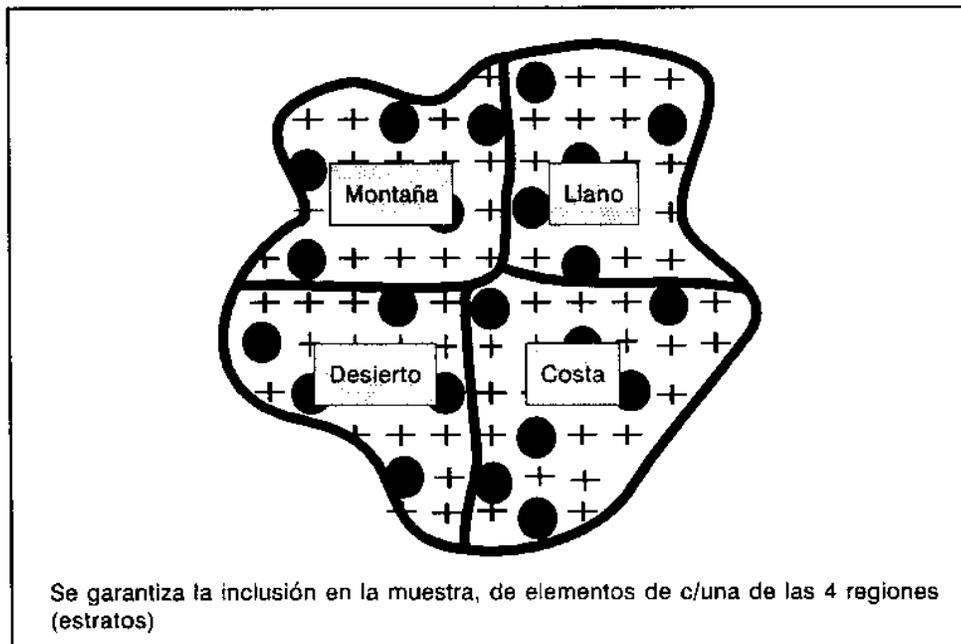
**TABLA NO. 4.3
EJEMPLO DE CONGLOMERADOS**

Conglomerado	Unidad de lista	Elemento	Aplicación
Manzana	Casa	Persona	Estimación del total de personas con hipertensión en la ciudad.
Condado	Hospital	Paciente	Estimación de la proporción de pacientes muertos en hospital en un estado.
Escuela	Clase	Estudiante	Estimación de la media de escolaridad, obtenida por los alumnos en un distrito.
Paquete de jeringas	Jeringa	Jeringa	Estimación de la proporción de jeringas con defectos.
Página de texto	Línea de texto	Palabra	Estimación del número total de palabras en un libro.
Conjunto de archivos	Folder	Cuenta única	Estimación del número total de cuentas vencidas.
Semana calendario	Día	Día	Estimación de la proporción de días con nivel de ozono por arriba del nivel permitido.

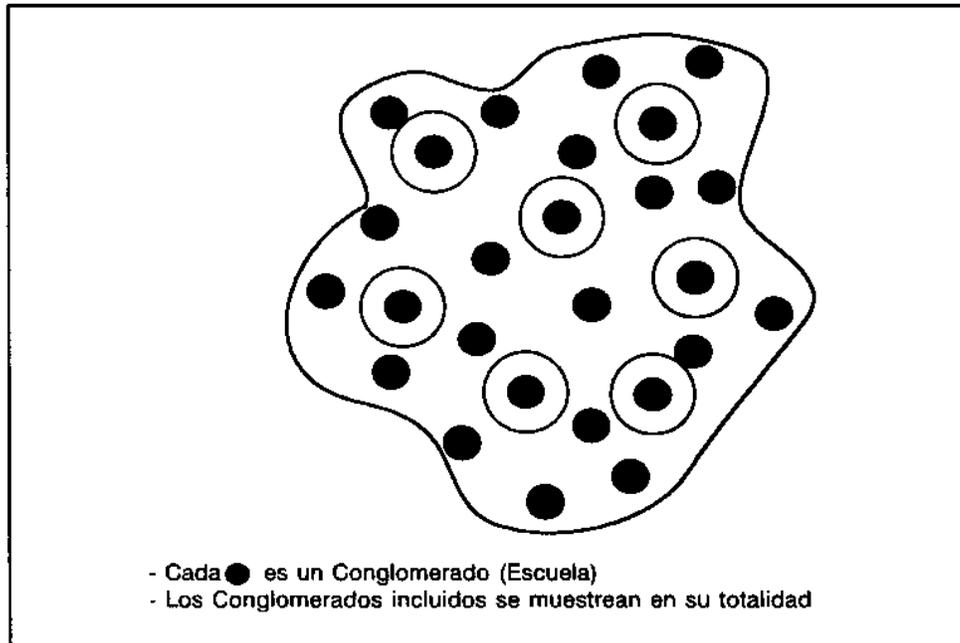
ESQUEMA 4.1.1
MUESTREO ALEATORIO SIMPLE Y MUESTREO SISTEMÁTICO



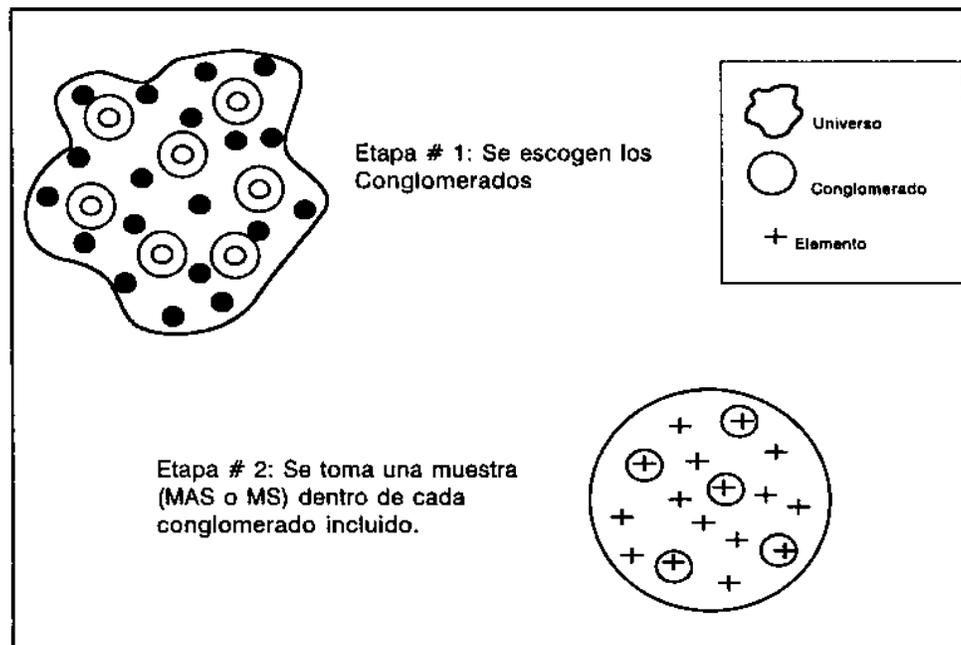
ESQUEMA 4.1.2
PAVARANDOCITO. MUESTREO ESTRATIFICADO



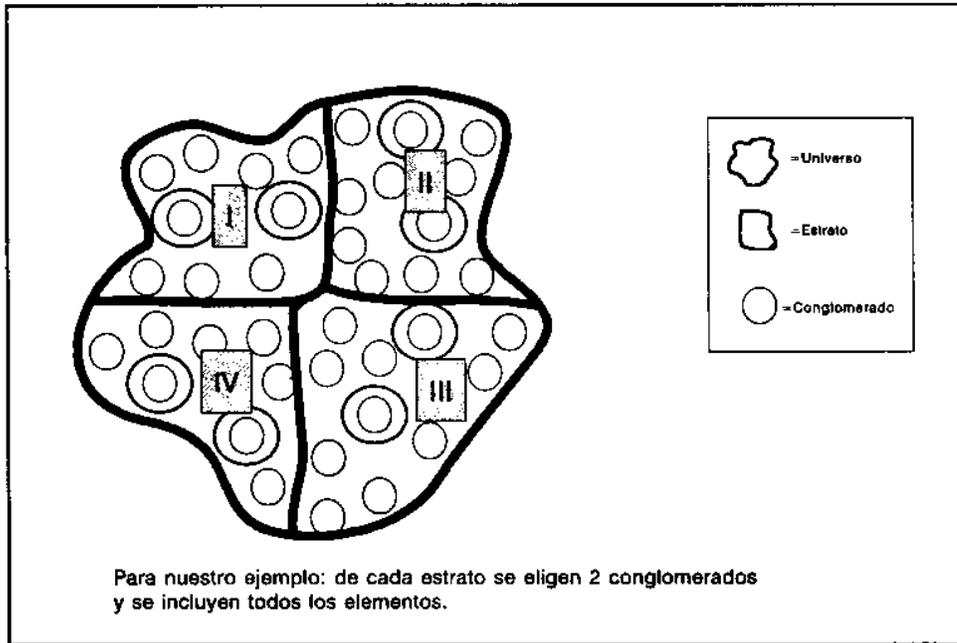
ESQUEMA 4.1.3
PAVARANDOCITO. MUESTREO DE CONGLOMERADOS (MONOETÁPICO)



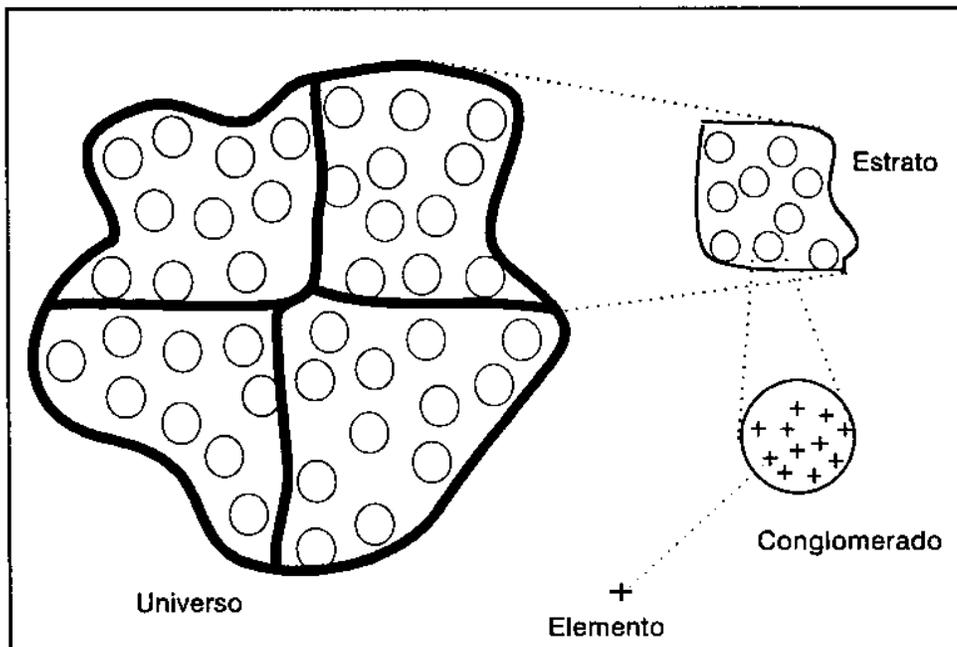
ESQUEMA 4.1.4
MUESTREO DE CONGLOMERADOS (BIETÁPICO)



**ESQUEMA 4.1.5
PAVARANDOCITO. MUESTREO MIXTO**



**ESQUEMA 4.1.6
REPRESENTACIÓN GRÁFICA DE LOS CONCEPTOS**



CAPÍTULO V

APLICACIONES CON MAS

Si bien ya hemos revisado los diferentes tipos de muestreo en el capítulo anterior, en este retomaremos los conceptos estudiados con los siguientes objetivos:

- Ilustrar el Cálculo del Error Estándar.
- Presentar las fórmulas para estimación de medias y proporciones.
- Tomarlo como punto de partida para las fórmulas de los otros tipos de muestreo.

Estrategia: Continuando con nuestro ejemplo, presentaremos la forma de obtención de los parámetros más usuales con el fin de poder extrapolarlo a los otros tipos de muestreo para los cuales se anexarán las fórmulas correspondientes.

Ejercicios: tendientes a aplicar las fórmulas de MAS, para afianzar los conceptos.

1. INTRODUCCIÓN

Cuando un investigador ya ha resuelto qué tipo de muestreo aplicará y cuál será su tamaño de muestra, aún se enfrenta a dos grandes problemas: ¿Cómo seleccionar la muestra? y ¿Cómo estimar los diversos parámetros poblacionales? El primer problema se tratará en el Capítulo VII y el segundo será tratado en este capítulo.

Si en este caso el investigador ya ha elegido el MAS, deberá tomar su muestra por el método que ya hemos estudiado, con la ayuda de un buen marco muestral, pero una vez terminada la etapa de campo y con todos los datos en sus manos, ¿cómo es que puede hacer inferencias poblacionales de una manera correcta y válida desde el punto de vista estadístico, científico?

Anteriormente establecimos que el objetivo de una encuesta por muestreo es hacer inferencias acerca de una población, a partir de la información contenida en una muestra. El objetivo de una encuesta por muestreo es, frecuentemente, estimar una media poblacional denotada por μ , o una proporción P. El investigador de Pavarandocito desea conocer el valor medio poblacional del biomarcador en estudio y/o el total o la proporción de niños escolares que están por arriba de ciertos límites aceptados internacionalmente para ese biomarcador.

Antes de referirnos a fórmulas concretas debemos retomar los conceptos mencionados en el Capítulo II acerca de la distribución muestral.

La distribución muestral es la distribución de los estimadores obtenidos en cada una de todas las muestras posibles de tamaño n de una población de tamaño N . Según el teorema del límite central esta distribución, si se cumple el supuesto de "adecuación" del tamaño, debe ser normal y como tal tiene una media y una medida de dispersión. La distribución muestral es la que nos sirve de puente para pasar de la muestra al universo, ya que a partir de los datos obtenidos en la muestra y utilizando las propiedades de la distribución muestral podemos construir fórmulas que nos permiten estimar los parámetros (ver esquema 5.1, página 63).

