

AGENCY FOR INTERNATIONAL DEVELOPMENT WASHINGTON, D. C. 20523 BIBLIOGRAPHIC INPUT SHEET	FOR AID USE ONLY
---	-------------------------

1. SUBJECT CLASSIFICATION	A. PRIMARY Economics
	B. SECONDARY Agricultural Economics

2. TITLE AND SUBTITLE Area Frame Sampling in Agriculture
--

3. AUTHOR(S) Houseman, E.E.

4. DOCUMENT DATE 1975	5. NUMBER OF PAGES 84 p.	6. ARC NUMBER ARC
---------------------------------	------------------------------------	-----------------------------

7. REFERENCE ORGANIZATION NAME AND ADDRESS Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C.

8. SUPPLEMENTARY NOTES (<i>Sponsoring Organization, Publishers, Availability</i>) (In U.S. Department of Agriculture Statistical Reporting Service, SRS No. 20) (Available in English and Spanish; Spanish, 107 p.: PN-AAC-140)
--

9. ABSTRACT

An over-all view of area frame sampling (a means of collecting agricultural data) including the construction of area sampling frames and the selection of area samples. Concepts and general principles of area sampling, rather than specific applications, are discussed. Sampling frames should be constructed in recognition of the fact that agriculture is composed of numerous subpopulations that must be sampled. A sample designed efficiently for one subpopulation might be of little value for another. Thus several sampling frames might be required or if a single sampling frame is to be constructed, it probably should be multi-purpose. The development of improved sampling frames is called for by: 1) the trend toward larger, more specialized farms, 2) the general demand for more accurate statistics, and 3) the need to keep sample sizes and costs as low as possible. Also, to some degree, sample size is related inversely to capability of controlling non-sample error, which also is in favor of efficient sampling to keep sample sizes as small as possible. Survey plans should include provision for studies of sampling variance, response errors, coverage errors, and costs. Such studies should provide a continuing basis for adjusting the allocation of resources in an effort to achieve maximum accuracy at a given cost.

10. CONTROL NUMBER PN-AAC-062 in English. PN-AAC-140 in Spanish	11. PRICE OF DOCUMENT
--	------------------------------

12. DESCRIPTORS	13. PROJECT NUMBER
	14. CONTRACT NUMBER RSSA/USDA 7-74 GTS
	15. TYPE OF DOCUMENT

EL MUESTREO POR AREAS EN LA AGRICULTURA

por

Earl E. Houseman
Estadístico Matemático

Statistical Reporting Service
United States Department of Agriculture

Washington, D.C. 20250

Junio 1976

Esta publicación es la versión en castellano de AREA FRAME SAMPLING IN AGRICULTURE, publicado en inglés por el Departamento de Agricultura de los EE. UU. de América en noviembre de 1975.

Traducción de Raymond Stadelman.

Traducción revisada por el Profesor Evelio O. Fabbroni, Centro Interamericano de Enseñanza de Estadística, en Santiago de Chile.

(N. del T.: Puesto que el Profesor Fabbroni tuvo oportunidad de corregir únicamente la penúltima versión de esta traducción, la responsabilidad de cualesquiera errores existentes en ésta recae sobre el traductor)

PROLOGO

Este es el segundo documento escrito por Earl E. Houseman bajo los auspicios de la AID, el SRS y el Centro Internacional de Programas Estadísticos de la Oficina del Censo, con los cuales coopera el SRS. El primer documento fue "Expected Value of a Sample Estimate", publicado por el SRS en septiembre de 1974 (y cuya traducción al castellano "Valor Esperado de un Estimador Muestral" apareció en enero de 1975). El señor Houseman se encuentra entre los primeros estadísticos que trabajaron en la aplicación del muestreo por áreas en la agricultura. También utiliza sus años de experiencia con el desarrollo y refinamiento de la metodología del muestreo por áreas actualmente usada por el SRS.

Este documento se produjo como parte de un esfuerzo continuo para suministrar materiales mejorados para la enseñanza y como referencia en el área de estadística agropecuaria para estudiantes extranjeros y para el desarrollo del personal que trabaja con estas agencias.

WILLIAM E. KIBLER
Administrador

PREFACIO

Esta publicación presenta un enfoque global del muestreo por áreas, incluyendo la construcción de marcos de muestreo y la selección de muestras. Varían mucho los recursos disponibles para la construcción de marcos para muestreo por áreas y las condiciones presentes durante la aplicación del muestreo. El objetivo de esta presentación es ofrecer ideas acerca de cómo hacer el muestreo por áreas y dar énfasis a los factores importantes que se relacionan con las decisiones que han de tomarse. Se describen los principios generales y no las aplicaciones específicas. Conceptos técnicamente válidos del muestreo ayudan a formar una base firme para cualquier encuesta por muestreo. Si los conceptos no cuadran con la encuesta, el estadístico debe buscar conceptos válidos que estén más de acuerdo con la realidad. Los procedimientos de encuesta se desarrollan de los conceptos. Por eso, un entendimiento pleno de los conceptos provee una base para tomar decisiones acerca de muchos problemas prácticos que ayudan a obtener buenos resultados en las operaciones. La tenencia de la tierra y los sistemas de producción agrícola difieren mucho entre los países y hasta entre las regiones de un país. Esto significa que los planes de muestreo tienen que adaptarse a las situaciones y propósitos individuales. En otras palabras, estudie bien la situación antes de adoptar los detalles de un plan que dio buenos resultados en otra situación.

En el desarrollo de un enfoque global del muestreo por áreas es necesario incluir muchas generalizaciones. El lector debe entender que frecuentemente es posible encontrar algunas contradicciones y excepciones. Muchas de las afirmaciones reflejarán metas, reconociéndose que los recursos o las condiciones a menudo son tales que no permiten el logro inmediato de la meta final. La pericia en el diseño de muestras, el conocimiento de las condiciones locales involucradas en la aplicación del muestreo por áreas, la experiencia con encuestas y la calidad y detalle de los mapas disponibles con indicación de caminos, topografía y uso de la tierra son factores importantes en el desarrollo y uso efectivo del muestreo por áreas.

Esta publicación se dirige a los estudiantes de muestreo y a personas que estén considerando el muestreo por áreas como un medio para la recolección de datos agropecuarios. Se supone que la mayoría de los lectores tendrán por lo menos un conocimiento elemental de la teoría del muestreo y algo de experiencia en agricultura. Sin embargo, lectores interesados en el tema pero que no han tenido adiestramiento formal en los métodos de muestreo, deben encontrar útil esta descripción del muestreo por áreas.

Earl E. Houseman

5.	<u>Ilustración Numérica</u>	33
5.1	Estimación por Dominios y el Segmento Ponderado	38
5.2	Variación de Muestreo	38
6.	<u>Discusión de las Tres Definiciones de Muestreo por Areas</u>	43
6.1	Segmento Cerrado contra el Abierto o el Ponderado	43
6.2	Segmento Abierto contra Segmento Ponderado	45
6.2.1	La variación de muestreo y costos	47
6.2.2	Error de cobertura	49
6.2.3	Combinación de métodos	52
7.	<u>Construcción de Marcos de Muestreo por Areas</u>	54
7.1	Introducción	54
7.2	Especificaciones de las Unidades de Marco	54
7.3	Información Auxiliar y su Utilización	57
7.3.1	Control del tamaño del segmento	58
7.3.2	Estratificación y la definición de unidades de marco	59
7.3.3	Selección de datos auxiliares sobre las unidades de marco	63
7.4	Mapas para la Construcción de Marcos	65
7.5	División de las Unidades de Marco en Segmentos	66
8.	<u>Construcción de Marco - Ilustración No. 1</u>	69
8.1	Una Encuesta sobre Areas en Cultivos	71
8.2	Una Encuesta sobre Datos Económicos	75
8.3	Una Encuesta sobre Ganado Bovino de Carne	79
9.	<u>Construcción de Marco - Ilustración No. 2</u>	81
9.1	Una Encuesta sobre Areas Cultivadas	82
9.2	Una Encuesta sobre Todas las Explotaciones Agrícolas	85
10.	<u>Resumen y Breve Mirada al Futuro</u>	86

EL MUESTREO POR AREAS EN LA AGRICULTURA

1. INTRODUCCION

El concepto del muestreo por áreas es muy sencillo. Se divide el área total que ha de estudiarse en N pequeños bloques, sin omisión ni traslape; se selecciona una muestra aleatoria de n bloques; se obtienen los datos de los datos de las unidades de información que estén en los bloques de la muestra; y, para estimar los totales de la población, se multiplican los totales muestrales por $\frac{N}{n}$. La sencillez de la idea contrasta fuertemente con lo complejo de su aplicación exitosa. Sin embargo, muchos de los problemas que se encuentran al aplicar el muestreo por áreas a la agricultura son característicos de las poblaciones estudiadas y por lo tanto son comunes a todos los métodos de recolección de datos, sea por muestreo o por censo. Los métodos usados en las encuestas varían mucho con respecto a su capacidad, real o potencial, de habérselas con los problemas que aparecen en la práctica.

El requisito mínimo para la aplicación del muestreo por áreas es poseer mapas en los cuales se pueden delimitar pequeñas áreas como unidades de muestreo que tengan linderos que puedan ser identificados con exactitud por un entrevistador en el terreno mismo. En las aplicaciones deben considerarse tres condiciones importantes: (1) las unidades de información han de definirse de tal manera que sirvan a los propósitos de la encuesta, (2) tiene que haber modo práctico de asociar las unidades de información con las unidades de muestreo y (3) el muestreo por áreas debe compararse favorablemente con métodos alternativos de muestreo también factibles.

1.1 Definiciones

Antes de proceder con la exposición, se repasarán algunos conceptos y definiciones:

Unidades de información son los elementos individuales o unidades que constituyen una población para fines de recolección de datos. No hay definición estándar de "unidad de información". Típicamente, se completa un cuestionario para cada unidad de información. En la exposición siguiente, el significado específico de "unidad de información" será usualmente "tract" (palabra en inglés que se definirá después) o "explotación agrícola".

Unidades de muestreo son las unidades en que se divide, con fines de muestreo, una población que ha de estudiarse. Son las unidades que están

expuestas a la selección aleatoria. Generalmente cada unidad de información se asocia con una sola unidad de muestreo. En el muestreo por áreas el número de unidades de información en una unidad de muestreo es variable.

Marco de muestreo es una lista completa (o especificaciones que establecerían una lista completa) de las unidades de muestreo que constituyen una población. Provee acceso a una población de modo que permite el muestreo probabilístico. Si cada unidad de información se asocia con una y solamente una unidad de muestreo y si hay M_i unidades de información asociadas con la i -ésima unidad de muestreo, la población consta de

$$M = \sum_{i=1}^N M_i$$

unidades de información, donde N es el número total de unidades de muestreo en la población.

El término "marco de muestreo" sugiere que un marco se usa solamente para fines de muestreo. Realmente, se necesita un marco también para llevar a cabo un censo, que supone recolectar datos de todas las unidades del marco. Por ejemplo, un procedimiento equivalente al muestreo por áreas ha sido usado desde hace mucho tiempo al levantar censos--quizá a partir de los primeros empadronamientos; se definen distritos de enumeración y cada uno es enumerado por uno o más enumeradores. La lista de distritos de enumeración es el marco por áreas que se usa al levantar el censo. De paso vale decir que una encuesta bien puede ser encuesta por censo o encuesta por muestreo; la única diferencia básica es que en el censo se procura enumerar todas las unidades del marco, en vez de enumerar sólo una muestra de ellas.

Segmento se refiere a un pedazo de terreno con linderos marcados en un mapa. En el muestreo por áreas, el área total correspondiente a la población por muestrear se divide en segmentos. Además de significar un pedazo de terreno, la palabra "segmento" se usa en vez de "unidad de muestreo por áreas". Cuando quiere decir "unidad de muestreo", "segmento" se refiere al agregado de unidades de información que constituyen una unidad de muestreo por áreas. Debe ser fácil distinguir por el contexto si "segmento" se refiere a un pedazo de terreno deslindado en un mapa o a una unidad de muestreo por áreas (grupo de unidades de información).

Eficiencia de muestreo se refiere a la variancia de muestreo de un plan (es decir, un método específico de muestreo y estimación) en comparación con la variancia de muestreo de otro. Generalmente se comparan las variancias de muestreo con la condición de que sean iguales las fracciones de muestreo o los costos. Salvo indicación en contrario, el uso de "eficiencia de muestreo" se referirá a

la comparación de métodos alternativos con la condición de usar la misma fracción de muestreo en ambos casos.

Muestreo por conglomerados es el término general para planes de muestreo en los cuales las unidades de muestreo son grupos (conglomerados) de unidades de información. Una unidad de muestreo por áreas es un "conglomerado" de unidades de información; es decir, las unidades de información asociadas con un segmento. En otras palabras, el muestreo por áreas es una clase de muestreo por conglomerados y es aplicable la teoría del muestreo por conglomerados.

Población de estudio (survey population en inglés) es la población que realmente se muestrea (o que se enumera completamente). Es definida por el marco de muestreo y los procedimientos para su uso. A veces es necesario distinguir entre la población de estudio y una población "objetivo".

Población objetivo (target population en inglés) es la población que uno podría desear estudiar si tuviera completa libertad de escoger, pero que, por razones prácticas, no es la misma población que realmente se muestrea. Por ejemplo, uno podría preferir estimar la producción total de un cultivo, pero decidir omitir algunas regiones donde las cantidades producidas son pequeñas.

Teóricamente, las estimaciones (inferencia estadística) derivadas de la muestra se refieren a la población de estudio muestreada y no a la población objetivo. El lector debe consultar los primeros cuatro capítulos del libro de Deming * que presentan una excelente exposición de marcos de muestreo y poblaciones, y un enfoque global de muestras y de la obtención de inferencias a partir de muestras. Los capítulos introductorios de otros libros sobre muestreo también tratan de los principios generales de muestreo y estimación.

Variación muestral es la variación de un estimador derivado de una muestra.

Eficiencia del diseño, a veces llamado "efecto del diseño", se refiere a la variación muestral correspondiente a cualquier diseño de muestra y estimador en comparación con la variación muestral correspondiente a otro diseño o estimador. El muestreo aleatorio simple a menudo se usa como la base para la comparación. En la exposición que sigue se usará a veces el término "eficiencia muestral" en lugar de "eficiencia del diseño".

* Deming, W. Edwards, "Sample Design in Business Research", John Wiley & Sons, 1960.

Error de cobertura se refiere a la omisión o duplicación de unidades de información, incluyendo la determinación incorrecta del área que constituye una unidad de información.

Error de respuesta se refiere a la inexactitud de los datos correspondientes a una unidad de información. Algunos estadísticos definirían errores de cobertura y de respuesta de manera algo diferente, pero estas definiciones son convenientes en la discusión del muestreo por áreas.

1.2 Primeros Desarrollos del Muestreo por Areas

Parece que las primeras ideas sobre el muestreo por áreas en los Estados Unidos se relacionaron con el muestreo no aleatorio (selección dirigida). Se buscó una selección de áreas aproximadamente del tamaño de divisiones civiles menores (DCM) o de distritos censales de enumeración (DCE) que sería una muestra permanente que permitiera la medición exacta de cambios anuales. Las DCM y los DCE eran unidades reconocidas que habían sido definidas en mapas, y había datos no publicados acerca de cada una, obtenidos en censos anteriores, que servirían de base para la selección de muestras. No fueron alentadores los resultados de las investigaciones de las DCM y los DCE como unidades de muestreo. Se consideró como demasiado grande el tamaño de muestra requerido para lograr niveles aceptables de variancia muestral. En aquella época poco se sabía de la relación entre el tamaño de las unidades de muestreo y la eficiencia muestral, pero algunos estudios indicaron que las unidades de muestreo probablemente deberían ser menores que las DCM.

Ahora sabemos que, en general, una unidad de muestreo tan grande como un DCE (que contiene de 75 a 100 ó más explotaciones agrícolas) es simplemente muy ineficiente. El grado de ineficiencia está relacionado con el tamaño de la unidad de muestreo (el número de unidades de información en cada unidad de muestreo) y con la importancia de la tendencia que tienen las explotaciones agrícolas de ser similares a sus vecinas. Puesto que los recursos agrícolas y el ambiente en la vecindad de un punto tienden a ser similares, muchas características de explotaciones vecinas han exhibido una fuerte tendencia a la similitud. Esto explica por qué, por ejemplo, una muestra del 2 por ciento de unidades grandes de muestreo tiene generalmente variancias de muestreo mucho más grandes que las de una muestra del 2% de unidades más pequeñas de muestreo que están más esparcidas. Es decir, datos muestrales de una muestra de 2.500 explotaciones, por ejemplo, vendrían de solamente 25 lugares si cada unidad de muestreo por áreas tuviera 100 explotaciones; pero si cada unidad de muestreo contuviera 5 explotaciones, habría 500 lugares en que se recogerían datos, y

las variancias de muestreo serían mucho más pequeñas.

Con respecto a encuestas agrícolas, la primera prueba importante del muestreo probabilístico por áreas en los Estados Unidos, utilizando pequeñas áreas como unidades de muestreo, ocurrió en el Estado de Iowa.* Se hicieron dos encuestas, una a fines de 1938 y otra a fines de 1939, usando como unidades de muestreo los cuartos de sección. (Un cuarto de sección es aproximadamente cuadrado, con media milla por lado y 160 acres de superficie.) En aquella época, el promedio del número de explotaciones agrícolas por cuarto de sección era aproximadamente 0,9. Para cada encuesta, la muestra representaba todo el Estado y era una muestra aleatoria estratificada geográficamente que constaba de unos 900 cuartos de sección muy esparcidos (equivalente a menos del 0,5% de la población).

Tomando en cuenta el tamaño pequeño de la muestra, los resultados de la encuesta eran muy alentadores. Los errores estándares relativos (coeficientes de variación) de estimaciones de características importantes generalmente fueron menos del 4 por ciento. Fue posible comparar las estimaciones derivadas de las muestras por áreas con otras fuentes de información, incluyendo un censo agrícola conducido por el Estado de Iowa y el censo federal de agricultura correspondiente a 1939. De la información obtenida acerca del error aleatorio del muestreo, de la experiencia ganada en el campo, y de comparaciones de las estimaciones muestrales con otras fuentes de información, se hizo evidente en esa época que mucha atención tendría que dirigirse en el futuro a la minimización de errores debidos a fuentes ajenos al muestreo. De esta experiencia y otras con el muestreo probabilístico empezó a desarrollarse una nueva perspectiva de los errores en las estimaciones derivadas de encuestas.

Un resultado de este ensayo de muestreo por áreas fue el desarrollo, finalizado en 1945, de un marco de muestreo por áreas para todos los Estados.**

* Jessen, Raymond J., "Statistical Investigation of a Sample Survey for Obtaining Farm Facts", Iowa State University Research Bulletin 304, junio 1942, Ames, Iowa.

** King, A.J. and Jessen, R.J., "Master Sample of Agriculture", Journal of the American Statistical Association, tomo 40:38-46, 1945.

2. ALGUNAS CARACTERISTICAS CLAVES DEL MUESTREO POR AREAS

2.1 Versatilidad

No tienen límite los usos posibles del muestreo por áreas. La población de estudio podría constar de unidades de información que sean hogares, personas, explotaciones agrícolas, plantas, animales, desmotadoras de algodón, vendedores de insumos agrícolas, tractores, parcelas de tierra, almacenes o cualquier otra unidad de información definible que pueda ser asociada con segmentos. Su adaptabilidad a usos específicos y su versatilidad son fuertes atributos del muestreo por áreas. Muchas necesidades de información han sido satisfechas cuando el muestreo por áreas fue la única manera disponible de seleccionar una muestra probabilística.

2.2 Cobertura

Conceptualmente, un marco de muestreo por áreas siempre está completo y actualizado con respecto a cualquier definición de unidad de información. Por ejemplo, una muestra por áreas de explotaciones agrícolas es una muestra de las explotaciones tal como existen y se definen en el momento de la encuesta. En otras palabras, al seleccionar una muestra aleatoria de $1/5$ de todos los segmentos en la población, se "espera" que la muestra contenga $1/5$ de las unidades de información en la población, sin importar cómo se definan las unidades de información. (La palabra "espera" se usa en el sentido de esperanza matemática.) Para aclarar más el punto, considere el estimador $\frac{N}{n}\sum x$. Son conocidos el número de segmentos, N , en la población y el número de ellos, n , en la muestra. El total muestral, $\sum x$, es el total de la característica X correspondiente a todas las unidades de información que están asociadas con la muestra de n segmentos. Por lo tanto, puede expandirse la muestra sin tomar en cuenta la definición de la unidad de información. Fíjese que no es necesario saber el número de unidades de información en la población para poder aplicar el muestreo por áreas. De hecho, de una muestra por áreas se puede estimar el número de unidades de información que contiene la población. Un estimador es $\frac{N}{n}(r)$, donde r es el número de unidades de información encontradas en la muestra de n segmentos.

