

COMISION ECONOMICA PARA AMERICA LATINA
COMITE DE COOPERACION ECONOMICA
DEL ISTMO CENTROAMERICANO
SUBCOMITE DE COORDINACION ESTADISTICA
DEL ISTMO CENTROAMERICANO

CCC/SC.2/GTEI/III/DI/4
9 de Junio de 1964.

Tercera Reunión del Grupo de Trabajo
sobre Estadísticas Industriales
San Salvador, El Salvador, 10 a 13 de Junio de 1964

GUIA PARA EL MUESTREO DE ESTABLECIMIENTOS PEQUEÑOS

Versión revisada para la Tercera Reunión del Grupo de Trabajo de Estadísticas Industriales, del documento Una guía para el muestreo de pequeños establecimientos, preparado por los señores Jack L. Ogus y Jack Ingram de la Oficina de Censos del Departamento de Comercio de los Estados Unidos, para el Seminario Latinoamericano sobre Censos económicos celebrado en San José, Costa Rica (febrero-abril de 1962)

INTRODUCCION

La experiencia ha demostrado que la recolección de datos de censos económicos para todos los establecimientos es una empresa costosa. Es además económicamente ineficiente porque cuando se enumeran todos los establecimientos pequeños, éstos por razón de su número considerable absorben costos de recolección y elaboración, ofreciendo en cambio, información relativamente insignificante. En contraste, los establecimientos grandes, aunque generalmente su número es reducido, constituye la mayor parte de la actividad industrial de un país, entanto que los pequeños establecimientos contribuyen con una parte relativamente pequeña de la de producción, ventas, remuneraciones, etc.

El método del muestreo permite balancear más efectivamente el costo de la información para cada establecimiento. El método es especialmente aplicable al problema de recopilar datos de establecimientos pequeños, que son los más abundantes y los que proporcionan menos información.

Para propósitos de ilustrar tal aplicación, los Sres. Jack L. Ogus y J. Jack Ingram de la Oficina del Censo de los EE. UU. prepararon un documento entitulado "Una Guía para el Muestreo de Pequeños Establecimientos". Se presentó por primera vez en 1963 durante un seminario regional auspiciado por la Agencia para el Desarrollo Internacional ^{1/} en San Jose, Costa Rica,

^{1/} Las agencias colaborando en la dirección del seminario fueron: El Instituto Interamericano de Estadística, la Escuela de Economía de la Universidad de Costa Rica, la Dirección General de Estadística y Censos de Costa Rica y la Oficina de Censos de los Estados Unidos.

P. 2 - Introduction

al que asistieron 28 técnicos de 14 países latinoamericanos. El programa del seminario se basó en la presentación de un caso de estudio que comprendía un plan de censos económicos para un país imaginario de América Latina llamado Providencia.

De acuerdo con el caso de estudio, Providencia desarrolló planes para el levantamiento de los censos de minería, manufactura, construcción, producción y distribución de electricidad, comercio por mayor, por menor y servicios. El personal de la DGE de Providencia decidió utilizar las técnicas de muestreo en la enumeración de los establecimientos debido a la reducción de los gastos de operación.

El plan de enumeración de Providencia comprendió:

1. Una enumeración completa de establecimientos, con 5 ó más personas ocupadas, en los campos de minería, manufactura, comercio al por menor y servicios.
2. Una muestra del 10% de establecimientos con menos de 5 personas ocupadas, en los campos arriba mencionados.
3. Una enumeración completa de establecimientos mayoristas.
4. Una enumeración completa de empresas dedicadas a la producción y distribución de electricidad, y a construcción.

La Guía para el Muestreo analiza estas decisiones de la DGE de Providencia. Además presenta y discute los procedimientos a usarse para la selección y control de una muestra de establecimientos; también indica cómo hacer estimaciones de la muestra, y cómo medir la confiabilidad de estas estimaciones. Finalmente se indican los criterios necesarios para desarrollar técnicas de muestreo más refinadas que las empleadas en el caso de Providencia.

P. 3 - Introduction

La opinión general de los participantes al seminario de Providencia, en Costa Rica, fué que la Guía para el Muestreo presentó una contribución **significante** a la metodología de censos económicos. La Guía para el Muestreo, por lo tanto, se presenta al Grupo de Trabajo en Estadísticas Industriales como un documento de referencia de cuyo examen podran derivarse conclusiones importantes sobre las ventajas que presenta la aplicación adecuada de las técnicas de muestreo en la obtención de datos censales. Por ejemplo, algunos países excluyen de sus censos de manufacturas los establecimientos con menos de cinco personas ocupadas. Estos países podrían considerar la posibilidad de ampliar el alcance de sus investigaciones censales mediante la obtención de una muestra de tales establecimientos.



Benjamin Gura, Director
Seminario Latinoamericano
sobre Censos Económicos
(SELACE)

III-4. GUIA PARA EL MUESTREO DE ESTABLECIMIENTOS PEQUEÑOS

Jack L. Ogus y J. Jack Ingram
U. S. Bureau of the Census

1. Consideraciones Generales

Para nuestros propósitos, consideramos los establecimientos agrupados en dos partes: una que comprende los establecimientos que ocupan cinco o más personas, y otra que abarca las que ocupan menos de cinco personas. Esta clasificación particular de establecimientos es arbitraria; sin embargo, por razones prácticas, que se discutirán más adelante, se ha considerado conveniente adoptar tal clasificación.

La experiencia en Providencia mostró que los establecimientos con menos de cinco personas ocupadas comprendía una parte considerable del número total de establecimientos. Se observó, igualmente, que el aporte de este grupo representaba una proporción relativamente pequeña del total de las actividades industriales, comerciales y de servicios. La Ilustración III-4-1, muestra el porcentaje de algunas características informadas por los establecimientos que ocupan menos de cinco personas.

Estas distribuciones ciertamente deberían influir en los planes censales ya que uno de los objetivos principales de la planificación es distribuir los fondos disponibles en forma tal que se obtenga la mayor información con el mínimo de gasto. Por ejemplo en los censos anteriores de Providencia se enumeraron todos los establecimientos; sin embargo, se obtuvo de los establecimientos con menos de cinco personas ocupadas información para un número de temas que el registrado para los establecimientos grandes. La experiencia ha sido similar en otros países. El problema se caracteriza en el sector de la industria manufacturera (ver Ilustración III-4-1)

La enumeración de todos los establecimientos sin, distinción de tamaño, supone la investigación de los establecimientos pequeños cuyo número alcanza aproximadamente al 60% o 70% del total.

Las anteriores consideraciones conducen a preguntar si se justifica invertir sumas apreciables de dinero en la enumeración integral de los establecimientos pequeños, cuando su contribución a la economía nacional, en términos de producción, ventas, remuneraciones, valor agregado, etc., es poco significativa.

Organizaciones internacionales han recomendado que la cobertura de los censos sea lo más completa posible. Sin embargo, en reconocimiento de la desproporción entre los costos de operación relativamente altos y la limitada importancia de los establecimientos pequeños, han recomendado que los países estudien la posibilidad de utilizar métodos de muestreo para obtener datos de pequeños establecimientos. Una tercera alternativa -excluso los establecimientos pequeños de la cobertura Censal y por lo tanto excluirlos de la enumeración- ha sido rechazada. En las siguientes secciones consideraremos brevemente las principales características de cada una de las normas anteriores.

2 Tratamiento de los Establecimientos con Menos de Cinco Personas Ocupadas.

Con referencia a los establecimientos con menos de cinco personas ocupadas, se señalan los sistemas siguientes: (a) Sistema de "eliminación," bajo el cual se declaran como fuera del campo de investigación de los censos económicos aquellos establecimientos con menos de cinco personas ocupadas; (b) Sistema de "cobertura total," bajo el cual se investigarán todos los establecimientos sin distinción de tamaño; y (c) Sistema de "Muestreo," en el cual se investigará un número limitado (muestra) de los establecimientos con menos de cinco personas ocupadas.

- 2.1 Sistema de Eliminación. Este sistema excluye de la cobertura censal todos los establecimientos con menos de cinco personas ocupadas. El uso de este sistema conducirá a: disminuir el costo de recopilación y elaboración de datos; eliminar la tarea de localizar a todos los establecimientos pequeños; reducir el volumen de información disponible sobre la economía, lo cual puede afectar los estudios de planificación económica futura del país; omitir de los resultados censales los datos correspondientes a establecimientos grandes que fueron erróneamente clasificados dentro del grupo de establecimientos con menos de cinco personas ocupadas; y a, excluir los establecimientos que posterior a la fecha de matrícula, registraron un aumento en el número de personas ocupadas superior a los límites de eliminación. Por otra parte, debemos señalar que el sistema de eliminación es relativamente fácil de aplicar.
- 2.2 Sistema de Cobertura Total. Al igual que el sistema de eliminación, este sistema de cobertura total es relativamente fácil de emplear. Se requiere un informe de cada establecimiento. No se necesitan reglas o procedimientos especiales para determinar los establecimientos que deben llenar el cuestionario. Las oportunidades de errores se reducen a un mínimo, ya que al menos en la etapa de empadronamiento, todas las unidades reciben el mismo tratamiento. Además, como se obtiene un volumen máximo de cuestionarios, existe una libertad óptima para variar los procedimientos de elaboración.

Si esta libertad de elegir los procedimientos de elaboración se aplica con juicio, el sistema de cobertura total proporcionará una cantidad máxima de información. Sin embargo, esta ventaja y las

conveniencias de este sistema se obtienen al precio de llevar a un máximo la carga de obtener datos de los establecimientos pequeños, además de aumentar substancialmente el costo en que incurre el gobierno al levantar los censos. Una parte de este costo reside en el empleo de numeroso personal administrativo y profesional en las labores de recolección, examen y tabulación de la gran cantidad de cuestionarios provenientes de pequeños establecimientos.

2.3 El Sistema de Muestreo. Este sistema combina algunas de las características de los dos sistemas anteriores. Su aplicación es más compleja que cualquiera de los otros dos y aumenta, por lo tanto, las oportunidades de cometer errores.. Proporciona cierta información sobre los establecimientos más pequeños, pero es casi seguro que no será toda la información que fuera de desear. Elimina la carga de obtener datos de una gran número de establecimientos, pero parece imponer esa carga de manera arbitraria en algunos de los pequeños. Su virtud principal, y una razón importante para el uso extenso de este sistema, es que, a diferencia de los otros sistemas, este puede lograr un equilibrio razonable entre el costo y esfuerzo que se emplea en establecimientos pequeños y el rendimiento en términos de información producida.

Esta ventaja del sistema de muestreo -equilibrio entre costos e información obtenida- también se aplica al costo para los informantes. Para la clase de establecimientos pequeños en conjunto, el sistema de muestreo reduce muy substancialmente la carga total de recoger los datos. Una enumeración completa distribuye la carga equitativamente

a cada establecimiento, pero esto no justifica el causar molestias innecesarias a muchos establecimientos cuyos informes no se necesitan. Además, la carga sobre aquéllos seleccionados no se impone de manera arbitraria. Cada uno de los establecimientos más pequeños tiene igual oportunidad de ser seleccionado en la muestra. Pueden escogerse muchas muestras diferentes, utilizando exactamente los mismos procedimientos, y son grandes las posibilidades de que cualquier establecimiento particular sea seleccionado para una muestra y no lo sea para otra. Aun cuando esta selección pudiera parecer injusta para un informante en particular, el hecho de que ha sido elegido al azar significa que el sistema es equitativo.

Las objeciones de que el sistema de muestreo es más difícil de administrar y de que aumenta las oportunidades de errores, son reales. Deben tomarse en cuenta cuando se considera un sistema de muestreo y particularmente cuando hay diversos planes de muestreo que se pueden seleccionar. Deberán llevarse registros y controles adecuados para evitar que se cometan serios errores, y esto eleva los costos. Sin embargo, el costo neto de un sistema de muestreo bien construido puede ser mucho menor que el costo de un sistema de cobertura total; al mismo tiempo se evita la limitación del sistema de eliminación. Estas consideraciones generales nos llevan a concluir que el sistema de muestreo es preferible para los establecimientos pequeños.

Por lo tanto, en el Censo de Providencia se decidió seleccionar una muestra de un décimo (10 por ciento) de los establecimientos dedicados a la minería, manufacturas, comercio por menor y servicios, que

cuentan con menos de cinco personas ocupadas. Veamos si esta decisión fue razonable.

En la Ilustración III-4-2 se muestran ciertos datos que describen los costos básicos de los diversos planes para el Censo de Providencia. Para presentar un cuadro completo, se incluyen cifras tanto de empadronamiento directo como de empadronamiento por correo. Ya que hemos decidido hacer la enumeración del Censo por correo, debemos prestar atención al costo de un sistema de este tipo.

Como se ha dicho anteriormente, el costo sólo pueden compararse significativamente en relación a los datos "producidos." Si fuéramos a considerar el ahorro en costo solamente, concluiríamos que el mejor plan para el Censo sería aquél cuyo costo fuera cero, lo cual significa que no se levantaría Censo alguno. La Ilustración III-4-3, combina los dos tipos de medidas que necesitamos. Extrae de la Ilustración III-4-2 los costos primarios que se necesitan, y también presenta índices del volumen de información utilizable proporcionada por los diversos planes censales. En cada caso se supone una cobertura total de los establecimientos con cinco o más personas ocupadas. Se habrá de notar que los datos proporcionados por los diversos planes de muestreo arrojan índices cercanos a 1.00. Esto demuestra la realidad y pone énfasis en el hecho de que las muestras proporcionan estimaciones para todo el universo con un margen de error que no está presente en la cobertura total. Los índices particulares de información se presentan

sólo como indicadores, pero las diferencias que existen entre ellos son un reflejo de las diferencias que pudieran esperarse para las diversas muestras y clases económicas.

Un examen de la Ilustración III-4-3 indica dónde el muestreo puede significar un ahorro monetario importante y el precio en términos de pérdida de información, cuando se compara con una cobertura total. De la misma forma, el cuadro muestra cómo se obtienen estos datos y su costo, cuando se compara con el Sistema de Eliminación Total.