2. USOS DEL ERROR ESTÁNDAR

La Desviación Estándar de la distribución de los estimadores muestrales se conoce como Error Estándar o error de muestreo ya que nos permite conocer la probabilidad de que una muestra se desvíe del verdadero parámetro, es decir, nos permite conocer la desviación de la media de una muestra de la verdadera media del universo, dentro de ciertos límites de confiabilidad que el investigador mismo define, utilizando las propiedades de la curva normal (Cap. III, pág. 20).

Los usos estadísticos del error estándar son básicamente tres, pero su importancia es muy amplia. Ellos son:

- Estimación por intervalos de parámetros.
- Pruebas de hipótesis.
- Cálculo del tamaño de muestra.

En este capítulo haremos referencia al primer punto y en el Capítulo VII, dedicado al tamaño de la muestra nos referiremos al tercero. El segundo uso va más allá de los alcances de este manual y puede ser consultado en prácticamente cualquier libro de estadística (Pagano 1992).

3. CÁLCULO DEL ERROR ESTÁNDAR

En el Capítulo III, vimos que la fórmula simple del Error Estándar, construida a partir de los datos muestrales, mediante derivaciones matemáticas que podemos obviar es:

$$\sigma_{\bar{x}} = \sqrt{\frac{\hat{S}^2}{n}} = \frac{\hat{S}}{\sqrt{n}} \quad (5.1)$$

Donde \hat{S}^2 es la varianza de la muestra y \hat{S} la desviación estándar de la misma, la corrección para población finita :

$$\frac{N - n}{N} \quad (5.2)$$

puede ignorarse en el caso de poblaciones grandes, puesto que la fracción tiende a uno y no altera el resultado del resto de la ecuación (Cap. III, página 20).

La fórmula representa la manera de obtener el error estándar en el caso del MAS y del MS. En los otros tipos de muestreo esta es la fórmula que varía fundamentalmente y por ende todos los cálculos que impliquen el error estándar en ellos.

Al final del capítulo (anexo 5.1, página 65) se anexan las fórmulas para cálculo del Error Estándar en el caso del muestreo estratificado y de conglomerados.

4. ESTIMACIÓN DEL INTERVALO DE CONFIANZA

El Error Estándar para datos continuos es calculado con el fin de estimar un intervalo de confianza para el parámetro, lo cual se hace con la construcción de un intervalo de la manera siguiente: con la media muestral $\pm z$ veces el Error Estándar, tendremos un Intervalo de Confianza (IC) para el parámetro, cuyo nivel de confianza corresponde al valor de zeta empleado (Cap. III, página 20).

$$\bar{x} - Z\sigma_{\bar{x}} \leq \mu \leq \bar{x} + Z\sigma_{\bar{x}} \quad (5.3)$$

Esto es válido únicamente si el tamaño de la muestra es suficientemente grande o bien la población original está distribuida normalmente, de otro modo, se debe usar la distribución de t.

Antes de proceder a ejemplificar los conceptos con nuestro ejemplo de los valores de plomo, veamos qué sucede cuando la característica de interés es una variable dicotómica y lo que queremos estimar es una proporción.

Las fórmulas se mantienen iguales desde el punto de vista conceptual, ya que lo único que modificamos es la fórmula de la varianza, que es equivalente a $p_x(1-p_x)$ en el caso de variables nominales. Por ello la fórmula del Error Estándar para proporciones es:

$$\sigma_p = \sqrt{\frac{p_x(1-p_x)}{n}} \quad (5.4)$$

Donde $(1-p) = q$, y la fórmula del IC es:

$$p_x - Z\sigma_p \leq P \leq p_x + Z\sigma_p \quad (5.5)$$

Ahora veremos la aplicación de estas fórmulas a casos concretos.

Ejemplo 5.1 *A continuación se presentan los datos de una muestra de tamaño 30, del universo de valores de plomo presentado en la tabla No. 4.2 del Capítulo IV, página 48:*

TABLA 5.1
MUESTRA DE MEDICIONES DE PLOMO ($\mu\text{g/dl}$) DE SANGRE
DE 30 MUJERES PUÉRPERAS DE LA CIUDAD DE MÉXICO
EXTRAÍDAS POR MUESTREO ALEATORIO SIMPLE

No.	Pb	No.	Pb	No.	Pb	No.	Pb	No.	Pb	No.	Pb
1	8.9	6	17.0	11	6.6	16	14.4	21	6.2	26	7.3
2	8.7	7	10.4	12	7.3	17	11.8	22	16.5	27	14.3
3	9.7	8	4.9	13	10.0	18	8.5	23	7.3	28	17.1
4	4.7	9	8.1	14	7.4	19	9.9	24	4.4	29	8.8
5	7.8	10	4.8	15	5.7	20	8.5	25	12.6	30	7.3

Estimando una media: Según las fórmulas de estadística descriptiva presentadas en el Capítulo II podemos decir que:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{276.9}{30} = 9.2 \quad (5.6)$$

y que $\hat{S}^2 = 13.177$ y por lo tanto $\hat{S} = 3.63$

$$\sigma_{\bar{x}} = \frac{3.63}{\sqrt{30}} = \frac{3.63}{5.48} = 0.662 \quad (5.7)$$

Bajo el supuesto de que la distribución del plomo es normal, el intervalo de 95% de confianza para la media de plomo (μ) es:

$$\bar{x} - Z\sigma_{\bar{x}} \leq \mu \leq \bar{x} + Z\sigma_{\bar{x}} \quad (5.3)$$

$$9.2 - 1.96(0.662) \leq \mu \leq 9.2 + 1.96(0.662)$$

$$7.9 \leq \mu \leq 10.5$$

Es decir la media de plomo en el universo se encuentra entre 7.9 y 10.5 $\mu\text{g/dl}$, con un 95% de confiabilidad.

Ejemplo 5.2 *Estimando una proporción. Deseamos conocer la proporción de mujeres que sobrepasan un nivel de 10 $\mu\text{g/dl}$ de plomo en sangre en el universo.*

Para ello debemos seguir los mismos pasos:

- *Calcular los estadísticos de interés en la muestra.*
- *Hacer el cálculo del Error Estándar para proporciones.*
- *Utilizando las propiedades del error estándar construir un intervalo de confianza.*

Podemos observar que en la tabla 5.1, hay 8 valores que sobrepasan los 10 $\mu\text{g/dl}$, los cuales son las observaciones marcadas con los números 6, 7, 16, 17, 22, 25, 27, 28. La observación número 13, cuyo valor es 10.0 no la tomaremos en cuenta por cuanto ya definimos operacionalmente como normales los valores hasta 10 $\mu\text{g/dl}$.

1.	8.9	6.	17.0	11.	6.6	16.	14.4	21.	6.2	26.	7.3
2.	8.7	7.	10.4	12.	7.3	17.	11.8	22.	16.5	27.	14.3
3.	9.7	8.	4.9	13.	10.0	18.	8.5	23.	7.3	28.	17.1
4.	4.7	9.	8.1	14.	7.4	19.	9.9	24.	4.4	29.	8.8
5.	7.8	10.	4.8	15.	5.7	20.	8.5	25.	12.6	30.	7.3

El primer paso es entonces calcular la proporción de valores altos en la muestra:

$$p = 8/30 = 0.27 \quad (5.8)$$

El segundo paso es calcular el Error Estándar estimado con la fórmula:

$$\sigma_p = \sqrt{\frac{(0.27)(0.73)}{30}} = 0.081 \quad (5.9)$$

Con los datos anteriores ya podemos construir nuestro intervalo de 95% de confianza:

$$p_x - 1.96 \sigma_p \leq P \leq p_x + 1.96 \sigma_p \quad (5.10)$$

$$0.27 - 1.96(0.081) \leq P \leq 0.27 + 1.96(0.081)$$

$$0.11 \leq P \leq 0.43$$

Lo cual expresado verbalmente quiere decir, que la proporción de mujeres con valores altos de plomo en la población se encuentra entre 0.11 y 0.43 con un 95% de confiabilidad, o lo que es lo mismo, el porcentaje de mujeres con valores altos de plomo en sangre en el universo se encuentra entre 11 y 43 % con un 95% de probabilidad.

Un intervalo de esta amplitud (poca precisión) nos permite ir intuyendo la importancia del tamaño muestral, en la precisión de los resultados.

Antes de dar por terminada esta ejemplificación es muy importante aclarar que si nosotros tenemos una idea confiable del valor del parámetro, ese es el valor que debemos usar en nuestras fórmulas. Por ejemplo si nosotros tenemos datos que permiten saber aproximadamente el valor de la varianza en el Universo, debemos usar ese valor en lugar de la S^2 de la muestra, igualmente con las proporciones.

Esto solamente podemos hacerlo si el valor es confiable, es decir si proviene de una encuesta seria o de aproximaciones previas de calidad.

Ejemplo 5.3 *El Secretario del Medio Ambiente de una ciudad con 20 millones de habitantes y serios problemas de contaminación ambiental, estima que aproximadamente el 80% de los vehículos que circulan en su ciudad cuentan con un convertidor catalítico. En una muestra de 500 vehículos tomada aleatoriamente por MAS se encuentra que 350*

vehículos poseen convertidor catalítico. Dé un intervalo de confianza del 95% para el porcentaje de vehículos con convertidor catalítico en dicha ciudad.

Nota: En este caso se tienen razones para confiar en la estimación del parámetro.

- *Primero los valores en la muestra: $n = 500$ $p = 350/500 = 0.7$ y por consiguiente $(1-p) = 0.3$.*
- *Cálculo del error estándar según las fórmulas ya conocidas. Aquí nos sentimos tentados a poner en nuestra fórmula:*

$$\sigma_p = \sqrt{\frac{(0.7)(0.3)}{500}} = 0.02 \quad (5.4)$$

Pero ello sería incorrecto, ya que tenemos un estimado de la varianza en el universo, es decir tenemos una idea del parámetro el cual sería en este caso (0.8) (0.2) y por ello el reemplazo correcto para la fórmula del Error Estándar sería como sigue:

$$\sigma_p = \sqrt{\frac{(0.8)(0.2)}{500}} = 0.018 \quad (5.4)$$

El Intervalo de Confianza para el 95% quedaría pues construido de la siguiente forma:

$$p_x - 1.96 \sigma_p \leq P \leq p_x + 1.96 \sigma_p \quad (5.5)$$

$$0.7 - 1.96(0.018) \leq P \leq 0.7 + 1.96(0.018)$$

$$0.665 \leq P \leq 0.735$$

Lo cual quiere decir que en la población, el porcentaje de vehículos con convertidor catalítico se encuentra entre 66.5 y 73.5 con un 95% de confiabilidad.

CONCLUSIONES

La estimación de parámetros se basa en la construcción de intervalos mixtos, utilizando para ello la distribución muestral a manera de puente para transitar de la muestra al universo mediante la utilización de las propiedades del Error Estándar.

La fórmula del error estándar varía de un tipo de muestreo a otro y se va complicando en sus términos a medida que se complica el tipo de muestreo, pero la esencia de la construcción del intervalo es la misma.

Siempre que tengamos alguna idea confiable del parámetro, éste debe ser usado en las fórmulas en lugar del valor muestral. Recordar que cuanto más conozcamos el universo más exitoso será nuestro muestreo.

La disponibilidad de paquetes computacionales en la actualidad, hace innecesario el cálculo manual, máxime en el caso de diseños complejos, sin embargo la aproximación manual a un método sencillo facilita una mejor comprensión de los cálculos computacionales complejos.

EJERCICIOS

1. Realice los mismos ejemplos de este capítulo utilizando el término de corrección para población finita. Compare y comente los resultados.
2. Utilice la muestra que tomó en el Capítulo IV y estime un intervalo para la Media de los valores de plomo en el universo. Construya intervalos para el 90, 95 y 99% de confianza. Haga los cálculos con factor y sin factor de corrección. Compare y comente.
3. Utilice la muestra que tomó en el Capítulo IV y estime un intervalo para el porcentaje de mujeres con valores "altos de plomo" en el universo. Construya intervalos para el 90, 95 y 99% de confianza. Haga los cálculos con factor y sin factor de corrección. Compare y comente.
4. El investigador de Pavarandocito cree, por su experiencia y por comentarios con los clínicos del lugar, que aproximadamente el 25% de los niños escolares tiene niveles elevados del biomarcador en estudio. En su muestra encontró que 120 niños tenían niveles por encima de lo aceptable. Construya un intervalo de confianza para el verdadero porcentaje de niños con niveles elevados del biomarcador con un 95 y con un 99% de confianza. Hágalo con factor de corrección y sin factor de corrección. Compare y comente.
5. Compare los resultados de estimación de media y proporción que usted obtuvo en su muestra, con los parámetros. Calcule los parámetros de la tabla 4.2. ¿Cuáles son sus conclusiones?

ESQUEMA 5.1
EL CAMINO DE LA INFERENCIA

Muestra	Distribución Muestral	Universo
<p>n - Tamaño</p> <p>\bar{x} - Media</p> <p>\hat{s} - Desviación estándar</p> <p>\hat{s}^2 - Varianza</p> <p>p - Proporción</p> <p>q - Complemento de p - $(1-p)$</p> <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 10px auto;">Estimadores Muestrales</div>	<p>\bar{x}_i - Media de cada una de las posibles muestras de tamaño n</p> <p>p_i - Proporción de cada una de las posibles muestras de tamaño n</p> <p>$\mu_{\bar{x}}$ - Media de "medias"</p> <p>μ_p - Media de "proporciones"</p> <p align="center">↓</p> <p>Estimadores insesgados</p> <p>$\sigma_{\bar{x}}$ - Error estándar para la media</p> <p>σ_p - Error estándar para la proporción</p> <p align="center">↓</p> <p>Ventajas: - Pueden estimarse a partir de la \hat{s} muestral o de la p muestral.</p> <p>- Permite la estimación de parámetros.</p> <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 10px auto;">Mediante un intervalo: Muestra y distribución muestral, se infiere para el Universo</div>	<p>N - Tamaño</p> <p>μ - Media</p> <p>σ - Desviación estándar</p> <p>σ^2 - Varianza</p> <p>P - Proporción</p> <p>Q - complemento de $P = (1-p)$</p> <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 10px auto;">Parámetros</div>

ANEXO 5.1

FÓRMULAS PARA EL CÁLCULO DEL ERROR ESTÁNDAR EN EL M.E. Y EN EL M.C.