El párrafo anterior indicó que un marco de muestreo por áreas está "conceptualmente completo", porque en la práctica el error de cobertura es probablemente mayor. Si se selecciona una muestra por áreas con la intención de usar $\frac{N}{n}$ como factor de expansión, tiene que hacerse con gran cuidado el trabajo de campo relacionado con la identificación de las unidades de información y su asociación con cada segmento en la muestra. Si la asociación de explo-

taciones agrícolas con los segmentos no se hace correctamente, la fracción de muestreo verdadera no será $\frac{n}{N}$. Por lo tanto, $\frac{N}{n}\sum x$ no será un estimador insesgado del total de la población.

2.3 Actualización

Un marco por áreas no se vuelve anticuado con respecto a la cobertura de la población; cambios en el uso dado a la tierra o en el número de unidades de información y su ubicación influyen en la variancia muestral pero no introducen sesgo. Pero algunos linderos de las unidades de muestreo pierden su identidad con el tiempo, lo que podría aumentar la posibilidad de sesgo debido a dudas acerca de la localización de los linderos. Hay dos motivos posibles para actualizar un marco por áreas: (1) mantener o lograr mejoras en la eficiencia del muestreo o (2) introducir mapas actualizados o nuevos para lograr mejores linderos de las unidades de muestreo. La actualización del marco por áreas puede hacerse por partes, según la necesidad.

2.4 Eficiencia

Las características de un marco de muestreo influyen mucho en la calidad de los resultados obtenidos de una encuesta. Deficiencias en el marco de muestreo pueden causar sesgos graves o una baja eficiencia de muestreo, o ambas cosas. Con el fin de minimizar los errores de cobertura, los estadísticos quisieran tener una lista actualizada y completa de todas las explotaciones (u otras unidades de información) para propósitos de muestreo. Pero las características agrícolas varían mucho; consecuentemente, para lograr el diseño de muestras eficientes para una amplia gama de propósitos, es importante tener información acerca de cada explotación en la lista. Por ejemplo, generalmente es útil tener una clasificación de explotaciones agrícolas según (1) tipo (por ejemplo, si es de ganado, árboles frutales u otros productos especificados), y (2) tamaño (preferiblemente una medida apropiada de tamaño para cada tipo). Obtener y mantener una lista completa de explotaciones agrícolas por tipo y tamaño es tarea mayor que debe considerarse como meta que ha de alcanzarse hasta el máximo factible.

Los atributos de un marco de lista (lista de operadores de explotaciones agrícolas) que lo hacen más eficaz para fines de muestreo también son valiosos para un marco de muestreo por áreas. Es decir, en el diseño de muestras por áreas, uno quisiera poseer información sobre el tipo y tamaño de cada segmento (unidad de muestreo) en la población. Pero la construcción de un marco de muestreo (sea de lista o por áreas) que permita un alto

nivel de eficiencia podría requerir una inversión grande, salvo que existiera información pertinente que pudiera ser incorporada fácilmente en el marco de muestreo. Pueden ser muy útiles los análisis técnicos de costos, variancias y sesgos en la valoración de las especificaciones alternativas factibles que podrían adoptarse para un marco de muestreo. Si no existe un buen acopio de experiencia, deben ensayarse las alternativas factibles adecuadamente antes de fijar las especificaciones finales y empezar la construcción del marco de muestreo. Realmente, los ensayos generalmente son aconsejables aunque haya experiencia que sirva de base.

2.5 Marcos por Areas como Complemento de Marcos de Lista

Una lista completa y actualizada de explotaciones agrícolas que contenga información pertinente acerca de las explotaciones es muy deseable para fines de muestreo. Posee fuertes ventajas con respecto a costo y eficiencia de muestreo. A pesar de esto, la cobertura de marcos de lista pronto se vuelve anticuada y a menudo es necesario el muestreo por áreas, a causa de deficiencias en los marcos de lista o su ausencia total. Como ya se ha dicho, el marco por áreas siempre está completo conceptualmente. Hay tres situaciones generales relacionadas con la aplicación del muestreo por áreas:

2.5.1 Marco de lista casi adecuado. Suponga que existe una lista de explotaciones agrícolas o una manera de formar una lista que defina una población de estudio que sea casi coincidente con la población objetivo. En este caso, la población de estudio definida por la lista podría aceptarse y una muestra seleccionada de la lista sería usada para la encuesta. Como manera de verificar la adecuación y completitud de la lista, podría usarse una muestra por áreas. Esto involucraría el cotejo de la lista con unidades de información de la muestra por áreas. Si la lista está completa, todas las unidades de información en la muestra por áreas deben estar en la lista. Pero el aparear unidades trae muchos problemas, porque una unidad de información no siempre se define y se identifica de la misma manera. Una discusión de los problemas de apareamiento está fuera del alcance de esta publicación.

Una consideración de costos, eficiencia de muestreo y un sinnúmero de factores técnicos podría conducir a una decisión de usar un marco de lista para el muestreo aunque la lista definiera una población de estudio que difiere en algo de la población objetivo. Un ejemplo podría ser una encuesta entre productores de trigo. Supongamos que exista una lista de productores que se cree adecuada pero cuya cobertura sea aconsejable investigar. Podría usarse el muestreo por áreas, pero sería necesario comunicarse con todos los agricultores en las unidades de muestreo por áreas para averiguar cuáles de ellos producen trigo.

Si la producción de trigo está muy esparcida y la proporción de productores de trigo es pequeña, razones económicas sugerirán que se tome la muestra de la lista. En este caso la encuesta podrá basarse sobre una muestra de la lista y una muestra pequeña por áreas podría usarse para obtener información acerca de la adecuación o calidad de la lista.

2.5.2 Marco de lista que cubre una parte de la población. Un marco de lista podría ser muy bueno aunque cubriera solamente una parte de la población sujeta a la encuesta. Si el marco de lista cubre una parte mayor o importante de la población y es satisfactorio aparte de ser incompleto, podría seleccionarse una muestra de él. Una muestra por áreas podría usarse para lograr la representación de aquella parte de la población que no esté incluida en la lista. Esto es un ejemplo de muestreo con marcos múltiples, o sea el uso simultáneo de dos o más marcos de muestreo. En algunas encuestas el muestreo con marcos múltiples posee ventajas importantes, pero esas ventajas son a menudo difíciles de lograr al estimar totales de la población, debido a la dificultad de determinar en la práctica qué unidades de información en la muestra por áreas se encuentran también en el marco de lista.

2.5.3 Inexistencia de un marco de lista adecuado. Puede ser que no exista un marco de lista y que no sea factible formar uno que provea un marco de muestreo satisfactorio, ni para una parte de la población. En tal caso, el muestreo por áreas ofrece la única posibilidad de seleccionar una muestra probabilística.

En las primeras dos situaciones (2.5.1 y 2.5.2) las unidades de información enumeradas en la muestra por áreas tienen que aparearse con las unidades de información en el marco de lista. Esta clase de muestreo por áreas se trata bajo la clasificación de muestreo con marcos múltiples, tema que está fuera del alcance de esta publicación. Nuestra exposición se limitará a la tercera situación (2.5.3).

3. TAMAÑO DEL SEGMENTO

3.1 Variación de Muestreo como Función del Tamaño del Segmento

"Tamaño del segmento" es un término general. Podría referirse, por ejemplo, al área de terreno dentro del segmento, al número de operadores de explotaciones agrícolas que viven dentro del segmento, a la cantidad de terreno regado o al área de terreno sembrado con árboles frutales. Pero en esta sección "tamaño del segmento" se referirá al número de explotaciones agrícolas que se encuentran "dentro" del segmento. Una explotación está "dentro" de un segmento si su sede está dentro del segmento. Esto se tratará en la sección 4.3 en relación con el método del segmento abierto.

Entre los factores que han de tomarse en consideración al definir segmentos se incluyen: variación de muestreo, costos, problemas asociados con los linderos de los segmentos, el detalle topográfico que aparece en los mapas disponibles y el método de asociar explotaciones con segmentos. Consideraciones de costo a menudo han dado origen a fuertes impresiones intuitivas que favorecen unidades de muestreo que son más grandes que lo debido. Esto indudablemente surge del hecho de que, para un costo fijo, más explotaciones pueden ser incluidas en la muestra cuando las unidades de muestreo son grandes. Se tratará del tamaño óptimo de segmento después de hacer un breve resumen de la situación respecto a la relación entre variación de muestreo y tamaño de segmento.

Para hacer hincapié en la diferencia entre las variaciones de segmentos grandes y las de segmentos pequeños, se presentan en la tabla 1 algunos resultados de un análisis inédito de datos recolectados en una encuesta agrícola en el Estado de Wisconsin. En esta encuesta las explotaciones agrícolas fueron enumeradas por "municipios" (en inglés, townships, que son las divisiones políticas más pequeñas en el Estado). Así es posible calcular las variaciones del muestreo por áreas cuando las unidades de muestreo son los municipios y comparar los resultados con las variaciones que resultan cuando las unidades de muestreo son las explotaciones individuales.

El promedio del número de explotaciones por municipio fue 69,5 y hubo un total de casi 102,000 explotaciones agrícolas en el Estado. Las columnas (2), (3) y (4) de la tabla 1 se explican en las notas al pie de la tabla. Se incluyó la columna (5) para hacer hincapié en un punto importante que se tratará después. Las columnas (6) y (7) muestran las razones entre las variaciones de muestreo para municipios y las variaciones para explotaciones agrícolas--se tratarán en los párrafos que siguen.

Tabla 1. Variancia relativa de municipios como unidades de muestreo comparada con la de explotaciones individuales 1/

Item	Porcentaje de explotaciones que informaron 2/	Promedio de explotaciones informantes por municipio 3/	Variancia relativa entre todas las explotaciones 4/	Variancia relativa entre explotaciones que informaron 5/	Variancia entre municipios relativa a variancia entre explotaciones 6/	
					Método 1	Método 2
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Tierra agrícola	100,0	69,5	0,73	0,73	-	-
Alfalfa	70,2	48,9	1,28	0,59	53,7	19,4
Maíz	66,5	46,2	3,33	1,88	26,0	14,2
Pastos	61,4	42,7	3,04	1,49	40,2	29,5
Vacas lecheras	58,7	40,8	1,34	0,37	43,6	10,9
Ganado de carne	26,4	18,4	16,1	3,50	9,0	7,2
Heno para ensilaje	15,9	11,1	13,2	1,25	8,1	4,9
Ganado vendido	7,5	5,2	163,9	11,36	3,7	3,5
Soya	4,1	2,8	73,4	1,99	9,5	9,3
Guisantes	3,1	2,2	134,2	3,24	5,0	4,6
Ovejas	2,7	1,9	138,2	2,76	2,1	1,9
Trigo (primavera)	1,2	0,8	488,0	4,74	4,1	4,1
Papas	0,7	0,5	789,0	4,76	3,2	3,2
Habichuelas	0,2	0,2	1.501,0	2,45	3,9	3,9

1/ Esta tabla se formó de un análisis inédito de datos de la encuesta agrícola de 1970 en Wisconsin.

2/ Porcentaje de explotaciones para las cuales $X_i > 0$, donde X_i es el valor de X para la i -ésima explotación.

3/ Número de explotaciones en el Estado para las cuales $X_i > 0$, dividido por el número total de municipios en el Estado.

4/ La variancia relativa entre todas las explotaciones es $\frac{F}{\bar{X}^2(F-1)} \sum (X_i - \bar{X})^2$, donde F es el número total de explotaciones en el Estado.

5/ La variancia relativa entre explotaciones que informaron es la variancia relativa de X entre las explotaciones que informaron sobre el ítem; es decir, las explotaciones para las cuales $X_i = 0$ no son incluidas en el cálculo de \bar{X} ni de la suma de cuadrados, $\sum (X_i - \bar{X})^2$.

6/ Vea el texto.

Al comparar las variancias de muestreo para municipios con las variancias de muestreo para explotaciones, se supuso el uso de muestreo aleatorio simple. Para los municipios se computaron las variancias correspondientes a dos estimadores distintos. El primer estimador fue un promedio por municipio:

$$x_1' = \frac{T}{t} \sum x_1$$

donde T es el número de municipios en el Estado, t es el número de municipios en la muestra y x_1 es el total de la característica X para el i-ésimo municipio en la muestra. El segundo estimador es un estimador por la razón:

$$x_2' = F \frac{\sum x_1}{\sum f_1}$$

donde F es el número total de explotaciones agrícolas en el Estado y f_1 es el número de explotaciones en el i-ésimo municipio de la muestra. Se incluyó el segundo estimador, x_2' , porque elimina de la variancia de muestreo la variancia entre municipios que está asociada con la variación en tamaño (número de explotaciones) de los municipios.

El estimador para una muestra aleatoria simple de explotaciones fue:

$$x_3' = \frac{F}{f} \sum x_j$$

donde f es el número de explotaciones en la muestra y x_j es el valor de la característica X correspondiente a la j-ésima explotación en la muestra.

Queremos comparar las variancias de muestreo para municipios y para explotaciones, suponiendo que las tasas de muestreo son iguales; es decir, cuando $f = 69,5t$. Así, pues, la columna (6) representa la variancia de x_1' dividida por la variancia de x_3' , suponiendo que $f = 69,5t$. De modo similar, la columna (7) es la variancia de x_2' dividida por la variancia de x_3' .

Por lo tanto, la primera entrada en la columna (6), por ejemplo, quiere decir que en el caso de la alfalfa, la variancia de muestreo para municipios, usando el primer estimador, x_1' , es 53,7 veces tan grande como la variancia de muestreo para explotaciones. Las columnas (6) y (7) pueden interpretarse también en términos de los tamaños de muestra necesario para lograr igual precisión (es decir, igual error de muestreo). Tomando como ejemplo el primer estimador y el caso de alfalfa, una muestra aleatoria simple de 100 explotaciones tiene la misma precisión que una muestra de 5.370 explotaciones cuando el municipio es la unidad de muestreo. Requeriría una muestra de aproximadamente 77 municipios para formar una muestra de 5.370 explotaciones agrícolas. La diferencia es mucho menor en el caso de otras características.

Fijese que la variancia de muestreo para municipios relativa a la variancia de muestreo para explotaciones individuales está relacionada con la proporción de explotaciones que informan tener el ítem (compare las columnas (6) y (7) con la columna (2)). Para algunos ítemes hay un promedio de menos de una explotación informante por municipio. (Vea la columna (3)). Si el tamaño del municipio se mide por el número de explotaciones que informan, entonces un municipio es una unidad de muestreo "pequeña" en el caso de algunos ítemes--los que están al pie de la lista en la tabla 1. La producción de estos ítemes es muy esparcida; en el caso de ellos, el municipio como unidad de muestreo tiene menos pérdida de eficiencia, como se ve en las últimas dos columnas de la tabla 1. Los resultados indican claramente que hay una pérdida grande en la eficiencia de muestreo cuando las unidades de muestreo por áreas tienen un número grande de explotaciones informantes, pero hay que tomar en cuenta otras consideraciones.

3.2 Variancia de Muestreo como Función del Porcentaje de Explotaciones que Informan

Las columnas (4) y (5) de la tabla 1 se incluyeron porque reflejan una situación general importante que tiene que reconocerse en el muestreo. Basado en el muestreo aleatorio simple de todas las explotaciones, se ve en la columna (4) que la variancia relativa de varios ítemes está estrechamente relacionada con la proporción de explotaciones que informan sobre el ítem. (Se da una definición de "variancia relativa" en la nota 4/ al pie de la tabla 1.)

En la columna (5), como se explica en la nota, se ve la variancia relativa cuando se eliminan de los cálculos de la variancia todos los valores de X_i que sean iguales a cero. Es la variancia relativa entre las explotaciones agrícolas que informaron sobre el ítem. Hay poca o ninguna relación entre las variancias en la columna (5) y los porcentajes de las explotaciones que informaron, columna (2).

La relación entre la variancia relativa de todos los valores de X (incluyendo cero) y la proporción de explotaciones que informan ha sido demostrada en la teoría de muestreo *. De hecho, la relación entre las columnas (4) y (5) es:

* Hansen, Hurwitz y Madow, "Sample Survey Methods and Theory", tomo 1, p. 122, John Wiley & Sons, 1953.

$$V_4^2 = \frac{V_5^2 + (1-P)}{P} \quad (1)$$

donde V_4^2 es la variancia relativa entre todas las explotaciones (columna (4)),

V_5^2 es la variancia relativa entre todas las explotaciones que informaron (columna (5)), y

P es la proporción de explotaciones que informaron; es decir, los valores en la columna (2) expresados como fracciones decimales en vez de porcentajes.

Suponga que se selecciona una muestra aleatoria simple de f explotaciones agrícolas y que x_3^j es el estimador del total de la población. La variancia relativa de x_3^j es:

$$\frac{V_4^2}{f} = \frac{V_5^2 + (1-P)}{fP}$$

suponiendo que la corrección por población finita, $\frac{F-f}{F}$, sea suficientemente pequeña para ser ignorada. Hemos notado que V_5^2 varía relativamente poco entre ítemes. Por eso, el valor de P es un factor de gran importancia en la determinación de la variancia relativa de x_3^j , el estimador derivado de la muestra.

La ecuación (1) también se aplica al muestreo por áreas, suponiendo una muestra aleatoria simple de segmentos. Suponga que hay N segmentos en la población y que N' es el número de segmentos en la población para los cuales X_1 es mayor que cero, donde X_1 es el total de X para el i-ésimo segmento de la población. Entonces

$P = \frac{N'}{N}$, y V_5^2 es la variancia relativa de X_1 entre los N' segmentos para los cuales X_1 es mayor que cero. Suponga que se selecciona una muestra aleatoria simple de n segmentos. La variancia relativa del total estimado,

$\frac{N}{n} \sum x_1$, es $\frac{V_5^2 + (1-P)}{nP}$, suponiendo que la corrección por población finita

sea suficientemente pequeña para ser ignorada. Sin entrar en una explicación completa, esto quiere decir que sería indeseable definir una población de segmentos en la cual la proporción de segmentos "cero" fuera más que un bajo porcentaje de todos los segmentos.

Muchos productos son producidos en menos del 20 por ciento de las explotaciones, y la ecuación (1) indica la gran variancia de muestreo cuando ese porcentaje es bajo. Esto señala la necesidad reconocida de usar lo que a menudo se llama "muestreo para propósitos especiales"; es decir, el desarrollo de marcos de muestreo y el diseño de muestras que sean eficientes con respecto a ciertos ítemes o propósitos. En esta publicación no es posible analizar las varias implicaciones de esto con respecto al muestreo de poblaciones agrícolas.

En breve, indica la inclusión, al grado que sea posible, de información en los marcos de muestreo acerca de quiénes producen los distintos ítemes, o información detallada sobre dónde se producen.

3.3 La Definición de Segmentos para Minimizar la Variancia de Muestreo

La variancia de muestreo es función de la variación entre segmentos. Por lo tanto, un objetivo en la definición de segmentos debe ser el de hacer tan pequeña como sea posible la variación entre segmentos. Es bien conocido, como se indicó en la sección 3.1, que la variancia de muestreo está relacionada con el promedio del tamaño de los segmentos y con la variación en el tamaño de los segmentos. Con respecto a variación en el tamaño de los segmentos, el objetivo es hacer los segmentos casi iguales en "tamaño", donde la medida de tamaño es una variable estrechamente relacionada con las variables que han de medirse en la encuesta. Si no es factible igualar los tamaños de los segmentos pero existe una medida pertinente de tamaño, la estimación por la razón ofrece una posibilidad de reducir la variancia de muestreo que está asociada con variación en el tamaño de los segmentos.