Por ejemplo, es evidente que el muestreo de pequeños establecimientos dedicados a comercio por menor y servicios produce un ahorro substancial de dinero con sólo pérdidas moderadas de datos. Por otra parte, si se eliminan de la investigación todos los establecimientos de este grupo se pierde la mitad de la información sobre comercio por menor y servicios, y al mismo tiempo el costo total en dólares sólo se reduce en forma mínima, si se compara con un sistema de muestreo. Dicho de otra forma, la cobertura total de los pequeños establecimientos de comercio por menor y servicios sería muy costosa. El muestreo puede ahorrar una cantidad substancial de este alto costo sin sacrificar demasiada información. Si se eliminan de la investigación todos los establecimientos pequeños, sólo se podría ahorrar una suma adicional moderada y se sacrificará un volumen substancial de datos. Por lo tanto, hemos concluido que en este caso la decisión de hacer uso de muestreo es correcta. (Estudiaremos la selección del tamaño de la muestra en la Sección 7).

Con respecto al comercio por mayor, la Ilustración III-4-3 señala un ahorro máximo posible de dolares, \$7,268, que se logra al eliminar de la investigación todos los establecimientos de comercio por mayor con menos de cinco personas ocupadas; dicho ahorro no es significativo, mientras que la pérdida de información si es substancial. Aun cuando la muestra del 20 por ciento recuperaría una parte importante de la información perdida, el ahorro de dinero en comparación con el costo de la cobertura total se reduciría sólo en una cantidad mínima. Por lo tanto, la decisión administrativa de enumerar todos los establecimientos dedicados al comercio por mayor está substancialmente justificada.

Además del comercio por mayor, las actividades de construcción, producción de electricidad, transportes, almacenaje y comunicaciones han sido seleccionadas para empadronamiento total.

Finalmente, tenemos al grupo que comprende la explotación de minas e industrias manufactureras. Veamos si los datos de la Ilustración III-4-3 apoyan la decisión administrativa de obtener una muestra del 10 por ciento de los establecimientos manufactureros y mineros, con menos de cinco personas ocupadas. Estos pequeños establecimientos contribuyen con sólo el 5 por ciento del total del valor agregado de los grupos industriales en cuestión. No cabe duda que la enumeración de estas unidades a un costo de 24,000 dólares, no sería aconsejable desde el punto de vista presupuestal. El muestreo en este caso es más indicado. La muestra mínima -de hecho, eliminación total- pudiera considerarse aún mejor. Tanto los costos como el volumen de información envueltos son tan pequeños que es difícil hacer

una elección sobre estas bases. Bien podría decidirse eliminarlos todos en algunos censos futuros, e investigarlos por muestra en otros. El deseo de ampliar nuestra experiencia con las técnicas de muestreo nos conducen a la decisión de obtener una muestra de los pequeños establecimientos mineros e industriales en este censo. El costo es pequeño y el riesgo de perjudicar los totales en forma significativa es casi nulo.

Y bien, ¿por qué tomar una muestra del 10 por ciento en lugar de, digamos, el 5 por ciento? Nuestra necesidad de obtener más experiencia con los métodos de muestreo nos da la respuesta. Esta es la primera vez que estaremos utilizando el muestreo en un censo económico en Providencia. Por lo tanto, queremos evitar complejidades innecesarias. La utilización de la tasa de muestreo del 10 por ciento que se seleccionó anteriormente para los grupos de comercio por menor y servicios es más sencilla que el uso de una tasa diferente en cada grupo.

Podrían diseñarse planes de muestreo matemáticamente superiores pero más complicados. Sin embargo, le estamos dando mucha importancia a las condiciones operativas en que se llevará a cabo este censo. Las técnicas refinadas de muestreo son inútiles si los requisitos de la teoría no se pueden observar totalmente en la práctica. Siempre es preferible utilizar especificaciones moderadas que se pueden ejecutar con más confiabilidad que utilizar procedimientos teóricamente más eficientes pero cuya aplicación es dudosa.

En resumen: Para este censo económico de Providencia, hemos dividido nuestro universo en dos clases para muestreo. Establecimientos que incluimos con certeza o sea un "muestreo con tasa de 1 en 1" y

establecimientos para muestreo con tasa de 1 en 10. Este es un plan muy sencillo y se adoptó principalmente en virtud de su simplicidad. Planes de muestreo más complicados, que den más flexibilidad a nuestra definición de la "clase con certeza", serían más eficientes en términos de la información obtenida por peso gastado. Sin embargo, aún este plan tan sencillo arroja un ahorro significativo y demuestra la utilidad de la técnica del muestreo.

3. Como se Utilizará la Muestra

Como resultado de la utilización de este plan de muestreo tan sencillo habremos aprendido a seleccionar una muestra, a controlar la muestra a hacer estimaciones basadas en la muestra, y a medir la confiabilidad de las estimaciones. En primer lugar, esta experiencia servirá como base para tomar el mismo tipo de muestra en el futuro. En segundo lugar, servirá como base para la utilización en el futuro de técnicas de muestreo más refinadas con el propósito de satisfacer mejor nuestras necesidades.

La muestra que hemos seleccionado aquí, proporcionará estimaciones de características seleccionadas para todos los establecimientos con menos de cinco personas ocupadas. Se harán estimaciones para los totales nacionales en minería manufactura comercio por menor y servicios. Estas se combinarán con los totales de los establecimientos con cinco o más personas ocupadas en la preparación de estimaciones para todos los establecimientos. Las estimaciones detalladas serán iguales a las arriba descritas.

La Ilustración III-4-5 es una lista de las características que habrán de estimarse para las industrias mineras y manufactureras; las correspondientes al comercio por menor y servicios se presentan en la Ilustración III-4-5. Es posible que las estadísticas sobre algunas de las características presentadas en la lista no se publiquen, si se decide que las estimaciones basadas en la muestra no son lo suficientemente confiables.

Los mismos datos de la muestra se utilizarán para determinar la confiabilidad de las estimaciones. El método se explica en la Sección 6.

Después del censo, la muestra puede utilizarse: (a) como un conjunto de establecimientos que habrá de utilizarse continuamente en la recopilación de datos económicos en encuestas continuas, (b) como un conjunto de establecimientos que habrá de utilizarse en investigaciones económicas especiales, (c) para tabulaciones suplementarias que no se contemplan en los planes originales de publicación censal.

4. Selección de la Muestra

Los procedimientos específicos de tabulación mecánica para la selección de la muestra se presentan en otro documento ^{1/}. El propósito de esta sección es dar algunas ideas sobre las técnicas allí descritas.

La muestra de establecimientos con menos de cinco personas ocupadas se seleccionará utilizando el siguiente equipo IBM: Clasificadora 083 con el Aditamento 978 para el recuento de tarjetas, Perforadora 024, la

Reproductora 513 con Aditamento para Numeración Consecutiva. Comenzando al azar, seleccionaremos del Directorio de Establecimientos, cada décimo establecimiento con menos de cinco personas ocupadas. El procedimiento para la selección de las tarjetas correspondiente a estos establecimientos se presenta a continuación. Se pone énfasis especial en aquellos pasos directamente relacionados con la muestra.

4.1 Clasifique por Grupo de Actividad Económica de 1 dígito todas las Tarjetas de Datos en el Registro de Establecimientos (Columna 15).

Las tarjetas quedarán clasificadas en los siguientes grupos:

<u>Código</u>	<u>Descripción</u>
1	Explotación de Minas y Canteras
2,3	Industrias Manufactureras
4,	Construcción, Electricidad, Comercio por Mayor
5	Electricidad
6	Comercio
7	Servicios
8, en blanco	
X o V	Códigos incorrectos

4.2 Separe a un lado, para uso posterior, las tarjetas con códigos 4, 5 ó 6 (Construcción, Electricidad etc. y Comercio por Mayor

Hemos decidido que estos establecimientos se empadronarán en su totalidad; por lo tanto, no hay necesidad de separarlos para propósitos de muestreo.

4.3 Combine las tarjetas restantes y clasifíquelas en dos grupos según tamaño:

4.31 Tarjetas correspondientes a establecimientos con menos de cinco personas ocupadas y,

4.32 Tarjetas correspondientes a establecimientos con cinco o más personas ocupadas.

4.4 Deje a un lado, para uso futuro (no para muestreo) las tarjetas identificadas en 4.32 arriba.

4.5 Clasifique por provincia las tarjetas en el numeral anterior

El propósito de esta clasificación, y la descrita en 4.6 abajo, es preparar las tarjetas de manera que se facilite seleccionar la muestra, controlar la muestra, y hacer estimados. Tarjetas pueden caer dentro de provincias "impropias" (provincias que no existen). Después de agotar todos los recursos para asignar estas tarjetas a las provincias pertinentes, distribuya al azar entre las provincias propias las que no se pudieron asignar debidamente. Esta distribución se deberá aplicar al grupo de industrias de 1 dígito más bien que a todas esas tarjetas en un grupo.

4.6 Clasifique las tarjetas para establecimientos con menos de 5 personas ocupadas por código industrial de 3 dígitos (cols. 15-17).

En esta clasificación es posible que las tarjetas caigan en industrias "impropias" (que no existen). Esto ocurrirá como resultado de errores en codificación, perforación u operación de equipo. Cuando ésto suceda, debe hacerse todo esfuerzo de comparar la tarjeta con la información de donde se perforó para obtener el código correcto. Tarjetas para las cuales es imposible obtener el código correcto deben separarse en grupo aparte para propósitos de muestreo.

Las tarjetas están ahora en orden por grupo industrial de 3 dígitos y cada grupo por provincia.

- 4.7 Enumere las tarjetas consecutivamente, empezando con 1 para cada grupo industrial de 1 dígito (cols. 67-71). Así habrá 4 grupos de tarjetas empezando con el número de serie 00001 representando cada censo para muestreo.

Usaremos estos números de serie para seleccionar la muestra, como veremos más tarde.

- 4.8 Asigne un número aleatorio para cada censo de la muestra: Para cada censo seleccione un número aleatorio entre el 1 y el 10, inclusive; entonces seleccione la primera tarjeta que tenga tal número en la Columna (71) y cada décima tarjeta en adelante.

Quando el número aleatorio sea 10, seleccione tarjetas con "0" en la Columna 71. Ya que estamos seleccionando una muestra de un establecimiento en 10, nuestro primer paso después de asignar los números de serie es seleccionar cinco números aleatorios entre 1 y 10, inclusive.

El quinto número aleatorio se usará para seleccionar una muestra del grupo de tarjetas antes separadas como "impropias,"

La fuente más conveniente para obtener estos números es de un Cuadro de Números Aleatorios. Vea Ilustración III-4-6.

Sin importar el método que se use, asigne el primer número aleatorio a minería el segundo a manufactura etc., asignando el quinto a los códigos industriales "impropios." Si se repite un número aleatorio debe asignarse como se seleccionó sin tratar de sustituirlo.

Insistimos que la selección debe ser al azar porque uno de nuestros objetivos --además de hacer estimaciones de las características--, es determinar la confiabilidad de las estimaciones. Sólo podemos hacer tal cosa si la estimación se basa en un muestreo al azar. (Aquí definimos una muestra aleatoria como una en la que cada establecimiento que forma la muestra lo hace con una probabilidad conocida).

4.81 El número de serie de cada establecimiento se perforó en las Columnas (67-71). Al seleccionar la muestra de 1 en 10, seleccionaremos todo establecimiento que contenga el número aleatorio seleccionado en la columna (71) para el grupo industrial de 1 dígito. Por lo tanto, clasificaremos todas las tarjetas con el número aleatorio seleccionado en la Columna (71), recordando que igualamos el número aleatorio "10" a "0" en la Col. 71).

El procedimiento anterior es sencillo y directo ya que se aplica a una muestra de 1 en 10 en virtud del diseño de la tarjeta que permite diez dígitos en cualquier columna. Los siguientes ejemplos deben ser útiles para idear instrucciones de muestreo cuando la tasa sea diferente de la utilizada en Providencia.

Ejemplo A: Intervalo de muestreo de 1 en 5 (Demostrado en la Ilustración III-4-7a). Seleccione dos números al azar: Uno del 0 al 4, inclusive, y otro del 5 al 9, inclusive. Seleccione para la muestra cada establecimiento con uno de estos dígitos como el dígito en la posición de las unidades de su número de serie. ($2/10 = 1/5$).

Ejemplo B: Intervalo de muestreo de 1 en 20 (Demostrado en la Ilustración III-4-7b). Siga el mismo procedimiento descrito arriba para

- (1) Para poder asegurar de que no se han incluido establecimientos que no deben estar en la muestra, pasamos por la clasificadora el grupo de tarjetas seleccionadas para cada grupo industrial. Clasificamos en la Columna (71). Todas las tarjetas del grupo deben caer en el bolsillo correspondiente correspondiente al "Número Aleatorio". En segundo lugar, pasamos por la clasificadora el grupo no seleccionado. No deberán caer tarjetas en el bolsillo correspondiente si las tarjetas que cayeron o no cayeron, deben o no estar en la muestra. Las tarjetas que no tienen un número perforado en la Columna (71) bien pueden caer en un bolsillo "Imposible". Asigne 1 a la primera tarjeta, 2 a la segunda, 3 a la tercera, o a la décima, 1 a la 11, 2 a la 12, etc. Seleccione de estas tarjetas las que tengan el número aleatorio industrial aplicable en la Col 71. Devuelva las tarjetas no-seleccionadas al grupo no de muestra.
- (2) Aún cuando se hayan seguido explícitamente las instrucciones para la selección de la muestra hasta este punto, es necesario mantener controles para asegurar que se ha seleccionado el número correcto de establecimientos en la muestra.

Estamos seleccionando una muestra del 10 por ciento (1 en 10). Por lo tanto, el primer requisito es que la muestra seleccionada sea una décima parte del número total de establecimientos en el grupo industrial. Por ejemplo, el total de establecimientos con menos de cinco personas ocupadas en industrias manufactureras en 16,864. Por lo tanto, deberán existir 1,686.4 establecimientos en la muestra del diez por ciento. Haremos

estimaciones para cada industria de 2 dígitos y controlaremos el tamaño de la muestra en ese nivel. El segundo requisito es que para cada uno de los grupos industriales de 2 dígitos se seleccione una muestra de una décima parte de los establecimientos en la correspondiente industria de 2 dígitos. Por ejemplo, el número total de establecimientos con menos de cinco personas ocupadas en "Industrias Manufactureras: Alimentos" (Código 20) es 3,811. Por tanto, se esperan 381.1 establecimientos en la muestra. Por supuesto, no podemos incluir una fracción de un establecimiento en la muestra. De esta manera, en el primer ejemplo se esperarían 1,686 o 1,687 establecimientos y en el segundo 381 o 382.