MUESTREO ESTRATIFICADO

Error estándar (de la media)

$$\sigma_{\bar{x}'} = \sqrt{\frac{1}{N^2} \sum_{i=1}^k \left(\frac{N_i - n_i}{N_i - 1} \right) N_i^2 \left(\frac{\sigma_i^2}{n_i} \right)} \quad (5.11)$$

- N_i - Número de elementos de un estrato i .
- N - Número total de elementos en el universo.
- n_i - Número de elementos en la muestra, en un estrato.
- σ_i^2 - Varianza de la variable en el i -ésimo estrato.
- k - Número de estratos.

Error estándar para una proporción

$$\sigma_{\hat{p}'} = \sqrt{\frac{1}{N^2} \sum_{i=1}^k \left(\frac{N_i - n_i}{N_i - 1} \right) N_i^2 \left(\frac{p_i q_i}{n_i} \right)} \quad (5.12)$$

- p_i - Proporción de casos en el i -ésimo estrato
- q_i - $1 - p_i$

Nota: En el muestreo estratificado el número de elementos en un estrato (N_i) puede ser pequeño y por ello se usa en el denominador $N_i - 1$, a menos que se esté muy seguro que el factor de corrección se aproxima a uno.

MUESTREO DE CONGLOMERADOS (MONOETÁPICO)

Error estándar de la media

$$\sigma_{\bar{x}} = \sqrt{\left(\frac{N-n}{N_n M^2}\right) \sum_{i=1}^n \frac{(x_i - \bar{x}_m)^2}{n-1}} \quad (5.13)$$

- N** - Número de conglomerados en la población.
n - Número de conglomerados seleccionados.
m_i - Número de elementos en el conglomerado i, i = 1, ..., N.
M - $\sum_{i=1}^n m_i$ - Número de elementos en la población.
 \bar{M} - $\frac{M}{N}$ - Tamaño promedio del conglomerado en la población.
x_i - Total de todas las observaciones en el i-ésimo conglomerado.

Error estándar para una proporción

$$\sigma_{\hat{p}} = \sqrt{\left(\frac{N-n}{N_n M^2}\right) \sum_{i=1}^n \frac{(a_i - \hat{p}_m)^2}{n-1}} \quad (5.14)$$

- a_i** - Número total de elementos en el conglomerado i que poseen la característica de interés.

$$\hat{p} = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n m_i}$$

Cálculo de la media

Para cada conglomerado: $\bar{x} = \frac{X}{M}$

Para el total: $\bar{x} = \frac{X}{N}$

CAPÍTULO VI

TAMAÑO DE LA MUESTRA

1. INTRODUCCIÓN

Uno de los primeros aspectos que más nos preocupa cuando estamos planeando realizar una encuesta por muestreo es qué tan grande será la muestra que necesitamos para obtener estimadores a partir de ella, que sean lo suficientemente confiables a fin de cumplir los objetivos del estudio.

Para tal efecto y antes de iniciar cualquier encuesta por muestreo, debemos conocer diversos factores que influyen en la determinación de un tamaño muestral mínimo adecuado que nos permita obtener estimaciones útiles sobre aquellos parámetros que se desean estudiar. En primera instancia, se presentarán todos los cálculos asumiendo muestreo aleatorio simple.

2. FACTORES QUE INFLUYEN EN EL TAMAÑO DE LA MUESTRA

2.1 Precisión

En primer término, nos referiremos al grado de precisión que requiere el estimador y que es equivalente a la mitad del intervalo que se pretende construir; mientras más precisión se imponga más corto será el intervalo que se obtenga y por consiguiente mayor será el tamaño de la muestra.

Ejemplo 6.1 Supóngase que la característica en estudio es el cadmio en cabello de niños de una localidad "x". A partir de un estudio previo se encontró el rango aproximado está entre 0.002 y 2.544 $\mu\text{g/g}$ de cabello, es decir una amplitud de 2.542. Por las propiedades de la curva normal, sabemos que el rango equivale aproximadamente a seis desviaciones estándar ($2.542/6 = 0.424$), por lo cual en este caso la varianza (s^2) sería $0.424^2 = 0.180$

Si la característica de interés es por ejemplo el nivel medio de plomo en sangre en niños, podemos consultar los registros de estudios anteriores en la misma región, en un proceso análogo al mencionado anteriormente. Si no es posible realizar este tipo de inferencias es necesario recurrir al estudio piloto.

Considere que la primera especificación es la creación de un intervalo del 95% de confianza basado en la media muestral más o menos 0.05 $\mu\text{g/g}$. Esta restricción conocida como precisión requiere que el 0.05 $\mu\text{g/g}$ sea igual a $1.96 \sigma_{\bar{x}}$

Asumiendo inicialmente MAS, e ignorando el factor de corrección para poblaciones finitas, tenemos que el $\sigma_{\bar{x}} = \sqrt{\frac{S^2}{n}}$ donde S^2 es la varianza poblacional, y n' es la estimación inicial a partir del cual se aproxima el tamaño muestral, es decir, sin considerar el factor de corrección para poblaciones finitas.

Entonces, sustituyendo el error estándar de la media, encontramos:

$$0.05 = \frac{1.96 (\sqrt{0.180})}{\sqrt{n'}}$$

y resolviendo para n' , encontramos que:

$$n' = \frac{(1.96^2)(0.180)}{0.05^2} = 276 \text{ personas} \quad (6.1)$$

Ejemplo 6.2 *Suponga usted que se conducirá una encuesta sobre una población para estimar la proporción de personas que usan loza vidriada para consumir sus alimentos; considere que la primera especificación es la creación de un intervalo del 95% de confianza basado en la proporción muestral más o menos 2%. Es decir, una precisión del 2% igual a $1.96\sigma_p$.*

Asumiendo inicialmente MAS, e ignorando el factor de corrección para poblaciones finitas, tenemos que el $\sigma_p = \sqrt{\frac{PQ}{n'}}$, donde P es la proporción poblacional que usan loza vidriada, Q es la proporción que no lo usa, es decir, $Q = 1-P$ y n' es la estimación inicial a partir de la cual se aproxima el tamaño muestral, sin tomar en cuenta el factor de corrección para poblaciones finitas.

Por la precisión fijada, tenemos que $1.96\sigma_p = 0.02$, sustituyendo el error estándar encontramos que:

$$1.96\sigma_p = 1.96 \sqrt{\frac{p(1-p)}{n'}} = 0.02$$

y resolviendo para n' encontramos que:

$$n' = \frac{(1.96)^2 P(1 - P)}{(0.02)^2} \quad (6.2)$$

De principio, podrían obtenerse diversos tamaños muestrales, usando diferentes valores en la precisión y manteniendo constantes el nivel de confianza del 95% ($Z = 1.96$), una $P = 0.35$ y una $Q = 0.65$.

TABLA NO. 6.1
TAMAÑOS MUESTRALES PARA DIFERENTES VALORES
DE PRECISIÓN, MANTENIENDO CONSTANTES EL NIVEL
DE CONFIANZA (95%), UNA $P = 0.35$ Y UNA $Q = 0.65$

Precisión	Cálculo del tamaño muestral	Tamaño de n'
0.01	$n' = \frac{(1.96)^2 (.35)(.65)}{(0.01)^2}$	$n' = 8740$
0.02	$n' = \frac{(1.96)^2 (.35)(.65)}{(0.02)^2}$	$n' = 2185$
0.05	$n' = \frac{(1.96)^2 (.35)(.65)}{(0.05)^2}$	$n' = 350$
0.10	$n' = \frac{(1.96)^2 (.35)(.65)}{(0.10)^2}$	$n' = 88$

2.2 Proporción poblacional

De la fórmula anterior, podemos notar, que necesitamos tener alguna idea de la proporción poblacional, que puede ser aproximada mediante estudios anteriores, a través de alguna muestra piloto o simplemente y desde un punto de vista muy conservador, utilizando un valor máximo obtenido para PQ que será cuando $P = Q = 0.5$

Si la P sugerida está alrededor de 0.2 a 0.35 la elección más conservadora sería $P = Q = 0.35$, sustituyendo para n' tenemos que:

$$n' = (1.96)^2 (0.35)(0.65) / (0.02)^2 = 2185$$

Si este tamaño muestral fuera pequeño comparado con el tamaño poblacional, dejaríamos este como el tamaño muestral requerido; sin embargo, si este no es el caso, debemos tomar en cuenta el factor de corrección y corregir el tamaño muestral mediante:

$$n = \frac{n'}{1 + \frac{n'}{N}} = 1907 \quad (6.3)$$

Ejemplo 6.3 *Se puede hacer notar, que el máximo tamaño muestral será cuando P y Q son iguales, es decir $P=Q=0.5$, como veremos en la siguiente tabla, en la que se manejarán diversas P's y Q's manteniendo constantes la precisión y el nivel de confianza.*

TABLA NO. 6.2
TAMAÑOS MUESTRALES PARA DIFERENTES P's Y Q's,
MANTENIENDO CONSTANTES UNA PRECISIÓN DE 0.02
Y UN NIVEL DE CONFIANZA DEL 95% (Z= 1.96)

P	Q	Cálculo del tamaño muestral	Tamaño de la muestra (n')
0.1	0.9	$n' = \frac{(1.96)^2(.1)(.9)}{(0.02)^2}$	n' = 96
0.2	0.8	$n' = \frac{(1.96)^2(.2)(.8)}{(0.02)^2}$	n' = 1537
0.3	0.7	$n' = \frac{(1.96)^2(.3)(.7)}{(0.02)^2}$	n' = 2017
0.5	0.5	$n' = \frac{(1.96)^2(.5)(.5)}{(0.02)^2}$	n' = 2401
0.8	0.2	$n' = \frac{(1.96)^2(.8)(.2)}{(0.02)^2}$	n' = 1537

De la tabla anterior se desprende que el mismo tamaño muestral corresponde para P = 0.2 y Q = 0.8 que para P = 0.8 y Q = 0.2 ya que estas proporciones podrían corresponder para el primer caso a la proporción de las personas que usan loza vidriada y para el segundo a la proporción de los que no la usan y ya que uno es complemento del otro, da lo mismo hacer el cálculo para cualquiera de las dos proporciones.

2.3 Nivel de Confianza

Hasta aquí, hemos utilizado frecuente un nivel de confianza del 95%, es decir, nuestros intervalos de confianza o la determinación del tamaño muestral usa un valor de Z = 1.96, el cual determina este nivel de confianza.

Sin embargo, podríamos estar interesados en otros niveles de confianza, lo cual modificaría directamente en el tamaño muestral. Otros niveles de confianza usualmente utilizados son:

- Z= 1.645 para un nivel de confianza del 90%
- Z= 2 para un nivel de confianza del 95.45%
- Z= 2.33 para un nivel de confianza del 98%
- y Z= 2.575 para un nivel de confianza del 99%.

Ejemplo 6.4 Siguiendo el ejemplo anterior, podemos calcular algunos tamaños muestrales para cada uno de estos niveles, manteniendo constantes la precisión igual a 0.02, la P= 0.35 y la Q=0.65

TABLA NO. 6.3
TAMAÑOS MUESTRALES PARA DIFERENTES NIVELES
DE CONFIANZA Y MANTENIENDO CONSTANTES LA
PRECISIÓN IGUAL A 0.02, LA P=0.35 Y LA Q=0.65

Nivel de Confianza	Valor de Z	Cálculo del tamaño muestral	Tamaño de la muestra aprox.
90%	1.645	$n' = \frac{(1.645)^2(.35)(.65)}{(0.02)^2}$	n' = 1540
95%	1.96	$n' = \frac{(1.96)^2(.35)(.65)}{(0.02)^2}$	n' = 2185
95.45%	2.0	$n' = \frac{(2)^2(.35)(.65)}{(0.02)^2}$	n' = 2275
98%	2.33	$n' = \frac{(2.33)^2(.35)(.65)}{(0.02)^2}$	n' = 3088
99%	2.575	$n' = \frac{(2.575)^2(.35)(.65)}{(0.02)^2}$	n' = 3772

2.4 Otros tipos de diseños muestrales

Cuando nuestro diseño es otro que un muestreo aleatorio simple, será necesario modificar este tamaño muestral multiplicándolo por el efecto del diseño en cuestión con respecto al aleatorio simple. Este efecto de diseño no es más que la razón de la varianza del error estándar obtenido con el diseño complejo entre la varianza que se obtendría con un muestreo aleatorio simple.

Si se pensara utilizar un muestreo estratificado, se requeriría un tamaño muestral más pequeño que con el aleatorio simple, debido a las ganancias en precisión que es inherente al proceso de estratificación, es decir el efecto de diseño tendría un valor menor a 1; sin embargo, esta ganancia es pequeña cuando estimamos un porcentaje con una estratificación proporcional, de tal manera que la reducción del tamaño muestral será modesta.

Ejemplo 6.5 *Si el error estándar obtenido para el consumo de loza vidriada a partir de un muestreo aleatorio simple resulta ser 0.0102 (luego entonces la varianza del error estándar es de 0.000104) y de otros estudios anteriormente realizados, se sabe que el uso de loza vidriada es más común para el área rural, donde se encontró con un muestreo estratificado, una varianza del error estándar de 0.000091, el efecto de diseño resulta ser de:*

$$def = \frac{0.000091}{0.000104} = 0.875$$

por lo que si se decide efectuar un muestreo estratificado y el tamaño muestral para muestreo aleatorio simple fue de 1907, se necesitaría un tamaño de $(1907)(0.875) = 1669$ personas en total de ambos estratos; y podría ser el tamaño muestral para cada estrato proporcional al tamaño del estrato.

Por otro lado, cuando aplicamos un diseño de conglomerados multietápico, ya sea con probabilidad proporcional al tamaño u otro tipo de selección, implícitamente el proceso conlleva una homogeneización dentro de los conglomerados y por consiguiente una disminución en la precisión del intervalo, por lo que en este caso el efecto de diseño será mayor de 1 y por lo tanto el tamaño muestral deberá incrementarse con respecto al aleatorio simple.

Ejemplo 6.6 *En una prueba piloto con muestreo de conglomerados en dos etapas se encuentra un error estándar de 0.01724 (es decir la varianza del error estándar es de 0.0002974), y continuando con la varianza del*

error estándar para muestreo aleatorio simple de 0.000104 tenemos que:

$$\text{def} = \frac{0.0002974}{0.000104} = 2.859$$

por consiguiente, si calculamos un tamaño muestral para el MAS de 1907 personas y estuviéramos pensando en usar un muestreo bietápico por conglomerados, requeriríamos un tamaño muestral 2.859 veces mayor que el del aleatorio simple; es decir, $(1907)(2.859) = 5453$ aproximadamente.