Con respecto al promedio del tamaño de los segmentos, y considerando únicamente la variancia de muestreo, el objetivo sería, generalmente, definir los segmentos de tal manera que haya una unidad de información en cada uno. Por ejemplo, si la encuesta propuesta se refiere solamente a explotaciones ganaderas, el objetivo será definir los segmentos para que hubiera una explotación ganadera en cada uno. Pero generalmente es muy limitada la información disponible para la definición de segmentos. Por esto, el grado de éxito en el logro del objetivo de tener segmentos de tamaños iguales es limitado por la naturaleza de la información pertinente que exista.

3.4 El Tamaño Optimo de Segmento

Aunque puede enumerarse una muestra aleatoria de 500 segmentos con 4 explotaciones cada uno a menos costo que enumerar una muestra aleatoria de 2,000 segmentos con 1 explotación cada uno, en el último caso el error de muestreo será más pequeño. El tamaño óptimo de segmento podría ser de dos o tres explotaciones, dependiendo de la relación entre costo y variancia. La experiencia acumulada sugiere el uso de segmentos muy pequeños; es decir, pequeños en términos del número de unidades de información tal como estas sean definidas para la encuesta. De hecho, un óptimo universal es difícil determinar y definir en la práctica, especialmente cuando se calculan estimaciones para muchas características y para varios dominios además de calcularlas para la población entera. Podría ser difícil evaluar una diferen-

cia de una o dos unidades de información en el promedio del tamaño de los segmentos. Sin embargo, suponiendo que se mantiene constante el costo de la encuesta, a medida que se aumenta el tamaño de segmento, llega uno al punto donde la variancia de muestreo aumenta rápidamente. Pequeñas desviaciones del óptimo carecen de importancia, pero las grandes podrían causar una pérdida grave de eficiencia de muestreo. Por lo tanto, procure establecer como objetivo un tamaño de segmento que esté cerca del óptimo, salvo que la falta de detalle topográfico para deslindar segmentos dicte lo contrario. En los Estados Unidos, considerando variancias y costos, la experiencia ha demostrado que el tamaño óptimo de segmento, para muchos propósitos, es menor que el mínimo práctico dictado por problemas asociados con linderos de segmentos y las limitaciones en el detalle topográfico que aparece en los mapas. *

El tamaño óptimo de segmento, tal como se trató en el párrafo anterior, se refería a la variancia de muestreo y no al error cuadrático medio, que es una combinación de variancia de muestreo y sesgo. Esto nos trae a asuntos de sesgo asociado con linderos de segmentos. La razón entre el perímetro de un segmento y su área es función de su tamaño y forma. La razón es mayor en el caso de segmentos pequeños, en comparación con los grandes; por esto, se espera que el efecto de cualquier sesgo asociado con ambigüedad acerca de los linderos del segmento será relativamente mayor en el caso de segmentos pequeños. Además, a medida que se disminuye el tamaño del segmento, se hacen más escasas las prominencias topográficas adecuadas para deslindar los segmentos. Por esto, en términos del error cuadrático medio, el tamaño óptimo de segmento podría ser más grande que el óptimo basado únicamente en la variancia de muestreo. Sobre este punto existe poca o ninguna información cuantitativa. Pero la experiencia indica claramente que debe darse alta prioridad al deslinde de segmentos por linderos que sean identificables efectivamente por los enumeradores en el campo. La decisión sobre el tamaño medio de segmento a menudo se reduce a la determinación del promedio más pequeño que sea práctico con la cantidad de detalle topográfico disponible.

* Houseman, Earl E. y Trelogan, Harry C., "Progress Toward Optimizing Agricultural Area Sampling". Proc. de la 36^a sesión del Instituto Estadístico Internacional, Sydney, 1967.

4. DEFINICIONES DE UNIDADES DE MUESTREO POR AREAS

4.1 Introducción

No es posible deslindar segmentos de tal manera que ninguna explotación agrícola traslape los linderos de los segmentos. Esto es la raíz de muchos problemas encontrados al asociar explotaciones con segmentos. Al hacer frente a tales problemas, los estadísticos han desarrollado tres métodos principales de aplicación del muestreo por áreas: el de segmento cerrado, el de segmento abierto y el de segmento ponderado. Estos tres métodos se refieren a tres maneras distintas de definir una unidad de muestreo por áreas. Pero, antes de explicar estos métodos, tenemos que definir "tract", concepto que juega un papel importante en cada uno de los tres métodos.

Un tract (palabra en inglés más o menos equivalente a "parcela" o "lote") es una subdivisión de un segmento que está bajo una sola administración. Puede ser una explotación agrícola entera, una parte de una explotación o un área de tierra no agrícola. Es decir, un tract es determinado por la definición de "explotación agrícola" y por los linderos de un segmento. Una explotación agrícola contiene uno o más tracts.

Con una excepción (que será tratada después), la aplicación rigurosa del muestreo por áreas exige que cada segmento en la muestra sea dividido en tracts y que todo el terreno dentro de los linderos de un segmento sea tomado en cuenta, como se indica en la figura 1. Esto es necesario para minimizar el error de cobertura. La descripción de los siete tracts en la figura 1 no se presenta como ilustración de la información que se necesitaría obtener en una encuesta verdadera. La información que hay que recoger y los detalles del procedimiento varían con el método de aplicación del muestreo por áreas. Puesto que en las páginas que siguen se harán referencias frecuentes a la figura 1, se sugiere a los lectores que la estudien ahora.

En los usos iniciales del muestreo por áreas se empleaba el segmento abierto, pero las dificultades encontradas en la práctica condujeron al uso del segmento cerrado cuando no era necesario que las unidades de información fueran explotaciones agrícolas. En las encuestas en que las unidades de información tengan que ser explotaciones agrícolas, sólo son aplicables los métodos de segmento abierto y segmento ponderado.

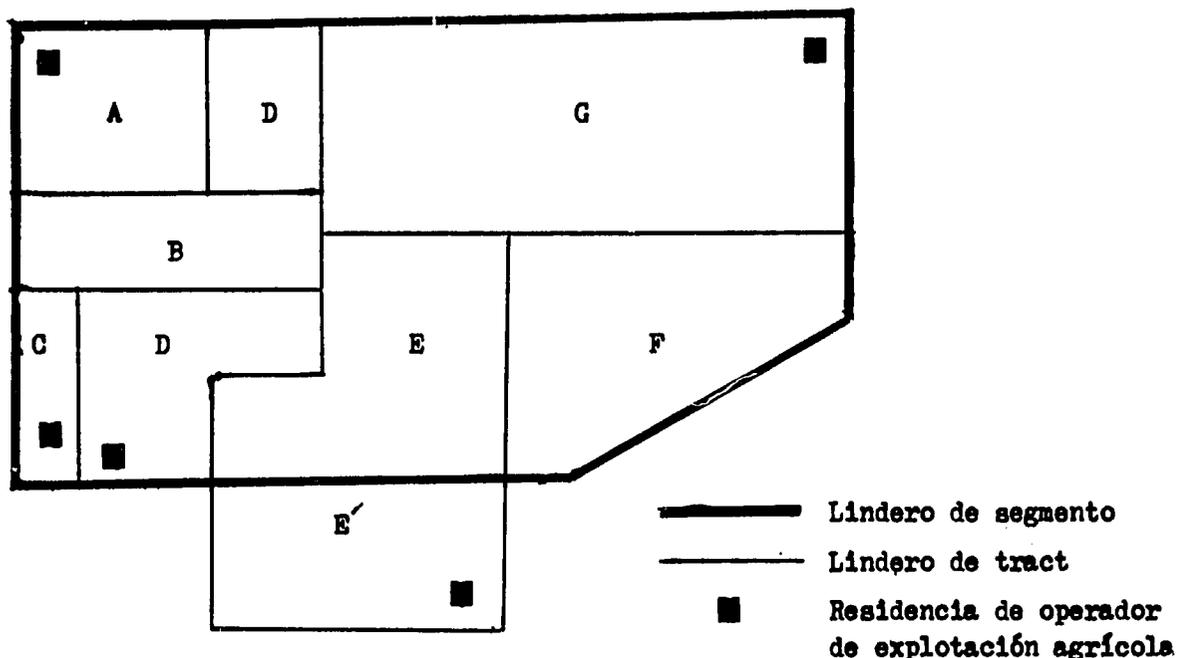


Figura 1. La división de un segmento en tracts.

Sigue una descripción de los tracts y explotaciones agrícolas relacionados con el segmento de la figura 1:

<u>Tract</u>	<u>Explotación</u>	<u>Descripción</u>
A	1	El tract A es una explotación entera; el operador vive en su explotación.
B	2	El tract B es una explotación agrícola, pero el operador no vive en su explotación ni dentro del segmento.
C	3	El tract C es un tract no agrícola. Uno de los dos hermanos que operan una explotación agrícola vive en este tract. Ninguna parte de su explotación está dentro del segmento. De acuerdo con las reglas que especifican a una persona como "operador" de una explotación, el hermano que vive en el tract C es el operador. De modo que una explotación (la 3) ha sido asociada con este tract, aunque ha sido clasificado como no agrícola porque no se hacen operaciones agrícolas en él.
D	4	El tract D se compone de dos parcelas o lotes separados pero dentro del segmento. Es operado por una persona que vive en el segmento y no tiene tierra fuera de él.
E	5	Los tracts E y E' forman la explotación número 5. Esto es ejemplo de un lindero de segmento que cruza una explotación y la divide en dos tracts. El operador vive en E'.
F	6	El tract F es parte de la explotación número 6. El resto de la explotación es un tract ubicado a varios kilómetros de este segmento. El operador vive fuera del segmento.
G	7	El tract G es parte de la explotación número 7. El operador vive en el segmento y en su explotación.

4.2 Método de Segmento Cerrado

Una fuerte ventaja del método de segmento cerrado es su sencillez. La idea es recolectar datos sobre ítemes o actividades específicos dentro de los linderos de los segmentos en la muestra. Por ejemplo, si se requiere información sobre el uso de la tierra, se recolectan datos sobre el uso dado a todo el terreno dentro de los linderos del segmento muestral. En caso de querer información sobre ganado, la meta es obtener información sobre todo el ganado que esté dentro de los linderos del segmento en el momento de la entrevista. Los tracts (tal como se definieron antes) son las unidades de información, salvo que sea más apropiada alguna otra definición de "unidad de información". Con referencia a la figura 1, el "segmento cerrado" se compone de todos los tracts A hasta G, inclusive. Si no se va a recolectar información acerca de tracts no agrícolas, se podría decir que el segmento cerrado se compone de seis tracts: A, B, D, E, F y G. (El tract D se compone de dos parcelas.)

En donde sea aplicable, el segmento cerrado posee una ventaja importante en comparación con los métodos de segmento abierto y segmento ponderado, porque se elimina la ambigüedad acerca de lo que es una explotación agrícola--una ambigüedad que origina errores de cobertura a causa de la omisión total o parcial de explotaciones o la duplicación de datos correspondientes a ellas. Para estudiar el uso de la tierra, incluyendo las áreas dedicadas a cultivos, generalmente se ha comprobado que el método de segmento cerrado es muy superior a los métodos de segmento abierto y de segmento ponderado, especialmente cuando existen fotografías aéreas que ayudan a identificar los linderos de los tracts. Casi todos los agricultores en los Estados Unidos conocen las áreas de sus campos y por lo tanto son capaces de informar correctamente sobre el hectaraje de los campos que están dentro de un segmento. Si el operador de un tract dentro de un segmento no puede ser entrevistado, los cultivos en el tract pueden identificarse y el hectaraje puede estimarse mediante el uso de fotografías o por otros medios. Por lo tanto, los errores de respuesta y de cobertura son relativamente pequeños. Además, la variancia de muestreo con el segmento cerrado es generalmente más pequeña que la variancia correspondiente al segmento abierto.

Desafortunadamente, en el caso de muchos ítemes, los agricultores no son capaces de suministrar datos exactos con respecto a ciertas partes de sus explotaciones--es decir, respecto a ciertos tracts dentro del segmento, como es requerido por el método de segmento cerrado. Por ejemplo, un operador de una explotación probablemente sabrá el número de jornales pagados y su valor; pero, si su explotación traslapa el lindero del segmento, es posible que no

sepa cuántos jornales se gastaron en un tract dentro del segmento. El problema que encara el productor al informar sobre un tract dentro de un segmento, en vez de informar sobre la explotación total, varía en dificultad, siendo relativamente fácil en el caso de área de terreno bajo cultivo, pero prácticamente insoluble en el caso de datos económicos, como compras de maquinaria agrícola o ventas de productos.

Los linderos de segmentos deben seguir prominencias topográficas, pero eso no siempre es posible y a veces la topografía cambia. Un entrevistador encuentra a veces casos en que una parte del lindero del segmento cruza una parcela. Tales casos podrían manejarse en una de dos maneras: (a) obtener información para toda la parcela y en la oficina decidir aleatoriamente omitir toda la parcela o incluirla toda; o preferiblemente (b) si existe base suficiente, estimar la proporción de la parcela que está en el segmento y multiplicar el total de la parcela por esa proporción. Se podría dar al entrevistador instrucciones para hacer tales determinaciones en el campo, pero esto es menos deseable que hacer la determinación en la oficina sobre la base de datos suministrados por el entrevistador. El personal de oficina debe ser capacitado a tal grado que sea menos propenso a introducir sesgo cuando es necesario hacer una determinación.

Puesto que el ganado camina de un lugar a otro, ocurren algunos problemas peculiares al ganado. Por ejemplo, aunque el lindero entre el tract E y el tract E' es visible en el terreno, podría ser posible que el ganado lo cruce, pasando de un tract al otro. En ese caso, el agricultor podría ignorar en el momento de la entrevista la localización exacta de su ganado con respecto al lindero del segmento. Este caso se trataría con el uso de técnicas como las sugeridas en el párrafo anterior. Los métodos de segmento abierto y segmento ponderado (que se tratarán más adelante) son también utilizables.

4.3 Método de Segmento Abierto

La idea general del método de segmento abierto es formular reglas prácticas que asocien cada explotación en la población con un solo segmento. Para lograr esto, se define para cada explotación agrícola un punto único de referencia que se llama "sede". Entonces, una explotación pertenece al segmento en que se encuentre su sede. Conceptualmente, la probabilidad que tiene una explotación de ser incluida en la muestra está clara. Es la misma que la probabilidad de seleccionar el segmento en que se encuentra la sede de la explotación.

Ha habido dos modos generales para identificar y delimitar una explotación: por el operador, y por la explotación.

4.3.1 Modo por operador. Este modo consiste en visitar cada segmento de la muestra en busca de operadores de explotaciones agrícolas. La residencia de un operador es, por definición, la sede de la explotación. Hay que visitar cada residencia dentro del segmento y averiguar si un operador vive en la casa. Se completa un cuestionario para la explotación de cada operador encontrado, sin importar dónde se ubique la explotación. Con referencia a la figura 1, las explotaciones numeradas 1, 3, 4 y 7 están en la muestra, porque los operadores de ellas viven dentro del segmento. No se recogerá información de las demás explotaciones.

La aplicación del modo por operador requiere la formulación de reglas que aseguren, por definición, una correspondencia de uno a uno entre operadores de explotaciones y las explotaciones mismas. Esto es necesario porque es posible aceptar a más de una persona como operador de una explotación dada. Un buen ejemplo de esto es una explotación operada en común por dos hermanos que viven en residencias distintas. Con el modo por operador, sería fácil que la explotación se contara dos veces (que es igual a tener dos oportunidades de ser seleccionada para la muestra), salvo que se siguieran fielmente reglas que definan a uno de los dos hermanos como operador de la explotación. Por ejemplo, con referencia a la figura 1, dos hermanos operan la explotación número 3. Uno de los hermanos vive fuera del segmento y el otro vive en el tract C dentro del segmento. Por definición, el hermano que vive en el tract C es el operador. Por lo tanto, la explotación número 3 se asocia con el segmento en la figura y no con el segmento en que vive el otro hermano.

Puesto que hay muchos casos en que más de una persona operan una explotación, es necesario formular un pequeño cuestionario que se usará al entrevistar en cada residencia dentro del segmento. Las preguntas tienen que ser formuladas con esmero y para permitir averiguar si alguien que vive en la residencia es operador de una explotación, de acuerdo con una definición de una explotación agrícola y una definición de operador de explotación que establezcan una correspondencia de uno a uno entre explotaciones y sus operadores.

Además del problema de la posibilidad de omitir o duplicar a causa de la ambigüedad acerca de la correspondencia entre operadores y explotaciones, existe otro problema importante que se encuentra a menudo en la práctica del modo por operador. Es el problema de encontrar a todos los operadores en un segmento que tiene muchas residencias no agrícolas (residencias no ocupadas por operadores agrícolas, como en las áreas urbanas). Puesto que es una labor mayor visitar todas las residencias en un segmento que contiene

muchas residencias no relacionadas con explotaciones agrícolas, puede ser necesario usar una técnica especial. Hay dos posibilidades:

(1) Que los entrevistadores visiten las residencias según su juicio, en un esfuerzo para localizar a todos los operadores. Es decir, en las residencias que visitan, hacen preguntas para averiguar si los residentes o los vecinos son operadores. Esta técnica no se considera satisfactoria, porque es muy fácil omitir a operadores.

(2) Formular un plan para seleccionar una submuestra aleatoria de residencias en el segmento. Por ejemplo, el segmento se podría dividir en segmentos más pequeños, de los cuales uno sería seleccionado para una muestra. (No olvide la necesidad de ponderar los datos a causa del submuestreo.) Podría preferirse usar desde el principio segmentos más pequeños en áreas residenciales, y usar fracciones de muestreo más pequeñas en tales áreas. Acuérdesse que en este caso se tiene un área con gran proporción de residencias no agrícolas. Otra cosa es un pueblo en que la mayoría de los residentes son operadores de explotaciones.

La dificultad de lograr la identificación completa de operadores que viven en segmentos muestrales dentro de áreas densamente pobladas, donde es baja la proporción de operadores de explotaciones agrícolas, combinada con la dificultad de aplicar reglas para establecer una correspondencia de uno a uno entre explotaciones y operadores, ha resultado en la adopción por parte de los estadísticos del modo "por explotación" en tales situaciones. El modo por operador no requiere la división del segmento en tracts, pero el modo por explotación sí lo requiere.

4.3.2 Modo por explotación. Esta técnica involucra la identificación y deslinde de una explotación y luego la determinación del operador u otro informante que sea capaz de dar información acerca de la explotación. La diferencia entre los dos modos es principalmente asunto de procedimiento: o primero se busca al operador y luego se identifica su explotación, o primero se busca la explotación y luego a su operador. En la práctica los resultados pueden ser distintos, aunque la definición de una explotación es igual para los dos modos; es decir, los errores de cobertura pueden ser diferentes. Además el modo escogido podría influir en la definición de segmentos. Esto se explicará en la sección sobre construcción de marcos.

Con el modo por explotación, la tarea es identificar las explotaciones que tienen sus sedes dentro del segmento muestral y completar un cuestionario para cada una. Generalmente no es suficiente darles a los entrevistadores mapas con los segmentos deslindados y mandarles completar cuestionarios para las

explotaciones cuyas sedes estén dentro de los segmentos, aunque se les dé también definiciones completas de "explotación agrícola" y de "sede". La experiencia ha demostrado que el éxito con el modo por explotación depende de la identificación completa y rigurosa de todas las explotaciones que tienen tierra dentro del segmento y luego la determinación de cuáles de estas explotaciones tienen sus sedes dentro del segmento. Parece necesario que los entrevistadores sigan, como mínimo, un procedimiento de tres pasos con la ayuda de un formulario especialmente diseñado:

Paso 1. Tomar en cuenta todo el terreno dentro de cada segmento en la muestra, mediante la división de cada segmento en tracts, como se ve en la figura 1.

Paso 2. Anotar, en un formulario especial, cada explotación que corresponde a un tract identificado en el paso 1 y obtener respuestas a las preguntas que, en el formulario, están destinadas a establecer el área de cada explotación. La idea es obtener respuestas a preguntas que determinan claramente los linderos, el área y la identidad de cada explotación de manera única.

Paso 3. Determinar la ubicación de la sede de cada explotación. Las preguntas que sean necesarias dependerán de la definición de "sede".