El número de establecimientos en la muestra puede variar por uno, en cualquier dirección, del número esperado, dependiendo del número aleatorio seleccionado y del remanente, cuando el número total de establecimientos se divide por el intervalo de muestreo (en nuestro caso 10). Esto es cierto ya esté el control en el grupo industrial o en la industria de 2 dígitos. En este procedimiento permitiremos que el tamaño de la muestra varíe por 3, en cada dirección, del número esperado ya sea al nivel de 1 dígito o al nivel de 2 dígitos. Existen sistemas más complicados para determinar la variancia permisible de un total esperado. Sin embargo, para nuestros propósitos esta simple regla será satisfactoria.

En las Ilustraciones III-4-8a a la III-4-8d se presentan hojas de control de la muestra para cada grupo industrial con su componente de industrias de 2 dígitos.

En los casos en que el tamaño de la muestra varía más que la variancia permisible (3) en cualquier dirección, deberán efectuarse las siguientes comprobaciones en el orden que sigue:

- (1) Comprobar si hay errores de transcripción en la hoja de control;
- (2) Comprobar si hay errores de cómputo en las cifras en cualquiera de las últimas tres columnas de la hoja de control;
- (3) Comprobar si hay error en el cómputo del número de establecimientos que caen en la muestra para el grupo industrial y/o en las industrias de 2 dígitos;
- (4) Comprobar si hay error en el cómputo original del número de establecimientos en el grupo industrial y/o en las industrias de 2 dígitos;
- (5) Comprobar si hay errores de perforación en los números de serie;
- (6) Comprobar si hay error en la aplicación del procedimiento de selección de la muestra.

Si la diferencia excede la tolerancia especificada, después de que se han corregido los errores encontrados, se necesitará hacer el muestreo de nuevo.

5. Preparación de Estimaciones Basadas en la Muestra

En las Ilustraciones III-4-4 y III-4-5 se proporcionan listas del máximo de características que pueden estimarse. El método para hacer

estimaciones basadas en una muestra al azar sencilla, como la que tenemos aquí, es esencialmente el mismo que se utiliza para determinar los valores de las características de poblaciones con un 100 por ciento de cobertura. Sin embargo, en el caso de la muestra de establecimientos, debemos multiplicar la suma por una ponderación igual al intervalo de muestreo. Se hace lo mismo con una cobertura del 100 por ciento, pero en este caso, la ponderación es 1, ya que el intervalo de muestreo es uno.

Para mantener las anotaciones sencillas, sólo se pretende estimar las características al nivel nacional de 2 dígitos, relegando a un proceso mecánico la suma de las estimaciones de 2 dígitos, para la obtención de totales para los grupos industriales principales. Denominaremos como N el número total de establecimientos con menos de cinco personas ocupadas y como N_i , el número correspondiente de establecimientos que se dedican a una actividad de 2 dígitos en particular. En ambos casos nos referimos a la lista total de establecimientos en el Directorio que tienen menos de cinco personas ocupadas. En el caso de la muestra, designaremos los números correspondientes como n y n_i , respectivamente. En forma similar, denominaremos el valor total de una característica determinada, para todos los establecimientos, con una X , y el valor total para la actividad de 2 dígitos con X_i . También utilizaremos el símbolo X_{ij} para designar el valor de la característica para un establecimiento en particular.

El subíndice i se refiere al grupo industrial y el subíndice j a un establecimiento de la muestra. Los dos subíndices juntos, ij , indican el

establecimiento de orden j en la muestra perteneciente al grupo industrial i . En las estimaciones rigen las siguientes relaciones fundamentales:

$$N = \text{suma de } N_i, \text{ o } N = \sum_{i=1}^L N_i^*$$

$$n = \sum_{i=1}^L n_i$$

$$X_i = \sum_{j=1}^{N_i} X_{ij}$$

$$x_i = \sum_{j=1}^{n_i} X_{ij}$$

$$X = \sum_{i=1}^L X_i = \sum_{i=1}^L \sum_{j=1}^{N_i} X_{ij}$$

Nótese que las mayúsculas N , X_{ij} , X se refieren a totales del universo, mientras que las minúsculas n_i y x_i se refiere a totales de la muestra.

La muestra incluye una fracción, $f = n/N$, de todos los establecimientos N en el universo. De esta manera, podemos esperar que la muestra incluya aproximadamente esa misma fracción del valor total de la característica X , o sea fX . Así, si se multiplica la suma de la muestra por el recíproco de la fracción de muestreo o sea N/n , el resultado será la

* Con todo detalle, esta anotación significa:

El número total de establecimientos N es igual a la suma, comenzando con el primer grupo ($i = 1$) y terminando con el último grupo (L), de los establecimientos en cada grupo (N_i).

expansión al nivel del universo. El factor N/n , que es igual al intervalo de muestreo, a veces se denomina la ponderación de la muestra, W .

Para los totales resultados de la expansión, es conveniente tener símbolos similares a X_i y X . Para este propósito utilizaremos X' . Las letras mayúsculas representan totales al nivel del universo, y los símbolos prima indican que son estimaciones. Así, para cada industria de 2 dígitos tenemos las siguientes expresiones equivalentes:

$$X'_i = Wx_i = \frac{N}{n} \sum_{j=1}^{n_i} X_{ij} = \sum_{j=1}^{n_i} WX_{ij}$$

También tenemos las relaciones: $X' = \sum_{i=1}^L X'_i$.

Se notará que nos hemos referido consistentemente al "valor de la característica determinada para el establecimiento en el orden j de la muestra perteneciente al grupo industrial de 2 dígitos i ". Esto significa que cada establecimiento seleccionado para la muestra forma parte de la misma y permanece dentro de ella. A veces un establecimiento que aparece en el Directorio de Establecimientos con menos de cinco personas ocupadas informará al Censo que cuenta con cinco o más personas. Estos casos deberán retenerse en la muestra y tratarse como cualquier otro. Por otra parte, establecimientos que aparecen en el directorio con cinco o más personas ocupadas pueden informar al Censo que tienen menos de cinco. Estos no deberán incluirse en la muestra, pero sí deberán retenerse dentro del grupo con cobertura de 100 por ciento para su tabulación posterior.

De la misma forma, es posible que la clasificación industrial original de algunos establecimientos sea diferente de la indicada en el

Cuestionario de Matrícula. Para las tabulaciones del Censo deben incluirse en el grupo industrial respectivo con su ponderación apropiada. Por ejemplo, un establecimiento de servicios que fue seleccionado como parte de la muestra de establecimientos de comercio por menor debe tabularse en el grupo de servicios, pero todas sus cifras deberán expandirse usando su ponderación respectiva, W .

Utilizando métodos mecánicos de tabulación, las estimaciones para X'_i pueden derivarse de varias formas. Entre ellas están:

- (a) Reproduzca, para cada establecimiento en la muestra, un número de tarjetas equivalente al intervalo de muestreo, p.e., si la muestra de 1 en 20 (5 por ciento), deberá perforarse la primera tarjeta y reproducirse diecinueve tarjetas adicionales; si la muestra es de 1 a 10 (10 por ciento) se perforará la primera tarjeta y se reproduciran nueve tarjetas adicionales. Para la tabulación se deberán tabular las tarjetas como si cada una de ellas correspondiera a un establecimiento verdadero. Las sumas tabuladas representarán las estimaciones. Este método no es conveniente para determinar los errores de muestreo correspondientes de los totales calculados. Para este propósito no es posible utilizar el grupo de tarjetas reproducidas. Para la estimación del error de muestreo, habrá de mantenerse separadamente o extraerse de entre las tarjetas reproducidas, el grupo de tarjetas correspondientes a la muestra y éstas se elaborarán por separado para estimar el error de muestreo.
- (b) Tabule una sola tarjeta para cada establecimiento de la muestra. Multiplique la suma de la muestra por la ponderación de la misma

(igual al intervalo de muestreo = $\frac{N}{n}$). Por ejemplo, si la muestra es de 1 en 20, multiplique por veinte; si la muestra es de 1 en 10, multiplique por diez. Para el Censo de Providencia se ha elegido este segundo método. (Si se selecciona una muestra de 1 en 10 puede utilizarse un tercer método en que sólo se tabula el grupo seleccionado de tarjetas de muestra y se multiplica por 10, moviendo la suma un dígito a la izquierda).

- (c) Una variante del segundo método es reproducir las tarjetas de muestra multiplicando los datos en cada campo por 10, moviendo los datos un dígito a la izquierda y a continuación perforando ceros en la posición de las unidades que quedará en blanco en cada campo. Este método general también se puede aplicar cuando usan ponderaciones de 100, etc.
- (d) Cuando se disponga de una multiperforadora o calculadora, y se aplican a la muestra diversas ponderaciones se puede perforar inicialmente la ponderación para cada establecimiento en todas las tarjetas correspondientes al establecimiento en cuestión. A continuación pueden calcularse los valores expandidos multiplicando cada campo de datos X_{ij} por la ponderación W_{ij} .
- Debe notarse que cualquiera que sea el procedimiento mecánico utilizado, el método estadístico para la estimación del valor de la característica es el mismo. Los métodos (c) y (d) tienen la ventaja de que cualquier total calculado que se desee puede derivarse simplemente mediante la tabulación de los valores ponderados de las tarjetas apropiadas.

Ejemplo: (Este ejemplo corresponde a una sola característica; sin embargo, en caso de que se perforen varias características en una sola tarjeta, las tabulaciones y sumas de la muestra para todas las características pueden obtenerse en una sola pasada de las tarjetas).

Los valores para el Total de Remuneraciones Semanales para cada uno de los 200 establecimientos con menos de 5 personas ocupadas clasificados en Industria 36 (Industria Manufacturera: Maquinaria, exceptuando maquinaria eléctrica) del Directorio de Establecimientos se muestran en la Ilustración III-4-9. Se consideró conveniente asignar un número de serie a cada establecimiento.

Primero se selecciona una muestra del diez por ciento de este grupo y se calcula el Total de Remuneraciones Semanales de los establecimientos con menos de 5 personas ocupadas en la Industria 36. Entonces agregamos este estimado al Total de Remuneraciones Semanales para establecimientos con cinco o más personas ocupadas, llegando así a la estimación para el total de establecimientos en la Industria 36.

Es verdad que en el Censo conoceremos sólo las nóminas semanales de los establecimientos seleccionados en la muestra que proporcionaron la información. El propósito de incluir aquí todo el grupo es proporcionar un mayor entendimiento del método de hacer estimaciones y la medición de su confiabilidad.

El número aleatorio utilizado para seleccionar la muestra de establecimientos manufactureros del Directorio de Establecimientos fué nueve. Por lo tanto, se seleccionó el noveno y cada décimo

establecimiento subsiguiente en orden sucesivo. Estos establecimientos se identifican con un asterisco en la Ilustración III-4-9 y se enumeran por separado en la Ilustración III-4-10. El método utilizado para estimar las características de los establecimientos con menos de cinco personas ocupadas se muestra en la Parte (1) de la Ilustración III-4-10. El total de la muestra es \$2,126. Nuestra estimación del Total de Remuneraciones Semanales de la población total de establecimientos con menos de 5 personas ocupadas es \$21,260 (10 x \$2,126). Si nos referimos a la Ilustración III-4-9 podemos observar que mediante el empadronamiento total de todos los establecimientos con menos de 5 personas ocupadas el Total de Remuneraciones Semanales habría sido de \$22,091. En este caso particular, la diferencia es de \$831.

Para propósitos de ilustración asumiremos que el Total de Remuneraciones Semanales de los establecimientos con cinco o más personas ocupadas es de \$924,000. La Parte (2) muestra el método de calcular el Total de Remuneraciones Semanales para todos los establecimientos en Providencia. Nótese que esto se hace sólo sumando las remuneraciones totales semanales correspondientes a los establecimientos con cinco o más personas ocupadas al estimado de establecimientos con menos de 5 personas ocupadas. La suma es \$945,260. Esta suma difiere por \$831 de la que se habría obtenido con un empadronamiento total.

Hasta ahora hemos asumido que se obtendrá contestación de todos los establecimientos en la muestra. ¿Qué sucedería si no fuera? En nuestro caso se recordará que la muestra fué seleccionada del

Directorio de Establecimientos (aquéllos existentes a la fecha de abril de 1963 y los que se registraron posteriormente porque comenzaron sus actividades después de esta fecha). De esta forma, cada establecimiento en la muestra se encontraba activo durante 1963. Algunos establecimientos seleccionados en la muestra pueden haber suspendido operaciones, pueden haber cambiado de ubicación, pueden estar operando bajo nueva administración, o puede que, por error, su cuestionario no llegue a su destino. Los establecimientos incluidos en la muestra, pero que han desaparecido plantean el mismo problema que la ausencia de algún establecimiento en la parte del Censo que no es por muestra. Si se puede obtener, de alguna manera, un informe para la parte del año en que el establecimiento estuvo en operación, no hay problema. Por lo tanto, primero debe tratarse de obtener este informe. Si esto no es posible, debemos utilizar el método de duplicación que se describe a continuación. Sin embargo, en lugar de reproducir la tarjeta seleccionada para representar el establecimiento que falta, debe crearse una tarjeta que contenga la mitad de las cantidades informadas para cada característica.

Otro problema que surge tanto en la cobertura total como en el empadronamiento por muestra es aquél que surge debido a un cambio de propietario entre el momento de constituir el Directorio de Establecimientos y el levantamiento del Censo. Estos pueden identificarse debido a que contienen marcas en las casillas (1) ó (2) del Tema 5 del Cuestionario Censal. Por lo general se obtendrá al menos un informe censal para este tipo de establecimiento. Sin embargo, frecuentemente

no se sabe con precisión qué período del año está incluido en el informe. Por ejemplo, un comprador puede informar sobre las actividades de todo un año, combinando sus cifras con las de su antecesor; puede sólo dar las cifras que representan el período en que él fue el propietario; si el vendedor del establecimiento recibe el cuestionario por correo, puede contestar cualquiera de los períodos anteriores. Por lo tanto, es necesario utilizar la información sobre "dirección" anotada en el Tema 5 para obtener los datos adicionales o para asegurar que los recibidos estén completos.

