

# Capítulo 7. DISEÑO DE MUESTREO

El diseño de fuentes de información es el tercer paso de una investigación econométrica. El investigador debe distinguir dos tipos de fuentes de información que puede aplicar a su trabajo econométrico; las fuentes de información primaria y las fuentes de información secundaria. Las primeras justifican el estudio de las técnicas de muestreo. Hasta aquí el investigador conoce qué problema debe investigar y cómo plantear su tabla de datos. Debe tener en cuenta que se trata de un problema económico que para investigarlo debe fundamentarse en la teoría de la inferencia estadística desarrollada en los capítulos anteriores. Generalmente, el investigador posee datos de fuentes de información secundaria, que debe recolectar, procesar y organizar como se verá en la cuarta parte de este libro. Si el investigador no cuenta con información secundaria para su trabajo, deberá realizar un relevamiento aplicando los métodos y tipos de muestreo que se desarrollan en este capítulo.

---

## 7.1. Fuentes secundarias de información

El proceso de investigación econométrica, que permite la modelización empírica de los fenómenos económicos observados, no es un simple proceso de ajuste que se realiza de forma automática dada una teoría y un conjunto de números, sino que constituye el instrumento que permite el progresivo aumento en la acumulación de conocimiento sobre la realidad económica.

“No obstante, explican Espasa y Cancelo (1993) es cierto que, siendo la Economía una ciencia para explicar una realidad que se va observando continuamente, el trabajo empírico no se toma en serio en la profesión –véase la queja unánime que sobre este punto hacen Hendry, Leamer y Poirier et al (1990), sección 2.2 – ni mucho menos constituye el camino por el que mayoritariamente los economistas van acumulando opiniones. Las razones pueden ser muchas, como la discrepancia entre la teoría de cómo hacer el trabajo empírico y el tipo de trabajo empírico que realmente se hace; o también, porque no se dispone de patrones claros para juzgar el trabajo empírico” (p.p. 24-25).

Así, elevar la disertación de la metodología econométrica a un nivel en el que haya patrones, sobre los cuales se pueda concluir que un determinado trabajo empírico está bien hecho y es conveniente, constituye uno de los principales objetivos de este libro y de los economistas cuantitativos de la actualidad.

En cuenta de ello, el trabajo económico más sólido es el que se basa en la correcta aplicación del proceso de investigación econométrica, el cual aparece respaldado por los datos. Estos, en general, se encuentran en fuentes secundarias de información, la cual deberá ser chequeada antes de ser aplicada. Un buen “chequeo” de la información proveniente de fuentes

secundarias debe incluir un análisis pormenorizado de las variables que se incluyen, tendencia, integración, ploteo, entre otros.

Con frecuencia, el investigador parte de la idea de que una encuesta es la única manera de recolectar datos para un proceso de investigación econométrica. En realidad, solo debe utilizarse una encuesta, y por lo tanto un diseño de información primaria basado en un diseño de muestreo, si los datos no pueden recolectarse empleando fuentes de información más eficientes. En consecuencia, el primer paso en el diseño de las fuentes de información consiste en determinar si los datos ya existen.

Los *datos secundarios* se pueden definir como aquellos datos publicados que se recolectaron con propósitos distintos de las necesidades específicas de la investigación econométrica que se está desarrollando.

Esta información secundaria puede provenir de dos fuentes principales. Las fuentes sindicalizadas y las fuentes bibliográficas. Las primeras son provistas por empresas que se dedican al relevamiento y sistematización de bases de datos a requerimiento de clientes; estas fuentes son caras y raramente se usan en un trabajo econométrico. Las bibliográficas incluyen un conjunto de divulgaciones que circulan públicamente; a sobremano luego de la cultura web. La principal ventaja de los datos secundarios es el ahorro de tiempo y dinero en comparación con las fuentes de datos primarios. Por supuesto, habrá que ver el grado en que estos datos se ajustan a las necesidades de información del proyecto y la exactitud de los mismos.

### **Organización y homogeneización de los datos.**

Puesto que los datos secundarios se recolectan con propósitos diferentes a los del proyecto de investigación que se está llevando a cabo, rara vez satisfarán adecuadamente las necesidades del mismo. Esto da lugar a que deban realizarse ajustes, los que están influenciados por los siguientes factores:

- Las unidades de medición
- La dimensión geográfica que miden
- La actualidad de la publicación

Con respecto a las unidades de medición, al trabajar con datos reales provenientes de fuentes secundarias o del contexto económico, es habitual en Econometría tener variables medidas con diferente periodicidad. Por ejemplo, el PBI medido en forma trimestral y el Índice de Precios medido mensualmente.

Al incorporarlas en un modelo econométrico donde la unidad de observación es el tiempo, las variables deben referir a la misma unidad de tiempo. Por ejemplo, hay situaciones en las que es necesario contar con el dato del PBI medido en forma mensual. Para esto es necesario construir la serie mensual a partir de los datos trimestrales.

Un criterio que permite construir la serie mensual supone que las variaciones dentro del trimestre se producirán de manera constante. Los pasos a seguir consisten en obtener el coeficiente de variación trimestral y, a este convertirlo en su equivalente mensual. Luego, se asigna el valor de un trimestre al último mes del trimestre y se le aplica a este el coeficiente de actualización mensual para obtener el mes siguiente.

**Ejemplo.** La Figura 7.1 muestra el PBI medido en millones de pesos corrientes entre el primer trimestre de 1998 y el cuarto trimestre de 1999. Para convertir la serie en datos mensuales se hará lo siguiente:

- 1) Se le asigna al último mes del trimestre el valor del trimestre (Figura 7.2)
- 2) Se calcula la variación entre dos trimestres (Figura 7.3)

$$\frac{\text{junio}'98}{\text{marzo}'98} = \frac{312129}{282764} = 1.10384985$$

- 3) Se calcula la variación mensual a partir de la trimestral (Figura 7.3)

$$(1.10384985)^{1/3} = 1.03348299$$

- 4) Se calculan los valores del PBI mensual (Figura 7.4)

$$\text{abril}'98 = \text{marzo}'98 * 1.03348299 = 292232$$

$$\text{mayo}'98 = \text{abril}'98 * 1.03348299 = 302017$$

$$\text{junio}'98 = \text{mayo}'98 * 1.03348299 = 312129$$

| Periodo | PBI    |
|---------|--------|
| 1998 I  | 282764 |
| II      | 312129 |
| III     | 305475 |
| IV      | 295425 |
| 1999 I  | 270746 |
| II      | 288830 |
| III     | 285087 |
| IV      | 289429 |

NOTA: PBI a precios de mercado en millones de pesos corrientes.  
FUENTE: Ministerio de Economía de la Nación. [www.mecon.gov.ar](http://www.mecon.gov.ar)

**Figura 7.1. PBI trimestral**

| Periodo    | PBI    |
|------------|--------|
| 1998 Marzo | 282764 |
| Junio      | 312129 |
| Setiembre  | 305475 |
| Diciembre  | 295425 |
| 1999 Marzo | 270746 |
| Junio      | 288830 |
| Setiembre  | 285087 |
| Diciembre  | 289429 |

NOTA: PBI a precios de mercado en millones de pesos corrientes.  
FUENTE: Elaboración propia. Calculado a partir de datos del Ministerio de Economía de la Nación. [www.mecon.gov.ar](http://www.mecon.gov.ar)

**Figura 7.2. PBI mensual**

| Periodo | PBI       | Variación trimestral | Variación mensual |
|---------|-----------|----------------------|-------------------|
| 1998    | Marzo     | 282764               |                   |
|         | Junio     | 312129               | 1.10384985        |
|         | Setiembre | 305475               | 0.97868189        |
|         | Diciembre | 295425               | 0.96710042        |
| 1999    | Marzo     | 270746               | 0.91646272        |
|         | Junio     | 288830               | 1.06679323        |
|         | Setiembre | 285087               | 0.98704082        |
|         | Diciembre | 289429               | 1.01523044        |

NOTA: PBI a precios de mercado en millones de pesos corrientes. Variaciones expresadas en tanto por uno.