4.3.3 Problemas en la definición de la sede de una explotación. Las especificaciones operacionales de "sede" tienen que formularse de tal manera que cada explotación tenga un solo punto que pueda llamarse "sede". Ejemplos de puntos que podrían considerarse como sede son: la residencia del operador de la explotación, el sitio en donde se guardan los archivos de la explotación, la esquina nordeste de la explotación, el lugar donde se guarda la maquinaria agrícola, y la entrada principal de la explotación. Hay algo de ambigüedad en la aplicación de cualquier definición de "sede". Una residencia y su localización con respecto al lindero del segmento son muy distintivas, pero el grado de éxito logrado al usar la residencia del operador depende, entre otras cosas, de la obtención de una correspondencia de uno a uno entre operadoras y explotaciones. El uso de la esquina nordeste ofrece a menudo ambigüedad, porque la configuración geométrica de explotaciones varía mucho. Es posible que la maquinaria agrícola se guarde en más de un lugar y la entrada principal no es siempre determinable. Por esto, el mayor problema con el método de segmento abierto es la falta de sencillez y posibles ambigüedades en las especificaciones de la sede.

Con el modo por operador (sección 4.3.1) la residencia del operador de la explotación es el punto que lógicamente se define como sede de la explotación. Como se indicó en el párrafo anterior, el problema principal encontrado en la práctica se relaciona con la tenencia de la explotación y la identificación del operador. Si la tenencia de la explotación (o de la tierra) es tal que basten reglas sencillas para especificar a una sola persona como el único operador de la explotación, entonces el modo por operador (y el uso de la residencia del operador como sede de la explotación) podría ser la mejor técnica para la encuesta. Pero si los detalles de tenencia u organización de la explotación son complejos, o si es necesario examinar a muchas personas para identificar a los operadores de explotaciones agrícolas, podría ser más efectiva otra técnica.

Con el modo por explotación (sección 4.3.2) la residencia del operador también podría ser definida como sede de la explotación. En este caso, las preguntas en el paso 3 se dirigirían a la determinación exacta de quién es el operador. Entonces se averiguaría la ubicación de la residencia del operador para determinar si la explotación está "dentro" del segmento. Sin embargo, los procedimientos operacionales tienen que ser desarrollados y probados en detalle.

La explotación número 3 en la figura 1 suministra un ejemplo de la clase de detalle que tiene que considerarse en el proceso de formular especificaciones e instrucciones para los entrevistadores. Suponga que se utiliza el modo por explotación y que la sede de la operación se define como la residencia del operador. Según las especificaciones, la explotación número 3 está "dentro" del segmento ilustrado en la figura 1 porque la sede (residencia del operador) está dentro de este segmento. Pero, ¿incluirán correctamente esta explotación en la muestra los procedimientos de segmento abierto y modo por explotación, si es que el segmento de la figura 1 resulta seleccionado para la muestra? Recuerde que el tract C se describió como tract no agrícola. Si se incluyen solamente tracts agrícolas en la lista requerida en el paso 2 (vea la página 23), la explotación número 3 se omitiría en vez de ser incluida como es debido. La explotación número 3 ilustra un problema encontrado en el modo por explotación, pero que no se encuentra en el modo por operador. El problema es cómo tomar en cuenta la explotación cuando el operador no vive en su explotación y su residencia es por definición la sede de la explotación.

Una solución es incluir siempre la residencia del operador (el terreno en que está) como parte de la explotación. Esto demandaría procedimientos de dividir los segmentos en tracts de manera que el tract C (o un pequeño

lote en que está la residencia del operador) fuera identificado como parte de la explotación número 3. Para estar seguro de que la residencia del operador fuera siempre incluida como parte de una explotación, sería necesario visitar todas las residencias dentro de un segmento de la muestra con el propósito de identificar todas las residencias de operadores e incluirlas en las explotaciones. Este procedimiento nos conduce al modo por operador.

Una solución alternativa requiere definir condiciones distintivas que permitan una clara determinación de si un operador vive en su finca o no vive en su finca. Los operadores que viven en sus explotaciones agrícolas se llaman a veces "operadores residentes"; los que no viven en sus explotaciones se llamarían "operadores no residentes". En breve, el plan es como sigue: En el caso de explotaciones con operadores residentes, defina la residencia del operador como la sede. En el caso de explotaciones con operadores no residentes, algún punto distinto de la residencia del operador se definiría como la sede. Este plan ha sido usado en muchas encuestas; pero, con el modo por explotación, una definición mejor o generalmente aceptada de "sede" no ha aparecido. La búsqueda de una definición operacional satisfactoria continúa y probablemente continuará dondequiera y cuandoquiera se emplee el método de segmento abierto.

La siguiente definición de "sede" es una ilustración de algunos de los intentos que se han hecho. Representa un esfuerzo inicial de establecer una definición operacional de "sede" para un área donde una alta proporción de los operadores vivían en sus explotaciones. Se supone el uso del modo por explotación, y en áreas con muchas residencias no agrícolas requiere la búsqueda de explotaciones agrícolas en vez de operadores. Además, su aplicación requiere especificaciones operacionales (no incluidas aquí) para averiguar si un operador vive en su explotación agrícola. Tales especificaciones tienen que incluir una definición de "operador de explotación agrícola" que establezca una correspondencia de uno a uno entre operadores y explotaciones. La siguiente definición de "sede" no es necesariamente recomendable. Se presenta como una ilustración de criterios que podrían usarse en una definición operacional:

(1) Si el operador de la explotación vive en ella, su residencia es la sede.

(2) Si el operador no vive en su explotación, pero hay una sola casa ocupada en la explotación, esa casa es la sede.

(3) Si el operador no vive en la explotación y hay más de una casa ocupada en la explotación, la casa ocupada de mayor valor es la sede.

(4) Si no hay casas ocupadas en la explotación pero hay otros edificios, el edificio de mayor valor es la sede.

(5) Si no hay edificios en la explotación, la "entrada principal" de la explotación es la sede.

(6) Si ningún punto puede identificarse como la entrada principal, la sede será aquella esquina de la explotación que quede más al oeste y más al norte (en ese orden).

Como alternativa se podrían combinar las partes (2), (3), (4) y porciones de las partes (5) y (6) como sigue:

Si el operador no vive en su explotación y hay uno o más edificios en la explotación, el edificio de mayor valor es la sede.

Si no hay edificios en la explotación, la esquina de la misma que quede más al oeste y más al norte (en ese orden) es la sede.

Con referencia a la figura 1, no se dio suficiente información para ilustrar la aplicación de la definición dada arriba. Sin embargo, la definición provee alguna indicación de la posible complejidad que puede tener. Se debe buscar una definición sencilla, fácil de aplicar y tan libre de error como sea posible.

En la práctica, cualquier definición tiene que interpretarse con respecto a las muchas situaciones en que se encontrará. ¿Qué quiere decir "en la explotación"? ¿Qué es un edificio? ¿Qué es una explotación? ¿Quién es el operador? Afortunadamente, para la mayoría de las explotaciones las respuestas a estas preguntas están muy claras, pero hay muchos casos en los cuales la ambigüedad causa errores de cobertura. Se necesita tener mucha experiencia para poder desarrollar definiciones e instrucciones completas y bien adaptadas que sirvan para desarrollar programas de adiestramiento y procedimientos de supervisión del trabajo de campo que den resultados de alta calidad. Hay que ensayar y evaluar alternativas. Lo oneroso es la cantidad de detalles necesarios para hacer frente a las muchas situaciones que se presentan. No pase por alto la necesidad del equilibrio en el juicio. Por ejemplo, es posible enfocar tanto la atención en hacer completas las instrucciones que se pierda el énfasis sobre los puntos más importantes.

4.3.4 Observaciones generales. La experiencia en general con el método de segmento abierto demuestra una fuerte tendencia hacia la cobertura insuficiente. Por ejemplo, suponga una muestra por áreas del 5 por ciento. El número de explotaciones que determinan los entrevistadores como pertenecientes a la muestra tiende a ser menos del 5 por ciento. Aun con experiencia y con mucho énfasis sobre la definición correcta de las explotaciones y su debida asociación con los segmentos, es difícil reducir el error de cobertura a un nivel insignificante. De paso conviene decir que el error de cobertura varía de una característica a otra en la misma encuesta. Por ejemplo, hay muchas pequeñas operaciones agrícolas que presentan problemas de ambigüedad con respecto a si deben clasificarse o no como explotaciones. El resultado de la clasificación tendrá más efecto sobre una estimación del número de explotaciones que sobre una estimación de hectáreas en cultivo.

En síntesis, la ambigüedad sobre la sede de la explotación y la cuestión de ser o no ser una operación agrícola una explotación agrícola son fuentes importantes de errores de cobertura. Dichos errores pueden evitarse con el uso del método de segmento cerrado, cuando este es aplicable. Sin embargo, cuando se requiere que la unidad de información sea la explotación, se pueden considerar dos métodos posibles que no requieren el uso de la sede:

(i) El primero de estos métodos consiste en completar un cuestionario para cada explotación que está parcial o totalmente dentro de los segmentos de la muestra (refiérase al paso 2 en la página 23). Este método se llama "segmento ponderado", porque es necesario ponderar los datos; será explicado en detalle en la siguiente sección.

(ii) El otro método de evitar el problema de la sede no es generalmente factible. En él se da a cada explotación agrícola apuntada en la lista del paso 2 una probabilidad condicional de estar en la muestra, siendo esta probabilidad igual a la proporción de la explotación que esté dentro del segmento en la muestra; no se obtienen datos sobre el operador. No es factible, en opinión del autor, dejar que los entrevistadores hagan las determinaciones de las probabilidades. Por lo tanto, es necesario mandar las listas del paso 2 a la oficina para que se hagan allí las determinaciones aleatorias. Esta necesidad de enviar las listas a la oficina aumenta el costo y el tiempo requeridos para llevar a cabo la encuesta, en comparación con lo que serían si los entrevistadores procedieran con el paso 3 y las entrevistas necesarias. Además, la variancia muestral sería muy grande.

4.4 Método de Segmento Ponderado

El método de segmento ponderado requiere la recolección de datos de cada explotación que esté parcial o totalmente dentro de un segmento muestral. Los datos para cada explotación entonces se ponderan por la proporción de la totalidad de la explotación que se encuentre dentro del segmento.

Las reacciones iniciales de los estadísticos hacia el segmento ponderado a menudo han sido desfavorables por varias razones. Una es el hecho de que los datos tienen que ser ponderados. Otra es que solo la mitad, más o menos, de las explotaciones agrícolas en la lista del paso 2 en la página 23 tendrán sus sedes dentro de los segmentos en la muestra. Entonces, para un número dado de segmentos muestrales, el segmento ponderado incluirá el mismo número de explotaciones y cuestionarios que incluiría el segmento abierto. Si se considera la variancia muestral, suponiendo un número fijo de explotaciones en la muestra, la primera reacción también podría ser desfavorable en comparación con la variancia con otros métodos. Además, no se evitan las ambigüedades acerca de qué es una explotación. Sin embargo, el método de segmento ponderado es prometedor y merece investigación a fondo. En comparación con el segmento abierto el ponderado evita los problemas asociados con el establecimiento de sedes y parece tener mayor potencial para la reducción del error de cobertura. También, como veremos adelante, posee una variancia muestral mucho menor por segmento que la del segmento abierto. Este punto y otros se verán más claramente al presentar en detalle el método de segmento ponderado.

Es conveniente presentar el método de segmento ponderado en relación con una población entera de segmentos en vez de una muestra de segmentos. En efecto, cada explotación agrícola en la población se prorratea entre todos los segmentos en que se encuentra. Es decir, con referencia a un segmento dado, los datos para cada explotación que está parcial o totalmente dentro del segmento son multiplicados por la proporción de la explotación que está dentro del segmento. Por lo tanto, al sumar todos los datos de todos los segmentos, el resultado es el total correcta de la población. Esto se demostrará en una ilustración numérica en las páginas que siguen. Se sugiere que el lector pase a la ilustración numérica en la página 33 si encuentra dificultad con la formulación algebraica que sigue. No incluimos descripciones matemáticas correspondientes a los segmentos cerrado y abierto, porque la teoría del muestreo por conglomerados, tal como se presenta en los textos sobre muestreo, es suficiente.

4.4.1 Descripción algebraica del segmento ponderado. Suponga que A_j es la cantidad de terreno agrícola en la j -ésima explotación en la población, donde $j = 1, \dots, F$, y F es el número total de explotaciones en la población. Sea A_{1j} la cantidad de terreno agrícola de la j -ésima explotación que está dentro del i -ésimo segmento de la población, donde $i = 1,$

\dots, N . Entonces $P_{1j} = \frac{A_{1j}}{A_j}$ es la proporción de la j -ésima explotación

que está en el i -ésimo segmento. Si toda la j -ésima explotación está en el i -ésimo segmento, $P_{1j} = 1$. Si ninguna parte de la j -ésima explotación

está dentro del i -ésimo segmento, $P_{1j} = 0$. También $\sum_{j=1}^N P_{1j} = \sum_{j=1}^N \frac{A_{1j}}{A_j} = 1$, y

$\sum_{j=1}^{FN} P_{1j} = F$, donde F es el número de explotaciones agrícolas en la población

de la encuesta. (Recuerde que P_{1j} es proporción, no probabilidad.)

Suponga que X_j es el valor de alguna característica X correspondiente

a la j -ésima explotación. Entonces $\sum_{j=1}^F X_j$ es el total de X para la población.

El total de X para el i -ésimo segmento se define como

$$X_1 = \sum_{j=1}^F P_{1j} X_j \quad (2)$$

Excluyendo la posibilidad de errores de información, X_1 es un valor único para el i -ésimo segmento. Al sumarse todos los valores de X_1 en todos los segmentos de la población, el resultado es igual al total de la población. Así, pues,

$$\sum_{i=1}^N X_1 = \sum_{i=1}^N \sum_{j=1}^F P_{1j} X_j = \sum_{j=1}^{FN} P_{1j} X_j$$

Observe que

$$\sum_{i=1}^N P_{1j} X_j = X_j \text{ porque } \sum_{i=1}^N P_{1j} = 1.$$

Entonces, sigue que $\sum_{i=1}^N X_1 = \sum_{j=1}^F X_j$, lo que indica que $\sum_{i=1}^N X_1$ es el total correcto.

La ecuación (2) puede escribirse en una forma más conveniente para trabajar con datos muestrales. Sea $k = 1, \dots, f_1$ el subíndice para explotaciones asociadas con el i -ésimo segmento. "Asociadas con" se refiere a todas las explotaciones que están parcial o totalmente dentro del segmento. Sean X_{1k} el valor de X para la k -ésima explotación en el i -ésimo segmento

y P_{1k} la proporción de la k -ésima explotación asociada con el i -ésimo segmento que está dentro del i -ésimo segmento. Entonces X_i puede escribirse como sigue:

$$X_i = \sum_k^{f_1} P_{1k} X_{1k} \quad (3)$$

(Parecía al autor algo más fácil usar la ecuación (2) en vez de la ecuación (3) para mostrar que la suma de los X_i es igual al total correcto.)

4.4.2 Estimadores y sus variancias. Puesto que hay un valor único, X_i , para cada segmento en la población, es aplicable la teoría de muestreo por conglomerados. Puede usarse cualquier plan de muestreo adecuado para la selección de una muestra de segmentos.

Por su sencillez y para ilustrar cómo se podrían hacer estimaciones a partir de una muestra, suponga una muestra aleatoria simple de n segmentos. Sea x_{1k} el valor de X para la k -ésima explotación asociada con el i -ésimo segmento en la muestra. El cuestionario tiene que suministrar un valor numérico de A_{1k} y de A_k para que se pueda calcular $p_{1k} = \frac{A_{1k}}{A_k}$, donde p_{1k} es la proporción de la k -ésima explotación que está dentro del i -ésimo segmento. (Conviene recordar que A se definió como tierra agrícola. Es necesario investigar otras posibles medidas de la proporción de una explotación que está dentro de un segmento.) Entre los estimadores de interés se incluyen:

Estimador del total de X en la población:

$$\hat{X} = \frac{N}{n} \sum_{1k}^{nf_1} P_{1k} x_{1k} \quad (4)$$

Estimador del número total de explotaciones, que se obtiene al poner $x_{1k} = 1$:

$$\hat{F} = \frac{N}{n} \sum_{1k}^{nf_1} P_{1k} \quad (5)$$

Estimador del promedio del valor de X por explotación:

$$\frac{\hat{X}}{\hat{F}} = \frac{\sum p_{1k} x_{1k}}{\sum p_{1k}} \quad (6)$$

Podría simplificarse la notación en los estimadores al usar un solo subíndice para explotaciones en la muestra, pero se necesitan los subtota- les por segmentos para estimar el error de muestreo.

Sean $x_1 = \sum_k p_{1k} x_{1k}$ y $p_1 = \sum_k p_{1k}$. Entonces, suponiendo el muestreo

aleatorio simple, las fórmulas para estimar la variancia de las estimaciones pueden escribirse como sigue:

$$\text{Var}(\hat{X}) = \frac{N(N-n)}{n} \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \right) \quad (7)$$

$$\text{Var}(\hat{P}) = \frac{N(N-n)}{n} \left(\frac{\sum_{i=1}^n (p_i - \bar{p})^2}{n-1} \right) \quad (8)$$

$$\text{Var}\left(\frac{\hat{X}}{\hat{P}}\right) = \left(\frac{\hat{X}}{\hat{P}}\right)^2 [\text{Var}(\hat{X}) + \text{Var}(\hat{P}) - 2\text{Cov}(\hat{X}, \hat{P})] \quad (9)$$

donde $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$, $\bar{p} = \frac{\sum_{i=1}^n p_i}{n}$

y $\text{Cov}(\hat{X}, \hat{P}) = \frac{N(N-n)}{n} \left(\frac{\sum_{i=1}^n (x_i - \bar{x})(p_i - \bar{p})}{n-1} \right)$

Las fórmulas citadas son aplicables aunque una parte de la misma explotación esté en más de un segmento; es decir, una parte ponderada, $p_{1k} x_{1k}$, se incluye en cada segmento en que se encuentra.

4.4.3 Estimación por la razón. Si existe una medida del tamaño de cada segmento, es posible usar la estimación por la razón. Por ejemplo, podría conocerse el área total de la población y podría ser factible obtener el área, y_1 , de cada segmento en la muestra. Si los segmentos varían mucho en tamaño y X_1 está correlacionado con Y_1 , un estimador por la razón del total de X puede tener una variancia menor. El estimador, \hat{X}_1 , sería

$$\hat{X}_1 = Y \left(\frac{\hat{X}}{\hat{Y}} \right) \quad (10)$$

donde $Y = \sum_1^N Y_1$, el área total de los N segmentos, \hat{X} está dado por la ecuación

$$(4) \text{ e } \hat{Y} = \frac{N}{n} \left(\sum_1^n y_1 \right)$$

La variancia estimada de X_1 está dada por:

$$\text{Var}(\hat{X}_1) = Y^2 \left(\frac{\hat{X}}{\hat{Y}} \right)^2 [\text{Var}(\hat{X}) + \text{Var}(\hat{Y}) - 2\text{Cov}(\hat{X}, \hat{Y})]$$

donde

$$\text{Var}(\hat{X}) = \frac{N(N-n)}{n} \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \right),$$

$$\text{Var}(\hat{Y}) = \frac{N(N-n)}{n} \left(\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} \right)$$

$$\text{Cov}(\hat{X}, \hat{Y}) = \frac{N(N-n)}{n} \left(\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} \right)$$

También podría usarse, con modificaciones apropiadas, un estimador por la razón como la ecuación (10), en el caso del segmento cerrado. Con el segmento abierto, el uso de estimación por la razón probablemente no involucraría el área de los segmentos. Antes de decidir usar un estimador por la razón es importante considerar las condiciones en que sería mejor que el estimador especificado por la ecuación (4). Además, con referencia a la ecuación (10), no pase por alto el hecho de que las condiciones deben ser tales que el valor esperado de \hat{Y} esté muy próximo a $\sum_{i=1}^N Y_i$; si no está, hay un sesgo en la expansión de la muestra. Para ilustrar, suponga que el área total utilizada en la ecuación (10) para expandir la razón, $\left(\frac{\hat{X}}{\hat{Y}}\right)$, se obtiene de un reconocimiento geodésico del área total. El área total determinada por el reconocimiento geodésico podría no ser igual a $\sum_{i=1}^N Y_i$, que es el valor esperado de \hat{Y} , porque el reconocimiento geodésico no obtuvo el área total mediante la suma de medidas de las áreas de los segmentos en la población. De hecho, la experiencia muestra que distintas maneras de medir la misma cosa generalmente no dan resultados idénticos y la diferencia a veces es importante. Esto no quiere decir que Y_i tiene que ser una medición sin error. Podría existir error considerable en los valores de Y_i . Las dos cosas importantes son (1) que el valor esperado de \hat{Y} esté cerca de $\sum_{i=1}^N Y_i$ y (2) que Y_i esté relacionado con X_i de una manera que reduzca la variancia muestral. (Vea el tema de estimación por la razón en los textos sobre muestreo.)