2.5 Tasa de no respuesta

Otro factor a considerarse en el cálculo del tamaño muestral es la tasa de no respuesta, por ejemplo, si la tasa de respuesta es del 75%, el tamaño muestral necesario para alcanzar el requerido será para este ejemplo $1907/0.75 = 2543$. Por supuesto, este es un ajuste para alcanzar el tamaño muestral y de ninguna manera es una garantía para evitar un sesgo de no respuesta.

Del mismo modo, es importante tomar en consideración los costos de la encuesta relacionándolos con los niveles de precisión, pues resulta poco práctico efectuar encuestas a bajo costo para obtener intervalos muy amplios para obtener información.

3. OTRAS CONSIDERACIONES

Mientras que los ejemplos anteriores muestran los aspectos principales para el cálculo de la muestra, esto está muy simplificado, pues en la práctica las encuestas cumplen con varios objetivos, conteniendo un número sustancial de estimadores y estos a su vez, cubrirán un rango grande de subclases como serían regiones diferentes, diversos grupos de edad o múltiples grupos educacionales u ocupacionales, lo cual implica realizar varios cálculos para tamaños muestrales y se aceptará aquél que permita hacer divisiones finas para el análisis por subclases, por lo que en general, podemos decir que mientras mayor sea el tamaño muestral alcanzaremos una mayor precisión y se podrá conducir un análisis más detallado.

El tamaño muestral idóneo dependerá de los factores mencionados que junto con el aspecto financiero será aquel que nos permita realizar interpretaciones útiles de los intervalos obtenidos, por lo que resulta poco práctico efectuar cálculos con poca precisión o niveles de confianza

demasiado bajos debido a un financiamiento limitado, ya que el alcance de nuestras inferencias será pobre.

4. TAMAÑO DE MUESTRAS PARA LA COMPARACIÓN DE DOS MEDIAS O DOS PROPORCIONES EN DOS POBLACIONES EN ESTUDIOS TRANSVERSALES

Hasta aquí se ha incluido el diseño de encuestas por muestreo refiriéndose básicamente a una sola población. Cuando se trata de estudios comparativos que involucran dos poblaciones como en el caso de los diseños epidemiológicos transversales, de casos y controles, estudios de cohorte, ensayos clínicos, etc., es necesario hacer consideraciones referentes a los tipos de error (tabla No.6.4).

TABLA 6.4
TIPOS DE ERROR SEGÚN LA SITUACIÓN VERDADERA DE LA HIPÓTESIS NULA

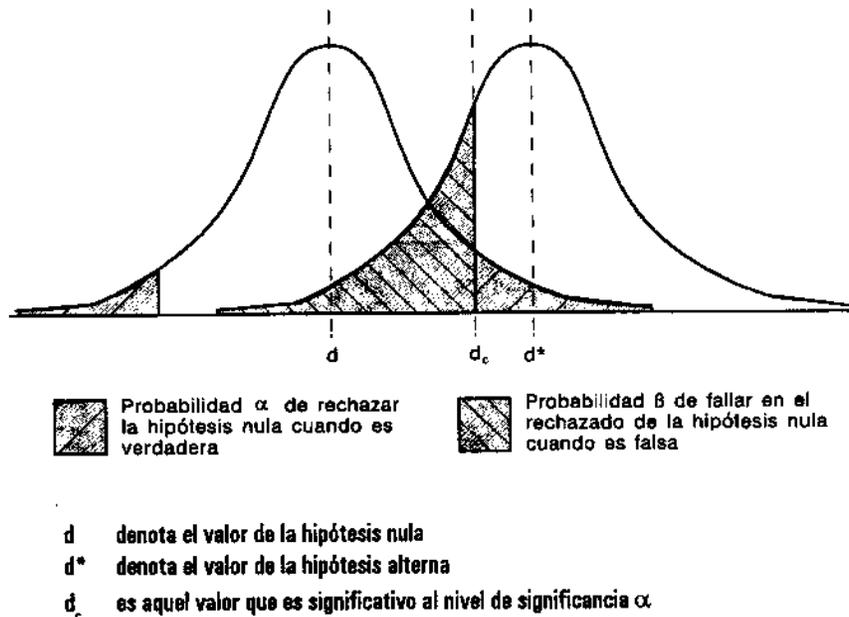
		Decisión hecha	
		H_0 no rechazada	H_0 rechazada
Situación	H_0 verdadera	Sin error ($1 - \alpha$)	Error Tipo I (α)
real	H_0 falsa	Error tipo II (β)	Sin error (Poder = $1 - \beta$)

4.1 Poder estadístico de la muestra ($1-\beta$) que es la capacidad de la misma para detectar una diferencia de los parámetros de las poblaciones cuando ésta existe y se expresa en términos de porcentaje, como el complemento del Error tipo II o Beta que no es más que la probabilidad de aceptar una hipótesis nula cuando ésta es falsa.

4.2 Nivel de significancia (α) que corresponde a la probabilidad de rechazar una hipótesis nula (h_0) cuando ésta es verdadera.

La figura 6.1 ilustra los principios básicos de estos conceptos. Si se postula una hipótesis nula verdadera, esto implica que no hay diferencia de los parámetros entre las dos poblaciones estudiadas, es decir la diferencia de los parámetros es igual a 0 ($d=0$), α sería la probabilidad de rechazar h_0 cuando es verdadera y d_c denota el valor de la diferencia que es significativo desde d a un nivel de significancia α . Por otro lado, cuando h_0 es falsa, esto es, existe diferencia entre los parámetros poblacionales, el valor real de la diferencia de las medias o de las proporciones estudiadas es diferente de 0 ($d^* \neq d$), entonces el área que queda a la izquierda de d_c representa la probabilidad de aceptar la hipótesis nula cuando ésta es falsa (Error tipo II o β) y el área a la derecha corresponde a la probabilidad de rechazar la hipótesis nula cuando ésta es falsa representando el poder de detectar la diferencia (o asociación).

FIGURA 6.1
DISTRIBUCIÓN MUESTRAL DE DIFERENCIA DE MEDIAS O PROPORCIONES
CUANDO LA HIPÓTESIS NULA ES VERDADERA Y CUANDO ES FALSA



Por consiguiente d_c puede ser expresado como $d + Z_{\alpha/2} \sigma_d$ o equivalentemente $d^* + Z_{\beta} \sigma_d$, donde:

$Z_{\alpha/2}$ es el valor de Z correspondiente a la posición de d_c según el nivel de significancia seleccionado, sobre la distribución alrededor de d y donde la diferencia de los parámetros puede ser positiva o negativa.

Z_β denota la posición de d_0 según el nivel de β seleccionado, sobre la distribución alrededor de d^* y σ_d y σ_d son los errores correspondientes de cada distribución.

Puesto que los valores de $Z_{\alpha/2}$ y Z_β son ampliamente utilizados, se presentan aquellos más usuales y para diferentes niveles.

TABLA 6.5
VALORES DE $Z_{\alpha/2}$ Y Z_β PARA DIFERENTES NIVELES DE ERROR.

Niveles de Error	$Z_{\alpha/2}$	Z_β
0.01	2.58	2.33
0.02	2.33	2.05
0.05	1.96	1.645
0.10	1.645	1.28

De este modo, tenemos la siguiente igualdad:

$$d + Z_{\alpha/2} \sigma_d = d^* + Z_\beta \sigma_d.$$

Debido a que los errores estándar son generalmente cercanos en valor, y mediante un poco de álgebra podemos resolver esta ecuación para obtener una buena aproximación del tamaño muestral para un estudio transversal:

Si se trata de diferencia de medias :

$$n = \frac{(Z_{\alpha/2} + Z_\beta)^2 \sigma^2 (r + 1)}{(d^*)^2 r}$$

Si es diferencia de proporciones:

$$n = \frac{(Z_{\alpha/2} + Z_\beta)^2 \bar{p} (1 - \bar{p}) (r + 1)}{(d^*)^2 r}$$

donde:

d^* Es el valor de la diferencia en proporciones o medias (es decir, la magnitud de la diferencia que el investigador desea detectar).

- n El número de individuos expuestos estudiados.
- r El cociente del número de individuos no expuestos estudiados, entre el número de individuos expuestos estudiados.
- σ Desviación estándar en la población para una variable continua.
- p_1 Proporción de individuos expuestos quienes tienen la consecuencia (enfermedad).
- p_0 Proporción de sujetos no expuestos quienes tienen la consecuencia (enfermedad).

$$\bar{p} = \frac{p_1 + rp_0}{1 + r} = \text{promedio ponderado de } p_1 \text{ y } p_0$$

Ejemplo 6.7 Para ilustrar el cálculo del tamaño muestral para diferencia de proporciones considere un estudio para evaluar si fumar durante el embarazo está asociado con un incremento en la prevalencia de bajo peso al nacer del recién nacido. Se conoce de estudios anteriores que por cada mujer que fuma durante el embarazo hay alrededor de 3 que no lo hacen, por lo tanto, en un estudio donde los sujetos se seleccionan aleatoriamente sin considerar los hábitos de fumar, encontramos una $r = 3$, una proporción de fumadoras del 26% y una proporción de no fumadoras del 75%. Además las encuestas reportan una prevalencia de bajo peso al nacer (es decir, recién nacidos de 2500 gr o menos) de 7%. Suponga que la diferencia a detectarse que se considera de interés entre los grupos expuestos en la proporción de recién nacidos con bajo peso al nacer es de 8% ($d^* = 0.08$).

Si el nivel de significancia es del 5%, entonces el valor correspondiente de Z es 1.96 y el poder es de 90% el valor que corresponde de Z es 1.28. Utilizando la fórmula para diferencia de proporciones tenemos que:

$$n = \frac{(1.96 + 1.28)^2(0.07)(0.93)(3 + 1)}{3(0.08)^2} = 142.4$$

Se estima que aproximadamente 143 recién nacidos de fumadoras tienen que ser incluidos en la muestra junto con 429 recién nacidos de no fumadoras, es decir, un tamaño global de 572 recién nacidos.

Ejemplo 6.8 *Suponga usted que se requiere calcular un tamaño muestral para realizar un estudio cuyo propósito es comparar el nivel de plomo en hueso entre los sujetos que usan loza de barro vidriado y aquellos que no lo usan. Por estudios anteriores, se conoce que la media de plomo entre aquellas personas que no lo usan es alrededor de 9.26 µg/g de hueso y la media de plomo entre aquellas personas que lo usan es de 16.01. Además se sabe que la varianza de plomo en hueso es de 103.18 y que por cada persona que usa loza vidriada hay otra que no lo usa, por consiguiente $r=1$. Si se quisiera detectar una diferencia de 3 unidades con un nivel de significancia del 5% y una potencia del 95%, el tamaño muestral requerido para este ejemplo sería calculado como sigue:*

$$n = \frac{(1.96 + 1.645)^2 103.18(2)}{(3)^2} = 238.4$$

De este modo, el tamaño requerido, resulta ser de 239 personas que usan loza vidriada y otro tanto que no lo usa, es decir, se requiere un tamaño global de 478 personas, bajo las especificaciones señaladas.

Los ejemplos anteriores se reducen a presentar el cálculo del tamaño muestral para estudios transversales, es decir, cuando las características bajo estudio son obtenidas en un solo momento; sin embargo, existen otros diseños epidemiológicos como casos y controles que requieren considerar definiciones específicas para un adecuado cálculo muestral.

5. ERRORES FRECUENTES EN LA ESTIMACIÓN DEL TAMAÑO MUESTRAL

5.1 El tamaño de la muestra depende principalmente del tamaño del universo. Es una práctica común elegir un tamaño muestral como un porcentaje de la población, por decir, el 10%. Esta decisión no es justificada, pues como se revisó, está más influenciada por la precisión, el nivel de confianza y la proporción de la característica más que por el tamaño de la población.

La eficiencia de la muestra depende más bien de su número absoluto y no del valor proporcional con respecto al universo.

5.2 Hacer toda una etapa de muestreo probabilístico y luego no hacer los cálculos correspondientes para el error estándar y para la construcción del Intervalo de Confianza, limitándose a la descripción del estimador muestral.

5.3 Hacer reemplazos no planeados puede alterar la estructura probabilística del muestreo al alterar la probabilidad de ciertas unidades, por lo que si se llegaran a efectuar deberá considerarse la probabilidad de cada elemento de quedar en la muestra.

5.4 Aplicar un diseño complejo para seleccionar la muestra y analizar como en un muestreo aleatorio simple. Este ha sido un error que se comete frecuentemente por la facilidad de las fórmulas del MAS, las cuales no toman en cuenta las variaciones de diseño, provocando errores en el cálculo del intervalo de confianza y por consiguiente errores en su interpretación.

EJERCICIOS

1. Calcule el tamaño muestral del ejemplo de plomo en mujeres, si la proporción encontrada hubiese sido del 30%. Utilice los datos de los tres investigadores incluidos en el cuadro. Comente cómo afecta este cambio el tamaño muestral.
2. Calcule un tamaño de muestra para la población de Pavarandocito, sabiendo que la desviación estándar del biomarcador en cuestión es de 564 Unidades Internacionales. Haga el cálculo con un nivel de confianza del 95% y del 99%. Precise usted el error que aceptaría y déjelo fijo en ambos cálculos.
3. Que ocurriría con la n anterior si el diseño es complejo y se espera una tasa de no respuesta del 20%.

CAPÍTULO VII

MARCOS MUESTRALES

M.C. Patricia Nájera Aguilar
Lic. Oscar Méndez Carniado
Lic. Leticia Cantoral Uriza
Tec. Tito Alejandro Badillo

1. INTRODUCCIÓN

En los estudios poblacionales la vivienda constituye la unidad de muestreo más accesible completa y eficiente.

La elaboración de un estudio que tenga a la vivienda como unidad última de muestreo es un proceso que consiste básicamente, en una selección de áreas geográficas definidas. Esta selección se hace subdividiendo y seleccionando áreas por etapas hasta seleccionar manzanas o conjuntos de manzanas y conjuntos equivalentes en zonas rurales.