FUENTE: Elaboración propia en base a datos de PBI del Ministerio de Economía de la Nación. [www.mecon.gov.ar](http://www.mecon.gov.ar)

**Figura 7.3. Tasas de variación del PBI**

| Periodo | PBI trimestral | Variación trimestral | Variación mensual | PBI mensual |
|---------|----------------|----------------------|-------------------|-------------|
| 1998    | Marzo          | 282,764              |                   | 282764      |
|         | Abril          |                      |                   | 292232      |
|         | Mayo           |                      |                   | 302017      |
|         | Junio          | 312,129              | 1.10384985        | 312129      |
|         | Julio          |                      |                   | 309895      |
|         | Agosto         |                      |                   | 307677      |
|         | Setiembre      | 305,475              | 0.97868189        | 305475      |
|         | Octubre        |                      |                   | 302088      |
|         | Noviembre      |                      |                   | 298738      |
|         | Diciembre      | 295,425              | 0.96710042        | 295425      |
|         | Enero          |                      |                   | 286958      |
|         | Febrero        |                      |                   | 278734      |
| 1999    | Marzo          | 270,746              | 0.91646272        | 270746      |
|         | Abril          |                      |                   | 276645      |
|         | Mayo           |                      |                   | 282672      |
|         | Junio          | 288,830              | 1.06679323        | 288830      |
|         | Julio          |                      |                   | 287577      |
|         | Agosto         |                      |                   | 286329      |
|         | Setiembre      | 285,087              | 0.98704082        | 285087      |
|         | Octubre        |                      |                   | 286527      |
|         | Noviembre      |                      |                   | 287974      |

NOTA: PBI a precios de mercado en millones de pesos corrientes. Variaciones en tanto por uno.

FUENTE: Elaboración propia en base a datos trimestrales suministrados por el Ministerio de Economía de la Nación. [www.mecon.gov.ar](http://www.mecon.gov.ar)

**Figura 7.4. Cálculo de PBI mensual a partir del PBI anual**

**Ejemplo.** La Figura 7.5 tiene información sobre el Índice de Precios al Consumidor del Gran Buenos Aires, elaborado por el INDEC en forma mensual; para transformarlo en trimestral, una alternativa es calcular el promedio de los índices de tres meses.

| Año  | Mes | IPC mensual | Año  | Trimestre | IPC trimestral |
|------|-----|-------------|------|-----------|----------------|
| 1998 | 1   | 101.6677    | 1998 | I         | 101.861333     |
| 1998 | 2   | 102.0222    |      |           |                |
| 1998 | 3   | 101.8941    |      |           |                |
| 1998 | 4   | 101.9056    |      | II        | 101.921333     |
| 1998 | 5   | 101.8320    |      |           |                |
| 1998 | 6   | 102.0264    |      |           |                |
| 1998 | 7   | 102.3453    |      | III       | 102.3485       |
| 1998 | 8   | 102.3659    |      |           |                |
| 1998 | 9   | 102.3343    |      |           |                |
| 1998 | 10  | 101.9599    |      | IV        | 101.7939       |
| 1998 | 11  | 101.7178    |      |           |                |
| 1998 | 12  | 101.7040    |      |           |                |

NOTA: Índice de Precios al consumidor del Gran Buenos Aires. Nivel general serie base 1999=100  
FUENTE: Elaboración propia en base a datos del INDEC. [www.indec.gov.ar](http://www.indec.gov.ar)

**Figura 7.5. Cálculo del Índice de Precios trimestral a partir del mensual**

## Problema de exactitud

Una seria limitación de los datos secundarios tiene que ver con la dificultad de evaluar la exactitud, lo cual tiene que ver con si los datos son exactos para el propósito de la investigación que se está realizando. Para evaluar la exactitud habrá que chequear:

- Fuente
- Propósito de la publicación
- Evidencia relacionada con la calidad

Una regla fundamental en la utilización de datos secundarios es la obtención directamente de la fuente original y no de la fuente adquirida. Esta última es la que obtiene los datos de una fuente original y los sistematiza en una base de datos. Esto tiene que ver con que la fuente original muestra con detalle el proceso de recolección y análisis de los datos; además, es más detallada y de mayor exactitud que la fuente adquirida. Los errores de transcripción, la falta de

reproducción de los pies de tabla y de otros comentarios textuales, pueden influir en la exactitud de los datos.

El segundo criterio que debe tenerse en cuenta en la determinación de la exactitud de los datos secundarios es la evaluación del propósito de la publicación. Son sospechosas las publicaciones realizadas por una fuente, gubernamental o no, que desea poner énfasis a su favor ante situaciones comprometidas, o bajo condiciones que sugieren una controversia, o una forma que revela un intento forzado hacia la “franqueza”, o para poner en controversia deducciones de otros datos.

La evidencia relacionada con la calidad tiene que ver con los detalles del diseño de la investigación. Cuando no se revelan los detalles puede ocurrir que la organización que relevó los datos primarios tiene algo que ocultar. El investigador tiene que poder evaluar puntos de la investigación original tales como: el plan de muestreo, el procedimiento de recolección de datos, las técnicas de elaboración de cuestionarios y los procedimientos de análisis de datos.

### **Fuentes gubernamentales de datos**

La mayor fuente de datos estadísticos son los estados nacionales y, en menor medida, la de los estados provinciales. En consecuencia, la investigación econométrica requiere un completo conocimiento de los datos del gobierno en el espacio de referencia del estudio.

En general, la oficina de censos de los institutos de estadística oficiales son las fuentes de datos secundarios más importantes. Están disponibles también datos de los ministerios o de los bancos centrales. Hoy en día, en internet se pueden encontrar todo tipo de datos económicos para realizar investigaciones adecuadas.

Los datos de censos económicos, por ejemplo, están disponibles a diferentes niveles de desagregación. Los datos de censos poblacionales incluyen niveles que van desde la nación hasta poblaciones menos importantes.

Además de las fuentes gubernamentales, existen numerosas publicaciones que contienen datos aplicables a un amplio número de modelos econométricos.

## **7.2. ELEMENTOS DE MUESTREO**

El diseño de muestreo es una de las formas que el investigador posee para realizar la investigación econométrica por medio de fuentes de información primaria. Otras formas de llevar a cabo el diseño de estas fuentes son:

- la observación directa, examinar una situación sin modificarla;
- la experimentación, construir una situación controlada por el investigador;

- el estudio de huellas o rastros, observación diferida de determinadas consecuencias del fenómeno analizado.

Estas formas de diseño no son taxativas, existen otras como las historias de vida, el análisis de contenido, medida de actitudes, evaluación de programas y simulación por computadora.

Muestreo aparece en el tercer paso de una investigación estadística y es aquella teoría que establece los procedimientos que permiten generalizar sobre la población a través del estudio de una parte de la misma; es decir, a través del estudio de la muestra.

Cada observación, o elemento tomado de la población, contiene cierta cantidad de información acerca del parámetro o parámetros de interés. Ya que la información cuesta dinero, el investigador debe determinar cuánta información debe comprar: poca información impedirá realizar buenas estimaciones; mientras que, mucha información ocasiona un despilfarro de dinero. La cantidad de información depende del número de elementos muestreados y de la cantidad de variación de los datos, ambos fenómenos pueden ser controlados a través del diseño de la encuesta y el tamaño de la muestra.

Esta introducción a los métodos estadísticos para la selección de muestras aleatorias en marcos poblacionales determinados, no se dedica al estudio de métodos específicos de muestreo; sino que trata de elaborar cuáles serán los objetivos de la aplicación de metodologías estadísticas a problemas de investigación. Constituyen las respuestas al porqué, cómo y dónde, algunas razones y motivaciones para usar buenos métodos de muestreo, una visión panorámica de los problemas básicos y de los métodos para resolverlos, y una indicación de la manera que el concepto de muestreo de poblaciones cabe dentro de los métodos de muestra y dentro de la búsqueda general de conocimiento científico.