Cuando se han dado los pasos necesarios para asegurarse de que se han recibido todos los cuestionarios de los establecimientos y se encuentra que todavía faltan algunos, deberá determinarse el número que falta para cada industria de dos dígitos. De la muestra de establecimientos para la industria de 2 dígitos correspondiente, seleccione, al azar, el número de tarjetas que faltan. Usando tarjetas de color diferente, reproduzca una sola vez cada una de las tarjetas seleccionadas, dejando sin perforar los datos sobre identificación (Columnas 1-11). En caso de establecimientos que hayan dejado de operar, tómese, para duplicación, la mitad de los valores informados en la tarjeta seleccionada. Devuélvase la tarjeta original al grupo, junto a la tarjeta reproducida. Procédase a hacer las estimaciones en la forma descrita anteriormente.

6. Determinación del Error de Muestreo *

Se ha seleccionado una muestra de establecimientos con menos de cinco personas ocupadas. Es una muestra sistemática, comenzando con un número al azar, o efectivamente, una muestra aleatoria. Es tan solo una de las muchas muestras posibles del mismo tamaño que podrían haberse seleccionado de la población total de este tipo de establecimientos. Las estimaciones variarán de muestra a muestra y es probable que una estimación hecha con cualquiera de ellas sea diferente del valor obtenido mediante la enumeración y tabulación del valor de la característica para todos los establecimientos con menos de cinco personas ocupadas. Sin embargo, de acuerdo con las condiciones que hemos establecido (una muestra aleatoria sencilla con una ponderación igual al recíproco de la fracción de muestreo) el promedio de todas las estimaciones posibles basadas en la muestra será igual al valor calculado del empadronamiento completo de todos los establecimientos. Damos el nombre de error de muestreo a la diferencia entre la estimación basada en la muestra y el valor del universo.

La utilidad de la estimación basada en una muestra puede aumentarse en buena medida, si se conoce la magnitud de esa diferencia. En una situación real no podemos saber la magnitud exacta del error de muestreo ya que la estimación proviene de una sola muestra y se desconoce cuál de las muchas posibles estimaciones tenemos. Sin embargo, de los datos de muestreo podemos obtener una medida de la dispersión de todas estas estimaciones.

* La Sección 5, arriba, y esta sección son simplificaciones para el cómputo de estimaciones basadas en una muestra al azar sencilla y para estimar errores de muestreo. La exposición es limitada en virtud de limitaciones de espacio. Para un estudio más completo de la teoría y la técnica, véase Hansen, M.H., Hurwitz, W.N., y Madow, W.G., Sample Survey Methods and Theory, Vol. I: Methods and Applications, Nueva York, John Wiley and Sons, Inc., 1959, Caps. 4 y 5.

Esta medida se denomina error estándar de las estimaciones basadas en muestra. Como es natural, se relaciona con la dispersión de los valores individuales que componen el universo del que se obtuvo la muestra.

Esta idea puede ser ilustrada por los siguientes universos sencillos:

$$\begin{aligned} \text{Universo A: } X_1 &= \$1,000, X_2 = \$1,000, X_3 = \$1,000, \\ X_4 &= \$1,000 \end{aligned}$$

$$\begin{aligned} \text{Universo B: } X_1 &= \$1,000, X_2 = \$10,000, X_3 = \$100,000 \\ X_4 &= \$1,000,000 \end{aligned}$$

Cada establecimiento en el Universo A tiene el mismo valor, \$1,000. La dispersión es cero. Cada muestra de, digamos, $n = 2$ seleccionada de este universo, arrojará la misma estimación del total,

$$X' = \frac{N}{n} \sum_{i=1}^n X_i = \frac{4}{2} (\$1,000 + \$1,000) = \$4,000$$

Los valores individuales en el Universo B, sin embargo, son muy diferentes. Por lo tanto, en este caso, las diferentes muestras posibles de $n = 2$, producen estimaciones bien diferentes del total, pues se puede obtener cualquiera de las siguientes muestras y estimaciones correspondientes:

$$\begin{aligned} X_1 \text{ y } X_2 &: X' = \$22,000 \\ X_1 \text{ y } X_3 &: X' = \$202,000 \\ X_1 \text{ y } X_4 &: X' = \$2,002,000 \\ X_2 \text{ y } X_3 &: X' = \$220,000 \\ X_2 \text{ y } X_4 &: X' = \$2,020,000 \\ X_3 \text{ y } X_4 &: X' = \$2,200,000 \end{aligned}$$

Dejaremos como ejercicio el cómputo del promedio de todas estas posibles estimaciones de la muestra, y la comparación del resultado con el total real del universo.

Para todos los universos en que podemos interesarnos, la medida más útil de dispersión es la variancia. Es un promedio ajustado de los cuadrados de las diferencias entre los valores individuales y el promedio del universo. En símbolos:

$$\text{Variancia} = s^2 = \sum_{j=1}^N \frac{(X_j - \bar{X})^2}{N} \frac{N}{N-1},$$

donde $\bar{X} = \sum_{j=1}^N \frac{X_j}{N}$ = al valor promedio de la característica específica en todo el universo. La expresión $\sum_{j=1}^N (X_j - \bar{X})^2/N$ también

se considerará como promedio. Se multiplica por el factor de ajuste $(N/N - 1)$, porque el universo N es de tamaño finito. La raíz cuadrada de la variancia, que indicamos con el símbolo S , se denomina desviación estándar.

Ya se ha observado que la estimación de un total de una muestra dada, es sólo una estimación proveniente de un universo que contiene todas las estimaciones posibles. Este universo también tiene una variancia:

$$s_X^2 = \left(\frac{N - n}{Nn} \right) N^2 S^2 = \left(\frac{N}{n} - \frac{n}{n} \right) \frac{N^2}{N} S^2 = (W-1)N S^2,$$

donde n es el número de la muestra, y $W = N/n$ es la "ponderación de la muestra".

Cuando la muestra es de una población infinita, la variancia del promedio estimado es

$$\sigma_{\bar{X}'}^2 = \frac{\sigma^2}{n}$$

Cuando la muestra es de una población finita, la fórmula comparable para la variancia del promedio estimado es

$$s_{\bar{X}'}^2 = \frac{N - n}{N} \frac{s^2}{n}$$

El factor $\frac{N - n}{N}$ es un ajuste que refleja el hecho de que estamos incluyendo un porcentaje finito del universo en la muestra.

Para un total estimado X' , multiplicaríamos el promedio estimado \bar{X}' por N , i.e., $X' = N \bar{X}'$. Entonces como la variancia de un constante por un variable es igual al cuadrado del constante por la variancia del variable,

$$s_{X'}^2 = s_{N\bar{X}'}^2 = N^2 s_{\bar{X}'}^2 = \frac{N(N - n)}{n} s^2 = N(W - 1) s^2$$

Esta fórmula muestra explícitamente la relación entre $s_{X'}^2$, o sea la variancia del total calculado, y la variancia del universo s^2 . Si substi-

tuimos $\sum_{j=1}^N \frac{(x_j - \bar{X})^2}{N-1}$ por s^2 , obtendremos la igualdad

$$s_{X'}^2 = \left(\frac{N}{N - 1} \right) (W - 1) \sum_{j=1}^N (x_j - \bar{X})^2$$

Esta forma es más sencilla para propósitos de estimaciones pero dado que implica la suma de todo el universo, no se puede aplicar cuando sólo se dispone de los datos de una muestra. Anteriormente se apuntó que la suma

de una muestra se puede expandir a los niveles del universo con sólo multiplicarla por la ponderación W apropiada. Aplicando esta idea aquí, podemos escribir

$$s_{x'}^2 = \frac{N}{N-1} W(W-1) \sum_{j=1}^n (X_j - \bar{X})^2$$

donde se reemplaza el símbolo $S_{x'}^2$ por $s_{x'}^2$ para mostrar que ahora se está utilizando una estimación de la muestra.

Todavía queda un problema. Esta fórmula implica un promedio \bar{X} del universo que no se conoce. Podemos calcular \bar{X} fácilmente dividiendo la suma de la muestra por el número de la muestra, es decir, la estimación de la muestra $\bar{X}' = \sum_{j=1}^n X_j/n$. Sin embargo, si sustituimos \bar{X}' por \bar{X} , se tenderá a subestimar la variancia. Para compensar este sesgo, debemos multiplicar por el factor $(n/n-1)$, con lo cual se obtiene la fórmula:

$$s_{x'}^2 = \left(\frac{n}{n-1}\right) W(W-1) \sum_{j=1}^n (X_j - \bar{X}')^2 \quad \text{de}$$

donde
$$\sum_{j=1}^n (X_j - \bar{X}')^2 = \sum_{j=1}^n \left[X_j^2 - 2X_j \bar{X}' + (\bar{X}')^2 \right] =$$

$$\sum_{j=1}^n X_j^2 - 2\bar{X}' \sum_{j=1}^n X_j + \sum_{j=1}^n (\bar{X}')^2$$

Notando que $\bar{X}' = \frac{\sum_{j=1}^n X_j}{n}$, $n \bar{X}' = \sum_{j=1}^n X_j$, y $\sum_{j=1}^n (\bar{X}')^2 = n(\bar{X}')^2$,

esta expresión se convierte en $\sum_{j=1}^n X_j^2 - 2(\bar{X}') (n\bar{X}') + n(\bar{X}')^2$

$$= \sum_{j=1}^n X_j^2 - 2n(\bar{X}') + n(\bar{X}')^2 = \sum_{j=1}^n X_j^2 - n(\bar{X}')^2$$

Como resultado de esta operación $s_{x'}^2$ puede expresarse de la siguiente manera:

$$\begin{aligned} s_{x'}^2 &= \left(\frac{n}{n-1}\right) = W(W-1) \left[\left(\sum_{j=1}^n X_j^2 \right) - n(\bar{X}')^2 \right] \\ &= \left(\frac{n}{n-1}\right) W(W-1) \left[\left(\sum_{j=1}^n X_j^2 \right) - \left(\frac{\sum_{j=1}^n X_j}{n} \right)^2 \right] \\ &\doteq W(W-1) \left[\left(\sum_{j=1}^n X_j^2 \right) - \left(\frac{\sum_{j=1}^n X_j}{n} \right)^2 \right] . \end{aligned}$$

El símbolo "≈" significa aproximadamente igual y, como puede verse, las últimas igualdades se aproximan mucho ya que tanto el factor $(N/N-1)$ como el $(n/n-1)$ se acercan a la unidad.

Es importante notar que en esta sección nos hemos estado refiriendo a valores para todo el universo o para toda la muestra, y se ha utilizado el símbolo X_j para indicar el valor de cualquier establecimiento, a diferencia de la connotación X_{ij} que se utilizó antes para indicar el valor de cualquier establecimiento en una clase particular. La mayoría de nuestras estimaciones se van a referir a totales para grupos industriales particulares, o posiblemente a otras clasificaciones de tipo geográfico,

por ejemplo. En este contexto, si el establecimiento en cuestión no se está tabulando se considera que el valor de X_j es igual a cero.

Aplicando esta observación a la fórmula anterior para $s_{x'}^2$, tenemos que todos los valores X_j son cero excepto el que se refiere al grupo industrial i , de tal manera que podemos escribir $\left(\sum_{j=1}^n X_j\right)^2/n$ así:

$$\left(\sum_{j=1}^{n_i} X_{ij}\right)^2/n. \text{ De igual manera, ya que } X_j^2, \text{ también será cero cuando } X_j$$

sea cero, sólo aparecerán los cuadrados de los valores para los estable-

cimientos incluidos en el grupo especificado, $\left(\sum_{j=1}^n X_j^2\right)$ que también se

puede escribir así: $\left(\sum_{j=1}^{n_i} X_{ij}^2\right)$. De esta manera, ambas sumas en nuestra

variancia se limitan a los establecimientos incluidos en la categoría que se está tabulando. Sin embargo, debemos poner énfasis en que el número

de muestra n se refiere al tamaño total de la muestra. Por lo tanto,

$$\left(\sum_{j=1}^{n_i} X_{ij}\right)^2/n \text{ será muy pequeño en comparación con } \sum_{j=1}^{n_i} X_{ij}^2 \text{ y puede descartarse.}$$

Esto produce una sobreestimación de la variancia compensando la simplicación previa que nos llevó a subestimarla. Tomando en cuenta las consideraciones anteriores, llegamos, finalmente, a la fórmula aproximada para la variancia del total calculado que nos es de gran utilidad:

$$s_{x'}^2 = W(W-1) \sum_{j=1}^{n_i} X_{ij}^2$$

Como esta estimación de la variancia, podemos calcular su raíz cuadrada, $s_{x'}$, para encontrar el error estándar de la estimación y conocer el grado de confiabilidad de nuestra estimación. Si computamos la amplitud

entre los límites $X_i' - s_{x_i'}$ y $X_i' + s_{x_i'}$, podemos decir que la probabilidad de que se incluya el valor del universo X_i es de dos en tres. Si se utilizan límites más separados, las probabilidades serán correspondientemente mayores. Por ejemplo, si se utiliza el doble del error estándar, la probabilidad será aproximadamente de 95 en 100, y es prácticamente seguro que dentro de los límites definidos utilizando tres veces el error estándar se incluirá el valor X_i .

Hasta ahora sólo hemos considerado el universo de establecimientos con menos de cinco personas ocupadas. Sin embargo, debe notarse que el error estándar de la estimación se aplica al conjunto de establecimientos con menos de cinco personas ocupadas y aquellos con cinco o más. Esto es verdad porque no existe un error de muestreo para establecimientos con cinco o más personas ocupadas. En el Censo se incluirán todos los establecimientos de tal tamaño. De hecho, el universo se divide en dos partes, una de las cuales no contribuye al error de muestreo. En otras palabras, se aplica el mismo error estándar para los establecimientos más pequeños que para el total.

Frecuentemente resulta conveniente expresar la confiabilidad de la estimación de la muestra en términos relativos en lugar de hacerlo en términos absolutos. Esta medida, que se denomina error relativo de muestreo, o para mayor brevedad, error relativo (E.R.), consiste simplemente en el error estándar $s_{x_i'}$ dividido por el total calculado X_i' . Cuando el mayor interés reside, como aquí, en el total combinado de las clases de menos de cinco y de cinco o más personas ocupadas, el error relativo resulta particularmente apropiado. Nótese de nuevo que el valor de $s_{x_i'}$

para los establecimientos más pequeños también se aplica a la suma de ambas clases; por lo tanto, el error relativo del total combinado se encuentra dividiendo s_{x_i} por ese total combinado.