4.4.4 Probabilidades de selección desiguales. El método de segmento ponderado no está limitado al muestreo de segmentos con probabilidades de selección iguales. Con probabilidades de selección desiguales los estimadores (4) y

(5) serían reemplazados por los siguientes:

$$\hat{X} = \sum_{i=1}^n R_i \sum_{k=1}^{f_i} P_{ik} x_{ik} \quad (11)$$

y

$$\hat{F} = \sum_{i=1}^n R_i \sum_{k=1}^{f_i} P_{ik} \quad (12)$$

donde R_i es la recíproca de la probabilidad de que se seleccione el i -ésimo segmento para la muestra. Pero las fórmulas citadas anteriormente para la variancia ya no serán aplicables. Deben usarse fórmulas que correspondan al diseño de la muestra.

4.4.5 Estimación por dominios. En muchas encuestas se desean obtener estimaciones por dominios. "Dominio" es un término general que se refiere a una parte de la población; por ejemplo, una clase de explotaciones, tales como ganaderas, o explotaciones con más de 200 hectáreas de tierra cultivable. Las fórmulas para estimación y variancia dadas en la sección 4.4.2 son todavía aplicables si hacemos la siguiente modificación. Sean $x'_{ik} = x_{ik}$ y $P'_{ik} = P_{ik}$ si una explotación pertenece al dominio; $x'_{ik} = 0$ y $P'_{ik} = 0$ si la explotación no pertenece al dominio. Ponga x'_{ik} y P'_{ik} en lugar de x_{ik} y P_{ik} . Entonces la ecuación (4) es un estimador del total del dominio, la ecuación (5) suministra una estimación del número de explotaciones en el dominio, la ecuación (6) da una estimación del promedio por explotación en el dominio, y el uso de x'_i y P'_i en vez de x_i y P_i en las ecuaciones (7), (8) y (9) provee estimaciones de las variancias de muestreo de las estimaciones correspondientes al dominio.

5. ILUSTRACION NUMERICA

Para ilustrar y comparar los tres métodos de aplicación del muestreo por áreas se construyó una población hipotética de 25 segmentos, 47 tracts y 30 explotaciones agrícolas. La mayoría de los datos para esta ilustración se copiaron de una lista de datos sobre tracts y explotaciones en una muestra por áreas que fue tomada en una región dedicada al engorde de ganado. Se escogió un número desproporcionadamente grande de explotaciones con ganado y maíz para esta ilustración.

La tabla 2 muestra los datos sobre explotaciones y tracts por segmentos. En la primera columna, el número a la izquierda del punto identifica tracts dentro de segmentos (vea la sección 4.1 para la definición de "tract"). Tracts que tienen el mismo número de explotación (vea la columna (5)) componen una explotación agrícola. Un asterisco (*) agregado a un

Tabla 2. Datos de tracts y explotaciones por segmentos

No. de segmento y de tract	Datos sobre tracts			Datos sobre explotaciones agrícolas					
	Tierra agrícola	Ganado	Maíz	No. de la explotación	Tierra agrícola	Ganado	Maíz	Otros tracts en la explotación	Proporción de la explotación que está en el tract
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1.1	160	37	64	1*	160	37	64	ninguno	1,000
2.1	150	246	43	2*	1,260	246	203	3.3, 7.1, 9.1	0,119
2.2	312	26	122	3*	576	26	262	3.1	0,542
2.3	18	6	0	4*	18	6	0	ninguno	1,000
3.1	264	0	140	3	576	26	262	2.2	0,458
3.2	80	24	32	5	400	91	90	6.2	0,200
3.3	320	0	160	2	1,260	246	203	2.1,7.1,9.1	0,254
4.1	237	93	114	6*	400	93	159	9.2	0,592
4.2	90	0	0	7*	90	0	0	ninguno	1,000
5.	No hay tracts agrícolas en el segmento No. 5								
6.1	160	23	43	8*	160	23	43	ninguno	1,000
6.2	320	67	58	5*	400	91	90	3.2	0,800
6.3	4	0	0	9*	4	4	4	ninguno	1,000
7.1	630	0	0	2	1,260	246	203	2.1, 3.3, 9.1	0,500
7.2	120	0	116	10*	120	0	116	ninguno	1,000
8.1	159	27	25	11	320	82	25	14.2	0,497
8.2	236	82	104	12*	236	82	104	ninguno	1,000
9.1	160	0	0	2	1,260	246	203	2.1, 3.3, 7.1	0,127
9.2	163	0	45	6	400	93	159	4.1	0,408
9.3	80	0	0	13	4,400	777	320	10.1, 11.1, 12.1, 13.1	0,018
10.1	630	320	160	13	4,400	777	320	9.3,11.1,12.1,13.1	0,143
11.1	1,275	437	160	13*	4,400	777	320	9.3,10.1,12.1,13.1	0,290
12.1	1,800	0	0	13	4,400	777	320	9.3,10.1,11.1,13.1	0,409
13.1	615	0	0	13	4,400	777	320	9.3, 10.1, 11.1, 12.1	0,140
13.2	140	26	39	14*	140	26	39	ninguno	1,000

Tabla 2. (Continuación)

No. de segmento y de tract	Datos de tracts			No. de la explotación	Datos de explotaciones				Proporción de la explotación que está en el tract
	Tierra agrícola	Ganado	Maíz		Tierra agrícola	Ganado	Maíz	Otros tracts en la explotación	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
14.1	55	7	11	15*	55	5	11	ninguno	1,000
14.2	161	55	0	11*	320	82	25	8.1	0,503
14.3	160	0	120	16*	160	0	120	ninguno	1,000
14.4	86	6	16	17	160	25	56	15.1	0,538
15.1	74	19	40	17*	160	25	56	14.4	0,462
16.	No hay tracts agrícolas en el segmento No.16								
17.1	360	100	170	18	366	100	170	18.1	0,984
17.2	2	2	0	19*	2	2	0	ninguno	1,000
17.3	81	0	23	20*	81	0	23	ninguno	1,000
17.4	74	0	0	21*	74	0	0	ninguno	1,000
17.5	320	145	120	22*	320	145	120	ninguno	1,000
18.1	6	0	0	18*	366	100	170	17.1	0,016
18.2	480	0	116	23*	640	0	116	19.1	0,750
19.1	160	0	0	23	640	0	116	18.2	0,250
19.2	160	28	0	24*	160	28	0	ninguno	1,000
19.3	80	201	19	25*	300	201	118	20.1	0,267
20.1	220	0	99	25	300	201	118	19.3	0,733
21.1	320	0	0	26*	320	0	0	ninguno	1,000
22.1	160	19	86	27	360	63	186	23.1,24.2	0,445
22.2	120	0	60	28*	120	0	60	ninguno	1,000
23.1	80	44	0	27*	360	63	186	22.1,24.2	0,222
24.1	280	46	80	29*	280	46	80	ninguno	1,000
24.2	120	0	100	27	360	63	186	22.1,23.1	0,333
25.1	400	0	160	30*	400	0	160	ninguno	1,000
Totales	12.082	2.106	2.645						30,000

NOTA: Cada explotación en la tabla 2 está marcada con un asterisco (*) que señala el tract en que está ubicada la sede de la explotación.

número de explotación indica el tract en que se encuentra la sede de la explotación. Por ejemplo, la explotación número 3 se compone de los tracts 2.2 y 3.1, y su sede está en el tract 2.2.

En síntesis, los tres métodos para definir unidades de muestreo por áreas requieren la recolección de datos como sigue:

Segmento cerrado. En una encuesta que utiliza el segmento cerrado se recolectarían datos para tracts dentro de los segmentos.

Segmento abierto. Al usar el segmento abierto, se recolectarían datos para todas las explotaciones cuyas sedes estén dentro de los segmentos en la muestra.

Segmento ponderado. En una encuesta que utiliza el segmento ponderado se recolectarían datos para cada explotación que esté parcial o totalmente dentro de un segmento muestral.

Como ejemplo específico de los datos que se recolectarían al usar cada uno de los tres métodos, suponga que se han seleccionado para una muestra los segmentos 5, 7 y 19. Dependiendo del método empleado, uno de los siguientes grupos de datos (vea la tabla 2) sería obtenido:

Segmento Cerrado

Número del segmento	Número del tract	Datos correspondientes a los tracts		
		Tierra agrícola	Ganado	Maíz
5	-	-	-	-
7	1	630	0	0
7	2	120	0	116
19	1	160	0	0
19	2	160	28	0
19	3	80	201	19

Segmento Abierto

Número del segmento	Número de la explotación	Tierra agrícola en el segmento	Datos correspondientes a explotaciones		
			Tierra agrícola	Ganado	Maíz
5	-	-	-	-	-
7	10	120	0	116	
19	24	160	28	0	
19	25	300	201	118	

Segmento Ponderado

Número del segmento	Número de la explotación	Tierra agrícola en el segmento	Datos correspondientes a explotaciones		
			Tierra agrícola	Ganado	Maíz
5	-	-	-	-	-
7	2	630	1,260	246	203
7	10	120	120	0	116
19	23	160	640	0	116
19	24	160	160	28	0
19	25	80	300	201	118

Puesto que cada uno de los 47 tracts en la población se asocia con un solo segmento, está claro que el total de todos los segmentos cerrados en la población tiene que ser igual al total de la población. De modo similar, puesto que cada una de las 25 explotaciones se asocia con un solo segmento, los totales de los segmentos abiertos tienen que sumar al total de la población. Es menos obvio, pero también los totales de los segmentos ponderados (después de "ponderarse" los datos) tienen que igualar el total de la población. Considere el segmento número 19. Tres explotaciones agrícolas, 23, 24 y 25, están parcial o totalmente dentro del segmento. Las proporciones de estas explotaciones que están dentro del segmento son:

<u>Explotación</u>	<u>Proporción</u>
23	$\frac{160}{640} = 0,250$
24	$\frac{160}{160} = 1,000$
25	$\frac{80}{300} = 0,267$

Estas proporciones son los valores de P_{ik} que aparecen en la ecuación (3) y los de p_{ik} en los estimadores, ecuaciones (4), (5) y (6), para el método de segmento ponderado. La última columna de la tabla 2 contiene los valores de P_{ik} . Observe que los valores de P_{ik} suman 1 para cada explotación. Usando como ejemplo el segmento 19, los totales del segmento ponderado son:

$$\begin{aligned} \text{Ganado} & (0,250)(0) + (1,000)(28) + (0,267)(201) = 81,7 \\ \text{Maíz} & (0,250)(16) + (1,000)(0) + (0,267)(118) = 60,5 \\ \text{Tierra agrícola} & (0,250)(640) + (1,000)(160) + (0,267)(300) = 400 \\ \text{Número de explotaciones} & (0,250) + (1,000) + (0,267) = 1,517 \end{aligned}$$

Estos totales y los totales de segmento ponderado correspondientes a todos los demás segmentos aparecen en la tabla 3. También se muestran los totales de segmentos para los segmentos cerrado y abierto. Observe que los totales de segmento ponderado para tierra agrícola (400 para el segmento número 19) son iguales a los totales correspondientes a los segmentos cerrados. Por esto, no se muestran en la tabla 3 los totales de segmento ponderado para tierra agrícola.

5.1 Estimación por Dominios y el Segmento Ponderado

Algunos analistas han solicitado confirmación de la aplicabilidad del segmento ponderado a los estudios analíticos. Puesto que el valor de X para una explotación es multiplicado por la proporción de la explotación que está dentro del segmento, podrá parecer, a primera vista, que estamos tratando de fracciones de explotaciones en vez de explotaciones enteras. Pero eso no es el caso. La situación es similar a la ponderación de datos muestrales cuando se usan varias fracciones de muestreo. Este punto fue tocado brevemente en la sección 4.4.5. La técnica que se esbozó es usada comúnmente por los estadísticos como un medio general corto de especificar un procedimiento para hacer tanto estimaciones por dominios como estimaciones para la población entera.

Para ilustrar, suponga que las explotaciones numeradas 2, 7, 12, 17 y 22 forman un dominio y se quieren hacer estimaciones para este dominio. De la tabla 2 pueden obtenerse los totales y los promedios correspondientes a las 5 explotaciones en este dominio. Los resultados son:

<u>Item</u>	<u>Total</u>	<u>Promedio</u>
Tierra agrícola	2,066	413,2
Ganado	498	99,6
Maíz	483	96,6

El lector puede comprobar que los estimadores dados por las ecuaciones (4), (5) y (6) y el procedimiento esbozado en la sección 4.4.5 son apropiados para estimar estos totales y promedios. Trate los 25 segmentos como una muestra; es decir, haga los cálculos como si los 25 segmentos fueran una muestra de una población más grande. El tomar los 25 segmentos como muestra elimina el error de muestreo aleatorio simple y los resultados deben concordar exactamente con los totales y promedios dados arriba para las cinco explotaciones.

5.2 Variación de Muestreo

Puesto que la variación de muestreo es función de la variación entre totales de segmentos, es importante estudiar la tabla 3 y su derivación de la tabla 2. Examine la variación entre segmentos con respecto a los tres métodos. Para cultivos y otros ítemes que son limitados por la cantidad de terreno, el método

Tabla 3. Totales de segmentos (cerrado, abierto, ponderado)

Número del segmento	Tierra agrícola		Número de explotaciones		Ganado			Maíz		
	Cerrado	Abierto	Abierto	Ponderado	Cerrado	Abierto	Ponderado	Cerrado	Abierto	Ponderado
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	160	160	1	1,000	37	37	37,0	64	64	64,0
2	480	1.854	3	1,661	278	278	49,4	165	465	166,2
3	664	0	0	0,912	24	0	92,6	332	0	189,5
4	327	490	2	1,592	93	93	55,1	114	159	94,1
5	0	0	0	0	0	0	0	0	0	0
6	484	564	3	2,800	90	114	95,8	101	133	115,0
7	750	120	1	1,500	0	0	123,0	116	116	217,5
8	395	236	1	1,497	109	82	122,7	129	104	116,4
9	403	0	1	0,553	0	0	83,2	45	0	96,4
10	630	0	0	0,143	340	0	111,1	160	0	45,8
11	1.275	4.400	1	0,290	437	777	225,3	160	320	92,8
12	1.800	0	0	0,409	0	0	317,8	0	0	130,9
13	755	140	1	1,140	26	26	134,8	39	39	83,8
14	462	535	3	3,041	68	89	61,7	147	156	173,7
15	74	160	1	0,462	19	25	11,5	40	56	25,9
16	0	0	0	0	0	0	0	0	0	0
17	837	477	4	4,984	247	147	245,4	313	143	310,3
18	486	1.006	2	0,766	0	100	1,6	116	286	89,7
19	400	460	2	1,517	229	229	81,7	19	118	60,5
20	220	0	0	0,733	0	0	147,3	99	0	86,5
21	320	320	1	1,000	0	0	0	0	0	0
22	280	120	1	1,445	19	0	28,0	146	60	142,8
23	80	360	1	0,222	44	63	14,0	0	186	41,3
24	400	280	1	1,333	46	46	67,0	180	80	141,9
25	400	400	1	1,000	0	0	0	160	160	160,0
Total	12.082	12.082	30	30,000	2.160	2.160	2.160,0	2.645	2.645	2.645,0

de segmento cerrado impone un máximo sobre el total del segmento. Obviamente, el área en maíz, por ejemplo, no puede ser mayor que el área de tierra agrícola dentro del segmento. Pero, en el caso del segmento abierto, el área máxima en maíz que podría estar "dentro" de un segmento puede ser, por lo menos, tanto como el área en maíz que tiene la explotación en la población que más maíz cultiva.

Observe en la tabla 3 la variación entre segmentos de la cantidad de tierra agrícola y compare el segmento abierto con el cerrado. Para características que son altamente correlacionadas con la cantidad de tierra agrícola, el segmento cerrado tendrá variancias de muestreo mucho más pequeñas que las del segmento abierto, suponiendo que la cantidad de terreno en segmentos puede ser controlada efectivamente en el procedimiento de deslindar segmentos. Se esperaría que las diferencias en variancia entre el segmento abierto y el cerrado fueran menores en el caso del ganado, porque el número de cabezas es menos limitado por la cantidad de terreno, en comparación con cultivos.

Para características correlacionadas con la cantidad de tierra agrícola, el método de segmento ponderado, como el segmento cerrado, impone algo de control sobre los valores máximos de los totales para segmentos. Por ejemplo, el área en maíz correspondiente a un segmento después de la ponderación de los datos no puede sobrepasar la cantidad de tierra agrícola en el segmento. Es decir, con referencia a la ecuación (3), si X es el área en cualquier cultivo dado, el total del segmento ponderado, X_1 , no puede exceder el área total del segmento. Recuerde que la variancia de muestreo para el segmento ponderado involucra la variancia entre los X_1 .

Como otro ejemplo de cómo difieren el segmento ponderado y el abierto con respecto a la variancia de muestreo, refiérase a la tabla 2 y la explotación número 13. Partes de esta explotación se encuentran en cinco segmentos. La explotación tiene 4,400 acres de tierra agrícola y 777 cabezas de ganado. El método de segmento abierto asigna al segmento número 11 todo el ganado, sin importar dónde se encuentren. Esta sola explotación tiene un impacto importante en la variancia de muestreo del segmento abierto. El método de segmento ponderado reduce, en este caso, la variancia de muestreo mediante la "división" de la explotación en partes. Sin importar dónde se encuentre el ganado, la ponderación usada en el método de segmento ponderado tiene el efecto de distribuir el ganado entre los cinco segmentos como sigue:

<u>Segmento</u>	<u>Ganado</u>
9	14
10	111
11	225
12	318
13	109
Total	<u>777</u>

Fíjese que cuanto mayor es el número de segmentos en que se divide una explotación, mayor es la probabilidad de que la explotación se encuentre en la muestra.

La tabla 4 muestra la variancia relativa entre segmentos para cada uno de los tres métodos. Las variancias se calcularon con los datos mostrados en la tabla 3. Aunque esta situación numérica no da una base para generalizar, los resultados en la tabla 4 no son contrarios a la experiencia general. Como se esperaba de la exposición anterior y como demuestran varios estudios, el segmento abierto tiene variancias mucho más grandes que las del segmento cerrado.

Tabla 4. Variancia relativa entre totales de segmentos

Item	Variancia relativa *		
	Cerrado	Abierto	Ponderado
Tierra agrícola	0,68	3,55	0,68
Número de explotaciones	xxxx	0,87	0,84
Ganado	2,12	3,71	0,97
Maíz	0,73	1,21	0,48

* $\frac{\sum(x_1 - \bar{x})^2}{\bar{x}^2(N-1)}$ donde X_1 es un total de segmento en la tabla 3.

Puesto que una explotación es igual a un tracto o es más grande, una muestra de n segmentos que emplea el segmento ponderado obtiene datos para una mayor proporción de la población, en comparación con el empleo del segmento cerrado. Pero, después de la ponderación de los datos, el "tamaño" del segmento ponderado con respecto al área de tierra agrícola es igual al "tamaño" del segmento cerrado. Por esto, la parte de la variancia entre segmentos (variancia de muestreo) que puede asociarse con la variación en tamaño de segmentos parece ser aproximadamente igual, ya que se emplee el segmento ponderado o el cerrado. Además la ponderación de los datos del segmento ponderado tiene un efecto de promediar. Por lo tanto, es razonable esperar que las variancias de muestreo para el segmento ponderado sean algo menores que las variancias de muestreo para el segmento cerrado. No

obstante, hay que tomar en cuenta los costos.

Es de interés comparar la variancia relativa entre las 30 explotaciones en el ejemplo numérico, que se indica en la última columna de la tabla 5, con la variancia relativa entre segmentos. Para fines de comparación, la variancia relativa entre segmentos tiene que convertirse al equivalente de una explotación. El segmento abierto tiene un promedio de $\frac{30}{25} = 1,2$ explotaciones por segmento; en el caso del segmento ponderado, el promedio de explotaciones (no ponderado) es $\frac{47}{25} = 1,88$. Para convertir las variancias en la tabla 4 al equivalente de una explotación, multiplique las variancias de segmento abierto por 1,2 y las variancias de segmento ponderado por 1,88. Esto da los resultados para ambos segmentos mostrados en la tabla 5.