El diseño de muestra tiene dos aspectos: un proceso de selección, que consiste en la regla y operaciones mediante las cuales se incluyen en la muestra algunos miembros de la población, y un proceso de estimación para calcular los estadísticos de la muestra, que son estimaciones muestrales de los parámetros (valores) de la población.

En la teoría del muestreo son importantes las siguientes definiciones:

*Elemento:* objeto sobre el cual se toman las mediciones.

*Población:* conjunto de elementos acerca de los cuales se desea hacer alguna inferencia, es el universo de referencia.

*Unidades de muestreo:* conjunto no superpuesto de elementos de la población que cubren la población completa. Si bien en el estudio necesito encuestar individuos, es cierto que un hogar significa un conjunto de individuos (es decir un conjunto de elementos) y el proceso de seleccionar hogares, y dentro de los hogares seleccionar el elemento, puede resultar más eficiente siempre y cuando la persona no sea

encuestada dos veces. Es posible que el número de elementos y el número de unidades de muestreo coincidan, esto es así cuando se muestrean individuos en lugar de hogares.

*Marco:* es una lista de unidades de muestreo. La selección del elemento (individuo) puede hacerse directamente del marco, esto es si se poseen listas de individuos. También pueden darse marcos múltiples: primero seleccionando viviendas y dentro de las viviendas individuos.

*Muestra:* conjunto de unidades seleccionadas de un marco o de varios marcos. De la muestra se obtendrán los datos objeto de la investigación que se utilizará para describir la población y realizar estimaciones sobre ella.

---



---

**Ejemplo.** Una importante firma industrial alimenticia quiere conocer la proporción de habitantes de cierta población que han consumido una nueva línea de polvos para helados de reciente aparición en el mercado.

Elemento será la persona que habita en esa población y que fue seleccionada por algún método de muestreo para que respondiera acerca de su conocimiento o no del producto.

Población son todos los habitantes de la población que superen un cierto límite de edad a definir por el investigador.

---



---

Perez Lopez (2005) manifiesta que “Al hablar de métodos de muestreo nos referimos al conjunto de técnicas estadísticas que estudian la forma de seleccionar una muestra lo suficientemente representativa de una población cuya información permita inferir las propiedades o características de toda la población cometiendo un error medible y acotable. A partir de la muestra, seleccionada mediante un determinado método de muestreo, se estiman las características poblacionales (media, total, proporción, etc) con un error cuantificable y controlable. Las estimaciones se realizan a través de funciones matemáticas de la muestra, denominadas estimadores, que se convierten en variables aleatorias al considerar la variabilidad de las muestras. Los errores se cuantifican mediante varianzas, desviaciones típicas o errores cuadráticos medios de los estimadores, que miden la precisión de estos. La metodología que permite inferir resultados, predicciones y generalizaciones sobre la población estadística, basándose en la información contenida en las muestras representativas previamente elegidas por métodos de muestreo formales, se denomina inferencia estadística”.

### 7.3. Selección de Muestras

El muestreo estudia los métodos para seleccionar y observar una parte de la población con el fin de hacer inferencias acerca de toda la población. Una muestra puede tener varias ventajas sobre un censo completo:

- a) economía;
- b) rapidez y oportunidad;
- c) posibilidad de hacerse (si la observación es destructiva, el empleo de un censo no es práctico);
- d) calidad y precisión (en algunas situaciones, no hay dinero suficiente para pagar el personal adiestrado y los supervisores necesarios para realizar un buen censo, o aun para obtener una muestra grande).

Los censos completos poseen ventajas especiales en algunas situaciones:

- a) se pueden obtener datos para unidades pequeñas;
- b) la aceptación pública es más fácil de alcanzar para datos completos;
- c) la colaboración y la respuesta del público se pueden obtener más fácilmente.

### **Criterios del diseño de la muestra**

- 1) Orientación hacia la meta. El diseño completo, tanto en la selección como en la estimación, debe orientarse a los objetivos de la investigación, hechos a la medida del diseño de la encuesta y ajustados a las condiciones de la encuesta.
- 2) La medibilidad es una característica de los diseños que permite calcular, a partir de la propia muestra, estimaciones válidas o aproximaciones de su variabilidad de muestreo. Esto se suele expresar en las encuestas con los errores estándares, pero a veces, pueden utilizarse otras expresiones de la función de verosimilitud o de la distribución de muestreo. Esta es la base necesaria para la inferencia estadística que sirve como puente, científico y objetivo, entre el resultado de la muestra y el valor desconocido de la población.
- 3) La practicidad se refiere a los problemas que deben resolverse para llevar a cabo el diseño esencialmente como se propuso. Una muestra probabilística no puede crearse por suposición, ni estará dada, como sucede en los problemas teóricos. El método de los muestreadores de cuota a sus entrevistadores: "vayan y obtengan una muestra aleatoria", es sumamente impráctico; ni el entrevistador ni el que lo envía pueden hacerlo. Se requiere de cuidado para traducir el modelo de selección teórico a un conjunto de instrucciones de oficina y campo. Estas instrucciones deben ser simples, claras, prácticas y completas. Por ejemplo, para identificar un segmento de muestra, no se le debe pedir al entrevistador que localice una línea marcada arbitrariamente en un mapa; sus deberes deben confinarse a localizar calles y direcciones.

- 4) La economía se refiere a cumplir los objetivos de la encuesta con un costo mínimo y al grado en que se alcanza este objetivo.

En líneas generales, hay dos formas de obtener una muestra: de manera informal y casual, o bien, de manera probabilística.

#### **7.4. Muestras informales y casuales**

Las muestras no probabilísticas constituyen un problema en la inferencia pues no hay una manera de estimar qué tan representativas son esas muestras seleccionadas. Los procedimientos usados en este tipo de muestreo son:

- muestras de juicio: el entrevistador selecciona a cualquier sujeto que desee.
- muestras de cuotas: son muestras de juicios pero con previa asignación de cuotas por sexo, edad, clase social, raza, entre otros, que tratan de simular características conocidas de la población.
- trozo de pastel: en este caso el entrevistador no interviene en el proceso de selección pues consiste en una autoselección, personas que responden a un cupón, que concurren a un centro de exhibición, el público de un teatro en particular, por ejemplo.

Los peatones pueden ser interrogados en cuanto a sus opiniones de un nuevo producto. Si la respuesta de todos en la población es uniforme, todos ellos lo odian o lo aman, tal enfoque puede ser satisfactorio.

#### **7.5. Muestras Probabilística**

Todos los miembros de la población tienen una probabilidad conocida de estar en la muestra. Una muestra probabilística tiene las ventajas de:

- permitir al investigador demostrar la representatividad de la muestra.
- permitir un planteamiento explícito en cuanto a la cantidad de variación que será introducida, porque se usa una muestra en lugar de un censo de la población.
- hacer posible la identificación más explícita de las probables desviaciones.

El objetivo de un muestreo probabilístico es hacer una inferencia acerca de la población con base en la información contenida en la muestra. Existen dos factores que pueden afectar la información contenida en la muestra:

- El tamaño de la muestra
- La cantidad de variación en los datos (que puede ser controlada por el *método* de selección de una muestra)

Las unidades de muestreo contienen los elementos y se usan para seleccionarlos en la muestra. En el muestreo de elementos, cada unidad de muestreo contiene solamente un elemento; pero en el muestreo de conglomerados cualquier unidad de muestreo, llamada conglomerado, puede contener varios elementos.

---

---

**Ejemplo.** Una muestra de estudiantes se puede obtener de una muestra de aulas; o una muestra de viviendas de una muestra de manzanas de la ciudad.

---

---

Una misma encuesta puede usar diferentes clases de unidades de muestreo; en muestreo polietápico se usa una jerarquía de unidades de muestreo o conglomerados, de manera que el elemento pertenezca únicamente a una unidad de muestreo en cada etapa.

---

---

**Ejemplo.** Puede tomarse una muestra de los habitantes de una Región al seleccionar sucesivamente los municipios, las manzanas, las viviendas y, finalmente, las personas.