La Ilustración III-4-11 muestra el cómputo del error estándar del total calculado y el error relativo para la Industria 36.

En Providencia $s_{x_i}^2$ se computará con máquinas calculadoras de escritorio. Esto requiere la tabulación por lista de las tarjetas correspondientes a los establecimientos en la muestra y su uso como hoja de trabajo básica. El trabajo se facilita si se dispone de equipo de tabulación, que efectúe operaciones de multiplicación. En ese caso, las sumas ponderadas de los cuadrados de todos los establecimientos con menos de cinco personas ocupadas para una industria y una característica dada se perforan en tarjetas sumarias. A continuación, las tarjetas sumarias se tabulan por lista y más tarde se utilizan como lista de las variancias. Por lo tanto, esto significa que para obtener s_{x_i} , sólo se necesita la raíz cuadrada de los números.

7. Análisis Adicional de Selección del Diseño de la Muestra

En la Sección 2 examinamos los diversos planes de cobertura de que disponíamos, y se dieron razones para las selecciones principales. En aquel punto se supuso que había una forma conveniente y compacta para comparar la eficacia de los diversos planes, tomando en consideración tanto los costos como el volumen relativo de información que proporcionaría cada uno de los diversos sistemas. En realidad, la exposición se simplificó demasiado. Tomando en cuenta los puntos discutidos en las Secciones 3 a 6 se puede estudiar el problema más a fondo. En lugar de los "índices de

información", en cierta medida artificiales, de la Ilustración III-4-3, examinaremos la pregunta más decisiva de la confiabilidad de las estimaciones para los grupos industriales individuales; la pregunta de la relación entre confiabilidad y tamaño de la muestra --y por tanto, costo-- y la cuestión adicional entre errores de muestreo vs. errores que no son de muestreo. Incidentalmente, contestaremos la pregunta de por qué se seleccionó, en última instancia, una muestra del 10 por ciento en lugar de algún otro porcentaje.

Una posible definición del objetivo básico de un plan de muestreo para un censo consiste en la obtención de un grado igual de confiabilidad para todas las estimaciones a un nivel dado de publicación, por ejemplo, para todos los totales de los grupos industriales al nivel nacional. Ese objetivo implica que se considera a esos totales igualmente importantes. Para propósitos de planeamiento y de análisis económico en general, lo más probable es que este concepto se descarte rápidamente. Una idea más atractiva consiste en dar la misma importancia a los valores agregados iguales. La estimación de la actividad económica total en una provincia dada debe tener aproximadamente el mismo grado de confiabilidad que la estimación del total para un grupo industrial al nivel nacional, si ambos totales son más o menos iguales cuantitativamente. Asimismo, sería deseable que nuestra estimación fuese más confiable para aquellas industrias u ocupaciones de mayor importancia económica que para aquellos grupos poco importantes. Otra manera de expresar este concepto es que deseamos que los errores relativos de muestreo, s_{x_i} / x_i , sean aproximadamente iguales para agregados económicos más o menos iguales, y que varíen en sentido inverso al tamaño del total que se está calculando.

Aun en caso de una enumeración completa, los totales del Censo a diversos niveles de detalle siguen más o menos este patrón. Esto es así porque es inevitable que se presenten muchos errores, algunos positivos y otros negativos, en los datos individuales de los establecimientos. Estos errores ocurren igualmente cuando se trata de una cobertura total y cuando se trata de una muestra. Nos referimos a ellos como errores sistemáticos de elaboración o errores que no son de muestreo, para diferenciarlos de aquellos errores puramente de probabilidad debidos al muestreo. Aun cuando por lo general no se compensarán exactamente a todos los niveles de detalles, éstos tenderán a convergir más en un valor fijo al combinarse un número cada vez mayor de informes. Este valor del error neto se denomina sesgo total (total bias), B , cuando se combinan todos los datos de una enumeración completa. Asimismo, el sesgo promedio, \bar{B} , es el sesgo total dividido por el total de establecimientos en el universo. Para la casilla de una publicación que incluye N_i establecimientos, podemos esperar que los errores sistemáticos, en promedio afecten el total en $N_i \bar{B}$. Sin embargo, para agregados dados, como por ejemplo, datos para todos los establecimientos en una casilla particular de publicación, el efecto neto de errores de elaboración puede variar mucho del efecto promedio esperado. Esta variación tenderá a disminuir a medida que se amplía el agregado para incluir un número cada vez mayor de establecimientos, conducta que resulta similar a la de los errores de muestreo.

Es difícil, y en la práctica a menudo imposible, medir los errores que no son consecuencia del muestreo. Sin embargo, es muy importante reconocer su existencia. Afectan directamente a la selección entre una cobertura completa y una muestra.

Cuando se utiliza el muestreo es posible poner más cuidado en la elaboración de cada informe y en el control de los errores que no son consecuencia del muestreo, que cuando se ha de trabajar con el volumen enorme de informes de un Censo completo. Por tanto, con el muestreo podemos reducir el "sesgo" y también la variabilidad asociada con dichos errores. Estas ventajas, al menos, contrarrestan en parte el grado de confiabilidad que se pierde con el muestreo.

Para las casillas muy pequeñas, ni la enumeración completa ni el muestreo producirán resultados de alta calidad. Para las casillas grandes, los errores puramente accidentales o las variaciones debidas al muestreo pueden ser muy moderadas y comparan favorablemente con la variabilidad mayor de la elaboración y el "sesgo" de una enumeración completa pero menos eficientemente controlada. Además, para repetir una observación anterior, los totales importantes son de más utilidad y tienen mayor valor que los pequeños.

Nuestro plan de muestreo para Providencia pone énfasis sobre totales importantes, y toma en cuenta el principio de que se desea más confiabilidad. Primero estudiaremos el problema de relacionar el error de muestreo con el tamaño de la estimación, y más tarde examinaremos nuestras decisiones con respecto a la definición de totales importantes.

Hasta ahora sólo se ha descrito nuestro principio para relacionar cuantitativamente las estimaciones y sus errores correspondientes de muestreo. Para poder aplicar el principio debemos relacionar los errores relativos con sus estimaciones por medio de una fórmula específica. La función de esta fórmula debe ser clara cuando se conozcan los usos específicos de

los datos. Sería inútil fijar los objetivos de muestreo de las estadísticas generales de un censo en términos de un propósito único. La única utilidad de esta observación negativa es sugerir que se busque una relación general. Después de luchar con este problema en otros proyectos hemos llegado a la conclusión de que el error relativo debe estar en relación proporcional inversa a la raíz cuadrada de la importancia del tema que se está calculando, es decir,

$$\text{E.R.}_{X_i} = k/\sqrt{X_i} .$$

De acuerdo con esta fórmula, un tema que tiene 4 veces la importancia de otro, sólo debe tener un error relativo igual a la mitad del segundo. Si tiene 9 veces la importancia de otro tema, su error relativo debe ser igual a la tercera parte del segundo, etc. Puede observarse que esta fórmula está de acuerdo con la idea de que la exactitud debe aumentar o disminuir con los cambios en la importancia de la estimación. Utilizando los valores $\sqrt{X_i}$ en lugar de los valores X_i se moderan las diferencias en la exactitud de temas de importancia variable. Los analistas de los datos han aceptado que este tipo de moderación es acertada. Además, si la fórmula se sigue fielmente, tiene la ventaja de que las estimaciones a todos los niveles van a contener la misma relación. Por ejemplo, si se obtienen las estimaciones de \$36 millones y \$64 millones y su suma, los errores relativos respectivos (E.R.) deberán ser proporcionales a 1/6, 1/8, y 1/10.

En vista de los obstáculos con que nos enfrentamos en Providencia no se pudo aplicar totalmente este principio. Puede resultar alentador mencionar que en nuestra experiencia existen obstáculos de uno u otro tipo que afectan

cualquier plan de muestreo y provocan alguna desviación de la meta estadística "ideal". Como el muestreo no se puede controlar con toda precisión, aunque no exista este tipo de obstáculos no puede lograrse una concordancia exacta con el objetivo establecido para cada estimación. En la práctica no se considera que esta regla (o cualquier otra que pueda adoptarse) es un requisito rígido que debe satisfacerse con precisión para cada estimación. Por lo contrario, se interpreta como una guía general de nuestros objetivos, y se trata de **diseñar** la muestra de tal manera que en promedio se observe la relación deseada.

En este punto podría sentarse la objeción de que existen situaciones especiales conocidas que hacen que algunas cifras sean más importantes que otras. Por ejemplo, si se está considerando la ampliación de algún tipo de servicios, los datos relativos a este tipo de servicio pueden tener una importancia estratégica mucho mayor que la que demuestra sus ingresos monetarios (o cualquier medida común que se esté utilizando como índice de importancia). Este tipo de situación puede adaptarse fácilmente ponderando las cifras de ingresos de los diversos grupos, y trabajando con las cifras ponderadas al diseñar la muestra. En nuestra fórmula, simplemente se re-define cada X_i para que se refiera a la cifra de ingresos o ventas ajustadas, etc.

Esta técnica tiene su mayor utilidad cuando es posible emplear varias tasas de muestreo, porque así se puede asignar también un índice de importancia X_{ij} a cada establecimiento; estos se pueden agrupar en intervalos de tamaño de acuerdo con estos valores; y pueden asignarse tasas de muestreo a cada intervalo de tamaño en proporción a los puntos medios de los intervalos (\bar{X}_i).

Fundamentalmente, la teoría en que se basa este método implica un supuesto de que la variancia asociada con cada establecimiento es proporcional al cuadrado de su tamaño. En la práctica se ha encontrado que este supuesto funciona bastante bien cuando los límites de los valores X_{ij} 's son valores reales comparables como por ejemplo las ventas. Se obtienen también buenos resultados cuando las X_{ij} 's son valores ponderados independientemente, pero en este caso, el método tiene una propiedad adicional: Este ajusta automáticamente las tasas de muestreo asignadas a cada intervalo de clase de modo que reflejen los diversos grados de importancia asignados a las unidades en los diversos grupos industriales, etc. (Cuando sea posible, los establecimientos que requieren una ponderación adicional deberán separarse en intervalos de clase y antes de incorporarse en el muestreo. De hecho, el universo deberá dividirse en sub-universos o estratos, cada uno de los cuales entrará independientemente en el muestreo.)

El muestreo de Providencia no presenta problemas en cuanto al uso de clases múltiples o ponderación diferencial, en virtud de los fuertes obstáculos administrativos con que tuvimos que enfrentarnos. El muestreo sólo se aplicó a unos cuantos grupos: minería, manufactura, comercio por menor y servicios. Nos limitamos también al muestreo del intervalo de clase relativamente limitado que incluye los establecimientos con menos de cinco personas ocupadas; y fue deseable restringir el diseño a una sólo tasa de muestreo, además de la cobertura de 100 por ciento ya definida.

Si la parte "sin certeza" de nuestro universo hubiera cubierto un horizonte más amplio, es decir, hasta 10, 20 ó 50 personas ocupadas, se habrían definido varios intervalos de clase. Aquí no fue necesario efectuar estas

operaciones. La subdivisión del intervalo de clase de menos de cinco personas ocupadas habría arrojado beneficios casi nulos. Para un horizonte tan reducido se pueden descartar, sin mayores consecuencias, las diferencias dentro del intervalo de clase al diseñar la muestra.

Las posibles diferencias en la variación entre los grupos comerciales y los industriales que han de entrar en el muestreo presentan mayores dificultades. Como se explicó en la Sección 6, el error estándar de una estimación s'_x es una función de la variancia del universo S^2 o de la desviación estándar S . ~~Esta misma relación básica existe para grupos estándar S .~~ Esta misma relación básica existe para grupos comerciales individuales, provincias, etc., es decir, el error estándar del total calculado para un grupo i , s'_{x_i} , depende de S_i^2 , o sea la variancia del universo del grupo i . Si la variabilidad del grupo i es relativamente alta, será relativamente difícil obtener una estimación exacta de ese grupo y a la inversa.

En la Sección 6 se estudiaron diversas formas de estimar el error estándar s'_{x_i} aplicables sólo cuando se muestrea independientemente el universo completo de todos los grupos. Si se utilizan las cifras de muestreo para calcular el grado de confiabilidad logrado, las fórmulas indicadas pueden aplicarse con confianza. Su simplicidad permite utilizarlas de modo ventajoso, aún en sistemas refinados de muestreo en que, teóricamente, la relación entre s'_{x_i} y S_i debe definirse de manera explícita.