Tabla 5. Variancia relativa por explotación

Item	Variancia relativa entre segmentos, por explotación		Variancia relativa entre explotaciones
	Abierto	Ponderado	
Tierra agrícola	4,26	1,28	3,89
No. de explotaciones	1,04	1,58	xxxx
Ganado	4,45	1,82	4,40
Maíz	1,45	0,90	0,90

Tal como se esperaba, debido a la correlación dentro de segmentos, las variancias entre segmentos abiertos (tabla 5) son mayores que las variancias entre explotaciones individuales. Con referencia al segmento ponderado, el efecto de la correlación dentro de segmentos fue neutralizado por el hecho de que el segmento ponderado tuvo el efecto de dividir las grandes explotaciones en unidades más pequeñas. Por lo tanto, como se ve en la tabla 5, el resultado neto fue que (aún sobre la base de variancia por explotación) la variancia para el segmento ponderado era menor que la variancia entre explotaciones individuales. Esta ilustración numérica no provee una base para generalización; sin embargo, los resultados no son contrarios a lo que se esperaba.

6. DISCUSION DE LAS TRES DEFINICIONES DE UNIDADES DE MUESTREO POR AREAS

La magnitud de las diferencias entre los tres métodos de definición de las unidades de muestreo por áreas depende de las condiciones locales. A un extremo, los tres métodos podrían ser idénticos. Por ejemplo, suponga una situación en que cada operador vive en su explotación agrícola y cada explotación es una sola parcela pequeña. Si ninguna de las explotaciones traslapa linderos de segmento, los métodos cerrado, abierto y ponderado serán idénticos. Pero las explotaciones varían muchísimo en tamaño y tipo; algunas se componen de más de un tract y las condiciones de tenencia y administración presentan ambigüedades acerca de qué constituye una explotación y quién es el operador. Parece que ningún método es universalmente mejor que otro.

Al comparar los tres métodos, tenemos que considerar el carácter de la población por muestrear, la clase de datos que se recogerán, la aplicabilidad de los conceptos en que se basa cada método, la variancia de muestreo, el error de cobertura, el error de respuesta y los costos. Se necesita mucha experiencia como base para formar juicios prácticos sobre la selección de métodos. En esta publicación no es factible ir más allá de una breve discusión de conceptos y una que otra indicación de las circunstancias en que un método se esperaría ser mejor que otro. Son muy limitados los estudios documentados sobre la comparación de métodos alternativos y los procedimientos para aplicar el muestreo por áreas.

6.1 Segmento Cerrado contra el Abierto o el Ponderado

Puesto que el segmento cerrado se limita a encuestas en que los tracts son unidades de información aceptables, la comparación del segmento cerrado con el abierto o el ponderado tiene que limitarse a tales encuestas.

Inicialmente, por lo menos en los Estados Unidos, se usaba el método de segmento abierto. Pero los problemas de errores de cobertura, especialmente los problemas de identificar explotaciones y asociarlas con segmentos, condujeron a los estadísticos a buscar una alternativa mejor. Se ensayó el segmento cerrado y resultó, donde fue aplicable, muy superior al segmento abierto con respecto a la variancia de muestreo y el error de cobertura, especialmente cuando se usaban fotografías en la enumeración de los segmentos. Como resultado, se desarrolló una fuerte tendencia hacia el uso del segmento cerrado al grado máximo que fuera posible. Aunque el error de cobertura para el segmento cerrado es relativamente bajo, el error de respuesta es un factor que limita su aplicabilidad. El error de

respuesta varía desde cero en el caso de áreas bajo cultivo hasta representar un problema de alguna importancia en el caso de inventario de ganado y aún hasta hacer prácticamente imposible la aplicación del método en el caso de ciertas características con respecto a las cuales el agricultor no puede informar sobre un tract. Por ejemplo, generalmente no es práctico recolectar datos por tracts sobre características como costos de producción o ventas de productos agrícolas. Tales datos a menudo se llaman "datos económicos" y generalmente están asociados con la explotación como un negocio y no con un tract.

Hendricks, Searles y Horvitz han comparado los segmentos cerrado, abierto y ponderado en el muestreo de áreas bajo cultivo. * Sus resultados, como algunos no publicados por el Statistical Reporting Service, demuestran que las variancias de muestreo son definitivamente más pequeñas al usar el segmento cerrado en vez del segmento abierto. Los resultados publicados por Hendricks et al. también muestran que las variancias para el segmento ponderado variaban desde igualdad con las del segmento cerrado hasta ser algo más pequeñas que estas. Para otras clases de datos, la comparación podría dar resultados muy diferentes.

El promedio de costos de campo por segmento cerrado depende en gran parte de la necesidad de entrevistar a los operadores de todos los tracts dentro del segmento. Para algunos tracts y tipos de datos, puede ser innecesario entrevistar a los operadores de todos los tracts. Por ejemplo, en una encuesta sobre áreas bajo cultivo, puede ser innecesario entrevistar a los operadores de tracts que sólo tienen árboles. Pero, si suponemos que hay que entrevistar a todos los operadores de tracts, el costo de campo para el segmento cerrado podría ser casi tanto como el costo de campo para el segmento ponderado. Esta afirmación se basa en la suposición de que el cuestionario es el mismo, excepto que en un caso de un tract y en otro caso de una explotación. En el caso del segmento ponderado, el promedio del tiempo requerido para la entrevista probablemente sería mayor, aunque en muchos casos un operador de una explotación agrícola puede responder más fácilmente con respecto a su explotación que con respecto a un tract. Sin embargo, el costo de dividir los segmentos en tracts y entrevistar a todos los operadores es una porción sustancial del costo total. Quizá en algunas encuestas la diferencia en el promedio de costo por segmento sería tan pequeña como el 10 por ciento. Así es que hay circunstancias en que

* Hendricks, W.A., Searles, D.T., y Horvitz, D.C.; capítulo II de "Estimation of Areas in Agricultural Statistics", Food and Agriculture Organization, Roma, 1965.

los métodos de segmento cerrado y segmento ponderado parecen ser competitivos (o casi competitivos) en términos de variancia de muestreo por unidad de costo. Por esto, puesto que los errores de cobertura y de respuesta tienden a ser fuentes mayores de error, existe una fuerte indicación de que, para algunas encuestas, el criterio más importante al escoger entre el segmento cerrado y el ponderado es la cuestión de qué método involucra el menor error de cobertura y de respuesta.

Una comparación similar entre el segmento cerrado y el abierto es más difícil, porque estos dos tienen menos en común. Sin embargo, en este punto de nuestra discusión, la cuestión parece resolverse en cómo se comparan los segmentos abierto y ponderado. Es decir, cuando no es aplicable el segmento cerrado, ¿qué alternativa, segmento abierto o segmento ponderado, es mejor? En la práctica, la tendencia ha sido usar el segmento cerrado al grado máximo posible, y usar el segmento abierto solamente cuando no es aplicable el cerrado; pero el segmento ponderado empieza a cobrar más interés.

Como se indicó anteriormente, el segmento cerrado no es aplicable cuando (1) los requisitos de la encuesta establecen que tienen que ser las explotaciones agrícolas las unidades de información, o (2) los errores de respuesta hacen necesario descartar el uso de tracts como unidades de información. En algunas encuestas es factible recolectar solamente una parte de los datos requeridos mediante el método de segmento cerrado. Entonces, para aprovechar el segmento cerrado, se ha usado simultáneamente en la misma encuesta y con la misma muestra de segmentos una combinación de los dos métodos (o cerrado y abierto, o cerrado y ponderado). ¿Qué combinación es mejor? Puesto que la respuesta depende en parte de cómo se comparan el segmento abierto y el ponderado, la discusión de este asunto se pospondrá hasta una sección posterior.

6.2 Segmento Abierto contra Segmento Ponderado

Los métodos de segmento abierto y de segmento ponderado son aplicables cuando las unidades de información son las explotaciones agrícolas.

En el caso del segmento abierto, la selección entre el modo por operador y el modo por explotación (descritos en secciones 4.3.1 y 4.3.2) es una consideración importante. El método de segmento ponderado utiliza únicamente el modo por explotación--el concepto del segmento ponderado y el modo por operador no son compatibles. Por esto, en la discusión que sigue, de segmento abierto contra segmento ponderado, se supone el uso del modo por explotación. Pero primero revisemos las condiciones que son favorables al

modo por explotación y al segmento abierto.

Recordará que con el modo por operador el objetivo es encontrar, dentro de los límites de cada segmento en la muestra, todas las residencias de operadores de explotaciones agrícolas. Las explotaciones correspondientes a operadores que tienen residencias en un segmento de la muestra están en la muestra. (Nota: no se incluyen en esta discusión las encuestas en que los hogares agrícolas son las unidades de información apropiadas.)

El modo por operador tendrá error mínimo de cobertura cuando (1) se puedan formular y aplicar con poca ambigüedad reglas sencillas que establezcan una correspondencia uno a uno entre operadores y explotaciones, (2) cada operador tenga una sola residencia y (3) la mayoría de residencias dentro de los segmentos en la muestra sean ocupadas por operadores de explotaciones agrícolas. En estas condiciones la tarea de encontrar a los operadores de explotaciones agrícolas no es factor costoso y la tendencia a pasar por alto alguna residencia de operador de explotación agrícola debe ser mínima. Si, además, es posible diseñar la muestra de manera que haya aproximadamente el mismo número de operadores de explotación agrícola en cada segmento, las condiciones son generalmente favorables al segmento abierto (usando el modo por operador), con respecto a error de cobertura y la variancia de muestreo.

Como se indicó anteriormente, las razones para considerar el modo por explotación como alternativa al modo por operador son (1) los problemas para encontrar a los operadores en segmentos poblados por muchas familias no agrícolas y (2) los problemas relacionados con el pareo de explotaciones con operadores. Conceptualmente, para cualquier muestra dada de segmentos, los dos modos dan la misma muestra de explotaciones, salvo que haya diferencia en la definición de "sede". Hay gran diferencia entre los procedimientos usados al aplicar los dos modos. En cualquier caso, el objetivo principal es lograr la identificación completa y exacta de todas las explotaciones agrícolas que tienen su sede en la muestra de segmentos. Generalmente, la omisión es más frecuente que la duplicación. El porcentaje de identificaciones incompletas puede variar desde cero hasta varias por ciento, dependiendo de los materiales disponibles para la encuesta y los procedimientos usados (y si los procedimientos se basan en conceptos válidos). La experiencia del organismo que conduce la encuesta y el énfasis puesto en la capacitación y supervisión de los enumeradores son también factores importantes que contribuyen al logro de cobertura completa y exacta. Ha habido mucha experiencia con el método de segmento abierto y se han ensayado muchos procedimientos distintos. Sin embargo, se necesitan mejores soluciones del problema de error de cobertura, lo que es una razón

importante para dirigir más atención al método de segmento ponderado.

El propósito principal de las dos secciones que siguen es indicar que el método de segmento ponderado tiene mucho mérito y que debe ser puesto a prueba ampliamente como alternativa que podría ser muy superior al segmento abierto, por lo menos en algunas circunstancias.

6.2.1 La variancia de muestreo y costos. Para resumir brevemente, el método de segmento ponderado requiere la división de cada segmento en la muestra en tracts y la entrevista del operador (u otra persona apropiada) de cada explotación agrícola que está total o parcialmente dentro de los linderos del segmento. Los datos recolectados corresponden a explotaciones agrícolas y no a tracts. El método de segmento abierto (modo por explotación) también requiere la división de cada segmento en tracts. Las explotaciones con sede dentro de los segmentos en la muestra son incluidas en la muestra y los operadores de tales explotaciones son entrevistados. Suponga que la sede se define de tal modo que siempre es un punto único dentro de los linderos de la explotación. Entonces, para cualquier muestra dada de segmentos, las explotaciones en la muestra que usa el segmento abierto son un subconjunto de las explotaciones que estarían en la muestra si se usara el método de segmento ponderado.

Como una ayuda en la exposición, serán de valor unos modelos sencillos de variancias y costos. Suponga una muestra aleatoria estratificada de segmentos y una fracción constante de muestreo. Haciendo caso omiso de la corrección por población finita, las variancias de los promedios muestrales por segmento pueden escribirse como sigue:

$$V(\bar{x}_a) = \frac{V_a}{n_a} \quad \text{y} \quad V(\bar{x}_p) = \frac{V_p}{n_p} \quad ,$$

donde

$V(\bar{x}_a)$ es la variancia de \bar{x}_a , la media muestral por segmento correspondiente al método de segmento abierto,

n_a es el número de segmentos en la muestra de segmentos abiertos,

V_a es la variancia entre segmentos abiertos dentro de los estratos y

$V(\bar{x}_p)$, n_p y V_p se definen similarmente para el método de segmento ponderado.

Para los modelos de costos, suponga que

$$C = C_f + n_a C_a \quad \text{y} \quad C = C_f + n_p C_p \quad ,$$

donde

C es el costo total de la encuesta (es lo mismo para ambos métodos),

C_f es la parte fija del costo total, que no está relacionada con el número de segmentos en la muestra,

C_a es el costo medio por segmento abierto, y

C_p es el costo medio por segmento ponderado.

Suponiendo que el costo total es fijo, los tamaños de muestra n_a y n_p se determinan a partir de los modelos de costos. Puede ser demostrado que la variancia con el segmento ponderado, $V(\bar{x}_p)$ será menor que la variancia con el segmento abierto, $V(\bar{x}_a)$, con tal que sea válida la desigualdad

$$\frac{V_p}{V_a} < \frac{C_a}{C_p}$$

Parece que, en general, V_p es mucho menor que V_a . Como se indicó anteriormente, hay buena razón para creer que la variancia en el caso del segmento ponderado es más o menos igual a la variancia correspondiente al segmento cerrado o menor que ella, y está bien establecido que, en general, la variancia de muestreo correspondiente al segmento cerrado (por lo menos con respecto a áreas bajo cultivo) es mucho menor que la variancia correspondiente al segmento abierto. De paso, los resultados publicados por Hendricks et al. mostraron que, para las áreas de siete cultivos, la variancia con el segmento ponderado fue, como promedio, cerca del 25 por ciento menor que la variancia con el segmento abierto. Para estimaciones de la diferencia entre dos años, usando una muestra apareada de segmentos, sus análisis indicaron que las variancias con el segmento ponderado fueron menos de la mitad de las variancias con el segmento abierto.

Para ver los costos comparados, consider el costo del segmento ponderado y los ahorros que se lograrían si se usara el método de segmento abierto. En el caso del segmento ponderado, los primeros dos pasos al final de la sección 4.3.2 se llevarían a cabo y se completaría un cuestionario para cada explotación anotada en el paso 2 por el hecho de tener algún terreno dentro de un segmento en la muestra.

Ahora supongamos que los procedimientos de campo incluidos en los pasos 1, 2 y 3 de la sección 4.3.2 son iguales tanto para el método de segmento abierto como para el de segmento ponderado. En los Estados Unidos aproximadamente la mitad de las explotaciones anotadas en el paso 2 por poseer terreno dentro de un segmento también tienen su sede dentro de los linderos del segmento. La mayor parte de la diferencia (aumento) en costos resultante del uso del segmento ponderado ocurre en la adquisición de datos para las explotaciones en la muestra cuyas sedes están fuera del segmento.

La necesidad de minimizar el error de cobertura requiere una cuidadosa aplicación de reglas para asociar las explotaciones con los segmentos y así determinar qué explotaciones están en la muestra. Para aplicar eficazmente los procedimientos del método de segmento abierto, probablemente será necesario comunicarse con algunos operadores que tienen su sede fuera del segmento. Esto podría ser necesario para resolver cualquier incertidumbre acerca del terreno dentro de la explotación y la sede de la misma. Suponga que f_p es el número de explotaciones agrícolas en una muestra de n segmentos usando el método de segmento ponderado y suponga que f_a es el número de explotaciones en la misma muestra de segmentos usando el método de segmento abierto. Puesto que f_a es aproximadamente $f_p/2$, parece claro que C_a tiene que ser bastante mayor que $C_p/2$, por dos razones: (1) los costos de dividir un segmento en tracts y de identificar todas las explotaciones que están total o parcialmente dentro de los segmentos son comunes a ambos métodos (este costo es parte de C_a y C_p , no de C_f) y (2) algunas explotaciones en f_p que no están en f_a tendrían que visitarse si se aplicara cuidadosamente el método de segmento abierto. No es posible juzgar de antemano exactamente cómo se compara C_a con C_p en toda situación relacionada con encuestas. Empero, aunque no sea válida la desigualdad citada, parece que C_a , en relación con C_p , es bastante grande para justificar el ensayo y comparación de los dos métodos, especialmente cuando se considera la necesidad de minimizar el error de cobertura.

6.2.2 Error de cobertura. Conviene dividir los errores de cobertura en dos categorías: (1) determinación incorrecta de la composición de explotaciones agrícolas individuales y (2) asociación incorrecta de explotaciones con los segmentos en la muestra. Estas dos clases de error no son independientes.

Con el segmento ponderado, la cobertura correcta depende de que se tome en cuenta todo el terreno dentro de un segmento y no omitir alguna explotación que esté parcialmente dentro del segmento. Los procedimientos de campo, los materiales para la encuesta y las instrucciones tienen que desarrollarse con eso en mente. Cada enumerador tiene que saber exactamente qué es una explotación agrícola y tener la capacidad de determinar su localización geográfica. Hay que recolectar datos para la totalidad de la explotación si ella tiene algún terreno dentro de un segmento de la muestra.

Con el segmento abierto, pero no con el ponderado, el enumerador tiene que saber determinar la sede de una explotación y su localización.

Resulta difícil y complejo tanto el desarrollo de especificaciones que definen "sede" como la capacitación de enumeradores de modo que adquieran un claro entendimiento de cómo manejar toda situación. El propósito de evitar los problemas de definición de "sede" y los errores de cobertura asociados es una razón de peso que a menudo conduce a los estadísticos a buscar una alternativa al método de segmento abierto. El uso del segmento ponderado evita los problemas de identificación y localización de la sede, pero eso no necesariamente quiere decir que los errores de cobertura serán menores.

Para desarrollar más ampliamente la comparación de los conceptos de segmentos abierto y ponderado, debemos considerar algunas situaciones distintas. Por ejemplo, suponga que hay dentro de un segmento un tract pequeño que no evidencia ninguna actividad agrícola. Suponga que por definición este tract es parte de una explotación agrícola y que el resto de la explotación se encuentra fuera del segmento. Puesto que el tract dentro del segmento no parece ser parte de una explotación agrícola, podría clasificarse fácilmente como no parte de una explotación, especialmente por un enumerador que no da atención completa a los detalles o que no entiende ampliamente los conceptos de la encuesta en relación con su cometido. En todo caso, suponga que el tract se clasifica equivocadamente como no parte de una explotación agrícola. Esto resultaría en una omisión bajo el método de segmento ponderado, pero la omisión sería una fracción (la proporción que está dentro del segmento) de la explotación y no la explotación entera. Con el uso del segmento abierto, esta clasificación equivocada omitiría indebidamente toda la explotación solamente si la sede estuviera en el segmento. De paso, este es un buen ejemplo que ilustra en parte por qué el segmento cerrado tiene bajo error de cobertura. Si un tract dentro de un segmento no tiene actividad agropecuaria que deba incluirse en la encuesta, no importa (con el segmento cerrado) si el tract es clasificado correcta o incorrectamente como parte de una explotación agrícola. Podría ser un ejercicio útil considerar cómo podría ocurrir error de cobertura en varios otros casos, pero no hay sustituto para la experiencia y el ensayo de alternativas en las condiciones reales de operación.

Estadísticos con experiencia en el muestreo por áreas tienen distintas opiniones acerca del valor potencial del método de segmento ponderado. El autor se cuenta entre los que creen que el método de segmento ponderado debe ser explorado y desarrollado al máximo. Es fácil describir circunstancias (quizá hipotéticas) en que sería claramente preferible el segmento abierto, especialmente si muchos de los datos que han de recolectarse se relacionan con características de los hogares de los operadores y otras personas asociadas a la explotación, en vez

de relacionarse con las explotaciones mismas. Pero fueron los problemas en la aplicación práctica del método de segmento abierto los que condujeron al desarrollo del segmento cerrado. El autor no espera que el error de cobertura del segmento ponderado sea tan pequeño como el del segmento cerrado, pero, como ya se ha dicho, hay características para las cuales la combinación de los errores de cobertura y de respuesta podría ser menor en el caso del segmento ponderado que en el caso del segmento cerrado. Además, se necesita un método mejor que el de segmento abierto cuando las unidades de información tienen que ser las explotaciones agrícolas.