---

---

Las unidades de listado se usan para identificar y seleccionar unidades de muestreo a partir de listas. A veces se necesitan procedimientos detallados para convertir listados en unidades de muestreo, como por ejemplo, para convertir un listado de direcciones en viviendas y hogares. Los problemas pueden ser serios si los elementos no se identifican unívocamente con los listados. Por ejemplo, una muestra de familias tomada de los listados de teléfonos puede involucrar serias dificultades.

Entre los métodos de selección de una muestra probabilística se encuentran:

- 1) Muestreo aleatorio simple
- 2) Muestreo sistemático
- 3) Muestreo estratificado
- 4) Muestreo por conglomerados

## 5) Muestreo por etapas múltiples

**Muestreo Aleatorio Simple**

El muestreo aleatorio simple es un enfoque en el cual cada miembro de la población, y por tanto cada muestra posible, tiene una probabilidad igual de ser seleccionado.

Las muestras aleatorias simples pueden ser seleccionadas mediante el uso de una tabla de números aleatorios. La manera de hacer inferencias es estimar ciertos parámetros de la población utilizando la información de la muestra.

Frecuentemente, el objetivo es estimar una media poblacional (parámetro  $\mu$ ) o un total poblacional (parámetro  $\tau$ ); otras veces se requiere estimar la proporción poblacional (parámetro  $\pi$ ).

Para lograr la estimación se usa el promedio muestral (estadístico  $\hat{\mu}$ ), el estimador del total poblacional (estadístico  $\hat{\tau}$ ) y la proporción muestral (estadístico  $\hat{\pi}$  o  $p$ ), donde:

$$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}; \quad \hat{\tau} = N\hat{\mu};$$

siendo,

$N$ , tamaño de la población

$n$ , tamaño de la muestra

$x_i$ , datos muestrales,  $i = 1, \dots, n$

En el muestreo aleatorio simple, el tamaño de la muestra -es decir, la cantidad de observaciones necesarias para estimar un parámetro poblacional con un límite para el error de estimación (asumir un riesgo determinado de cometer dicho error)- viene dado por

$$n = \frac{z^2 \sigma^2}{e^2}$$

**Muestreo Sistemático**

Consiste en esparcir sistemáticamente la muestra a lo largo de la lista de miembros de la población. Si la población tiene 10.000 individuos y se desea un tamaño de muestra de 1.000, cada décima persona es seleccionada para la muestra.

Aunque en casi todos los ejemplos prácticos tal procedimiento generaría una muestra equivalente a una muestra aleatoria simple, el investigador debe estar consciente de las regularidades dentro de la lista.

Se debe hallar primero la frecuencia de extracción de elementos, es decir, cada cuántos elementos se extrae uno. Esto se logra haciendo

$$\frac{N}{n} = K$$

Para determinar cuál es el primer elemento de la muestra, se selecciona de una tabla de números aleatorios un valor inferior a  $K$ . A este elemento se lo denomina arranque aleatorio ( $a$ ). El segundo individuo será el  $a+K$ , el tercero  $a+K+K$  y así sucesivamente.

La ventaja de este método de selección es su practicidad; la desventaja surge a partir de la determinación de los elementos  $a$  y  $K$ , donde algunas unidades de observación pasan a tener probabilidad cierta de ser seleccionadas y otras probabilidad nula.

### **Muestreo Estratificado.**

En el muestreo aleatorio simple, una muestra aleatoria se selecciona de una lista, o de un marco muestral, que representa a la población.

Al desarrollar un plan de muestreo, es aconsejable buscar subgrupos naturales que sean más homogéneos que la población total. Tales subgrupos se denominan "estratos".

Este tipo de muestreo es conveniente cuando la población o universo puede ser dividido en categorías estratos o grupos que reúnen cierto interés analítico y que por razones teóricas y empíricas presentan diferencias entre ellos.

- 1) La estratificación puede producir un límite más pequeño para el error de estimación que el que se generaría por una muestra aleatoria del mismo tamaño. Este resultado es particularmente cierto si las mediciones dentro de los estratos son homogéneas.
- 2) El costo por observación en la encuesta puede ser reducido mediante la estratificación de los elementos de la población en grupos convenientes.
- 3) Se pueden obtener estimaciones de los parámetros poblacionales para subgrupos de la población. Los subgrupos deben ser entonces estratos identificables.

---

**Ejemplo.** Se necesita información sobre las actitudes de los estudiantes hacia una nueva instalación atlética dentro de la Universidad. Se conoce que existen tres grupos de estudiantes con características diferenciadas:

A) los que viven en residencias estudiantiles tienen actitudes muy homogéneas hacia la instalación propuesta, la variación o la varianza en sus actitudes es muy pequeña.

B) residentes en la ciudad, son menos homogéneos.

C) residentes fuera de la ciudad, varían ampliamente en sus opiniones.

En tal situación, en lugar de permitir que la muestra provenga de la totalidad de los tres grupos aleatoriamente, será más prudente tomar un menor número de miembros del grupo de residentes y extraer más del grupo ajeno al campo. Se particiona la lista de los estudiantes en los tres grupos y se extrae una muestra aleatoria simple de cada uno de los tres grupos.

---

Una muestra estratificada puede ser

- 1) Proporcional, donde la fracción de muestreo es igual en cada estrato de la muestra que la existente en la población
- 2) No proporcional

La estimación de la media de la población, en el muestreo estratificado, es un promedio ponderado de las medias de las muestras encontradas en cada estrato:

$$\mu_X = \sum_j \pi_j \bar{X}_j$$

donde,  $\bar{X}_j$  = la media de la muestra para el estrato  $i$

$\pi_j$  = la proporción de la población en el estrato  $i$

---

**Ejemplo.** El propósito de la investigación es estudiar el rendimiento escolar según sea su extracción de clase social. Para ello, el investigador se sitúa en una escuela a la cual concurren 500 alumnos y de los cuales le informan la composición de clase social en la Figura 7.6.

| Población       |           | Muestra      |                 |
|-----------------|-----------|--------------|-----------------|
| Clases sociales | Elementos | Proporcional | No proporcional |
| Alta            | 50        | 5            | 25              |
| Media           | 300       | 30           | 25              |
| Baja            | 150       | 15           | 25              |
| Total           | 500       | 50           | 75              |

**Figura 7.6. Composición de clase social**

Si se quiere que el tamaño de la muestra sea del 10% de la población se aplica esa fracción de muestreo a cada estrato, dando lugar a un muestreo proporcional. Ahora bien en el estrato Clase Alta solo se tienen 5 casos y, puede ocurrir, que no alcancen para realizar ciertas estimaciones. Si

se establece que se necesitan 25 casos en cada clase para poder realizar el estudio se está en presencia del muestreo no proporcional, donde el tamaño muestral es de 75 y los resultados a obtener dentro de cada estrato se deben ponderar por el peso del estrato en la población. En el caso del muestreo proporcional esto no ocurre porque es autoponderado.

### ¿Cómo se determina la mejor aplicación del presupuesto de muestreo a diversos estratos?

Este problema clásico de muestreo fue solucionado en 1935 por Jerzy Neyman, a partir de la siguiente expresión:

$$n_i = \frac{\pi_i \sigma_i / \sqrt{c_i}}{\sum_i \pi_i \sigma_i / \sqrt{c_i}} n$$

donde,

- $n_i$  = el tamaño de la muestra para el estrato  $i$
- $\pi_i$  = la proporción de la población en el estrato  $i$
- $\sigma_i$  = la desviación estándar de la población en el estrato  $i$
- $c_i$  = el costo de una entrevista en el estrato  $i$
- $\Sigma_i$  = la suma a lo largo de todos los estratos
- $n$  = tamaño de la muestra

**Ejemplo.** La Figura 7.7 presenta información de la encuesta sobre el uso mensual de las cajeros automáticos. La población se encuentra estratificada por ingreso. El segmento de ingresos *altos* tiene la variación más alta y el costo de entrevista más alto. Los estratos de ingresos medios y bajos tienen el mismo costo de entrevista, pero difieren con respecto a la desviación estándar del uso de los procesadores bancarios.