Al planificar la muestra nuestro interés principal es saber cuál es la forma funcional exacta que relaciona s'_{x_i} con S_{x_i} . Esta relación se puede derivar de la variancia de un total calculado como se explica en la Sección 6. Teníamos que con

$$S_{X'}^2 = N(W-1) S^2,$$

$$\text{con } S^2 = \sum_{j=1}^N \frac{(X_j - \bar{X})^2}{N} \frac{N}{N-1} = \sum_{j=1}^N \frac{(X_j - \bar{X})^2}{N}, \text{ y } \bar{X} = \sum_{j=1}^N X_j / N,$$

$$\text{de manera que } S_{X'}^2 = (W-1) \sum_{j=1}^N (X_j - \bar{X})^2 = (W-1) \sum_{j=1}^N X_j^2 - (W-1) \left(\sum_{j=1}^N X_j \right)^2 / N$$

Al estudiar un grupo particular i , para tabulación (industria, comercio, provincia, etc.) resulta conveniente definir $X_j = 0$ si el establecimiento j estaba clasificado fuera del grupo i . De esta forma, sólo se incluyen los valores X_{ij} (o sea los valores de un establecimiento j que se encuentra en el grupo i). De los valores N en el universo total, todos menos N_i desaparecen de los términos $\sum_{j=1}^N X_j^2$ y $\sum_{j=1}^N X_j$. Si aplicamos el mismo método aquí, podemos escribir:

$$S_{X'_i}^2 = (W-1) \sum_{j=1}^N X_j^2 - (W-1) \left(\sum_{j=1}^N X_j \right)^2 / N = (W-1) \sum_{j=1}^{N_i} X_{ij}^2 - (W-1) \left(\sum_{j=1}^{N_i} X_{ij} \right)^2 / N$$

Multiplicando y dividiendo el primer término por N_i , podría escribirse como

$$N_i(W-1) \sum_{j=1}^{N_i} X_{ij}^2 / N_i. \text{ Ahora, a } S_{X'_i}^2 \text{ reste y sume } (W-1)N_i \bar{X}_i^2. \text{ Esto nos dá:}$$

$$S_{X'_i}^2 = N_i(W-1) \sum_{j=1}^{N_i} \frac{X_{ij}^2}{N_i} - (W-1) N_i \bar{X}_i^2 + (W-1) N_i \bar{X}_i^2 - (W-1) \left(\sum_{j=1}^{N_i} X_{ij} \right)^2 / N.$$

Combinando los primeros dos términos:

$$\begin{aligned}
 S_{X_i}^2 &= N_i (W-1) \left(\sum_{j=1}^{N_i} \frac{X_{ij}^2}{N_i} - \bar{X}_i^2 \right) + (W-1) N_i \bar{X}_i^2 - (W-1) \left(\sum_{j=1}^{N_i} X_{ij} \right)^2 / N \\
 &= (W-1) N_i S_i^2 + (W-1) N_i \bar{X}_i^2 - (W-1) (N_i \bar{X}_i)^2 / N \\
 &= (W-1) N_i S_i^2 + (W-1) N_i \bar{X}_i^2 (1 - N_i / N).
 \end{aligned}$$

El primer término de este resultado es exactamente igual a la fórmula con que comenzamos, habiéndose reemplazado N_i en lugar de N y S_i^2 en lugar de S^2 .

El término adicional

$$(W-1) N_i \bar{X}_i^2 (1 - N_i / N)$$

aparece porque hemos actuado como si se estuviera calculando el total de un grupo particular de una muestra enteramente al azar de todos los grupos. Sin embargo, se recordará que en la Sección 4 relativa a la selección de la muestra, incluimos la ordenación de todos los establecimientos por grupo industrial o comercial y por Provincia dentro de los grupos antes de efectuar el muestreo.

A no ser por unas cuantas clasificaciones erróneas, este procedimiento equivale a un muestreo independiente y separado de cada una de las clasificaciones primarias que han de tabularse. Si cada clasificación se muestrea independientemente, desaparecería el término adicional

$$(W-1) N_i \bar{X}_i^2 (1 - N_i / N)$$

que resulta de un muestreo al azar. Por esta razón se incluyeron controles adicionales de clasificación y de selección de muestra en los métodos de la Sección 4. Métodos similares de control en términos de clases primarias para tabulación deberán utilizarse, cuando sea posible, en otras situaciones de muestreo.

Con el sistema que se está utilizando en Providencia, la situación no sufrirá una distorsión importante si decimos que los errores estándares de nuestras estimaciones se describen por la sencilla relación:

$$s_{x'_i} \doteq S_i \sqrt{(W-1) N_i} .$$

Esta fórmula está de acuerdo con lo que se espera por intuición. El error de muestreo de las estimaciones aumentará o disminuirá de clase a clase según cambie la variabilidad inherente de las respectivas clases. Ya que $s_{x'_i}$ es una medida absoluta de error, tampoco es sorprendente que varíe según el tamaño de cada clase N_i .

Nuestro conocimiento de Providencia no es suficiente para decir cómo cambia S_i de un grupo industrial o comercial a otro. S_i se expresa en las mismas unidades que la característica a que se refiere de tal manera que, naturalmente, datos diferentes tendrán valores de S_i considerablemente diferentes. La magnitud de un S_i que se refiere a ventas, por ejemplo, será muy diferente de la magnitud de S_i correspondiente al número de empleados. El mismo fenómeno puede suceder entre clasificaciones industriales diferentes. Por ejemplo, el valor de las ventas de las tiendas de muebles al por menor con dos personas ocupadas, por lo general serán mayores, y tendrán un S_i más alto que una peluquería de dos personas ocupadas.

Sin embargo, afortunadamente, la experiencia con otro tipo de datos ha mostrado que la relación entre la desviación estándar del universo y el promedio, S_i/\bar{X}_i para unidades de aproximadamente el mismo tamaño, es bastante estable en temas y clasificaciones diferentes. Este coeficiente, S_i/\bar{X}_i se denomina coeficiente de variación y se designará con el símbolo $C.V._i$. Se supuso que dentro de cada una de las amplias categorías de servicios y comercio por menor, industrias mineras y manufactureras, los $C.V._i$'s eran aproximadamente iguales. El supuesto es útil cuando no existe información más específica. Será un error de importancia para sólo pocos casos. Los resultados que se basan en el supuesto no serán exactos, pero por lo general su grado de aproximación es bastante aceptable para propósitos prácticos cuando se está trabajando con clases reducidas, como sucede en este caso.

Nos interesa principalmente el error relativo de totales calculados $E.R._{X'_i} = S_{X'_i}/X_i$. Por lo tanto, si dividimos por X_i , la expresión que se había derivado para el estándar absoluto, tendremos:

$$\begin{aligned} E.R._{X'_i} &= S_{X'_i}/X_i = (S_i/X_i) \sqrt{(W-1) N_i} \\ &= (S_i/N_i \bar{X}_i) \sqrt{(W-1) N_i} = C.V._i \sqrt{(W-1) N_i}. \end{aligned}$$

Ya que se supone que $C.V._i$ es constante en todos los grupos y que $W-1$ también es constante, los errores relativos de nuestro diseño de muestra serán aproximadamente proporcionales a $1/\sqrt{N_i}$. Si además suponemos que los valores promedios \bar{X}_i para características dadas son aproximadamente iguales para todos los grupos, los errores relativos se ajustarían al objetivo sugerido

$E.R. X_i' \doteq k\sqrt{X_i}$. Este será el caso para el número de personas ocupadas. De hecho, nuestro diseño de muestra asigna ponderaciones reducidas a los valores de los grupos que tengan promedio de ventas más altas, etc. que a los otros. Si ésta hubiera sido una objeción de peso, había sido necesario modificar las tasas de muestreo y por tanto W , para algunos grupos.

Aun cuando suponemos que $C.V._i$ es bastante constante dentro de cada grupo principal, sería poco realista suponer que es igual para establecimientos de comercio por menor, y de manufactura, y por lo tanto, se ha tomado en cuenta en nuestro análisis. Al desarrollar los Indices de Información de la Ilustración III-4-3 (Sección 2), los $C.V._i$'s que se asignaron fueron:

Minería y Manufactura	2
Comercio por mayor	3
Comercio por Menor y Servicios	1

Estas son conjeturas bien informadas que se basan en el conocimiento general de la conducta de los diversos grupos y la experiencia en otras encuestas. Sin embargo, constituyen una guía general y útil para nuestro análisis.

Además de nuestros supuestos $C.V._i$'s, se contaba para cada grupo principal, con estimaciones de la fracción del valor total agregado o las ventas contribuidas por los pequeños establecimientos y con el recuento de pequeños establecimientos por industria o grupo comercial. De las últimas cifras derivamos promedios redondeados del número de pequeños establecimientos por grupo dentro de cada una de las categorías principales. Estos tres grupos de cifras, $C.V._i$, la fracción de las ventas totales, F_i , y el

promedio de pequeños establecimientos \bar{N}_i , en conjunto, se tomaron en cuenta para derivar los Índices de Información.

Como se dijo antes, los Índices de Información son medidas un tanto artificiales. En realidad, son el complemento de los errores relativos ($1 - E.R.X_i$) de los totales calculados que combinan los datos de cobertura total de los establecimientos mayores con las estimaciones de la muestra de la clase de menos de cinco personas ocupadas. Las medidas directas - los errores estándares proyectados de los totales combinados - tienen una interpretación más definida. Son las medidas mejores y más comparables que se pueden obtener de los resultados esperados de los diversos planes de cobertura, porque no nos interesan los datos de los pequeños establecimientos per se, sino los totales que representan al conjunto de establecimientos en una industria, comercio y servicios, provincia, etc.

La Ilustración III-4-12, resume el cálculo de los diversos errores relativos. La ilustración presenta a los errores relativos en términos de casillas de publicación de un tamaño dado (\bar{N}_i), en lugar de las casillas específicas de los cuadros básicos del Censo. Esto se ha hecho para examinar con mayor detalle las consecuencias de las muestras seleccionadas y también porque no existen límites fijos a los datos que pueden obtenerse de un Censo. Para hacer frente a todas las peticiones potenciales, se necesitarían recursos y tiempo ilimitados para recopilar y procesar las cifras. En el mundo real, los planificadores del Censo se enfrentan a fuertes presiones para mantener costos bajos y publicar los resultados a tiempo. También se debe reconocer la limitación de la exactitud en las cifras impuesta, inevitablemente, por los errores sistemáticos y la utilidad

decreciente de detalles cada vez más refinados. La decisión última implica siempre una transacción entre estas dos posiciones en conflicto.

El error estándar de la muestra es también el error estándar del total combinado para clase por tamaño de menos de cinco y de establecimientos mayores. El error relativo del estimado combinado entonces es

$$E.R._{X'} = \frac{s_{X'}(1)}{X_{(1)} + X_{(0)}} = \frac{s_{X'}(1)}{X}$$

Donde $s_{X'}(1)$ es el error estándar para la clase de la muestra, $X_{(1)}$ es el valor total de la clase de muestra, y $X_{(0)}$ es el valor total de la clase enumerada completamente. Multiplicando el numerador y denominador del error relativo por $X_{(1)}$ obtenemos:

$$E.R._{X'} = \frac{X_{(1)} s_{X'}(1)}{X_{(1)} X} = F_{(1)} \frac{s_{X'}(1)}{X_{(1)}}$$

donde $F_{(1)} = \frac{X_{(1)}}{X}$ es la fracción del total contribuido por la clase de donde se obtiene la muestra. Ya que $X_{(1)} = N_{(1)}\bar{X}_{(1)}$ podemos escribir:

$$\begin{aligned} E.R._{X'} &= F_{(1)} \frac{s_{X'}(1)}{N_{(1)}\bar{X}_{(1)}} = F_{(1)} \frac{s_{(1)}\sqrt{(W-1)N_{(1)}}}{\bar{X}_{(1)}N_{(1)}} \\ &= F_{(1)}C.V._{(1)}\sqrt{(W-1)/N_{(1)}} \end{aligned}$$

Esta es la misma fórmula que aparece a principios de la página para el total estimado de la clase de la muestra solamente, excepto por el

factor adicional $F_{(1)}$. Ya hemos visto por la discusión en la página 51 que para la clase de la muestra solamente nuestro procedimiento de muestreo aproximadamente conforma con nuestro objetivo expuesto de que

$$E.R.X_i \doteq k / \sqrt{X_i}.$$

Por lo tanto, si $F_{(1)}$ es razonablemente constante para todos los grupos de industrias, i , nuestro objetivo se satisfaría también para los estimados de los totales combinados.

Lo anterior se ejemplifica por las decisiones específicas tomadas en el Censo de Providencia. Para los segmentos considerados de minería, industria, comercio por menor y servicios, el costo aproximado de una muestra del 5 por ciento ascendería a \$10,000, \$20,000 para una del 10 por ciento, y \$40,000 para una muestra del 20 por ciento. Una cobertura total de estos mismos establecimientos costaría cerca de \$200,000 (Cuadro III-4-~~8~~). Desde el punto de vista del costo resultaba evidente que era deseable efectuar algún tipo de muestreo, pero el nivel más apropiado aun no se conocía. Es claro que la confiabilidad estaba implicada. Como un porcentaje tan alto de los totales correspondientes de minería e industria se estaba incluyendo ya por la enumeración completa de los establecimientos con cinco o más personas ocupadas, el problema de la exactitud podía ignorarse por lo que a ellos respecta. Para los grupos del comercio por menor y servicios, la exactitud era importante.

La confiabilidad de los totales calculados para cualquier grupo particular se puede alterar sólo con cambiar la tasa de muestreo. Los valores de F_i , $C.V._i$, y N_i son constantes universales que no se pueden controlar

La W en nuestra fórmula del error relativo será mayor o menor dependiendo de la selección de la tasa de muestreo y también se pueden calcular con fiabilidad mayor o menor cantidad de totales de grupos individuales.

Ahora podemos enfocar con mayor precisión nuestro problema. Estamos buscando un equilibrio adecuado entre el precio de una W pequeña y el volumen potencial de detalle que proporcionará. Nuestras cifras de costo, los datos de la Ilustración III-4-12, y el cómputo de pequeños establecimientos por grupos individuales de actividad en las Ilustraciones III-4-8c y III-4-8d pueden facilitar este juicio.

Para ver cómo se pueden interpretar estas cifras en conjunto debemos primero recordar que para cualquier tasa fija de muestreo o W , el error relativo de grupos diferentes variará con $1/\sqrt{N_i}$. Todos los demás factores se consideran constantes. Por lo tanto, los coeficientes de error relativo para dos grupos diferentes cualquiera, h e i , lo proporciona el simple coeficiente de

$$RE_{x'_h} / RE_{x'_i} = (1/\sqrt{N_h}) / (1/\sqrt{N_i}) = \sqrt{N_i/N_h}.$$

Por ejemplo, el N_i del comercio por menor en la Ilustración III-4-12 es 6,000. Las Ilustraciones III-4-8c y III-4-8d muestran que nuestro universo consiste de 53,000 pequeñas tiendas de comestibles, etc., y 24,000 pequeños restaurantes, etc. Puede verse fácilmente que el grupo de tiendas de comestibles tendrán un error relativo de aproximadamente un tercio y el grupo de restaurantes un error relativo de un medio del correspondiente a la Ilustración III-4-12. También podemos determinar rápidamente que los errores relativos de otros cinco grupos de comercio por menor y servicios se aproximarán a los errores tabulados.

En conjunto, estos grupos representan cerca del 90 por ciento del total de la actividad de comercio por menor y servicios en Providencia. La tasa de muestreo de 10 por ciento proporcionará estimaciones con errores relativos de aproximadamente 2 por ciento o menos. También proporcionará estimaciones para los grupos menos importantes. Los errores considerados en forma individual crecerán rápidamente ya que el número de pequeños establecimientos en estos grupos disminuye con rapidez. Sin embargo, pueden incorporarse las estimaciones de los grupos menores para formar varios totales que pueden ser útiles.