En estas áreas deberá llevarse a cabo la elaboración de un listado de viviendas que permita obtener un registro completo de unidades de muestreo que se encuentran dentro de un área; pues excluir alguna vivienda significa eliminar la posibilidad de incluirla en el estudio. Por ello es necesario seguir las reglas y recomendaciones que se exponen en este manual.

Existen muchos elementos que intervienen en el buen desarrollo del trabajo de campo de una investigación. Uno de ellos, es el levantamiento del Marco Muestral puesto que permite localizar y delimitar de las áreas seleccionadas cuáles serán incluidas en el estudio.

La elaboración y uso adecuado del marco muestral permite mejorar la operatividad de todas las actividades relacionadas a la selección y aplicación de instrumentos de recolección de información durante el trabajo de campo, en especial, en la eficacia del levantamiento, en la supervisión, la verificación, la ejecución del control de avances y en la representatividad de la muestra.

Los propósitos de este capítulo son: Familiarizar al personal de campo con el manejo de cartas, croquis y con la simbología contenida en estos materiales, tanto para su elaboración, como para su uso en el levantamiento de trabajo de campo de alguna encuesta y presentar un ejemplo concreto del diseño de un marco muestral en un área específica.

2. CONDICIONES CARTOGRÁFICAS

En términos generales el levantamiento de un marco muestral es el mismo tanto para las áreas rurales como para las urbanas. Sin embargo, cada tipo de zona presenta características específicas (conformación y ubicación de las viviendas, rasgos topográficos, etc.) que requieren un tratamiento particular para su registro en los formatos correspondientes.

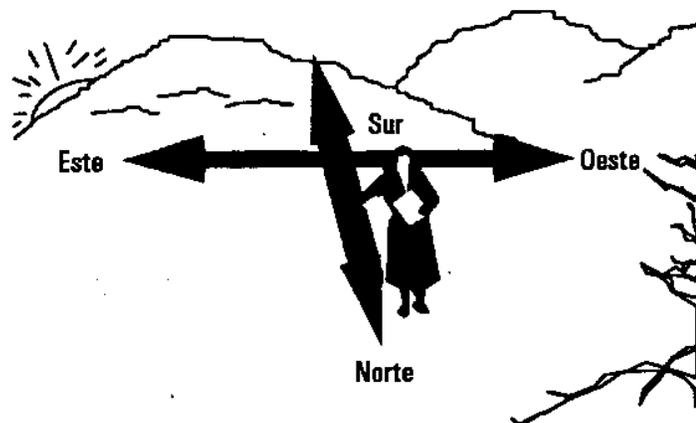
La elaboración del listado consiste en el registro de los datos de identificación de las unidades de vivienda o unidades muestrales, así como del diseño de un croquis donde se ubican dichas unidades. Estas tareas son fundamentales para que el listado de viviendas resulte operativo.

Al llegar al área el listador debe verificar sus límites señalados previamente en un mapa o croquis, mediante su verificación. Puede darse el caso de que los límites del área, en dicho recorrido de verificación, sean calles o caminos identificados por sus nombres los cuales están sujetos a modificaciones que repercuten de manera directa en los límites del área de listado. Una vez conocidas las características de los mapas y sabiendo distinguir, leer e identificar todos los detalles que hay en ellas, es muy importante la ubicación de la persona y la carta en el terreno de trabajo.

Para lograr una buena ubicación, antes que nada, debemos orientarnos correctamente: Sabemos que el sol siempre saldrá por el Este (Oriente), si nos ponemos de tal forma que nuestro brazo derecho esté del lado por donde sale el sol, el Norte nos quedará siempre hacia el frente, con esto logramos orientarnos en el terreno.

Para orientar la carta, bastará con poner su Norte — indicando con una flecha — hacia el frente de nosotros o Norte del terreno (ver fig. 7.1).

Figura 7.1



La forma de lograr la orientación en el mapa es por inspección y reconocimiento de los elementos geográficos. Esto requiere de un cuidadoso examen del mapa y del terreno para encontrar elementos lineales comunes a ambos lados como: caminos, vías de ferrocarril, líneas de transmisión o de conducción etc; haciendo coincidir los elementos simbolizados en el mapa con los del terreno, procurando reconocer el mayor número de detalles posibles para lograr una posición apropiada de la orientación carta-persona-terreno.

Deberá continuarse con la observación general de la zona de trabajo, tratando de ubicar en la carta puntos geográficos grandes (sierras, montañas, cerros, grandes cañones, barrancas, etc.), que tengamos a la vista buscando que coincidan las direcciones de estos puntos en el terreno con la carta.

Logrando esto, debemos dedicar nuestra atención a la observación de todos los detalles cercanos (sean naturales u obras del hombre), barrancas, arroyos, cerros pequeños, carreteras, vías de ferrocarril, terracería, brechas, tanques elevados, etc; ubicándolos de acuerdo a sus direcciones y estimando sus distancias por medio de pasos o en su defecto, haciendo un cálculo aproximado.

Por último, debemos verificar el punto exacto en el que nos encontramos en el terreno con la fotografía, verificar todos los detalles generales que hayan sido observados, debiendo estos coincidir en dirección y distancia.

IMPORTANTE: Cuando tratamos de ubicar la carta con respecto al terreno natural y no es posible ver una zona amplia, debemos movernos hacia los puntos que sí se ven en la carta (sin ir muy lejos), para checar su ubicación exacta, considerando direcciones y distancias. Esto deberá realizarse por lo menos cuatro veces para ubicar otros tantos puntos.

Una vez efectuado el recorrido de los límites y con la seguridad de encontrarse en el área de listado designado, al listador debe iniciar la elaboración del listado de viviendas. Lo importante es remarcar que, por ningún motivo se puede iniciar la elaboración del listado si antes no se verifica los límites del área. Al final del capítulo, para quien desee profundizar en el tema, se presentan las fuentes cartográficas más usuales (anexo 7.1, página 93).

3. DEFINICIONES

A continuación se presenta la definición de los términos que más comúnmente se usan en la elaboración de marcos muestrales.

3.1 Cartografía

Es el conjunto de métodos, técnicas y procedimientos necesarios para representar gráficamente un espacio geográfico determinado a través de: mapas, planos y croquis con el objeto de facilitar la planeación, organización y ejecución de los censos y las encuestas.

3.2 Localidad

Es todo lugar de un municipio ocupado por una vivienda o conjunto de viviendas que están cercanas unas de otras, generalmente tiene un nombre dado por la ley o la costumbre y se puede localizar por sus vías de acceso. Las localidades se dividen en urbanas y rurales.

La localidad urbana cuenta con 2,500 o más habitantes y la localidad rural tiene menos de 2,500 habitantes.

3.3 Municipio

Conjunto de localidades que conforman una unidad geopolítica y administrativa independiente.

3.4 Jurisdicción Sanitaria

Forma organizativa que por razones administrativas agrupan municipios para otorgar los servicios de salud.

3.5 Manzana

Es una superficie que forma parte de una localidad (urbana o rural) con límites perfectamente identificables como son calles, avenidas, andadores, ríos, barrancas, etc. Puede tener una o más viviendas agrupadas o dispersas, así como lotes baldíos y locales destinados para diferentes fines: escuelas, fábricas, oficinas, iglesias, etc.

3.6 Croquis de listado

Instrumento cartográfico diseñado para identificar, de manera general, la ubicación de las manzanas donde se encuentran las viviendas.

3.7 Listado

Es el registro completo de los datos que permitan identificar (ubicar) dentro de un área de trabajo seleccionada todos los locales que se utilicen como viviendas particulares en el momento de realizar el listado.

3.8 Vivienda

Es el espacio delimitado por paredes o techos de cualquier material de construcción, donde viven habitualmente una o más personas, es decir, en donde duermen, preparan sus alimentos, comen y se protegen del medio ambiente. Además, la entrada a la vivienda debe ser independiente, de tal modo que sus ocupantes puedan entrar o salir de ella sin pasar por el interior de otra vivienda. Las viviendas se dividen en dos clases: **particulares y colectivas**.

3.8.1 Vivienda particular. Está destinada al alojamiento de una o más familias, grupos de personas o una sola persona y puede estar habitada o deshabitada en el momento de realizar el listado. Estas son las viviendas que son registradas para conformar el conglomerado.

3.8.2 Vivienda colectiva. Es aquella destinada a servir de alojamiento habitual a personas relacionadas por un objetivo público o algún interés personal común, esto es, razones de salud, disciplina, enseñanza, religión, trabajo asistencial, alojamiento asistencial o militar. Algunas de las clases de viviendas colectivas son: pensión, casa de huéspedes, casa de asistencia, asilos, internados, conventos, cárceles, etc.

Es importante resaltar que las pensiones, casas de huéspedes y asistencia que tengan 5 o menos pensionados, serán consideradas como vivienda particular. En caso de las viviendas rentadas por estudiantes que comparten el pago de servicios sin importar el número siempre será considerada como particular.

Hasta aquí se deduce que cualquier local que se utilice para alojamiento habitual de personas, aunque no haya sido construido para ser habitado (bares, escuelas, bodegas, fábricas o talleres) debe ser considerado como vivienda.

3.9 Viviendas para el listado

En el momento de realizar el listado, las viviendas particulares pueden tener las siguientes condiciones:

3.9.1 Viviendas habitadas. Aquellas que en el momento de realizar el listado sirven de alojamiento habitual a familias, grupos de personas o una sola persona.

3.9.2 Viviendas deshabitadas. Aquellas que han sido construidas para servir de habitación pero que en el momento de realizar el listado no están habitadas por persona alguna.

En el momento de realizar el listado puede encontrar viviendas que se desconozca si están habitadas o no; **investigue con los vecinos para asegurarse si son una o más viviendas y si está(n) habitada(s) o no.**

Como hemos visto hasta aquí, las encuestas por muestreo requieren una selección de unidades de muestreo y que según la información disponible, el presupuesto y el tiempo nos determina la elección del tipo de muestreo a utilizarse.

Para los muestreos aleatorio simple, sistemático y estratificado, basta tener un listado de todas las unidades de muestreo para poder seleccionar la muestra. Sin embargo, si se carece de tal listado, o la población se encuentra muy dispersa, resulta más eficiente recurrir a los muestreos complejos, donde las unidades muestrales se van seleccionando en múltiples etapas y en donde sólo se requiere el listado de unidades muestrales que se van seleccionando en cada etapa.

Ejemplo 7.1 *Suponga que estamos interesados en estudiar a niños de 1 a 4 años para cuantificar el número cuadros diarreicos sufridos en un período, asociados con el nivel de exposición de un contaminante biológico en agua.*

Para esta investigación no se cuenta con un listado de niños y la población se encuentra muy dispersa y sólo se dispone de la información proporcionada por el Censo Nacional de Población y Vivienda y mapas por jurisdicción y planos por manzanas proporcionados por el Instituto Federal Electoral, por lo que se recurre a un muestreo complejo donde la unidad última de muestreo será la vivienda.

Aspectos generales

El área geográfica comprende tres jurisdicciones que presentan diferentes condiciones climáticas, políticas, socioculturales y de disponibilidad de servicios, por lo que se iniciará estratificando por jurisdicción. Estas a su vez, abarcan un total de 33 municipios. El

tamaño muestral fue estimado de 1 540 niños de 1 a 4 años. A través del estudio piloto se encontró que hay un niño de esta edad por cada dos viviendas visitadas.

Etapa 1. Selección de unidades en la primera etapa de muestreo (municipios).

Usando la información del censo y un mapa, se ordenan geográficamente de norte a sur los municipios de cada jurisdicción y se enlistan junto con el número de habitantes en cada municipio.

Se conforman conglomerados de 50 viviendas, conteniendo alrededor de 250 personas en cada uno de ellos (el censo reporta un promedio de 4.6 personas por vivienda) y se obtiene el número de conglomerados por municipio dividiendo el total de habitantes en el municipio entre 250, así para el primer municipio de la jurisdicción número 1 hay $6,313/250 = 25$ conglomerados en ese municipio y se incluyen en el listado junto con la suma acumulada de conglomerados, tal como se muestra en la tabla 7.1

**TABLA 7.1
LISTADO DE CONGLOMERADOS POR MUNICIPIO**

Jurisdicción y Municipios	Número de habitantes	Número de Conglomerados	Conglomerados acumulados ¹
Jurisdicción 1			
1.1 Municipio	6,313	25	1 al 25
1.2 Municipio	16,506	66	26 al 91 **
1.3 Municipio	60,492	242	92 al 333****
:	:	:	:
:	:	:	:
Jurisdicción 2			
2.1 Municipio	30,010	120	1,398 al 1,518
2.2 Municipio	20,009	80	1,519 al 1,598
:	:	:	:
:	:	:	:
Jurisdicción 3			
3.1 Municipio	4,921	20	2,007 al 2,026
3.2 Municipio	6,868	27	2,027 al 2,053
:	:	:	:
:	:	:	:
Total			3,282

¹ El número de asteriscos corresponde al número de conglomerados seleccionados en cada municipio.

Como cada conglomerado que se visitará tiene 50 viviendas, se aportará aproximadamente con 25 niños estudiados por lo que se requiere visitar $1,540/25 = 62$ conglomerados para cubrir el tamaño muestral requerido ($n = 1,540$)

Mediante un muestreo sistemático con probabilidad proporcional al número de conglomerados se elegirán los municipios que deberán ser visitados, de la siguiente manera:

Se calcula un intervalo de selección dividiendo el total de conglomerados entre el número a visitarse, $3,282/62 = 53$; se busca ahora un arranque aleatorio entre 1 y 53 y suponga que se encuentra el 39, el cual es identificado en la columna de conglomerados acumulados, que en este caso cae en el municipio 1.2 (ver tabla 7.1); a este 38 se le suma el intervalo de selección $38 + 53 = 91$ y se vuelve a buscar en la lista de conglomerados acumulados, encontrándose ubicado nuevamente en el municipio 1.2, al 91 se le vuelve a sumar 53 y tenemos $91 + 53 = 144$ el cual cae en el municipio 1.3 los cuales son identificados en la tabla con asteriscos (dos para el municipio 1.2 y cuatro para el municipio 1.3) y así sucesivamente hasta localizar los municipios en los que se encuentran los 62 conglomerados elegidos.