| Estrato (i) | Proporción ( $\pi_i$ ) | Desviación estándar ( $\sigma_i$ ) | Costo ( $c_i$ ) entrevista | $\pi_i \sigma_i / \sqrt{c_i}$ | $n_i$ |
|-------------|------------------------|------------------------------------|----------------------------|-------------------------------|-------|
| Bajo        | 0.3                    | 1                                  | 25                         | 0.06                          | 177   |
| Medio       | 0.5                    | 2                                  | 25                         | 0.20                          | 588   |
| Alto        | 0.2                    | 4                                  | 100                        | 0.08                          | 235   |

Figura 7.7. Uso de cajeros automáticos

Para asignar 1000 casos entre los diferentes estratos, se tiene en cuenta la  $\sum(\pi_i \sigma_i / \sqrt{c_i}) = 0.34$  y se calcula:

$$n_{bajo} = \frac{\pi_i \sigma_i / \sqrt{c_i}}{\sum(\pi_i \sigma_i / \sqrt{c_i})} * 1000 = \frac{0.06}{0.34} * 1000 = 177$$

$$n_{medio} = \frac{\pi_i \sigma_i / \sqrt{c_i}}{\sum (\pi_i \sigma_i / \sqrt{c_i})} * 1000 = \frac{0.20}{0.34} * 1000 = 588$$

$$n_{alto} = \frac{\pi_i \sigma_i / \sqrt{c_i}}{\sum (\pi_i \sigma_i / \sqrt{c_i})} * 1000 = \frac{0.08}{0.34} * 1000 = 235$$

Las cantidades muestrales de la última columna, presentan la división de la muestra de 1.000 personas en los tres estratos. Al estrato de ingresos altos le corresponden 235 personas que representan el 23.5% de la muestra; sin embargo, la proporción de personas de altos ingresos en la población es de 20.0%. Si se hubiera seleccionado una muestra aleatoria simple de tamaño 1000, 200 personas pertenecerían al estrato de altos ingresos y no sería un número suficiente para estimar los parámetros en la población.

---

La disponibilidad presupuestaria determina, en alguna medida, el tamaño de la muestra. Este es ajustado hacia arriba hasta que alcanza el límite presupuestal, de modo que el presupuesto debe ser calculado como:

$$\text{Presupuesto} = \sum_i c_i n_i$$

Conocido el tamaño de la muestra, se determina el error muestral y se decide si es excesivo o no. La fórmula del error muestral es:

$$\text{Error muestral} = z \sigma_{\bar{x}}$$

donde,

$$\sigma_{\bar{x}} = \sqrt{\sum_i \frac{\pi_i^2 \sigma_i^2}{n_i}}$$

Si el error muestral es excesivo se debe ampliar el presupuesto, lo que permite tomar un tamaño de muestra mayor; de no ser posible el proyecto debe desecharse.

## Muestreo De Conglomerados

En el muestreo de conglomerados, la población se divide nuevamente en subgrupos. En esta técnica se selecciona una muestra aleatoria de subgrupos y todos los miembros de los subgrupos forman parte de la muestra. Este método es útil cuando se pueden identificar aquellos subgrupos que sean representativos de la totalidad de la población.

---

**Ejemplo.** Se tomó una muestra de estudiantes universitarios de segundo año, que cursaban estadística en todas las facultades de Argentina, para estudiar la prevención del SIDA. Se contaba en ese momento con 200 comisiones de estadística, cada una de las cuales tenía, en promedio, 30 estudiantes.

El muestreo utilizado fue el de conglomerados: se seleccionaron 15 comisiones y, dentro en ellas, la totalidad de los alumnos. De modo que el tamaño de muestra fue de 450 alumnos.

Si se hubiera decidido tomar una muestra aleatoria simple reuniendo 450 alumnos en el total de 200 cursos de estadística, el costo sería significativamente mayor.

La gran pregunta, desde luego, es si los cursos son representativos de la población y la respuesta es "no necesariamente". Si los cursos de las áreas de ingresos superiores tienen diferentes opiniones acerca de la prevención del SIDA, que los cursos con estudiantes de ingresos más bajos, el supuesto que fundamenta al enfoque, no se mantendría.

---

La gran ventaja del muestreo de conglomerados es que su costo es más bajo. Los subgrupos o conglomerados son seleccionados de modo que, el costo para obtener la información deseada dentro del conglomerado, sea mucho más pequeño que si se obtuviera una muestra aleatoria simple.

## Diseños De Etapas Múltiples

Si se usan otros diseños muestrales, la lógica para generar el tamaño óptimo de la muestra aún se mantendrá; sin embargo, la fórmula puede complicarse.

---

**Ejemplo.** En un diseño por áreas, el primer paso puede ser el de seleccionar comunidades al azar. De este modo, el procedimiento puede ser escoger porciones de radios censales, luego manzanas y finalmente familias.

---

En tal diseño, la expresión para determinar el error estándar de  $\bar{X}$  se vuelve sumamente compleja. La situación consiste en repetir la totalidad del plan de muestreo y obtener dos, tres o cuatro estimaciones independientes de  $\bar{X}$ . Estas diferentes estimaciones pueden ser usadas para estimar el error estándar de  $\bar{X}$ .

Formalmente, el diseño en etapas múltiples bajo el supuesto que se conocen los tamaños de las unidades primarias, parte de considerar una población de individuos (llamadas unidades secundarias) de tamaño  $N$  repartida en  $M$  unidades primarias. Para  $k = 1, \dots, M$ , sea  $N_k$  el tamaño de la unidad primaria  $k$ ,  $N_k$  se supone conocido

$$\sum_{k=1}^M N_k = N$$

Sea  $Y$  una variable real de interés. Se denota  $\tau, \mu$  y  $\sigma^2$  el total, la media y la varianza de  $Y$  sobre la población;  $\tau_k, \mu_k$  y  $\sigma_k^2$  el total, la media y la varianza de  $Y$  sobre la unidad primaria  $k$  ( $k = 1, \dots, M$ ).

Se plantea la media y la varianza del total para cada unidad primaria de la siguiente manera:

$$\mu_\tau = \frac{1}{M} \sum_{k=1}^M \tau_k \quad \text{y} \quad \sigma_\tau^2 = \frac{1}{M} \sum_{k=1}^M (\tau_k - \mu_\tau)^2$$

Se tiene entonces que

$$\mu = \frac{M}{N} \mu_\tau$$

Por lo tanto, estimar  $\mu$  se reduce a estimar  $\mu_\tau$ .

Para ello se considera una muestra aleatoria simple  $M$  de probabilidades iguales sin remplazo de  $m$  unidades primarias. Luego, en cada unidad primaria de la muestra,  $k \in M$ , se toma una muestra aleatoria simple de probabilidades iguales sin remplazo de  $n_k$  individuos.

Se denotan  $\bar{y}_k$  y  $S_k^2$  la media y la varianza corregida de  $Y$  sobre esta muestra. Se estima entonces  $\mu_\tau$  mediante

$$\mu_\tau = \frac{1}{m} \sum_{k \in M} \hat{t}_k$$

Donde  $\hat{t}_k = N_k \bar{y}_k$

Y se deduce el estimador de  $\mu$ :

$$\hat{\mu} = \frac{M}{N} \frac{1}{m} \sum_{k \in M} \hat{t}_k$$

Es un estimador insesgado de  $\mu$  cuya varianza se estima sin sesgo mediante:

$$\hat{V}(\hat{\mu}) = \frac{M^2}{N^2} \left[ S_\tau^2 \left(1 - \frac{m}{M}\right) + \frac{1}{Mm} \sum_{k \in M} N_k^2 \frac{S_k^2}{n_k} \left(1 - \frac{n_k}{N_k}\right) \right]$$

Donde:  $S_\tau^2 = \frac{1}{m-1} \sum_{k \in M} (\hat{t}_k - \hat{\mu}_\tau)^2$

Ejemplo. Supóngase que se considera una población de 3.000.000 habitantes ( $N$ ) repartida en 3000 municipios ( $M$ ). Sea  $Y$  el consumo de un cierto producto en diciembre de 1995. Se extrae una muestra aleatoria simple con probabilidades iguales sin remplazo de 30 municipios, luego, para la extracción de las unidades secundarias se estudian dos métodos:

Primer método. En cada uno de los municipios de la muestra, se extrae una muestra aleatoria simple de habitantes con probabilidades iguales sin remplazo con una tasa de muestreo constante igual a 1/100. El número de unidades secundarias encuestadas en cada municipio es entonces aleatorio y su esperanza es igual a 10.