En este punto debe recordarse la interpretación del error estándar de estimación y el correspondiente error relativo. Si queremos especificar límites amplios que incluyan casi con seguridad el valor del universo, debemos multiplicar el error relativo por 3. Con una muestra del 10 por ciento los errores máximos anticipados en el muestreo son de aproximadamente el 2 por ciento para el grupo de tiendas de comestibles, 3 por ciento para el grupo de restaurantes, y 6 por ciento para todos los demás grupos de moderada importancia en el comercio por menor y servicios.

En comparación, la muestra de 20 por ciento reduciría estos porcentajes en más o menos un tercio, pero el costo aumentaría de \$20,000 a \$40,000. A la inversa, el muestreo de 5 por ciento aumentaría los errores en más del 40 por ciento con sólo un ahorro de \$10,000.

Las preguntas: "¿Son más perjudiciales los errores máximos de 8 ó 9 por ciento para los grupos de comercio por menor y servicios que los de 5 ó 6 por ciento?; ¿justifica la diferencia los \$10,000?; ¿valdría la pena gastar los \$20,000 adicionales para reducir los errores al 4 por ciento

aproximadamente?", no pueden contestarse de modo definitivo e inequívoco. En última instancia, la legislatura u otra autoridad que asigna los fondos para el Censo decidirá el tamaño del proyecto. La ventaja del tipo de análisis arriba presentado proporciona unas bases sobre hechos cuantitativos que puede utilizar el administrador a cargo de iniciar el presupuesto para desarrollar el programa del Censo.

Puede argumentarse que los llamados hechos se derivan en gran parte de conjeturas, supuestos y aproximaciones. Esto es cierto. También resulta a menudo cierto en el muestreo que las desviaciones moderadas de los valores exactos tienen poco efecto.

Nuestro problema inmediato en el muestreo en Providencia ejemplifica este punto. Después de una gran cantidad de supuestos sobre la conducta de las variancias, coeficientes de variación, porcentajes de contribución que los pequeños establecimientos hacen a los totales, etc., nuestras decisiones finales habrían sido iguales. Cuando mucho se habría reducido el presupuesto total del Censo en \$10,000 o se habría aumentado en \$20,000. Parte de esta falta de flexibilidad es una propiedad inherente del muestreo. En este caso particular, una buena parte fue resultado de las restricciones administrativas existentes. Restricciones similares se presentan en todo tipo de situaciones de muestreo, aun cuando pudiéramos esperar una mayor libertad de selección. En trabajos posteriores en el campo del muestreo, cuando sea posible reducir las restricciones y cuando se conozcan más a fondo los universos que habrán de estudiarse y las necesidades en cuanto a datos, podrán hacerse diferencias más marcadas e introducirse planes de muestreo más eficaces.

Ilustración III-4-1

Porcentaje de Algunas Características de los Establecimientos con menos de 5 Personas Ocupadas

Actividad económica	Porcentaje			
	de estable- cimientos	Personas ocupadas	Ventas e ingresos	Valor agregado
(1)	(2)	(3)	(4)	(5)
Minería	30	2	-	1
Manufactura	60	5	-	5
Comercio por mayor	65	49	24	-
Comercio por menor	88	62	50	-
Servicios	88	63	54	-

COSTO DE LA ETAPA DE ENUMERACION SEGUN COBERTURA Y SISTEMA ADOPTADO

Cobertura y enumeración	Costo			
	Total		Enumeración de Establecimientos con menos de 5 personas ocupadas ***	
	\$	Porcentaje ^{1/}	\$	Porcentaje ^{2/}
(1)	(2)	(3)	(4)	(5)
<u>Empadronamiento Directo *</u>				
• Todos los establecimientos	\$266,704	1.00	\$145,906	55
• Todos los establecimientos con 5 o más personas ocupadas y una muestra de 20% con menos de 5	164,570	62	43,772	27
• Todos los establecimientos con 5 o más personas ocupadas y una muestra de 10% con menos de 5	142,684	53	21,886	15
• Todos los establecimientos con 5 o más personas ocupadas y una muestra de 5% con menos de 5	131,741	49	10,943	08
<u>Censo Utilizando el Servicio de Correo ***</u>				
• Todos los establecimientos	263,108	99	199,891	76
• Todos los establecimientos con 5 o más personas ocupadas y una muestra de 20% con menos de 5	103,195	39	39,978	39
• Todos los establecimientos con 5 o más personas ocupadas y una muestra de 10% con menos de 5	83,206	31	19,898	24
• Todos los establecimientos con 5 o más personas ocupadas y una muestra de 5% con menos de 5	73,212	27	9,995	14

1/ Con relación al costo de enumeración directa e integral.

2/ Con relación a las cantidades de la columna (2).

Supuestos en que se basan los costos del empadronamiento directo

a. Costo unitario del empadronamiento de establecimientos con 5 o más personas ocupadas: \$3.00

b. Costo unitario del empadronamiento de establecimientos con menos de 5 personas ocupadas: \$1.00

c. Costo unitario del empadronamiento de establecimientos en la muestra con menos de 5 personas ocupadas: \$1.50

* Supuestos en que se basan los costos del Censo utilizando el Servicio de Correos:

Actividad	Proporción enviada por correo o investigadas sobre el terreno	Costo Unitario	
		5 o más	Menos de 5
primer envío	1 00 (con respuesta de 40%)	\$.50	\$.50
segundo envío	60 (con respuesta de 30%)	.30	.30
tercer envío	30 (con respuesta de 10%)	.30	.30
investigación sobre el terreno	20	4.00	3.00

* Incluye comercio por mayor

Quadro III-4-3

Costos e Indices de Información Asociados con Métodos de Empadronamiento
(Por Correo) para Establecimientos con Menos de 5 Personas Ocupadas

(Todos los establecimientos con 5 o más personas ocupadas se incluirán en el censo)

Método a aplicarse a establecimientos con menos de 5 personas ocupadas (todos los de 5 ó más serán empadronados)	Minería y manufactura			Comercio por menor y servicios			Comercio por mayor		
	Costo		Indice de información.	Costo		Indice de información	Costo		Indice de información
	Total	"Menos de 5"		Total	"Menos de 5"		Total	"Menos de 5"	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Incluya todos los establecimientos	\$42,740	\$23,652	1.000	\$192,474	\$165,677	1.000	\$11,518	\$7,140	1.000
Muestra de 20%	23,818	4,730	.993	59,932	33,135	.987	5,805	1,428	.936
Muestra de 10%	21,453	2,365	.990	43,365	16,568	.980	5,091	714	.904
Muestra de 5%	20,271	1,183	.985	35,081	8,284	.972	4,734	357	.860
Incluya ninguno	19,088	---	.950	26,797	---	.480	4,377	---	.760

Los Indices de Información anotados arriba se definen como sigue:

Indice de Información = $\frac{\text{Información útil producida por el método en la Columna (1)}}{\text{Información útil producida por el empadronamiento total}}$

Ilustración III-4-4

Características que Habrán de Estimarse de la Muestra
de Establecimientos

(Explotación de Minas y Canteras e Industrias Manufactureras)

1. Organización jurídica de los establecimientos: (Tema 4 del cuestionario de manufacturas)
 - a. Propietario individual
 - b. En sociedad
 - c. Corporación o sociedad anónima
 - d. Cooperativa
 - e. Empresa gubernamental
 - f. Otra

2. Número de personas ocupadas durante la semana que terminó lo más cerca del 15 de noviembre de 1963 por categoría de empleados: (Tema 6)
 - a. Propietarios trabajadores y socios activos
 - b. Familiares sin paga y otros trabajadores no remunerados
 - c. Trabajadores de producción (operativos)
 - d. Todos otros empleados
 - e. Total

3. Remuneración semanal durante la semana que terminó lo más cerca del 15 de noviembre de 1963 por categoría de empleados: (Tema 6)
 - a. Trabajadores de producción (operativos)
 - b. Todos otros empleados
 - c. Total

4. Nómina de pago anual por categoría de empleado: (Tema 7)
 - a. Trabajadores de producción (operativos)
 - b. Todos otros empleados
 - c. Total

5. Número promedio de trabajadores de producción por sexo: (Tema 8)
 - a. Total
 - b. Masculino
 - c. Femenino

6. Capacidad total de motores eléctricos de 1 HP o más: (Tema 10b)

7. Costo de materiales, combustibles, energía eléctrica, trabajo por contrato, etc. durante 1963: (Tema 14)
 - a. Materiales, partes, componentes, envases, provisiones, etc. usados
 - b. Productos comprados y revendidos sin transformación
 - c. Combustibles consumidos
 - d. Energía eléctrica comprada
 - e. Costo de trabajo manufacturero por contrato hecho por terceros
 - f. Costo de servicios de reparación y mantenimiento de maquinaria y planta manufacturera prestados por terceros
 - g. Costo total

(continúa)

Página 2
Ilustración III-4-4

8. Valor de producción e ingresos de este establecimiento durante 1963:
 - (Tema 17)
 - a. Productos elaborados en este establecimiento
 - b. Trabajo por contrato

ILUSTRACION III-4-5

CARACTERISTICAS A ESTIMARSE DE LA MUESTRA DE ESTABLECIMIENTOS

Comercio por menor y Servicios

1. Organización jurídica de los establecimientos: (Tema 4 del cuestionario de comercio al por menor y servicios)
 - a. Propietario individual
 - b. En sociedad
 - c. Corporación o sociedad anónima
 - d. Cooperativa
 - e. Empresa gubernamental
 - f. Otra

2. Número de personas ocupadas durante el período de pago que terminó lo más cerca del 15 de noviembre de 1963, por categoría de empleados: (Tema 6)
 - a. Propietarios trabajadores y socios activos
 - b. Familiares y otros trabajadores no remunerados
 - c. Empleados y vendedores
 - d. Total

3. Nómina de pago total durante 1963: (Tema 7)

4. Fuentes de ingresos por tipos: (Tema 8)
 - Ventas de mercancías a comerciantes para revender
 - Ventas de mercancías a consumidores domésticos, individuales y agricultores
 - Ingreso por servicios a clientes
 - Otro ingreso
 - a. Ventas de mercancías:
 1. A comerciantes para revender
 2. Para exportar
 3. A empresas comerciales e industriales, instituciones, y agencias gubernamentales para su propio uso
 4. A consumidores domésticos, individuales y agricultores
 5. Total de ventas (suma del 1 al 4)

 - b. Ingreso por servicios a consumidores

 - c. Otro ingreso

Ilustración III-4-6

Cuadro de 200 Números Aleatorios,
del 1 al 10

03 10 08 06	06 10 02 04	08 03 03 06
04 07 06 09	07 03 03 01	08 09 04 07
06 05 04 02	03 01 04 01	04 04 03 08
09 03 06 03	05 07 07 04	08 07 04 06
08 02 10 06	02 10 05 08	09 02 05 01
02 09 09 07	04 02 01 05	03 04 05 01
07 04 10 09	02 09 09 10	01 09 02 08
06 09 06 04	07 07 05 06	05 09 06 08
05 03 07 08	06 01 10 01	07 04 10 10
01 02 02 02	04 10 04 05	08 10 10 02
01 02 05 07	02 08 01 10	10 08 10 10
09 04 04 04	01 09 10 05	04 05 03 07
01 06 04 07	05 03 04 01	10 03 03 02
03 10 03 07	06 05 05 04	01 10 07 03
06 09 06 07	07 01 05 04	09 10 01 10
	01 01 10 03	
	09 02 02 08	
	09 09 09 02	
	07 06 09 04	
	06 04 03 01	

Ilustración III-4-7a

Selección para Muestra de 1 en 5Núm. de Serie

1	
2*	Números aleatorios: 09
3	y 02
4	Establecimientos en la
5	muestra están marcadas
6	con asterisco
7	
8	
9*	$8/40 = 1/5$
10	
11	
12*	
13	
14	
15	
16	
17	
18	
19*	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29*	
30	
31	
32*	
33	
34	
35	
36	
37	
38	
39*	
40	

Ilustración III-4-7b

Selección para Muestra de 1 en 20Núm. de Serie

1	
2	Número aleatorio: 8,
3	8 es par.
4	Establecimientos en la
5	muestra están marcados
6	con asterico
7	
8*	
9	$2/40 = 1/20$
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28*	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	

ILUSTRACION III-4-8a

Hoja de Contro de la Muestra

(Explotación de Minas y Canteras)

Establecimientos con Menos de Cinco Personas Ocupadas

Tasa de Muestra: 1 en 10

Código	Descripción	Número de establecimientos con menos de 5 TP	Tamaño anticipado de la muestra	Tamaño actual de la muestra	Dif (5) - (4) (+ o -)*
(1)	(2)	(3)	(4)	(5)	(6)
11	Carbón	100	10		
12	Minerales metálicos	100	10		
13	Petróleo crudo	-	-	-	
14	Piedra, arena y arcilla	100	10		
19	Otros minerales no metálicos y explotación de canteras	100	10		
	Total	400	40		

* La diferencia no debe ser mayor de ± 3.0

ILUSTRACION III-4-8b

Hoja de Control de la Muestra

(Manufactura)

Establecimientos con Menos de Cinco Personas Ocupadas

Tasa de la Muestra: 1 en 10

Código	Descripción	Número de establecimientos con menos de 5 TP	Tamaño anticipado de la muestra	Tamaño actual de la muestra	Dif. (5)-(4) (✓ o -)*
(1)	(2)	(3)	(4)	(5)	(6)
20	Productos alimenticios	3,800	380		
21	Bebidas	400	40		
22	Tabaco	1,200	120		
23	Textiles	400	40		
24	Calzado y prendas de vestir	3,000	300		
25	Madera y corcho	400	40		
26	Muebles y accesorios	1,700	170		
27	Papel y productos de papel	20	2		
28	Imprentas, editoriales e industrias conexas	800	80		
29	Cuero y productos de cuero	300	30		
30	Productos de caucho	30	3		
31	Substancias y productos químicos	550	55		
32	Productos derivados del petróleo y del carbón	0	0		
33	Productos minerales no metálicos	1,400	140		
34	Industrias metálicas básicas	300	30		
35	Productos metálicos	900	90		
36	Maquinaria, exceptuando maquinaria eléctrica	200	20		
37	Maquinaria	400	40		
38	Material de transporte	100	10		
39	Industrias manufactureras diversas	900	90		
	TOTAL	16,800	1,680		