Puede ser factible que por el tamaño de los municipios, puedan caer uno o más conglomerados o por el contrario, no caiga ningún conglomerado, respetando la probabilidad proporcional del tamaño del municipio.

Etapas 2. Selección de unidades en la segunda etapa de muestreo (localidades).

De acuerdo al número de unidades seleccionadas por municipio se procede a seleccionar las localidades en una forma similar a la anterior, agrupando localidades vecinas que por lo menos conformen un conglomerado, es decir, que contengan 50 viviendas. Posteriormente se enlistan para identificar las localidades que serán visitadas.

TABLA 7.2
LISTADO DE CONGLOMERADOS POR LOCALIDAD

Municipio y localidades	Número de habitantes	Número de Conglomerados	Conglomerados acumulados
Municipio 1.2			
Localidad 1.2.1	2,070	8	1 al 8
Localidad 1.2.2	103		
Localidad 1.2.3	412	2	2 al 10
Localidad 1.2.4	75		
Localidad 1.2.5	72		
Localidad 1.2.6	80		
Localidad 1.2.7	86	1	11
Localidad 1.2.8	1,048	4	12 al 15
Localidad 1.2.9	1,795	7	16 al 22
:	:	:	:
:	:	:	:
:	:	:	:

Etapa 3. Selección de unidades en la tercera etapa de muestreo (manzanas).

Al tener determinadas las localidades que serán estudiadas en los municipios se procede a seleccionar las manzanas que servirán para conformar los conglomerados de viviendas. Esta selección será mediante los planos proporcionados por el Instituto Federal Electoral, en los que aparezca la localidad fraccionada en secciones electorales, conteniendo el nombre de las calles, avenidas, privadas, etc., y puntos de referencias como iglesias, escuelas, plazas, mercados, etc. Así como un número progresivo de identificación asignado sistemáticamente en espiral a partir del centro de la localidad hacia las orillas.

Para la selección de manzanas se obtiene el número total de las mismas y se divide entre el número de conglomerados a visitarse, finalmente se eligen sistemáticamente mediante un arranque aleatorio.

Por último se identifica al conglomerado de viviendas que se refiere al registro sistemático de todos los datos que permitan identificar y ubicar los lugares que se utilizan como vivienda y que son identificados a partir de las viviendas seleccionadas.

Con tal fin se diseña una hoja de croquis de manzana en la que se incluye:

- Datos de Identificación: municipio, localidad y colonia.
- Número de identificación de la manzana.
- Número de conglomerado.
- Croquis. En el que aparecerá el nombre de las calles que lo delimitan, rasgos específicos (barrancas, callejones, escuelas, iglesias, comercios) y las viviendas identificadas por número.
- Listado de viviendas del conglomerado con los datos que identifican a cada vivienda numerada (calle, número exterior, número interior).

Realización de listados de viviendas

Con ayuda del plano por secciones se identifica el nombre de las calles que circundan la(s) manzana(s) y se verifica recorriendo la zona y tomando en cuenta las referencias como escuelas, mercados, iglesias, comercios, etc.

Para integrar los listados de vivienda del conglomerado se debe realizar un recorrido sistemático, por lo que es necesario estandarizarlo bajo los siguientes criterios:

- El recorrido de la(s) manzana(s) debe iniciarse en la esquina Noroeste o en el punto más próximo a esta esquina.
- El recorrido se hará de acuerdo a sentido en que giran las manecillas del reloj, es decir, si nos paramos de frente a la vivienda la que continúa la localizaremos a la izquierda.
- Se registran únicamente los locales que se ocupen como vivienda particular, teniendo cuidado de no omitir ninguna. Hay locales comerciales que pueden estar usando como vivienda.
- Si al completar el recorrido de la primera calle se listaron 50 viviendas o más, se termina el listado, en caso contrario se continúa con la(s) siguiente(s) calle(s).
- Cuando la manzana seleccionada no tiene viviendas o la cantidad no llega a 50, se toma la manzana que aparece al norte de la seleccionada y se continúa con el listado. De necesitarse más manzanas, se eligen de acuerdo a como giran las manecillas del reloj, tomando como centro la manzana seleccionada.

Con estos croquis de los conglomerados seleccionados, los encuestadores podrán ubicarse y recorrer todas las viviendas que lo constituyen a fin de recabar la información necesaria.

El ejemplo anterior, es sólo uno de los que en forma práctica se podrían utilizar, pero se debe tener en cuenta que existen otras múltiples combinaciones e información cartográfica y general que podría ser muy útil para conformar un marco muestral adecuado.

ANEXO 7.1

FUENTES CARTOGRÁFICAS PARA LA ENCUESTA (MAPAS Y CROQUIS)

Se requiere contar con una cartografía que contenga información suficiente para la identificación de viviendas a encuestar. Las fuentes cartográficas producidas en nuestro país, en primer lugar, la producida por el INEGI, y en segundo lugar, los croquis elaborados por trabajadores comunitarios o específicamente por el personal de apoyo muestral de cada investigación.

El material del INEGI es, principalmente, el de cartas topográficas escala 1:50000, con información confiable en cuanto a rasgos naturales, pero muchos de los rasgos culturales no están actualizados. Otro inconveniente es la escala, ya que la identificación de viviendas es muy difícil por la escasa aproximación al terreno.

La información por croquis de comunidad carece de escala y orientación adecuada además de que la distribución de viviendas y límites físicos son muy subjetivos y poco identificables en el terreno.

Por los problemas mencionados se plantea la necesidad de actualizar, verificar y, en algunos casos, modificar la información contenida en la cartografía o en su defecto producir nuevo material.

Interpretación del material cartográfico

- A) **Rasgos o Detalles Culturales.** Son aquellas obras hechas por el hombre, como son carreteras, caminos, escuelas, pueblos, ciudades, etc.
 - a) **Vías Terrestres.** Son obras hechas por el hombre para comunicar una población con otra, o algún sitio de interés, éstas pueden ser carreteras pavimentadas, terracerías, brechas, veredas o vías de ferrocarril.
 - b) **Líneas de Conducción.** Son las obras que sirven para llevar de un lugar a otro agua, petróleo, gas y energía eléctrica, así como comunicación telegráfica y telefónica.
 - c) **Otros rasgos culturales.** Son obras como puentes, minas, pozos de agua, pozos petroleros, ruinas, bancos de material (arena, grava, etc.), túneles, pasos a desnivel,

presas, etc., que pueden servir de auxiliares a otras obras, o bien, proporcionar servicios y recursos.

- d) **Asentamientos Humanos.** Es un grupo de casas que se encuentran en un lugar fijo, conservando un ordenamiento regular. En cartas para áreas rurales, los lugares más poblados como ciudades o pueblos grandes, se ven con las calles bien trazadas además tienen obras que dan servicios a la población como: hospitales, escuelas, panteones, etc.

Las casas aisladas aparecen en el mapa como pequeños cuadritos negros. Las casas son obras hechas por el hombre con el objeto de protegerse de las inclemencias del tiempo o desarrollar alguna otra actividad.

B) Rasgos o Detalles Naturales. Son aquellos en los que no interviene la mano del hombre como montañas, cerros, ríos, arroyos, pantanos, etc.

- a) **El Relieve.** Es la forma del terreno y se representa en el mapa por medio de líneas que reciben el nombre de curvas de nivel, que indican si es plano o montañoso, siguen su forma a determinada altura sobre el nivel del mar. Estas líneas son de 2 tipos:
 - i) **Curvas de Nivel Acotadas.** Son gruesas y siempre tienen un número que nos permite conocer la altura o profundidad del terreno con respecto al mar.
 - ii) **Curvas del Nivel Ordinarias.** Son líneas más claras y delgadas que se encuentran entre 2 curvas del nivel, son acotadas y sirven de auxilio a éstas para obtener valores intermedios.

Cuando las líneas se encuentran muy juntas indican que el terreno es más quebrado y cuando están muy separadas, que el terreno es más plano. Es muy importante no confundir las curvas de nivel con otras líneas como las que representan las corrientes de agua, veredas o brechas. Las corrientes de agua casi siempre cortan a las curvas de nivel y donde las corrientes de agua cortan tienen ligeros quiebres, las brechas o veredas son más rectas y unen puntos de interés para el hombre como pueblos, minas, ríos, etc.

Una curva de nivel siempre se cerrará, esto es, si seguimos una curva de nivel partiendo de un punto fijo, la curva siempre llegará al punto inicial. Estas curvas muchas veces se cortan en las orillas del mapa, pero si se juntan con los mapas siguientes, es posible encontrar la unión.

- b) **Rasgos o Detalles Hidrográficos.** Son las corrientes y depósitos de agua como ríos, arroyos, lagunas y terrenos sujetos a inundación. Las corrientes y depósitos de agua pueden ser perennes o intermitentes.

- i) Son perennes, cuando llevan agua en cualquier época del año.
- ii) Son intermitentes cuando llevan agua sólo en una época del año.

Ríos. Son cauces formados por el agua que baja de los arroyos. Los ríos y arroyos perennes, se ven en la carta como líneas continuas y los intermitentes como líneas cortadas, cuando el río o arroyo tienen nombre, está escrito sobre las líneas que los representan.

Corrientes que desaparecen. Cuando la línea que nos señala la corriente finaliza en una flecha, ésta indica que la corriente termina en ese lugar, que llegan a un lugar y no forman cauces.

Manantiales. Son brotes naturales de agua, que pueden escurrir o formar depósitos. Se representan en mapas con un punto y una línea ondulada.

C) **Escala.** Es la relación que hay entre cualquier distancia representada en el mapa y su correspondiente en el terreno real. Para saber la equivalencia de centímetros, en cualquier escala, sólo recorra dos dígitos de derecha a izquierda.

CAPÍTULO VIII

EL CUESTIONARIO

Es indudable que el puro tema del cuestionario merecería un manual aparte, sin embargo no podemos terminar éste sin hacer algunas consideraciones generales sobre el mismo.

Los objetivos de este capítulo son:

- Mencionar las características generales de un cuestionario.
- Ubicar la importancia de un buen diseño de cuestionario.

Estrategias: Se mencionarán las características generales del cuestionario y se presentarán dos ejemplos reales para su análisis.

Ejercicio: Elaborar un cuestionario sencillo, a manera de ejemplo, sobre un tema de interés.

Bajo la fórmula genérica de formularios, se incluye toda forma impresa destinada a la recolección de datos. En el caso específico de las encuestas, el formulario recibe el nombre de cuestionario.

1. CUESTIONARIO Es una forma impresa diseñada para ayudar en el proceso de recolección de la información y es por ello un instrumento de medición que se aplica a través de una secuencia de preguntas.

El cuestionario debe ser planeado cuidadosamente en tal forma que facilite la obtención de los datos. Debe ayudar a recoger la información de manera completa, eficiente, uniforme y evitando la recolección de datos inútiles o irrelevantes al estudio.

El cuestionario debe permitir recoger dos clases de datos:

- Generales, administrativos o de información.
- Datos sobre el problema en estudio.

Los primeros ayudarán a identificar las unidades en observación, e incluyen nombre, edad, sexo, residencia habitual, fecha de estudio, etc. A través de ellos puede medirse el rendimiento del personal y la distribución de algunas características generales.

En cuanto a los segundos se refieren al tema en estudio y deben estar en perfecto acuerdo con los objetivos planteados en el mismo.

Elaborar un cuestionario es, a nuestro modo de ver una de las actividades más difíciles y que requieren de mayor conocimiento sobre el tema que se esté estudiando. Hay sin embargo lineamientos generales que pueden orientar la construcción del mismo y que veremos a continuación. El lector interesado en profundizar en el tema puede consultar algunos de los libros que se mencionan en la bibliografía de los cuales recomendamos especialmente el texto de Babbie (Babbie, 1993).

2. ELABORACIÓN DEL CUESTIONARIO

Antes de elaborar el formulario deben considerarse: El propósito para el cual será utilizado y las circunstancias bajo las cuales se recogerá la información. Lo primero tiene importancia para decidir sobre los datos que se incluirán y lo segundo para la adopción del tamaño, forma y material más convenientes. Aún sin pretender dar reglas fijas, se recomienda:

2.1 Decidir los datos que se recogerán: Un cuestionario muy extenso conspira contra la buena recolección de los datos y de ahí que sólo deban recogerse aquéllos útiles y pertinentes, en ese sentido es útil:

- Hacer una lista de todos los datos "deseables" de recoger según la naturaleza del estudio.
- Considerar cuáles son "factibles" de recoger de manera fidedigna. Puede ser deseable conocer la exposición de la madre del sujeto, a sustancias tóxicas durante el embarazo, pero puede ser casi imposible si sólo nos basamos en el recuerdo del entrevistado.
- Considerar la relación costo-beneficio de cada una de las preguntas.

2.2 Decidir el orden en el cual se colocarán los datos en el formulario. La distribución debe ser de manera lógica y ordenada. En general debe comenzarse con los datos de identificación general para continuar con los relativos al asunto en investigación.

2.3 Dar un espacio adecuado a cada pregunta y utilizar en lo posible preguntas cerradas que permitan la elaboración de un cuestionario precodificado, el cual va a facilitar la codificación y la captura de los datos.

2.4 Cada pregunta debe tener todas las alternativas posibles de respuesta. Debe considerarse siempre la inclusión de rubros como: no sabe, no contesta, otros. Las opciones de respuesta deben ser excluyentes a menos que se señale la opción de escoger más de una respuesta.

2.5 El lenguaje debe ser claro, conciso e inequívoco en cuanto a su comprensión. Las preguntas negativas deben evitarse. Por ejemplo: ¿No ha asistido a cursos sobre uso de equipo de protección en el trabajo? Las preguntas "tendenciosas" o que inducen la respuesta también deben evitarse. Por ejemplo: ¿no es verdad que usar el equipo de protección en el trabajo nos protege?

2.6 Procurar el empleo de preguntas autoexplicativas. Las preguntas que requieren de "instructivo" ya sea para el entrevistador o el entrevistado han demostrado requerir más tiempo, dinero y mucho entrenamiento y supervisión.

2.7 Determinar quién recogerá la información, de quién, cómo, dónde y cuándo, permite decidir sobre la forma, tamaño, material y hasta el color de los formularios. Por ejemplo, el material depende del empleo. Si es un cuestionario amplio que se aplicará en el campo requiere un material más resistente que uno que solamente consignará un dato de laboratorio. El color puede ser indicativo de alguna circunstancia especial o motivar la atención del entrevistado.