Segundo método. En cada uno de los municipios de la muestra se extrae una muestra aleatoria simple con probabilidades iguales sin remplazo de 10 habitantes. (Aunque es útil a los fines prácticos, este procedimiento no es el mejor, ya que no se mantiene constante la probabilidad de un municipio a otro y la probabilidad de ser elegido para un individuo es diferente).

| $N_k$       | $\bar{y}_k$ | $s_k$ | $\frac{N_k \bar{y}_k}{10^3}$ | $\left(\frac{N_k \bar{y}_k}{10^3}\right)^2$ | $n_k$ | 1° METODO                                       | 2° METODO                                      |
|-------------|-------------|-------|------------------------------|---|-------|---|--|
|             |             |       |                              |   |       | $\frac{(N_k s_k)^2}{n_k} \times \frac{1}{10^6}$ | $\frac{(N_k s_k)^2}{10} \times \frac{1}{10^6}$ |
| 400         | 750         | 200   | 300                          | 90000                                       | 4     | 1600.0  | 640.0  |
| 400         | 800         | 180   | 320                          | 102400                                      | 4     | 1296.0  | 518.4  |
| 400         | 750         | 180   | 280                          | 78400                                       | 4     | 1296.0  | 518.4  |
| 500         | 850         | 220   | 425                          | 180625                                      | 5     | 2420.0  | 1210.0   |
| 500         | 750         | 180   | 375                          | 140625                                      | 5     | 1620.0  | 810.0  |
| 600         | 800         | 200   | 480                          | 230400                                      | 6     | 2400.0  | 1440.0   |
| 600         | 750         | 180   | 450                          | 202500                                      | 6     | 1944.0  | 1166.4   |
| 600         | 700         | 150   | 420                          | 176400                                      | 6     | 1350.0  | 810.0  |
| 600         | 750         | 120   | 450                          | 202500                                      | 6     | 864.0   | 518.4  |
| 700         | 800         | 200   | 560                          | 313600                                      | 7     | 2800.0  | 1960.0   |
| 700         | 700         | 130   | 490                          | 240100                                      | 7     | 1183.0  | 828.1  |
| 800         | 650         | 180   | 520                          | 270400                                      | 8     | 2592.0  | 2073.6   |
| 800         | 650         | 220   | 520                          | 270400                                      | 8     | 3872.0  | 3097.6   |
| 900         | 500         | 140   | 450                          | 202500                                      | 9     | 1764.0  | 1587.6   |
| 900         | 700         | 170   | 630                          | 396900                                      | 9     | 2601.0  | 2340.9   |
| 1000        | 650         | 190   | 650                          | 422500                                      | 10    | 3610.0  | 3610.0   |
| 1000        | 550         | 140   | 550                          | 302500                                      | 10    | 1960.0  | 1960.0   |
| 1000        | 650         | 180   | 650                          | 422500                                      | 10    | 3240.0  | 3240.0   |
| 1100        | 600         | 190   | 660                          | 435600                                      | 11    | 3971.0  | 4368.1   |
| 1100        | 650         | 150   | 715                          | 511225                                      | 11    | 2475.0  | 2722.5   |
| 1200        | 700         | 210   | 840                          | 705600                                      | 12    | 5292.0  | 6350.4   |
| 1200        | 700         | 220   | 840                          | 705600                                      | 12    | 5808.0  | 6969.6   |
| 1300        | 550         | 160   | 715                          | 511225                                      | 13    | 3328.0  | 4326.4   |
| 1400        | 600         | 150   | 840                          | 705600                                      | 14    | 3150.0  | 4410.0   |
| 1500        | 450         | 140   | 675                          | 455625                                      | 15    | 2940.0  | 4410.0   |
| 1600        | 550         | 180   | 880                          | 774400                                      | 16    | 5184.0  | 8294.4   |
| 1700        | 500         | 160   | 850                          | 722500                                      | 17    | 4352.0  | 7398.4   |
| 1800        | 550         | 170   | 990                          | 980100                                      | 18    | 5202.0  | 9363.6   |
| 1900        | 550         | 180   | 1045                         | 1092025                                     | 19    | 6156.0  | 11696.4  |
| 1900        | 600         | 200   | 1140                         | 1299600                                     | 19    | 7600.0  | 14440.0  |
| $\Sigma$    |             |       | 18710                        | 13144350                                    | 301   | 93870.0   | 113079.2                                       |
| $\Sigma/30$ |             |       | 623.67                       | 438145                                      |       | 3129.0  | 3769.3   |

Figura 7.8. Selección de municipios

En las tres primeras columnas de la Figura 7.8 se registran los resultados sobre las 30 muestras. Las demás columnas sirven para el cálculo de  $\hat{\mu}$  y  $V(\hat{\mu})$

Se tiene

$$\hat{\mu} = \frac{M}{N} \frac{1}{m} \sum_{k \in M} N_k \bar{y}_k \quad \text{con } \frac{M}{N} = \frac{1}{10^3}$$

Las tasas de muestreo del primer nivel y del segundo nivel son despreciables, pues se obtiene:

$$\hat{V}(\hat{\mu}) = \frac{M^2}{N^2} \left[ \frac{s_\tau^2}{m} + \frac{1}{Mm} \sum (N_k)^2 \frac{s_k^2}{n_k} \right]$$

Con  $n_k = N_k/100$  para el primer método y  $n_k = 10$  para el segundo.

Se tiene

$$\frac{s_\tau^2}{m} = \frac{1}{m-1} \left[ \frac{1}{m} \sum_{k \in M} (N_k \bar{y}_k)^2 - \left( \frac{1}{m} \sum_{k \in M} N_k \bar{y}_k \right)^2 \right]$$

En este ejemplo se obtienen los mismos resultados con los dos métodos porque el segundo término es despreciable:

Primer método

$$\hat{\mu} = 624$$

$$\hat{V}(\hat{\mu}) = \frac{1}{29} (438145 - 624^2) + \frac{1}{3000} \times 3129$$

$$\hat{V}(\hat{\mu}) = 1697$$

Segundo método

$$\hat{\mu} = 624$$

$$\hat{V}(\hat{\mu}) = \frac{1}{29} (438145 - 624^2) + \frac{1}{3000} \times 3769.3$$

$$\hat{V}(\hat{\mu}) = 1697$$

pues el intervalo de confianza al 95% de  $\hat{\mu}$  es idéntico para cada uno de los dos métodos.

$$\hat{\mu} \pm 1.96 \sqrt{\hat{V}(\hat{\mu})}$$

$$624 \pm 81$$

$$[543; 705]$$

**Ejemplo 2.** Se trata de un país de 22.200.000 habitantes distribuidos en 4000 municipios de acuerdo a la Figura 7.9. Se decide constituir una muestra de 2.220 habitantes; es decir, una tasa de muestreo de 1/10.000, utilizando, en cada categoría de municipios, un muestreo bietápico. Este consiste en una muestra de municipios, luego muestra de habitantes en los municipios extraídos, de tal manera que cada habitante tenga la misma probabilidad igual a 1/10.000 de pertenecer a la muestra.

Se puede procesar según dos métodos.

Primer Método

Se realiza un muestreo aleatorio estratificado proporcional de municipios con una tasa uniforme de 1/100, luego, en los municipios extraídos, un muestreo aleatorio simple de un habitante para cada 100 habitantes.