De paso, la experiencia ha demostrado que el error de cobertura varía considerablemente de una característica a otra dentro de la misma encuesta y la misma muestra. Eso es de esperarse si, por ejemplo, las pequeñas explotaciones se pasan por alto más frecuentemente que las grandes. El error de cobertura podría ser muy pequeño para los totales estimados de algunos ítemes como áreas en cultivos, pero muy grande para estimaciones del número de explotaciones, las que dependen mucho de la definición de "explotación agrícola" y de las ambigüedades en la aplicación de la definición. Como consecuencia, también son muy sensibles las estimaciones de promedios por explotación. En el caso del segmento abierto, el número de explotaciones por segmento, determinado por enumeradores, ha variado de una encuesta a otra, aunque no se cambió la definición de "explotación agrícola" ni el diseño de la muestra. Diferencias en (1) los propósitos de las encuestas, (2) los materiales usados en las encuestas, (3) los procedimientos operacionales y (4) el esfuerzo dedicado a encontrar todas las explotaciones que deben estar en la muestra, tienen su efecto sobre la magnitud del error de cobertura. No hay suficiente experiencia para saber si los resultados con el segmento ponderado serían más consistentes.

Se necesitan urgentemente investigación y análisis de los errores de cobertura. Tenemos que saber, por ejemplo, cómo se compara el promedio del error de cobertura (o sesgo debido al error de cobertura) en \bar{x}_p con el promedio del error de cobertura en \bar{x}_a , donde \bar{x}_p y \bar{x}_a son las medias muestrales por segmento correspondientes a los métodos de segmento ponderado y de segmento abierto. Con el uso del método de segmento ponderado, los datos agropecuarios para un segmento son ponderados (multiplicados por fracciones), lo que da un total para el segmento que es equivalente a la suma de las áreas de terreno en los tracts dentro del segmento. Esto quiere decir que la composición del sesgo debido al error de cobertura difiere de la del método de segmento abierto.

Una apreciación de un aspecto del error de cobertura puede expresarse

brevemente como sigue: Con una muestra aleatoria de n segmentos de una población de N segmentos, la fracción de muestreo es, teóricamente, $\frac{n}{N}$. La verdadera fracción de muestreo que se logra en la encuesta probablemente difiere algo de $\frac{n}{N}$ a causa del error de cobertura. Como se ha dicho anteriormente, la experiencia en las encuestas en que se usa el segmento abierto indica que es muy difícil lograr una fracción de muestreo verdadera que esté cerca de $\frac{n}{N}$. Posiblemente las operaciones con el segmento ponderado pueden controlarse con más éxito, en el sentido de que la fracción de muestreo lograda estará más próxima a $\frac{n}{N}$. Conceptualmente, con el segmento ponderado, el valor total de X para un segmento (vea la ecuación (2) en la sección 4.4.1) debe estar en un nivel que es equivalente a la suma de las áreas de terreno en los tracts dentro del segmento. Recuerde que, con el método de segmento cerrado, un total de una característica para un segmento está también en un nivel que es equivalente a la suma de las áreas de los tracts dentro del segmento. Necesitamos respuesta a la pregunta: ¿ofrece el método de segmento ponderado más valor potencial que el método de segmento abierto para la minimización del sesgo debido a errores de cobertura?

Tomando en consideración la experiencia ya adquirida, la mayor dependencia con relación al muestreo por áreas y los materiales mejorados para el muestreo por áreas, ha llegado la hora de explorar profundamente el método de segmento ponderado, especialmente en situaciones donde el de segmento abierto es menos factible. Los métodos de encuesta que se empleen no deben pasar por alto las posibilidades de una combinación de métodos, tal como la descrita en la próxima sección.

6.2.3 Combinación de métodos. En encuestas donde es posible recoger solamente una parte de los datos por el método de segmento cerrado, puede usarse en combinación con este método el de segmento abierto o el de segmento ponderado. ¿Qué combinación de métodos es mejor, cerrado y abierto, o cerrado y ponderado?

Parece que en todas las situaciones una muestra bien diseñada para aplicación del método de segmento ponderado sería también bien diseñada para uso del método de segmento cerrado. Con respecto a la combinación de abierto con cerrado, los principios para la definición de segmentos difieren entre el abierto y el cerrado. En algunas situaciones la misma muestra de segmentos no puede ser adecuada para ambos métodos, cerrado y abierto. Considere la situación donde casi todos los operadores de explotaciones agrícolas viven en poblados. En este caso, una muestra eficiente para los datos de segmento cerrado no se parecería en nada a una muestra eficiente para los datos de

segmento abierto, suponiendo que la sede de una explotación agrícola se define como la residencia del operador. (Si no está clara la diferencia, observe que una gran proporción de los segmentos en una muestra eficiente de segmentos cerrados se encontrarían en campo abierto donde está la tierra agrícola. Con una muestra de segmentos abiertos, queremos igualar el número de explotaciones "dentro" de los segmentos, lo que pondría una gran proporción de los segmentos muestrales en los poblados.) Además, en una muestra dada de segmentos en las circunstancias descritas, muy pocos operadores se entrevistarían tanto para los datos de tracts (segmento cerrado) como para los datos de la explotación (segmento abierto). Es decir, muy pocos de los tracts serían explotaciones enteras. Considerando las variancias de muestreo por dólar, quizá sería mejor utilizar dos encuestas con muestras distintas. Una podría diseñarse eficientemente para datos de segmento cerrado y la otra para datos de segmento abierto. Para la situación que acabamos de describir, parece que la combinación de cerrado con ponderado es claramente superior a la combinación de cerrado con abierto con respecto a cuestiones de diseño de la muestra y al hecho de que se usarían los mismos formularios en la recolección de datos sobre tracts y datos sobre explotaciones.

Al planificar una encuesta, considere con cuidado los costos por segmento para las combinaciones de cerrado-ponderado y cerrado-abierto. La diferencia en costos podría ser pequeña en relación con la variancia de muestreo más pequeña correspondiente a las estimaciones con el segmento ponderado.

Por último, hay un punto importante que hay que considerar con respecto al error de cobertura y de respuesta que no ha sido tratado y que a menudo se pasa por alto. La complejidad de la tarea del enumerador y su relación con la frecuencia de error es un factor clave. Es decir, los refinamientos adicionales que se hacen con el propósito de reducir errores podrían resultar en un aumento neto en el número de errores. ¿Cuál de las dos combinaciones (cerrado-abierto o cerrado-ponderado) es más fácil de entender para el enumerador? El cerrado y el ponderado tienen mucho en común y no es necesario confrontar problemas relacionados con la sede. Las explotaciones que corresponden a tracts en el segmento cerrado están en el segmento ponderado. Así, los mismos operadores se entrevistan para datos sobre tracts y datos sobre la explotación. En cambio, generalmente es más difícil para los enumeradores entender completamente los conceptos de la combinación de cerrado con abierto.

7. CONSTRUCCION DE MARCOS DE MUESTREO POR AREAS

7.1 Introducción

La construcción de un marco de muestreo por áreas se ve en esta publicación como una inversión mayor que se amortizará durante un período largo y después de muchas encuestas. Al terminar la construcción del marco, es aconsejable mantener personal que haga revisiones o mejoramientos en el marco y seleccione y prepare muestras cuando sea necesario. Un programa adecuado de mantenimiento y mejoramiento podría reducir o eliminar la necesidad de encontrar recursos para una reconstrucción completa del marco después de varios años. Monroe Y Finkner * han tratado de la construcción de un marco de muestreo por áreas para usarse en el muestreo de residencias.

Hay varias maneras de construir un marco por áreas, dependiendo de los recursos disponibles y los propósitos de la encuesta. Por esto, sólo se presentarán principios generales y algunas ilustraciones. Las personas responsables de la construcción de un marco de muestreo deben procurar utilizar al máximo las pericias disponibles sobre diseño de muestras y conocimiento de las condiciones locales que se encontrarán en la aplicación. Deben hacerse ensayos en pequeña escala antes de determinar las especificaciones finales para una inversión mayor.

7.2 Especificaciones de las Unidades de Marco

Para lograr economía en el diseño y selección de muestras por áreas, la "unidad de marco" ** es parte integral de un marco de muestreo por áreas. Una unidad de marco es un área de terreno que es más grande que un segmento, pero generalmente es más pequeña que la división política más pequeña.

Las partes esenciales de un marco de muestreo por áreas son (1) un conjunto de mapas en que se deslindan las unidades de marco, (2) una lista de las unidades de marco y (3) información sobre las unidades de marco, como el área de terreno o el número de hogares, que se utiliza para fines de diseño de muestras y para asignar un número de segmentos a cada unidad de marco. El

* Monroe, John y Finkner, A.L., "Handbook of Area Sampling", Chilton Company, 1959.

** Con los primeros marcos por áreas que se desarrollaron en los Estados Unidos, se usaba el término "unidad de conteo". Una unidad de conteo era más grande que una unidad de muestreo y se llamaba "unidad de conteo" porque las explotaciones indicadas en los mapas de carreteras se contaron para cada unidad de conteo. Aunque "unidad de conteo" tiene uso ampliamente extendido, el autor cree que debe ser rechazado, empleando en su lugar un término más general, como "unidad de marco".

número que se asigne puede variar según el propósito de la encuesta, el uso del segmento abierto, cerrado o ponderado, el detalle topográfico mostrado en los mapas e información disponible sobre el uso de la tierra o la agricultura dentro de la unidad de marco. Después de haber asignado un número de segmentos a cada unidad de marco y de haber formulado las especificaciones del diseño de la muestra, se selecciona una muestra de unidades de marco con probabilidades proporcionales al número de segmentos asignados. Cada unidad de marco seleccionada se divide entonces en el número de segmentos asignados y uno de ellos se selecciona aleatoriamente.

Hay dos preguntas principales que han de considerarse en el desarrollo de especificaciones para un marco: (1) ¿cómo deben definirse las unidades de marco? y (2) ¿qué información debe compilarse respecto a cada unidad de marco? Las dos preguntas no son independientes, pero se tratarán separadamente.

Los factores que influyen en las especificaciones para unidades de marco incluyen:

(1) Los linderos de unidades de marco deben ser detalles topográficos permanentes y reconocibles definitivamente. Los linderos de divisiones políticas menores (especialmente si se cambian frecuentemente o no siguen la topografía) a menudo no son satisfactorios como linderos. Las unidades de marco son las partes más permanentes de un marco por áreas y deben definirse con linderos que sean relativamente permanentes. Los datos relacionados con las unidades de marco, como número de residencias o el uso dado a la tierra, pueden revisarse a medida que se obtenga información nueva. Si hay áreas donde el uso de la tierra está cambiando rápidamente, podría ser suficiente la actualización de información sobre las unidades de marco en tales áreas.

(2) Las unidades de marco deben ser suficientemente grandes para acomodar especificaciones alternativas de segmentos que sean apropiadas para varias encuestas.

(3) Las unidades de marco deben proporcionar una solución económica para la selección de muestras por áreas. No es necesario dividir en segmentos una unidad de marco sino cuando ha de seleccionarse de ella un segmento para formar una muestra dada. En general, el trabajo requerido para seleccionar una muestra es mínimo cuando el número de unidades de marco es mucho más grande que el número de segmentos que se necesitan para una muestra. El número total de unidades de marco se relaciona inversamente con el tamaño medio de las unidades de marco. Hay que comparar los costos de definición de un número grande de unidades de marco pequeñas (en vez de

un número menor de unidades más grandes) con los costos de seleccionar muestras después de la construcción del marco.

El uso de unidades de marco también produce a veces ahorros en el costo de mapas o fotografías. Suponga que existen mapas relativamente baratos pero que son adecuados para deslindar unidades de marco, además de proveer documentación de los linderos de unidades de marco para uso de la oficina. Podría ser que tales mapas no suministraran suficiente detalle para dividir las unidades de marco en segmentos. Los mapas o fotografías más adecuados para este fin podrían ser muy caros. Podría ser suficiente limitar la compra de los mapas o fotografías más caros a los necesarios para dividir las unidades de marco de las cuales se seleccionará un segmento.

(4) Debe darse consideración a las varias clases de información que podría estar disponible y clasificada por unidades de marco, para usarla en el diseño de muestras. Esto podría influir en las especificaciones de la unidad de marco. Por ejemplo, para usar datos de un censo de agricultura, tal vez convendría que las unidades de marco coincidieran con los distritos de enumeración usados en el censo.

(5) Las poblaciones y subpoblaciones que han de someterse a encuesta son definidas generalmente en términos de extensión geográfica, además de su clasificación en unidades de información. Podría haber alguna ventaja en definir las unidades de marco de acuerdo con linderos geográficos que pudieran usarse en las especificaciones de poblaciones involucradas en la encuesta.

(6) Hay dos modos generales (y combinaciones de ellos) para establecer las especificaciones de las unidades de marco: (a) Un modo consiste en formular las especificaciones principalmente con referencia a tamaño (área de terreno) y detalles topográficos que sirvan de linderos. En este caso, el trabajo de definición de unidades de marco es mínimo. Después de definir las unidades de marco, se compila información apropiada para las unidades de marco con respecto al tipo de población que ha de ser muestreada y cómo han de definirse los segmentos. (b) Con el segundo modo, las especificaciones para las unidades de marco incluirían factores como el uso dado a la tierra, para lograr mayor homogeneidad dentro de las unidades de marco. Si la variación dentro de unidades de marco es pequeña, debe ser efectiva la estratificación de las unidades de marco para fines de muestreo. Además, podrían aplicarse distintos procedimientos a distintas clases de unidades de marco, lo que podría influir en la manera de definir las unidades de marco. Por ejemplo, las unidades de marco que abarcan áreas residenciales podrían tratarse de manera distinta de la que se emplearía en el caso de unidades de marco que incluyen solamente campo abierto. En

cualquier caso, sin importar las especificaciones para las unidades de marco, el resultado final es un conjunto definido de unidades de marco y alguna información sobre cada unidad que es útil y disponible para fines de muestreo. Los dos modos se relacionan con diferencias en los linderos físicos de las unidades de marco y diferencias en la manera en que se utiliza la información auxiliar. Empero, los objetivos son claros. Queremos detalles topográficos permanentes y visibles para los linderos de las unidades de marco y la utilización económica y efectiva de información auxiliar para reducir la variancia de muestreo. Los ajustes que hay que hacer se clarificarán a cierto grado en las secciones que siguen.

7.3 Información Auxiliar y su Utilización

Se llamará "información auxiliar" o "datos auxiliares" a la información o a los datos que están disponibles para su uso en el diseño de muestras. Hay una gran variedad de información auxiliar y hay muchas maneras de utilizar tal información en el diseño de muestras, siendo el objetivo general el de lograr el máximo de exactitud (suponiendo un costo fijo para la encuesta). En este momento quizá será útil un breve repaso de los principios claves involucrados en la aplicación de muestreo aleatorio estratificado de una sola etapa, en cuanto se relacionan con el muestreo por áreas.

Para minimizar la variancia de muestreo, el que diseña la muestra quiere definir estratos y unidades de muestreo por áreas (segmentos) de tal manera que la variancia entre las unidades de muestreo dentro del estrato sea tan pequeña como sea posible. Es decir, el diseñador de muestras se ocupa de (1) la selección de criterios para la estratificación y la afijación de la muestra en los estratos y (2) la definición de unidades de muestreo, incluyendo el control de la variación en tamaño de las unidades de muestreo. Dentro de los estratos, la variación entre unidades de muestreo será relativamente pequeña cuando las unidades de muestreo son casi iguales en "tamaño" y tienen características similares. El diseñador también busca un tamaño medio de unidad de muestreo que sea eficiente con respecto al error cuadrático medio para un costo dado. Estos detalles de diseño de muestras están relacionados con el propósito de la encuesta.

Como acabamos de indicar, hay típicamente dos maneras de utilizar los datos auxiliares al diseñar una muestra por áreas: una es para la estratificación, siendo el objetivo el de lograr homogeneidad dentro de los estratos; la segunda es el uso de una variable auxiliar como "medida de tamaño", con el objetivo de lograr segmentos de "tamaño" igual, donde la

medida de tamaño es una variable que está correlacionada con las variables que serán incluidas en la encuesta. Algunas clases de información son útiles para fines de estratificación, pero no son útiles como medidas de tamaño para controlar el tamaño de segmentos. (Ejemplos son localización geográfica, tipos de suelo, o mapas que indican tipos generales de áreas agropecuarias.) Hay características (e.g., área en cultivos) que pueden usarse o como medida de tamaño o como base para la estratificación. Generalmente, la misma variable auxiliar no se usaría como medida de tamaño a la vez que como criterio para la estratificación.

Teóricamente, no son independientes la selección de criterios para la estratificación de unidades de marco y la selección de una medida de tamaño que se usa para asignar números de segmentos a las unidades de marco y controlar la variación en tamaño de los segmentos. Cuando las opciones lo permiten, el autor generalmente prefiere dar primera prioridad a la selección de una medida de tamaño para controlar el tamaño de los segmentos y segunda prioridad a los criterios para la estratificación, con la debida consideración de la medida de tamaño, el estimador y los objetivos de la encuesta. No obstante, la oportunidad de considerar alternativas en el diseño es limitada por el grado hasta el cual se desarrolla el marco de muestreo por áreas (incluyendo datos auxiliares por unidades de marco) para acomodar varios objetivos de la encuesta.

7.3.1 Control del tamaño del segmento. Teóricamente, las maneras de controlar (reducir) la variancia de muestreo asociada con la variación en el tamaño de las unidades de muestreo incluyen: la estratificación de las unidades de muestreo según tamaño, la selección de unidades de muestreo con ppt (probabilidad proporcional al tamaño), estimadores de regresión o por la razón y el igualamiento del tamaño de las unidades de muestreo. En la discusión e ilustraciones que siguen, la atención se enfocará en el igualamiento de tamaños de segmentos. Sin embargo, el detalle que aparece en los mapas y los detalles topográficos, lo mismo que el tipo de información auxiliar que esté disponible, a menudo limitan el grado de igualamiento lograble. Si existe información pertinente para el control de la variación entre segmentos, pero la topografía limita severamente el igualamiento de tamaño de los segmentos, pueden considerarse los otros métodos mencionados. Con respecto a la estimación por la razón, recuerde la precaución anotada en la sección 4.4.3.

La selección de segmentos individuales con ppt no ha sido usada generalmente--involucra consideraciones técnicas que van más allá del alcance de esta publicación. De paso, la selección de segmentos con ppt no es lo mismo

que la selección de unidades de marco con probabilidad proporcional al número asignado de segmentos seguida por la división de cada unidad de marco seleccionada en su número asignado de segmentos, etc. Este procedimiento es un método que da a cada segmento la misma probabilidad de estar en la muestra. La estratificación se aplica a unidades de marco y no a segmentos individuales. En algunas circunstancias la estratificación puede ser útil en el control del tamaño de los segmentos. Esta situación se ilustrará después.

La selección de una medida de tamaño de segmento depende del propósito de la encuesta y del uso del segmento abierto, cerrado o ponderado. El control del tamaño del segmento involucra la asignación de un número conveniente de segmentos a cada unidad de marco y la división apropiada de las unidades de marco en segmentos. Por ejemplo, considere una encuesta sobre cultivos de frutas. Suponga el uso del método de segmento cerrado y la existencia de una medida aproximada del terreno usado para cultivo de frutas en cada unidad de marco. En este caso, el número de segmentos asignados a las unidades de marco sería proporcional a la cantidad aproximada de terreno usado para cultivo de frutas. La meta sería dividir una unidad de marco en el número asignado de segmentos de tal manera que cada segmento tenga aproximadamente la misma cantidad de terreno en cultivos de frutas. Este principio se usa en las ilustraciones que se presentarán más adelante.