2.8 Al planear el formato del cuestionario debe tenerse siempre en mente el método de análisis que se usará.

2.9 Es muy importante el uso correcto de los llamados "pases", tanto para el entrevistador, como para el entrevistado, según sea el método de aplicación que será empleado. Por ejemplo, si hay una serie de preguntas en relación al tabaquismo, habrá una inicial que permita distinguir a los fumadores. Para los no fumadores habrá un "pase" que permita omitir preguntas subsecuentes de detalles sobre la antigüedad y la frecuencia del hábito, mismas que sí se harán a aquéllos que hayan respondido ser fumadores y que pueden ser: ¿Desde hace cuántos años fuma? ¿Cuántos cigarrillos por día?

2.10 Finalmente un cuestionario debe siempre llevar un instructivo y/o párrafos introductorios. El instructivo debe ser lo más sencillo posible como ya se mencionó. Todo cuestionario debe ser sometido a una prueba piloto antes de su aplicación general.

Ejercicios:

- 1. Examine y comente el cuestionario que se anexa sobre Vigilancia Epidemiológica de plomo en el binomio madre-hijo (anexo 8.1).**
- 2. Examine y comente el tipo de preguntas que se anexa de un cuestionario sobre prevalencia de asma y enfermedades alérgicas en niños escolares (anexo 8.2) (19).**
- 3. Elabore un cuestionario breve y conciso en relación a un tema de su área de trabajo.**

EJERCICIO 1

ESTUDIOS DE PLOMO DURANTE LA LACTANCIA 1993 - 1995

CUESTIONARIO PARA LA VIGILANCIA EPIDEMIOLÓGICA DE EXPOSICIÓN Y NIVELES DE PLOMO EN SANGRE DURANTE EL EMBARAZO Y EL POSPARTO INMEDIATO

En este hospital estamos realizando un estudio sobre contaminación por plomo en mujeres y sus hijos.

Quisiéramos INVITARLA A PARTICIPAR EN EL ESTUDIO porque usted es una buena candidata para ello.

Su participación es voluntaria y consistirá en contestar preguntas a unos cuestionarios de su exposición a plomo y de su dieta. Además, le solicitaremos una muestra de sangre para medir sus niveles de plomo. Le daremos los resultados sin ningún costo.

Darle la carta de consentimiento.

¿ACEPTA USTED?

SI 1)

NO 2)

0. ¿Firmó la carta corta de consentimiento?

SI 1)

NO 2)

NÚMERO DE EXPEDIENTE DEL HOSPITAL _____

1. NOMBRE DE LA MADRE: _____

2. CLAVE DEL HOSPITAL _____

INPER 1) IMSS 3)

GEA GONZALEZ 2) OTRO 4)

3. NOMBRE DE LA ENTREVISTADORA: _____



<p>4. FECHA DE HOY DIA _____ MES _____ AÑO _____</p>	<p>DD/ MM/ AA <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></p>
<p>5. (DIAGNÓSTICO ACTUAL. ESTE RUBRO LO MARCA Y LO FIRMA EL MÉDICO) TACHE TODAS LAS OPCIONES QUE APLIQUEN (PERO CODIFIQUE MÁXIMO 2. SI SOLO HAY UN DIAGNÓSTICO, DEJE EN BLANCO EL SEGUNDO):</p> <p>NINGUNO (00)</p> <p>EMBARAZO MÚLTIPLE (01)</p> <p>ALTERACIONES MENTALES (02)</p> <p>CARDIOPATÍAS (03)</p> <p>NEFROPATÍAS (04)</p> <p>HISTORIA DE INFECCIONES URINARIAS REPETIDAS (05)</p> <p>HISTORIA DE FORMACIÓN DE CÁLCULOS RENALES (PIEDRAS DEL RIÑÓN) (06)</p> <p>HISTORIA DE DIABETES GESTACIONAL (07)</p> <p>HISTORIA DE PRE-ECLAMPSIA (08)</p> <p>LA SEÑORA TOMA ANTICONVULSIVOS (09)</p> <p>LA SEÑORA TOMA CORTICOIDES (10)</p> <p>ENFERMEDADES GASTROINTESTINALES GRAVES (ÚLCERA, COLITIS ULCERATIVA) (11)</p> <p>HIPERTIROIDISMO (12)</p> <p>HIPERTENSIÓN ARTERIAL: SISTÓLICA MÁS DE 140mm/Hg (13) DIASTÓLICA MÁS DE 90mm/Hg (14)</p> <p>NO TIENE LA INTENCIÓN DE DARLE PECHO A SU HIJO (15)</p> <p>OTROS _____ (16)</p> <p style="text-align: center;">especificar</p>	<p><input type="checkbox"/><input type="checkbox"/></p> <p><input type="checkbox"/><input type="checkbox"/></p>
<p>6. Domicilio actual de la madre CALLE _____ No. _____ COLONIA _____ DELEGACIÓN POLÍTICA _____ CÓDIGO POSTAL _____</p>	<p><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></p> <p><input type="checkbox"/><input type="checkbox"/></p>
<p>7. Teléfono de familiares o vecinos (subrayar de quién): _____ _____ del trabajo: _____</p>	<p><input type="checkbox"/><input type="checkbox"/></p>
<p>8. Edad gestacional (en semanas) _____ (semanas) (puede ser edad gestacional de embarazo o estimada por examen físico del recién nacido).</p>	<p><input type="checkbox"/><input type="checkbox"/></p>

9. Edad de la madre en años cumplidos: _____ (años)	<input type="checkbox"/>
10. ¿Cuántas veces ha estado embarazada? (contando este embarazo) _____ (embarazos)	<input type="checkbox"/>
11. ¿Cuántos años aprobó en la escuela? _____ (años aprobados)	<input type="checkbox"/>
12. ¿Cuál es su estado civil? CASADA 1) UNIÓN LIBRE 2) SOLTERA 3) SEPARADA 4) DIVORCIADA 5) VIUDA 6) NO CONTESTO 7)	<input type="checkbox"/>
13. Usa usted utensilios de barro como los de la fotografía, para preparar o guardar alimentos? SI 1) —————> Pase a la 15 NO 2)	<input type="checkbox"/>
14. ¿Cada cuando utiliza estos utensilios? DIARIO (1) CADA TERCER DÍA (2) SOLO EL FIN DE SEMANA (3) UNA VEZ AL MES (4) NO RECUERDO (5)	<input type="checkbox"/>
15. ¿Durante su embarazo tomaba usted leche? SI 1) NO 2) —————> Pase a la 17	<input type="checkbox"/>
16. ¿Cuántos vasos? UNO POR DÍA 1) MÁS DE UNO POR DÍA 2) UNO CADA TERCER DÍA 3) UNA VEZ POR SEMANA 4) NO RECUERDO 5)	<input type="checkbox"/>

17. ¿Durante su embarazo tomaba alguna vitamina que tuviera calcio?

SI 1)

NO 2) → Pase a la 19

18. ¿Cuántas pastillas?

UNO POR DÍA 1)

MÁS DE UNA POR DÍA 2)

UNA CADA TERCER DÍA 3)

UNA VEZ POR SEMANA 4)

NO RECUERDA 5)

19. Presión arterial _____ SISTÓLICA

_____ DIASTÓLICA

(La debe tomar la enfermera)

SISTÓLICA

DIASTÓLICA

20. PLOMO EN SANGRE: (_____)

o

MUCHAS GRACIAS POR SU COOPERACIÓN

ANEXO 8.2

EJERCICIO 2

PREGUNTAS SOBRE PREVALENCIA DE ASMA	
Folio:	□□□□
I. Identificación	
Nombre del Niño: _____	
II. Información de Salud	
25. En los últimos 12 meses, ¿su niño(a) ha tenido tos seca en forma repetida (que tosa por la mañana, por la tarde o por la noche) por más de tres semanas?	□□
() Sí	
() No (pase a la pregunta 31)	
26. ¿A qué hora tose más?	□□□
() Mañana	
() Tarde	
() Noche	
27. En los últimos 12 meses, ¿cuántas veces en promedio se ha despertado su niño durante la noche a causa de un ataque de tos?	□□□
() Nunca	
() Menos de una noche por semana	
() 1 o más noches por semana	
28. En los últimos 12 meses, ¿alguna vez su niño ha tenido un ataque de tos que impidiera su manera de hablar?	□□
() Sí	
() No	

<p>35. En los últimos 12 meses, ¿cuántas veces en promedio se ha despertado su niño durante la noche a causa de silbidos en el pecho?</p> <p>() Nunca () Menos de 1 noche por semana () 1 o más noches por semana</p> <p>36. En los últimos 12 meses, ¿alguna vez su niño ha tenido períodos de silbidos en el pecho que impidieran su manera de hablar?</p> <p>() Sí () No</p> <p>37. En los últimos 12 meses, ¿su niño ha tenido períodos de silbidos en el pecho después de haber jugado mucho o haber hecho ejercicio?</p> <p>() Sí () No</p> <p>38. En los 12 últimos meses su niño ha tenido período de silbido en el pecho debido a:</p> <p>(1) Clima frío (5) Humedad (9) Infección (2) Polvo al barrer (6) Humano de leña (10) Ejercicio (3) Plantas (7) Animales (4) Alimentos (8) Calor</p> <p>39. En los 12 últimos meses, ¿a su niño le ha silbado el pecho sin que tenga resfriado o catarro?</p> <p>() Sí () No</p>	<p><input type="checkbox"/><input type="checkbox"/></p> <p><input type="checkbox"/><input type="checkbox"/></p> <p><input type="checkbox"/><input type="checkbox"/></p> <p><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></p> <p><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></p> <p><input type="checkbox"/><input type="checkbox"/></p>
---	---

<p>40. Marque usted los meses en que su niño tuvo silbidos del pecho en los últimos 12 meses (puede señalar más de uno si es necesario)</p>	
(1) Enero (5) Mayo (9) Septiembre	□□□
(2) Febrero (6) Junio (10) Octubre	
(3) Marzo (7) Julio (11) Noviembre	□□□
(4) Abril (8) Agosto (12) Diciembre	
<p>41. ¿Alguna vez su hijo ha necesitado medicinas para silbidos en el pecho?</p>	□□
() Sí	
() No	□□
<p>42. ¿En los 12 últimos meses ha necesitado medicinas para silbido de pecho?</p>	□□
() Sí	
() No	
<p>Cuál(es) _____</p> <p>_____</p>	
<p>43. ¿Alguna vez le ha dicho un doctor que su hijo tiene asma?</p>	□□
() Sí	
() No	□□
<p>44. ¿Alguna vez le ha dicho un doctor que su hijo tiene bronquitis asmática?</p>	□□
() Sí	
() No	□□□
<p>45. ¿Alguna vez un médico le ha diagnosticado asma a algún miembro de su familia?</p>	□□□
La mamá del niño (sí) (no) (no sabe)	□□□
El papá del niño (sí) (no) (no sabe)	□□□
Algún hermano o hermana (sí) (no) (no sabe)	□□□

<p>46. ¿Alguna vez su hijo ha sido atendido en el servicio de urgencias por silbidos en el pecho?</p> <p>() Sí () No</p>	<input type="checkbox"/> <input type="checkbox"/>
<p>47. ¿Cuánto han interferido estos problemas de silbidos con las actividades diarias de su hijo?</p> <p>a. Alguna vez en el pasado</p> <p>() Nada () Moderadamente () Mucho</p> <p>b. En los 12 últimos meses</p> <p>() Nada () Moderadamente () Mucho</p>	<input type="checkbox"/> <input type="checkbox"/>
<p>48. Durante los últimos 12 meses, ¿cuántas veces ha faltado a la escuela por problemas de silbido de pecho?</p> <p>() Nunca () 1-3 () 3-5 () Más de 5</p>	<input type="checkbox"/> <input type="checkbox"/>
<p>49. ¿Ha cambiado su forma de ser, se ha vuelto enojón, triste, pelionero, sin amigos?</p> <p>a. Alguna vez en el pasado</p> <p>() Nada () Moderadamente () Mucho</p> <p>b. En los últimos 12 meses</p> <p>() Nada () Moderadamente () Mucho</p>	<input type="checkbox"/> <input type="checkbox"/>
	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

APÉNDICE 1

ESTUDIO DE CASO No. 1

1. ASPECTOS GENERALES

1.1 Problema de estudio

Determinar la prevalencia de asma y los factores de riesgo para la misma en niños de 6 y 12 años de edad.

1.2 Consideraciones

Es una población donde el 95% de los niños asisten a las escuelas.

Por normas de la Secretaría de Educación Pública, el ingreso a primer año escolar debe hacerse a los seis años.

Esto hace que los niños de este universo se encuentren ubicados en primero y sexto año escolar.

La Ciudad de México tiene una zona industrial con gran emisión de contaminantes al aire, los cuales se distribuyen en forma heterogénea a través de la ciudad.

2. PASOS A SEGUIR

2.1 Cálculo del tamaño muestral

Para ello es necesario tener una idea de la proporción del padecimiento en la población. Se consulta a los médicos alergólogos de la ciudad y se revisa la literatura existente, infiriendo que la prevalencia puede estar alrededor del 15%.

Los investigadores desean un Intervalo de Confianza del 95% ($z = 1.96$) y desean que la proporción estimada no se desvíe más de tres puntos porcentuales alrededor de la verdadera

proporción en el Universo ($d=0.03$). No les parece necesario hacer una corrección para población finita y estiman que la tasa de respuesta se encuentra alrededor del 80%.

Se aplica la fórmula que vimos en el Capítulo VI:

$$n = \frac{z^2 pq}{d^2}$$
$$n = \frac{3.84 * 0.15 * 0.85}{0.03 * 0.03}$$
$$n = 544$$

Debido a que se trata de un diseño muestral complejo es necesario realizar un ajuste tomando en consideración el efecto del diseño (4). Suponiendo que este efecto es igual a 1.68 calculamos un tamaño de:

$$544 * 1.68 = 914$$

a este número le agregan un 10% por lo mencionado acerca de la tasa de respuesta:

$$n = 914 + 92 = 1006$$

Los investigadores estiman que su presupuesto les alcanza para este número de niños por lo cual no hacen otro tipo de ajuste.