* La diferencia no debe ser mayor de ± 3.0

Ilustración III-1-8-c

Hoja de Control de la Muestra

(Comercio por Menor)

Establecimientos con Menos de Cinco Personas Ocupadas

Tasa de Muestra: 1 en 10

Código	Descripción	Número de establecimientos con menos de 5 TP	Tamaño anticipado de la muestra	Tamaño actual de la muestra	Dif. (5)-(4) (+ o-)*
(1)	(2)	(3)	(4)	(5)	(6)
6121	Abacerías y otras tiendas para la venta de productos alimenticios y bebidas alcohólicas	52,768	5,276.8		
6122	Farmacias	4,459	445.9		
6123	Tiendas de géneros textiles, prendas de vestir y calzado	8,649	864.9		
6124	Muebles y accesorios para el hogar	5,810	581.0		
6125	Ferreterías	2,537	253.7		
6126	Vehículos de motor y motocicletas	1,066	106.6		
6127	Estaciones de gasolina	1,497	149.7		
6128	Tiendas de departamentos y variedades	3,884	388.4		
6129	Comercio por menor, no clasificado en otra parte	6,992	699.2		
	TOTAL	87,662	8,766.2		

Ilustración III-4-8a

Hoja de Control de la Muestra

(Servicios)

Establecimientos con Menos de Cinco Personas Ocupadas

Código	Descripción	Número de establecimientos con menos de 5 TP	Tamaño anticipado de la muestra	Tamaño actual de la muestra	Dif. (5)-(4) (+o-) *
(1)	(2)	(3)	(4)	(5)	(6)
839	Servicios prestados a las empresas, no clasificados en otra parte	617	61.7		
841	Producción, distribución y exhibición de películas cinematográficas	200	20.0		
842	Teatros y servicios conexos	17	1.7		
843	Servicios de esparcimiento, exceptuando teatros y cinematógrafos	897	89.7		
852	Restaurantes, cafés, tabernas, etc.	23,742	2,374.2		
853	Hoteles y otros lugares de alojamiento	1,791	179.1		
854	Lavanderías y servicios de lavandería; limpieza y teñido	745	74.5		

(continúa)

Ilustración III-4-8-d

Hoja de Control de la Muestra

(Servicios)

Establecimientos con Menos de Cinco Personas Ocupadas

Código	Descripción	Número de establecimientos con menos de 5 TP	Tamaño anticipado de la muestra	Tamaño actual de la muestra	Dif. (5)-(4) (+o -)*
(1)	(2)	(3)	(4)	(5)	(6)
855	Peluquerías y salones de belleza	4,481	448.1		
856	Estudios fotográficos y fotografías comerciales	773	77.3		
859	Servicios personales no clasificados en otra parte	7	0.7		
	TOTAL	33,270	3,327.0		

* La diferencia no debe ser mayor de ± 3.0

ILUSTRACION III- 4-9

LISTA DE REMUNERACIONES SEMANALES PARA LOS 200
ESTABLECIMIENTOS CON MENOS DE 5 PERSONAS OCUPADAS EN LA INDUSTRIA 36

(Manufactura: Maquinaria, exceptuando maquinaria eléctrica)

Fuente de Información: Directorio de Establecimientos

Estable- cimiento Número	Total remun- eraciones semanales	Estable- cimiento Número	Total remun- eraciones semanales	Estable- cimiento Número	Total remun- eraciones semanales
1	103	41	50	81	121
2	208	42	84	82	45
3	151	43	40	83	137
4	210	44	176	84	87
5	75	45	143	85	83
6	177	46	99	86	134
7	32	47	168	87	63
8	61	48	5	88	90
9*	190	49*	132	89*	165
10	103	50	56	90	104
11	105	51	45	91	128
12	142	52	126	92	124
13	66	53	12	93	129
14	181	54	94	94	166
15	42	55	175	95	112
16	190	56	79	96	177
17	81	57	161	97	155
18	163	58	0	98	68
19*	119	59*	73	99*	75
20	195	60	79	100	52
21	48	61	72	101	123
22	6	62	78	102	139
23	133	63	146	103	72
24	95	64	148	104	45
25	150	65	106	105	59
26	41	66	108	106	159
27	47	67	81	107	142
28	39	68	115	108	159
29*	16	69*	88	109*	176
30	65	70	66	110	144
31	124	71	155	111	82
32	109	72	57	112	86
33	88	73	54	113	97
34	133	74	141	114	159
34	104	75	161	115	39
36	55	76	166	116	164
37	92	77	66	117	178
38	77	78	83	118	111
39*	93	79*	122	119*	78
40	55	80	81	120	94

(continúa)

Página 2
Ilustración III-4-9

Estable- cimiento Número	Total remu- neraciones semanales	Estable- cimiento Número	Total remu- neraciones semanales
121	117	161	131
122	78	162	87
123	178	163	132
124	121	164	146
125	120	165	91
126	99	166	73
127	171	167	116
128	165	168	111
129*	83	169*	90
130	111	170	142
131	163	171	92
132	136	172	119
133	182	173	134
134	100	174	95
135	132	175	103
136	156	176	87
137	135	177	145
138	100	178	129
139*	127	179*	86
140	146	180	103
141	129	181	106
142	132	182	75
142	132	183	144
144	141	184	112
145	76	185	127
146	134	186	133
147	128	187	107
148	145	188	143
149*	103	189*	91
150	142	190	89
151	142	191	105
152	115	192	116
153	94	193	104
154	130	194	132
155	79	195	143
156	135	196	98
157	121	197	90
158	145	198	111
159*	99	199*	120
160	124	200	91

$$N_i = 200$$

$$n_i = 20 = \frac{200}{10}$$

$$\sum_{j=1}^N X_{ij} = 22,091$$

$$\sum_{j=1}^{n_i} X_{ij} = 2,126$$

$$\bar{X}_i = 106.3$$

Ilustración III-4-10

Estimaciones de las Remuneraciones Totales para la Industria 36

(Manufactura: Maquinaria, exceptuando maquinaria eléctrica)

Estable- cimiento Número	Total remu- neraciones semanales x_{ij}	Industria 36: Manufactura de maqui- naria, exceptuando ma- quinaria eléctrica
(1)	(2)	
9	190	Tasa de muestra: 1 en 10
19	119	(1) Estimación del <u>Total de Remuneraciones</u>
29	16	<u>Semanales</u> para establecimientos con
39	93	menos de 5 personas ocupadas:
49	132	$X'_i = Wx_i$
59	73	$= \frac{200}{20} (2,126) = 10(2,126)$
69	88	$= \$21,260$
79	122	$X_i = \$22,091$ See III-4-9
89	165	(2) Estimación del <u>Total de Remuneraciones</u>
99	75	<u>Semanales</u> para todos los estableci- mientos
109	176	$X'_{(T)i} = X_{(0)i} + X'_i$
119	78	
129	83	
139	127	
149	103	
159	99	
169	90	
179	86	
189	91	
199	120	
Total	2,126 = x_i	

donde $X_{(0)i}$ es el valor de la característica para la industria de dos dígitos i

$$X'_{(T)i} = \$924,000^* + \$21,260$$

$$= \$945,260$$

$$X_{(T)i} = \$946,091$$

* Asuma que esta cifra pudo haberse obtenido si se hubieran enumerado todos los establecimientos con 5 ó más personas ocupadas en el grupo de industria 36.

ILUSTRACION III-4-11

Hoja de trabajo

Cálculo de Error Estándar del Total Estimado
y Error Relativo

Característica: Remuneración total Semanal.

Código de Industria	$\sum_{j=1}^{n_i} X_{ij}^2$	$W(W-1) \sum_{j=1}^{n_i} X_{ij}^2$	$\sqrt{(3)}$	Totales de la característica			Error relativo	
				Total esti- mado para grupo de menos de 5	Total para grupo de 5 ó más	Total (5) + (6)	Grupo de me- nos de 5 $\frac{(4)}{(5)}$	Total $\frac{(4)}{(7)}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
20							%	%
21								
22								
23								
...								
35								
36	256,062	23,045,580	4,801	21,260	924,000	945,260	22.5%	0.5%
...								
...								

* $W(W-1) = 10 \times 9 = 90$

Ilustración III-4-12

Cálculo del Error Relativo para Totales Estimados Correspondientes a Grupos Industriales Mayores y Algunas Tasas de Muestreo

Grupo Mayor	Establecimientos pequeños			E.E. $x_i = (F_i) (C.V._i)^{1/(w-1)} / N_i$		
	F_i	$C.V._i$	N_i	W=20	W=10	W=5
Minería y Manufactura	.05	2	800	.015	.010	.007
Comercio al por mayor	.24	3	500	.140	.096	.064
Comercio al por menor y servicios	.50	1	6000	.028	.020	.013

ILUSTRACION III-4-13

GLOSARIO DE TERMINOS Y NOMENCLATURA USADOS EN LA
GUIA DE LA MUESTRA

Muestra: Cualquier número n de establecimientos seleccionados de un universo de establecimientos N , $n \leq N$.

Muestra Aleatoria: Una verdadera muestra aleatoria es una en la que cada partida en el universo tiene igual ocasión o probabilidad de caer en la muestra, y cada combinación de muestra de partidas n tiene la misma oportunidad de ser seleccionada como cada dos combinaciones de muestra de unidades n . Algunas veces el término muestra aleatoria se usa intercambiabilmente con muestra de probabilidad.

Muestra Probabilística: Una muestra en que cada partida del universo se sabe que tiene una probabilidad conocida de caer en la muestra.

Estrato: Una de las divisiones únicas en las que se pueda dividir un universo para propósitos de seleccionar una muestra. Por ejemplo, si el universo de establecimientos se divide en dos partes: (a) aquéllos con cinco o más personas ocupadas, y (b) aquéllos con menos de cinco personas ocupadas, cada una de estas divisiones es un estrato.

Muestra Estratificada: Una muestra sacada de un universo que se ha dividido en estrata. Las probabilidades de selección de establecimientos en un estrato dado son independientes de las probabilidades de otros estrata; así, diferentes estrata pueden tener diferentes ponderaciones, W .

GLOSARIO DE TERMINOS

N = número total de establecimientos en el universo

n = número total de establecimientos en la muestra

X = universo total de cierta característica

x = muestra total de cierta característica

Página 2
Ilustración III-4-13

N_i , n_i , X_i y x_i se definen de la misma manera pero se refieren a un grupo particular de industria, provincia, u otra sub-clasificación.

X_j = valor de la característica bajo consideración para un establecimiento en particular, j .

$$N = \text{suma de } N_i, \text{ o } N = \sum_{i=1}^L N_i \quad *$$

$$n = \sum_{i=1}^L n_i$$

$$X_i = \sum_{j=1}^{N_i} X_{ij}$$

$$x_i = \sum_{j=1}^{n_i} X_{ij}$$

$$X = \sum_{i=1}^L X_i = \sum_{i=1}^L \sum_{j=1}^{N_i} X_{ij}$$

$f = n/N =$ la fracción de muestra o tasa de muestra

$W = 1/f = N/n =$ la ponderación de muestra

$X' = WX = \sum_{j=1}^n X_j =$ estimado de muestra del total del universo

$\bar{X} = \sum_{j=1}^N X_j / N =$ media del universo

* Esta anotación, explícitamente, significa:

El número total de establecimientos N es igual a la suma, empezando con el primer grupo ($l=1$) y terminando con el último grupo (L), del número de establecimientos en cada grupo (N_i).

$$\bar{X}' = X'/N = \sum_{j=1}^n X_j/n = \text{estimado de muestra de la medida}$$

$$S^2 = \frac{N}{N-1} \sum_{j=1}^N \frac{(X_j - \bar{X})^2}{N} = \text{la variancia del universo, un promedio de los cuadrados de las diferencias entre los valores individuales y la media del universo, ajustado para reflejar el hecho de que el universo es finito.}$$

$$S_{X'}^2 = \frac{(N-n)}{(N)} \frac{N^2 S^2}{n} = (W-1)NS^2 = \text{la variancia de estimados de muestra del total.}$$

$$s_{X'}^2 = \frac{W(W-1)n}{(n-1)} \sum_{j=1}^n \frac{(X_j - \bar{X}')^2}{(N-1)} = \text{fórmula exacta para estimar } S_{X'}^2 \text{ de los datos de la muestra.}$$

$$s_{x_i'}^2 = W(W-1) \sum_{j=1}^{n_i} X_{ij}^2 = \text{fórmula aproximada para computación práctica de un estimado de muestra de } S_{x_i'}^2 = \text{para un grupo dado, etc.}$$

$$S_{X_i'} = \sqrt{S_{x_i'}^2} = \text{error estándar del total estimado para un grupo dado, etc.}$$

$$S_{X_i'} = S_i \sqrt{(W-1)N_i} = \text{una fórmula aproximada para el error estándar absoluto de un total estimado para un grupo dado, etc.}$$

$$C.V._i = S_i/\bar{X}_i = \text{el coeficiente de variación, una medida relativa de la dispersión de un grupo de industrias, etc.}$$

$$R.E._i = \frac{S_i}{X_i'} = \text{error relativo del total estimado}$$

$$R.E._i = C.V._i \sqrt{(W-1)/N_i} = \text{una expresión para el error relativo del estimado de la muestra para un grupo particular en términos del coeficiente de variación, etc.}$$

$$F_i = \text{fracción del valor total sobre todas las clases de tamaños contribuidas por la clase de tamaño del que se está tomando la muestra.}$$

$$R.E._i = k/\sqrt{X_i'} = \text{guía sugerida para relacionar los errores relativos de varios estimados a la importancia de los estimados cuando se diseña la muestra.}$$

B = B /sesgo neto o error neto en el total del universo que resulta de errores fuera de la muestra. (errores de respuesta, codificación, perforación, etc.)

\bar{B} = B/N = cantidad promedio de sesgo por establecimiento en el universo.

B_i = $N_i \bar{B}$ = cantidad de sesgo esperada, por lo regular, en el total para un grupo particular de industrias, provincia, etc. que incluye establecimientos N_i .