Se obtiene entonces para cada muestra de municipios:

$$m_1 = 30 \qquad m_2 = 7 \qquad m_3 = 3$$

y, en promedio, para cada municipio de la muestra de cada categoría:

$$n_1 = 10 \qquad n_2 = 60 \qquad n_3 = 500$$

lo que da, como previsto, para cada categoría, un número medio de habitantes en la muestra de:

300, 420 y 1500 respectivamente

Con este método, los tamaños de las muestras de habitantes son aleatorios y el manejo del trabajo de los encuestadores se vuelve difícil. Además el número de "puntos de encuesta", que corresponden aquí al número de municipios de la muestra, es débil: solamente 40.

|  |   |   |
|--|---|---|
| MUNICIPIOS DE MENOS DE 2000 HABITANTES                         | MUNICIPIOS DE 2000 A 10000 HABITANES                          | MUNICIPIOS DE MAS DE 10000 HABITANTES                           |
| 3000 MUNICIPIOS<br>3.000.000 HABITANTES<br>(MEDIA: 1.000 HAB.) | 700 MUNICIPIOS<br>4.200.000 HABITANTES<br>(MEDIA: 6.000 HAB.) | 300 MUNICIPIOS<br>15.000.000 HABITANTES<br>(MEDIA: 50.000 HAB.) |

**Figura 7.9. Distribución de municipios en la población**

#### Segundo Método

Se decide encuestar a 10 habitantes por municipio de la muestra, lo que, cuando se toma en cuenta la duración del cuestionario, corresponde a un día de trabajo de un encuestador.

Hay pues que constituir una muestra de 222 municipios. Se constituye esta muestra de municipios con probabilidades desiguales, proporcionales al tamaño del municipio.

Se obtiene entonces, para la muestra de municipios:

$$m_1 = 30 \qquad m_2 = 42 \qquad m_3 = 150$$

y, como tenemos que encuestar 10 habitantes por municipio de la muestra, al final tenemos, tal como previsto:

300, 420 y 1500 habitantes encuestados en cada categoría

Con este método, se facilita la gestión del trabajo de los encuestadores y el número de puntos de encuestas crece (222).

Para un municipio, su probabilidad de pertenecer a la muestra de municipios es la misma dentro de una categoría pero distinta de una categoría a otra.

Tenemos  $\pi_i = 1/100$  para la primera categoría,  $\pi_i = 6/100$  para la segunda y  $\pi_i = 50/100$  para la tercera.

De hecho se trató, en el ejemplo anterior, de estimar la media de  $Y$  sobre el primer estrato; se obtuvo, con cualquiera de los dos métodos:

$$\hat{\mu} = 624 \quad \hat{V}(\hat{\mu}) = 1697$$

Se trata ahora de estimar  $\mu$  utilizando el estimador de estratificación.

Se supone que los resultados sobre los tres estratos son los de la Figura 7.10. Se obtiene pues, para el conjunto (última columna):

$$\hat{\mu} = \sum \frac{N_k}{N} \hat{\mu}_k \quad \hat{V}(\hat{\mu}) = \sum \left( \frac{N_k}{N} \right)^2 \hat{V}(\hat{\mu}_k)$$

|                        | Primer estrato | Segundo estrato | Tercer estrato | Conjunto   |
|------------------------|----------------|-----------------|----------------|------------|
| $N_k$                  | 3.000.000      | 4.200.000       | 15.000.000     | 22.200.000 |
| $N_k / N$              | 30/222         | 42/222          | 150/222        | 1          |
| $\hat{\mu}_k$          | 624            | 700             | 750            | 744        |
| $\hat{V}(\hat{\mu}_k)$ | 1700           | 1600            | 500            | 317        |

Figura 7.10. Muestreo de municipios

Pues la estimación de  $\mu$  por intervalo de confianza al 95%:

$$\hat{\mu} \pm 1.96 \sqrt{\hat{V}(\hat{\mu})} \quad 744 \pm 35 \quad [709; 779]$$

Se realizó aquí un muestreo de 3 etapas. Como se trata de una estratificación a la primera etapa, al final se tiene, en cada estrato, un muestreo bietápico: una muestra de municipios, luego, en cada municipio de la muestra, una muestra de habitantes.

## CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS

### Caso 7.1: Cooperativa de Alimentos

J.L., Gerente General de una Cooperativa de Alimentos, se dio cuenta recientemente que había perdido contacto con los patrones de compra de los miembros de la cooperativa. Simplemente, la cooperativa parecía mucho más grande ahora en comparación con los primeros años. El Gerente se preguntaba si podría hacer uso de algún tipo de dato que estuviese a su alcance con el propósito de ampliar su comprensión de los hábitos de compra de los miembros. Esperaba poder utilizar este conocimiento para planificar mejor la mezcla y el volumen de artículos que la cooperativa ofrecía.

J.L. y un pequeño grupo de voluntarios fundaron la cooperativa en 1974. Esta cooperativa había aumentado de diez miembros iniciales en enero de 1974 a 500 miembros en setiembre de 1990. La empresa estaba localizada en una vieja bodega del noroeste de Milán, Michigan. Milán era una comunidad de 7500 personas localizada en el sureste de Michigan, aproximadamente a 40 millas al suroeste de Detroit. La cooperativa obtuvo sus miembros de una gran cantidad de comunidades alrededor del Milán, incluyendo Ann Arbor y Monroe.

Antecedentes. El objetivo de la cooperativa era proveer productos alimenticios de alta calidad a un precio por debajo de los que se ofrecían en los supermercados locales. Para este objetivo, la cooperativa utilizaba las cajas de empaque como estanterías, solicitaba la colaboración de los mismos compradores para que marcaran sus precios en los artículos, vendía sólo las mejores marcas del mercado y, por lo general, no ofrecía los "lujos" relacionados con los supermercados tradicionales. Para poder comprar en la cooperativa, las personas tenían que ser socios. La cuota de afiliación era de \$25 anuales. Cualquier ganancia obtenida por la cooperativa durante un año, se devolvía a los socios en forma de crédito para sus compras. J.L. estaba convencida de que los socios compraban la mayor parte de sus alimentos en la cooperativa.

Preocupaciones de J.L. durante los primeros años de la cooperativa. J.L. se sentía orgullosa de conocer a todos sus miembros. Había gastado una cantidad considerable de tiempo en la tienda y sentía que conocía lo que las personas compraban y cuánto dinero estaban gastando. A medida que creció el número de socios, sus deberes administrativos la mantenían más tiempo en la oficina. Por tanto, ya no podía decir que conocía a todos los socios, ni que tenía idea de sus patrones de gastos. Deseaba conocer mejor estos aspectos de su negocio y pensó que tal vez parte de los datos que se habían recolectado con anterioridad sobre los miembros podrían proporcionarle respuestas.

### Datos disponibles

En junio de 1990 se utilizó un cuestionario para recolectar datos sobre los socios. Durante ese mes todos los miembros vinieron a la cooperativa por lo menos una vez. Por tanto existían datos sobre cada uno de los socios. Los datos consistían en las características demográficas de los miembros y en sus gastos semanales en alimentos.

Los datos se encontraban en las tarjetas que los socios habían llenado en el momento de la entrevista. J.L. tenía estas tarjetas en un archivador en su oficina. A continuación se encuentra una descripción del contenido de las tarjetas. Los valores reales de las tarjetas se hallan tabulados en las páginas siguientes.