2.2 Consecución o construcción de un marco muestral propio para los objetivos del estudio.

Mediante visita a la Secretaría de Educación Pública, los investigadores se enteran de que ésta posee una base de datos acerca de las escuelas primarias en la ciudad, tanto públicas como privadas.

Por actividades de gestoría interinstitucional logran conseguir una copia de esta base de datos, misma que contiene la siguiente información:

- Listado de escuelas primarias.
- Ubicación.
- Edad y sexo de los alumnos.
- Número de alumnos por grado escolar.
- Nombre del director de la escuela.

A partir de esta base, construyen un marco muestral propio que incluye los datos generales de la escuela y los referentes a los niños de primero y sexto de primaria.

De acuerdo a los datos del universo y a la opinión de expertos sobre la distribución de los factores de riesgo para asma en la ciudad, los investigadores deciden hacer un diseño mixto de muestreo: Primero dividir en estratos según los niveles de contaminación ambiental y posteriormente hacer un muestreo de conglomerados monoetápico, en cada uno de los estratos. Se definen como conglomerados las escuelas.

2.3 El marco muestral queda integrado por 137 escuelas (conglomerados), las cuales se ubican en un plano general de la ciudad, mismo que dividen en cuatro grandes estratos según la diferencia en factores de riesgo a consideración de un experto en el ramo. Esto permite tener una idea general de la distribución para proceder al muestreo.

2.4 Afijación del tamaño de n por estratos. Esto se hace en una forma muy sencilla ya que se precisa a qué porcentaje del universo corresponde la población de cada estrato y este mismo porcentaje se aplica a la muestra.

2.5 Determinación del número de conglomerados a muestrear en cada estrato.

Para ello se obtiene la media de alumnos por conglomerado (m) y se estima cuántos conglomerados hay que tomar para obtener la parte de n que corresponde a ese estrato. En este caso no se fija un número inflexible por conglomerado, sino que se decide tomar todos los niños incluidos en esa escuela, como corresponde a un muestreo de conglomerados.

2.6 Una vez decidido el número de conglomerados por estrato, estos se escogen aleatoriamente mediante MAS o ME.

2.7 Estrategias adicionales. Una vez seleccionadas las escuelas (conglomerados), se procede a la ubicación exacta de las mismas con un plano más detallado de la ciudad. Asimismo, cuando se realice la visita se deben llevar dos cartas previamente redactadas, una para el director de la escuela y otra para enviar a los padres junto con el cuestionario del estudio. Posteriormente se organiza una reunión con los maestros de primero y sexto de primaria y se entregan los cuestionarios correspondientes para su llenado por parte de los padres.

APÉNDICE 2

ESTUDIO DE CASO No. 2

1. ASPECTOS GENERALES

1.1 Problema de estudio

Ante la ausencia de una encuesta probabilística nacional que permita estimar la media de la concentración de plomo para los grupos de riesgo materno y de recién nacido, se plantea la necesidad de realizar, al menos una exploración de los valores.

Como no existe tampoco un programa de Vigilancia Epidemiológica de los niveles de plomo en este grupo, se propone el auto establecimiento del mismo.

1.2 Consideraciones

El binomio madre-hijo constituye un grupo de riesgo para los efectos de la intoxicación por plomo.

La detección de los niveles de plomo en este grupo, permitirá formarse un criterio acerca de lo que está ocurriendo en la población general.

La Ciudad en la cual se desea realizar el estudio, es muy populosa con una historia de tráfico vehicular intenso, con normatividad muy reciente en cuanto a la disminución de la concentración de plomo en las gasolinas.

2. PASOS A SEGUIR

2.1 Delimitación del Universo

Por consideraciones de carácter técnico se prefiere estudiar a los binomios madre-hijo, cuyo embarazo y parto hayan cursado normalmente. Se hace un análisis del número de partos normales que se atienden en la Ciudad y se observa que el 80% de ellos se concentra en nueve

grandes hospitales, el 20% restante se atiende de manera dispersa en unidades pequeñas y aún en casa.

Estos nueve hospitales pertenecen a servicios de Seguridad Social y población abierta en la siguiente forma:

- 01 al 04: Hospitales de Gineco-obstetricia del Instituto Mexicano del Seguro Social. Atienden población trabajadora en empresas privadas y sus derechohabientes (55%).
- 05 y 06: Hospitales de Gineco-obstetricia del Instituto de Seguridad Social para Trabajadores al Servicio del Estado y sus derechohabientes (10%).
- 07 al 09: Hospitales de Gineco-obstetricia de la Secretaría de Salud, los cuales atienden a población que carece de Seguridad Social, conocida como "Población Abierta" (15%).

Se define entonces como universo a:

- Mujeres con embarazo normo-evolutivo, residentes de la Ciudad x, derechohabientes de Seguridad Social o usuarios de los Servicios de Salud, que acuden para atención del parto a uno de los hospitales mencionados y cuyo trabajo de parto curse en forma normal.
- Niños hijos de las madres mencionadas en dicho parto.

2.2 Establecimiento de la Capacidad

Más que un cálculo del tamaño de muestra, lo que se establece es el número de muestras que es posible tomar, procesar, almacenar y leer en los laboratorios que poseen las técnicas necesarias y que están inscritos en un programa de control de calidad.

Este número de muestras posibles de leer, diariamente, se reporta proporcionalmente entre los nueve hospitales inscritos en el programa.

2.3 Elaboración del Programa

Se establecen claramente los objetivos, estrategias actividades etc., pero además muy concretamente los criterios de inclusión y exclusión para protocolizar el programa en forma homogénea en los nueve hospitales.

Ejemplos de Criterios:

Criterios de inclusión

- **Aceptar participar voluntariamente.**
- **Tener 35 semanas o más de embarazo.**
- **No presentar ninguna de las siguientes patologías (diagnosticadas):
Diabetes, Hipertensión, Neurópata, Epilepsia, Toxemia Gravídica.**
- **Vivir en el área metropolitana de la Ciudad.**
- **No tener embarazo múltiple.**
- **No tener impedimento físico o mental para comunicarse y/o cooperar.**
- **Criterio de Exclusión.**
- **Tener alguno de los diagnósticos mencionados.**
- **Incapacidad para la comunicación verbal.**

2.4 Selección de los Binomios

Este es un estudio de participación voluntaria, con una tasa de aceptación alta pero no del 100%, por ello no se propone la selección sistemática (no puede programarse).

Debido a las condiciones de apoyo que brindan las Instituciones participantes, la selección se restringe al horario matutino (7-14 hrs.), en días hábiles de lunes a viernes.

A todas las madres que cumplan los criterios de inclusión se les informa del estudio y se les propone la participación. A todas las pacientes que acepten se les aplica un cuestionario que contiene datos de información sociodemográfica y exposición a fuentes de plomo y se les toma una muestra de sangre venosa. Se toma además una muestra de sangre del cordón umbilical.

Cuando se cumpla la cuota de muestreo de cada día, se suspende el procedimiento (puede ser una diaria, 2, 3, 4, dependiendo del volumen de partos normales que atienda el hospital).

2.5 Concentración y análisis de la información

Las muestras se recogen por Hospital Gubernamental y se envían al laboratorio, los resultados se envían para su concentración en la dependencia gubernamental correspondiente (Dirección de Epidemiología, Ministerio de Salud del País), donde los técnicos e investigadores deberán realizar el análisis de los datos para su difusión a las autoridades y a las Instituciones participantes.

APÉNDICE 3

GUÍA PARA EL USO Y RESOLUCIÓN DE LOS EJERCICIOS

CAPÍTULO IV

1. Se debe profundizar en las ventajas y desventajas de cada uno de los tipos de muestreo revisados en el capítulo. Debe hacerse una discusión acerca de las ventajas y limitaciones de los tipos de muestreo no probabilístico en el área de Salud Ambiental.
2. El alumno debe explicar el procedimiento y anotar los 50 elementos con su valor correspondiente.
3. Se trata de que el alumno pueda repetir el procedimiento seguido en el capítulo así:
 - MAS: Explicar cómo lo hace, ejemplificando solamente con 10 elementos (puesto que $n = 800$, no se pide todo).
 - MS: Explicar cómo calcula la K y cómo busca el arranque aleatorio.
 - ME, MC y mixto: Describir esquemáticamente.
4. Describir marcos muestrales propios de su medio.
5. Desarrollar la discusión en grupo para analizar las similitudes y divergencias de las diferentes propuestas.

CAPÍTULO V

1. Al realizar este ejercicio debe destacarse la magnitud de la diferencia en los resultados, cuándo se usa el término de corrección para población finita y discutir en qué casos es útil usarlo.
2. Destacar cómo aumenta el intervalo de confianza (se pierde precisión) al incrementar el porcentaje de confiabilidad.

3. Tiene el mismo objetivo de la pregunta no. 2, pero para porcentajes (características cualitativas). El fenómeno es el mismo.
4. Reforzar la comprensión del significado del intervalo de confianza y ver en qué forma se afecta por el uso del factor de corrección.
5. Abrir la discusión en grupo sobre qué tan útiles fueron los datos muestrales para estimar el parámetro, utilizando para ello el "Universo" de la tabla 4.2

CAPÍTULO VI

1. Reforzar la comprensión de los factores que afectan el tamaño muestral, en forma similar a como se desarrolla en el capítulo, pero realizando cada alumno sus propios cálculos. En este ejercicio se ve el efecto del factor: Proporción de la característica en estudio.
2. En el mismo sentido que el ejercicio no. 1 para comprender el efecto del factor: Nivel de confianza.
3. Reforzar el concepto de la influencia del efecto de diseño y de la tasa de no respuesta.

CAPÍTULO VIII

1. Discutir el cuestionario a la luz de los elementos suministrados en el desarrollo del capítulo.
2. Analizar preguntas concretas sobre datos de salud.

Tanto en la pregunta no. 1, como en la no. 2 deben analizarse aspectos de fondo y de forma: Diseño, precodificado, espacios para las preguntas, planteamiento de la pregunta, comentario acerca de los pasos sugeridos o faltantes.

3. Se trata de evidenciar en la práctica las dificultades inherentes a la elaboración de un formulario y el conocimiento del tema que debe tenerse. Se sugiere trabajo en equipos chicos (2 a 3 participantes) y discusión grupal posterior.

BIBLIOGRAFÍA

1. Babbie ER. *Métodos de Investigación por Encuesta*. México: Fondo de Cultura Económica. Biblioteca de la salud, 1993.
2. Daniel WW. *Bioestadística. Base para el análisis de las Ciencias de la salud*. México: LIMUSA, 3a. ed. 1992.
3. Fisher AA, Laing JE, Stoeckel JE and Townsend JW. *Manual para el Diseño de Investigación Operativa en Planificación Familiar*. México: The Population Council. 2a ed. 1991.
4. Kalton Graham. *Introduction to survey sampling*. Series Quantitative Applications in the Social Sciences. London: SAGE publications 10th ed. 1991.
5. Kelsey JL, Thompson WD, Evans AS. *Methods in Observational Epidemiology*. Oxford University Press, 1986.
6. Kish L. *Survey Sampling*. New York: John Wiley, 1965.
7. Levin J. *Fundamentos de Estadística en la Investigación Social*. México: Harla. 2a. ed. 1977.
8. Levy P and Lemeshow S. *Sampling for Health Professionals Lifetime Learning Publications*. London, 1980.
9. Lwanga SK and Lemeshow S. *Sample Size Determination in Health studies. A Practical Manual*. Geneva: WHO. 1991.
10. Polit D and Hungler B. *Investigación en Ciencias de la Salud*. 3a ed. Washington, 1990.
11. Sheaffer R L , Mendenhall W and Ott L. *Elementos de Muestreo*. México: Grupo Editorial Iberoamericano. 1987.
12. Silva Aycaguer L. *Muestreo para la Investigación en Ciencias de la Salud*. España: Díaz de Santos, 1993.
13. Stata Reference Manual, 5th ed. Sta. Mónica, Ca. USA, 1992.
14. Stuart A. *Basic Ideas of Scientific Sampling*. New York: Hafner Publishing Company. 1968.
15. SUDAAN Survey Data Analysis Software. Version 5.31. November: Copyright Research Triangle Institute, 1990.
16. Taro Y. *Statistics, An Introductory Analysis*. New York: Hoper & Row Publishers. 1975.
17. UNEP and WHO. *Guidance on Survey Design for Human Exposure Assessment Locations (Health) Studies*. Nairobi, 1993.
18. USA. Department of Commerce. *Curso Suplementario para un Estudio de Caso Sobre Encuestas y censos*. Conferencias sobre muestreo AID. 1980.
19. ISAAC (International Study of Asthma and Allergies in Childhood). Manual Averland, NZ, 1992.

Otro material de consulta

1. Guidelines on studies in environmental epidemiology published under the joint sponsorship of the United Nations Environment programme, the International Labour Organization, and the World Health Organization. World Health Organization; Geneva, 1983. 351 p. (Environmental Health; criteria 27).
2. *Lead published under the joint sponsorship of the United Nations Environment Programme and the World Health Organization.* World Health Organization; Geneva, 1973. 160 p. (Environmental Health; Criteria 3).
3. Beaglehole R, Bonita R, Kjølström T. *Basic epidemiology.* Geneva: World Health Organization; 1993.
4. Briggs D, Corvalán C, Nurminen N. *Linkage methods for environment and health analysis.* Geneva: United Nations Environment Programme, United States Environmental Protection Agency, Environmental Health WHO Office of Global and Integrated. 1996. (WHO/EHG/95.26).
5. Human exposure assessment series. *Human exposure to lead.* Geneva: WHO Office of Global and Integrated Environmental Health, United Nations Environment Programme, United States Environmental Protection Agency.
6. Smith, Kirk R. Human Exposure Assessment Series. *The potential of Human Exposure Assessment for Air Pollution Regulation.* Geneva: WHO Collaborating Center for Studies of Environmental Risk and Development; 1995. (WHO/EHG/95.9).

ISBN 92 75 32200 7

ECO/OPS

Apartado postal 37-473

06696 México, D.F. MÉXICO

Tels.: (52-72) 71-10-86; 71-10-91; 71-10-92; 71-10-93

Fax.: (52-72) 71-10-90

internet: cpehs@servidor.unam.mx

y

ECO_Email@paho.org

Oficinas: Rancho Guadalupe sin número,
C.P. 52140. Metepec, Estado de México.