Con el propósito de poder conocer mejor a los socios, J.L. quiere saber inicialmente el promedio semanal de gastos en alimentos, puesto que no dispone de demasiado tiempo, quiere hacer esto sin tener que mirar las quinientas tarjetas. Sin embargo, también desea asegurarse de que el promedio que calcule sea exacto. Para esto contrata a un consultor y este le recomienda realizar una muestra probabilística sobre las 500 tarjetas que se encuentran tabuladas en la tabla. Las tarjetas tienen la siguiente información:

### Descripción de las variables

**A** = número de identificación de la unidad familiar; 1 - 500

**B** = gastos reales semanales en alimentos

**C** = número de personas en la unidad familiar; 1 - 9

1 = una persona

2 = dos personas

3 = tres personas

4 = cuatro personas

5 = cinco personas

6 = seis personas

7 = siete personas

8 = ocho personas

9 = nueve personas

**D** = ingreso familiar anual real

**E** = educación del jefe del hogar; 1 - 5

1 = menos del 8 grado

2 = entre los grados 9 y 11

3 = con título en secundaria

4 = algunos años de educación secundaria

5 = con título universitario

**F** = edad actual del jefe del hogar

**G** = gasto semanal en alimentos, codificados en 7 categorías; 1 - 7

- 1 = menos de \$ 15
- 2 = \$ 15 a \$ 29.99
- 3 = \$ 30 a \$ 44.99
- 4 = \$ 45 a \$ 59.99
- 5 = \$ 60 a \$ 74.99
- 6 = \$ 75 a \$ 89.99
- 7 = \$90 ó más

**H** = algún niño menor de 6 años en el hogar; 1 - 2

- 1 = no
- 2 = sí

**I** = algún hijo entre 6 y 18 años en el hogar; 1 - 2

- 1 = no
- 2 = sí

**J** = ingreso anual familiar, codificado en seis categorías; 1 - 6

- 1 = menos de \$ 3000
- 2 = \$ 3000 - \$ 5999.99
- 3 = \$ 6000 - \$ 9999.99
- 4 = \$ 10000 - \$ 14999.99
- 5 = \$ 15000 - \$ 24999.99
- 6 = \$ 25000 ó más

**K** = edad del jefe del hogar, codificado en siete categorías; 1 - 7

- 1 = menos de 25
- 2 = 25 - 34
- 3 = 35 - 44
- 4 = 45 - 54
- 5 = 55 - 64
- 6 = 65 - 74
- 7 = 75 ó más

Se le pide:

- 1) Seleccione una *muestra aleatoria simple* de tamaño 50. Construya la tabla con las observaciones seleccionadas. Calcule los estadísticos que permitan describir la información.
- 2) Seleccione una *muestra sistemática* de tamaño 50. Construya la tabla con las observaciones seleccionadas. Calcule los estadísticos que permitan describir la información.

- 3) Seleccione una *muestra estratificada proporcional* según la presencia de niños menores de 6 años en el hogar de tamaño 25. Construya la tabla con las observaciones seleccionadas. Calcule los estadísticos que permitan describir la información.
- 4) Seleccione una *muestra estratificada no proporcional* según la presencia de niños menores de 6 años en el hogar donde cada estrato sea de tamaño 25. Construya la tabla con las observaciones seleccionadas. Calcule los estadísticos que permitan describir la información.

| A   | B    | C | D     | E | F  | G | H | I | J | K |
|-----|------|---|-------|---|----|---|---|---|---|---|
| 1   | 12   | 1 | 2500  | 1 | 56 | 1 | 1 | 1 | 1 | 5 |
| 2   | 16.5 | 1 | 2800  | 1 | 70 | 2 | 1 | 1 | 1 | 6 |
| 3   | 18   | 1 | 2000  | 1 | 20 | 2 | 1 | 1 | 1 | 1 |
| 4   | 17   | 1 | 4500  | 1 | 60 | 2 | 1 | 1 | 2 | 5 |
| 5   | 46.5 | 1 | 8000  | 1 | 40 | 4 | 1 | 1 | 3 | 3 |
| 6   | 45   | 1 | 7000  | 1 | 51 | 4 | 1 | 1 | 3 | 4 |
| 7   | 15   | 1 | 3500  | 1 | 76 | 2 | 1 | 1 | 2 | 7 |
| 8   | 60   | 2 | 2800  | 1 | 20 | 5 | 1 | 1 | 1 | 1 |
| 9   | 15   | 2 | 2500  | 1 | 51 | 2 | 1 | 1 | 1 | 4 |
| 10  | 18   | 2 | 4000  | 1 | 32 | 2 | 1 | 1 | 2 | 2 |
| 11  | 22.5 | 2 | 5000  | 1 | 47 | 2 | 1 | 1 | 2 | 4 |
| 12  | 20   | 2 | 8000  | 1 | 35 | 2 | 1 | 1 | 3 | 3 |
| 13  | 97   | 2 | 5500  | 1 | 58 | 7 | 1 | 1 | 2 | 5 |
| 14  | 57   | 2 | 6000  | 1 | 27 | 4 | 1 | 1 | 3 | 2 |
| 15  | 39   | 2 | 3000  | 1 | 38 | 3 | 1 | 1 | 2 | 3 |
| 16  | 30   | 2 | 4000  | 1 | 40 | 3 | 1 | 1 | 2 | 3 |
| 17  | 42   | 2 | 3000  | 1 | 19 | 3 | 1 | 1 | 2 | 1 |
| .   | .    | . | .     | . | .  | . | . | . | . | . |
| .   | .    | . | .     | . | .  | . | . | . | . | . |
| .   | .    | . | .     | . | .  | . | . | . | . | . |
| .   | .    | . | .     | . | .  | . | . | . | . | . |
| .   | .    | . | .     | . | .  | . | . | . | . | . |
| 497 | 115  | 6 | 24000 | 5 | 36 | 7 | 2 | 2 | 5 | 3 |
| 498 | 75   | 7 | 28000 | 5 | 37 | 6 | 2 | 2 | 6 | 3 |
| 499 | 105  | 7 | 20000 | 5 | 39 | 7 | 2 | 2 | 5 | 3 |
| 500 | 75   | 8 | 33000 | 5 | 42 | 6 | 2 | 2 | 6 | 3 |

Tabla extraída de Kinnear, T. Taylor, J. *Investigación de Mercados. Un enfoque aplicado*. Mc.Graw Hill. 1993.  
Puedes bajar de la web el archivo Cooperativa de Alimentos.xls

## Problemas

Problema 1. En una compañía se quiere determinar los gastos médicos familiares anuales promedio de los empleados. La gerencia de la compañía quiere tener 95% de confianza de que el promedio de la muestra está correcto con aproximación  $\pm\$50$  de los gastos familiares reales.

Un estudio piloto indica que la desviación estándar se puede estimar en \$1400. ¿Qué tan grande es el tamaño de la muestra necesario?

Problema 2. Se desea estimar la suma promedio de ventas con aproximación de  $\pm \$100$  con 99% de confianza y se supone que la desviación estándar es de \$200, ¿qué tamaño de muestra se necesita?

Problema 3. Un grupo de estudio quiere estimar la facturación mensual promedio por luz eléctrica en el mes de julio en una ciudad grande. En base a estudios efectuados en otras ciudades se supone que la desviación estándar es de \$20. El grupo querría estimar la facturación promedio de julio con aproximación  $\pm \$5$  del promedio real con 99% de confianza ¿Qué tamaño de muestra se necesita?

## Bibliografía

- **Berenson, Mark y Levine, Daniel.** *Estadística Básica En Administración.* México: Prentice Hall, 1996.
- \_\_\_\_\_. *Estadística Para Administración Y Economía. Conceptos Y Aplicaciones.* México: Mc. Graw Hill, 1993.
- **Daniel, W.** *Bioestadística, Base Para El Análisis De Las Ciencias De La Salud.* México: Editorial Limussa, 1999.
- **Dixon, W.J. y Massey, F.J.** *Introduction to Statistical Analysis.* Nueva York: Mc Graw Hill, 1957.
- **Espasa, Antoni y Cancelo, José.** *Métodos Cuantitativos Para El Análisis De La Coyuntura.* Madrid: Alianza Editorial SA, 1993.
- **Hernández Sampieri, R.; Fernández Collado, C. y Baptista Lucio, P.** *Metodología De La Investigación.* México: McGraw Hill, 2010.
- **Kazmier, L y Diaz Mata, A.** *Estaística Aplicada a La Administración Y a La Economía.* México: McGraw Hill, 1993.
- **Kinncar, T.y Taylor, J.** *Investigación De Mercado. Un Enfoque Aplicado.* McGraw Hill, 1993.
- **Padua, J.** *Técnicas De Investigación Aplicadas a Las Ciencias Sociales.* México: Fondo de Cultura Económica, 1996.
- **Perez Lopez, César.** *Muestreo Estadístico. Conceptos Y Problemas Resueltos.* España: Editorial Pearson Prentice Hall, 2005.