7.3.2 Estratificación y la definición de unidades de marco. Como se ha dicho en el párrafo (6) de la sección 7.2, la información auxiliar podrá usarse de manera que influya en la definición de las unidades de marco. Un ejemplo destacado de esto es la clasificación de todo el terreno según su uso, seguida por el deslindar de las unidades de marco dentro de cada una de las clases de uso del terreno. Una alternativa es deslindar las unidades de marco con poca o ninguna consideración del uso del terreno y entonces estratificar las unidades de marco según el uso del terreno para fines de muestreo. Una de las primeras cuestiones que hay que resolver es la de si la información sobre el uso del terreno se tomará en consideración antes de deslindar las unidades de marco o después de ello. Al comparar las alternativas y hacer la elección, es importante distinguir entre ventajas operacionales y otras consideraciones como la eficiencia de muestreo o la posibilidad de incurrir en sesgos. Se pueden describir situaciones en que la elección sería asunto de conveniencia en los procedimientos en vez de eficiencia en el muestreo.

Se podrían definir clases de usos del terreno antes de deslindar

unidades de marco, con intenciones de (1) estratificar para lograr homogeneidad dentro de los estratos, (2) hacer coincidir los linderos de unidades de marco con áreas que podrían usarse como dominios de estudio o (3) formar clases de unidades de marco de manera que las unidades de marco dentro de una clase se trataran de modo similar, pero el tratamiento de una clase pudiera ser diferente del de otra clase. Influyen mucho en la selección de especificaciones para unidades de marco el tipo de uso del terreno, la topografía y los propósitos más probables del marco de muestreo. Quizá será útil una breve exposición de dos casos hipotéticos que involucran tipos muy diferentes de uso del terreno y de topografía.

Caso 1. Suponga que el área total de terreno para la cual hay que producir un marco por áreas puede dividirse fácilmente en cuatro clases según el uso dado a la tierra: tierra sembrada con árboles, tierra labrada, pastoreo y tierra no agrícola. Suponga que la distribución de los usos y la topografía son tales que (1) las clases de tierra pueden deslindarse de modo que los linderos de las clases sirvan de linderos de unidades de marco y que (2) el uso de la tierra dentro de la clase concuerda con la clase excepto una que otra parcela esparcida que no significa más del 10 ó 15 por ciento del área total de la clase.

En este caso, probablemente sería ventajoso deslindar las clases de uso de la tierra y las unidades de marco dentro de esas clases. Las unidades dentro de una clase serían bastante parecidas y el área de las unidades de marco podría servir como medida útil de tamaño para varios propósitos de muestreo. Es decir, una lista de unidades de marco por clase de uso de la tierra y el área de cada unidad de marco provee una base que es razonablemente satisfactoria para muestreo de propósito general y da una base que puede refinarse o desarrollarse más en caso necesario.

Como ilustración, suponga que hay que diseñar y seleccionar una muestra para una encuesta sobre cultivos anuales. Una de las primeras decisiones que hay que tomar es reconocer la extensión geográfica de la población que ha de muestrearse. Supongamos que dos clases de uso de la tierra, la de pastoreo y la no agrícola, pueden omitirse, pero la tierra sembrada con árboles tiene demasiado terreno en cultivos anuales para poder ser ignorada. Las dos clases de tierra según su uso, cultivos anuales y las arboledas, se muestrearían de modos distintos, como sigue.

Suponiendo que se va a usar el método de segmento cerrado o el de segmento ponderado, una medida apropiada del tamaño de un segmento es la cantidad de tierra en cultivos anuales. Es decir, la meta es definir los segmentos para que cada uno tenga cantidades aproximadamente iguales de tierra cultivada. En

la clase de tierras en cultivos anuales, una gran proporción de la tierra es cultivada y el área total de las unidades de marco es una medida adecuada de tamaño, en lugar de estimaciones de la cantidad de tierra cultivada por unidades de marco. Así es que, bajo las circunstancias, el hacer los números de segmentos asignados a las unidades de marco proporcionales al área total de las unidades de marco probablemente conducirá a segmentos que son tan iguales en tamaño como serían si los números asignados fueran proporcionales a la cantidad de tierra cultivada en las unidades de marco. La conversión de las áreas de unidades de marco a números de segmentos es asunto fácil después de haberse decidido el tamaño medio de los segmentos. Por ejemplo, suponga que el promedio de tamaño de los segmentos se establece en 300 acres. Una unidad de marco con área estimada de 1,400 acres sería dividida en cinco segmentos. Para fines de muestreo, las unidades de marco en la clase de tierra cultivada podrían estratificarse geográficamente o de acuerdo con cualquier otro criterio que estuviera disponible. La selección de las unidades de marco y la división de las selecciones en segmentos estarían de acuerdo con los principios ya explicados.

En la clase de la tierra sembrada con árboles, hay que dar consideración a cómo se distribuye geográficamente la tierra cultivada. Si la tierra cultivada está distribuida uniformemente entre las unidades de marco, la asignación de números de segmentos a las unidades de marco podría hacerse del mismo modo, pero el promedio del tamaño (área) de los segmentos sería mayor. Por ejemplo, si como el 10 por ciento de la tierra es cultivada y se ha decidido asignar 100 acres de tierra cultivada como promedio por segmento, el área total de segmentos sería 1,000 acres, como promedio. Entonces, una unidad de marco con área total de 5,000 acres sería dividida en cinco segmentos. Si la proporción de tierra cultivada varía mucho entre unidades de marco, el método que acabamos de describir podría usarse, pero hay que considerar una alternativa que diera una variancia de muestreo menor. Por ejemplo, quizá sería factible examinar fotografías y asignar segmentos a las unidades de marco en proporción aproximada a la cantidad aparente de tierra cultivada en cada unidad.

Si se hubiera escogido el método de segmento abierto para esta encuesta, habría que dar atención a la densidad de sedes en lugar de la cantidad de tierra cultivada al dividir las unidades de marco en segmentos.

Caso 2. En contraste con el caso 1, suponga que el área para la cual se ha de construir un marco tiene un sistema de uso de la tierra y topografía de tal naturaleza que es imposible deslindar clases basadas en el uso dado a la tierra dentro de las cuales las unidades de marco fueran

parecidas, salvo que se aceptaran linderos vagos para las unidades de marco. Un ejemplo es un área donde la mayoría de la tierra no se cultiva a causa de las condiciones del suelo o de la topografía y las partes cultivadas son mayormente pequeñas parcelas irregulares esparcidas. Es obvio que si se intentara deslindar clases amplias según el uso dado a la tierra y dentro de las cuales se deslindarían unidades de marco, habría que hacer algún sacrificio importante. Habría que sacrificarse o la homogeneidad de uso de la tierra dentro de una clase o la calidad de los linderos de las unidades de marco. Además, podría ser muy difícil y prolongada en ciertas circunstancias la tarea de deslindar clases según uso de la tierra antes de deslindar las unidades de marco.

Con un costo relativamente bajo se podrían deslindar unidades de marco con poca o ninguna atención al uso dado a la tierra. Se podría compilar para cada unidad de marco una aproximación de la cantidad de tierra dedicada a distintos usos, con el fin de usarla como (1) medida de tamaño para la asignación del número de segmentos a las unidades de marco o (2) criterio para estratificar las unidades de marco para fines de muestreo. De manera que es posible hacer uso efectivo de información sobre el uso dado a la tierra sin aplicarla al deslinde de unidades de marco y sin introducir linderos ambiguos para las unidades de marco. La sección 8.2 presenta una ilustración de esta clase de situación.

El autor considera muy crítica la selección de linderos para las unidades de marco. En muchos casos una parte del lindero del segmento será también lindero de una unidad de marco. La experiencia ha demostrado que linderos ambiguos de unidades de marco dan mucho trabajo en la aplicación del muestreo por áreas, especialmente después del transcurso de algunos años. Como ya se ha dicho, las unidades de marco deben considerarse como el aspecto más permanente de un marco de muestreo por áreas. La flexibilidad para servir los propósitos de varias encuestas no es necesariamente restringida por el modo de definir las unidades de marco. Sin importar la definición, las unidades de marco pueden ser estratificadas de varias maneras y pueden dividirse de varias maneras en segmentos para propósitos distintos; la información auxiliar puede ser actualizada o suplementada en cualquier momento. Los logros de eficiencia en el diseño de una muestra dependen de la pertinencia y exactitud de la información correspondiente a las unidades de marco individuales. Es decir, es la extensión de la información pertinente acerca de las unidades individuales de marco lo que conduce a la adaptabilidad del marco para varios propósitos de encuesta.

El deslinde según uso dado a la tierra o según otra clasificación hecha antes de deslindar las unidades de marco es, en efecto, una manera de compilar información acerca de las unidades de marco. En comparación con el modo sencillo de deslindar unidades de marco con consideración mínima con respecto al uso dado a la tierra, debe justificarse sobre la base de una utilización más efectiva de la información auxiliar (lo que, en general, le parece algo dudoso al autor) o de economía en las operaciones de construcción de un marco y de selección de muestras. En cualquier caso, no debe hacerse caso omiso del uso dado a la tierra cuando se deslindan las unidades de marco. Por ejemplo, áreas urbanas y otras no agrícolas podrían requerir consideración especial. Pero, considere cuidadosamente las alternativas antes de hacer una inversión grande en deslindar clases según uso dado a la tierra antes de deslindar las unidades de marco, especialmente si involucra el sacrificio de calidad de los linderos de unidades de marco.

7.3.3 Selección de datos auxiliares sobre las unidades de marco. La disponibilidad de datos auxiliares varía entre países y aplicaciones, desde casi cero hasta incluir información altamente pertinente y efectiva en el diseño de muestras para minimizar la variancia de muestreo. El diseñador de muestras constantemente se confronta con escoger entre alternativas que influyen en la eficiencia de muestreo y en el sesgo. Además, frecuentemente tiene la responsabilidad de hacer recomendaciones o tomar decisiones respecto a datos auxiliares que parecen valer la pena de adquirir para uso futuro en diseños de muestras. Como apoyo a tales decisiones y a programas permanentes de mejoramiento de planes de muestreo y operaciones, debe existir un programa permanente de investigación y análisis de los distintos componentes de error y componentes de costos en las encuestas que se llevan a cabo.

El área total de terreno probablemente encabezará cualquier lista de información auxiliar que ha de compilarse acerca de las unidades de marco. Puede aproximarse fácilmente a partir de mapas a escala y probablemente se usará en muchos planes de muestreo. Estimaciones de la cantidad de tierra en cada unidad de marco según el uso dado a la tierra podrían ser importantes, dependiendo de la clase de encuestas que se esperan hacer y las circunstancias mencionadas en la sección 7.3.2. La cantidad de tierra en cada unidad de marco según el uso dado a ella es generalmente un dato más útil (más efectivo para reducir la variancia de muestreo) en los métodos de segmento cerrado y segmento ponderado que en el método de segmento abierto.

Las posibles fuentes de información acerca de unidades de marco

incluyen: (1) datos de los censos si las unidades de marco coinciden con distritos de enumeración, (2) mapas de los usos dados a la tierra, si contienen suficiente detalle, (3) fotografías aéreas y (4) estimaciones visuales hechas por observación de las unidades de marco en el campo. Las estimaciones visuales de las proporciones de tierra en cada una de las clases según uso correspondientes a cada unidad de marco podrían multiplicarse por el área estimada de la unidad de marco para obtener medidas de la cantidad de tierra dedicada a los distintos usos, lo que podría ser útil para fines de muestreo. El área de una unidad de marco puede estimarse con el uso de un planímetro o una reja transparente, si existen fotografías o mapas de escala conocida.

Si se va a usar el método de segmento abierto para encuestas de todas las explotaciones agrícolas, será útil una indicación del número de explotaciones "dentro" de cada unidad de marco, suponiendo que contribuye al objetivo de igualar el número de explotaciones "dentro" de los segmentos. Para encuestas sobre hogares, sería muy importante tener información sobre el número de hogares por unidad de marco.

No debe obtenerse información sobre las unidades de marco, especialmente cuando el costo es elevado, salvo que haya buenas perspectivas de que se usará de manera efectiva para reducir la variancia de muestreo. El costo de obtención de datos auxiliares tiene que ser considerado con respecto a la reducción de variancias de muestreo que podrían lograrse mediante un diseño mejorado de muestra. ¿Cómo se compara este costo con el costo de lograr reducciones similares de la variancia de muestreo mediante el aumento del tamaño de la muestra? Una inversión en la obtención de datos auxiliares para mejorar la muestra de una sola encuesta podría ser desaconsejable. Pero si se conducen periódicamente encuestas sobre los mismos ítemes (o temas), puede ser ampliamente justificable una inversión substancial en la adquisición de datos auxiliares.

Necesidades especiales importantes deben considerarse cuidadosamente. Por ejemplo, suponga que la producción de una clase especial de árboles frutales o de viñedos es de mucha importancia comercial en un país. Información sobre la localización exacta del cultivo o aproximaciones de la cantidad del cultivo en cada unidad de marco podría ser crítica para la obtención de un grado satisfactorio de eficiencia en el muestreo. Puede parecer demasiado caro el trabajo de campo necesario para adquirir información auxiliar acerca de las unidades de marco, pero el costo de baja eficiencia de muestreo podría ser aún mayor. Es interesante notar que los recuentos censales de árboles frutales a veces han sido justificados por la necesidad de tener una buena base para la

obtención de muestras para pronósticos o estimaciones de producción. A veces puede obtenerse información sobre unidades de marco, muy eficaz en el diseño de muestras para encuestas actuales de propósito especial, a menos costo que mediante un censo.

La capacidad de diseñar muestras eficientes por áreas en la agricultura (especialmente en el muestreo para propósitos especiales) depende mucho de información acerca de dónde se producen los distintos cultivos o ítemes. Si no hay disponible ninguna información auxiliar para el diseño de muestras eficientes y si el costo de obtenerla es demasiado, considere la posibilidad de usar un plan de muestreo doble. Es decir, seleccione una muestra grande y recolecte datos sobre las características de explotaciones agrícolas en la muestra. Esto proveerá una base para seleccionar submuestras que sean eficientes para varias necesidades específicas. Además, no pase por alto la posibilidad de relacionar datos censales con un marco por áreas. Un censo con un cuestionario corto podría planificarse para dos propósitos: (1) proveer estadísticas acerca de ítemes claves para publicación y (2) suministrar datos auxiliares asociados con un marco de muestreo por áreas que permitiera muestreo y estimación más eficientes en encuestas actuales.

7.4 Mapas para la Construcción de Marcos

Puede ser útil reconocer dos categorías grandes de mapas: (1) mapas que proveen detalles topográficos útiles para el deslinde de unidades de marco y segmentos y (2) mapas que proveen información auxiliar útil para el diseño de muestras. Algunos ejemplos son mapas que muestran el uso dado a la tierra, áreas bajo riego, tipos de suelos u otra información que pudiera usarse para la estratificación o para asignar un número de segmentos a cada unidad de marco.

En la primera categoría los mapas más comúnmente usados son los de carreteras, fotografías aéreas y mapas topográficos. Los requisitos de escala y detalle de los mapas difieren bastante para (1) propósitos de deslindar unidades de marco y proveer documentación de oficina sobre los linderos de unidades de marco y para (2) propósitos de dividir unidades de marco en segmentos e indicar los linderos de segmentos muestrales para uso en el campo. Para el primer propósito se usan generalmente mapas de carreteras o mapas topográficos que muestran carreteras. Consideraciones de costos y de espacio para archivar documentos podrían hacer que las unidades de marco sean definidas o marcadas en mapas relativamente baratos (y tal vez transferidos a microfilme). Para el segundo propósito, los mapas del marco

(mapas en que se definen las unidades de marco) no son siempre adecuados. Puede ser necesario usar fotografías o mapas más detallados, o puede ser necesario adoptar técnicas como las descritas en la próxima sección. De paso, cuando los segmentos se deslindan en fotografías aéreas para uso en el campo por los enumeradores, las fotografías son una ayuda valiosa para lograr cobertura completa y exacta de los segmentos de la muestra, además de proveer una identificación positiva de los linderos de los segmentos.

7.5 División de las Unidades de Marco en Segmentos

La división de las unidades de marco en segmentos a menudo presenta una gran variedad de problemas. Podría ser factible dividir algunas unidades de marco con la utilización de los mapas de marco, pero las fotografías aéreas o mapas más detallados son generalmente muy útiles y a menudo necesarios. Cuando el detalle en los mapas no permite la división satisfactoria de una unidad de marco en su número asignado de segmentos, hay varias técnicas que pueden ser útiles. Algunas técnicas alternativas son:

(1) Que el enumerador enumere completamente la unidad de marco. Es decir, trate la unidad de marco como si fuera un segmento de la muestra y complete un cuestionario para cada unidad de información en la unidad de marco. Suponga que k es el número de segmentos asignados a la unidad de marco. Para fines de tabulación podría ser usada una submuestra de una de cada k unidades de información enumeradas. Si todas las unidades de información se incluyen en la tabulación, recuerde que hay que usar la probabilidad p como base para la ponderación, donde p es la probabilidad que tuvo la unidad de marco de ser incluida en la muestra.

(2) Antes de empezar la encuesta, prepare una lista de las unidades de información que están en las unidades de marco y seleccione una submuestra de unidades de información, usando una fracción de muestreo igual a $\frac{1}{k}$. En este caso se daría al enumerador una muestra de unidades de información en vez de un segmento.

(3) Visite la unidad de marco y divídala en k segmentos sobre la base de detalles topográficos observados en el campo. Tenga la seguridad de que los croquis y apuntes describan adecuadamente los segmentos. En la oficina, seleccione aleatoriamente un segmento.

La primera alternativa es más práctica cuando k es pequeño, digamos 2 ó 3. En general, la tercera alternativa parece ser preferible a la segunda cuando se usa el método de segmento cerrado o ponderado, o cuando la misma muestra ha de usarse repetidamente.

A menudo es factible, usando los mapas disponibles, dividir parcial, pero

no totalmente, una unidad de marco. Por ejemplo, suponga que una unidad de marco ha de dividirse en cinco segmentos. Podría ser más factible dividirla en dos partes y asignar tres segmentos a la primera parte y dos segmentos a la segunda parte. Se seleccionaría aleatoriamente una de las partes, dando a la primera parte una probabilidad de selección de $3/5$ y a la segunda parte una probabilidad de $2/5$. La parte seleccionada entonces podría tratarse de acuerdo con una de las alternativas mencionadas. El valor de k sería 2 ó 3, dependiendo de cuál de las dos partes fuera seleccionada. Esta técnica de dividir parcialmente una unidad de marco podría reducir el número de mapas o fotografías necesarios. Por ejemplo, puede ser factible dividir parcialmente una unidad de marco mediante el uso de únicamente un mapa de carreteras. Entonces, para completar la tarea de división de la unidad de marco en segmentos, solamente se necesitarían fotografías para una parte de la unidad de marco.

A veces se encuentra que la división de una unidad de marco en el número asignado de k segmentos es posible solamente si se aceptan linderos no deseables. Pero los detalles topográficos podrían ser de tal naturaleza que permitieran la división satisfactoria de la unidad de marco en $k-1$ segmentos. Esta situación presenta una decisión entre "formar" la división de la unidad de marco en k segmentos, o dividirla en solamente $k-1$ partes. Si se acepta la división en $k-1$ partes, siguen dos alternativas: (1) trate las $k-1$ partes como si fueran segmentos, seleccione una de ellas aleatoriamente y, para fines de estimación, cambie la probabilidad de selección de $p(\frac{1}{k})$ a $p(\frac{1}{k-1})$, donde p es la probabilidad que tuvo la unidad de marco de ser seleccionada; (2) numere las partes 1 a $k-1$ inclusive. Suponga que la parte 1 es la más grande. Asígnele dos segmentos y a las demás $k-2$ partes asigne un segmento. Entonces seleccione una parte con probabilidad proporcional a su número asignado de segmentos. Si se ha seleccionado una de las partes de 2 hasta $k-1$ inclusive, úsela como segmento. Tuvo una probabilidad de selección igual a $p(\frac{1}{k})$. Si se ha seleccionado la primera parte, podría aplicársele una de las tres técnicas descritas al principio de esta sección. El valor de k sería 2.

Durante el deslinde y selección de segmentos, tenga la seguridad de especificar detalles de procedimiento que eliminen la posibilidad de sesgos. Por ejemplo, es muy importante que el proceso de división de las unidades de marco en segmentos sea separado de (es decir, completamente independiente de) el proceso de hacer las selecciones aleatorias. Para ilustrar cómo se puede introducir sesgo, suponga que la instrucción dada al personal de oficina es la de dividir una unidad de marco en segmentos y seleccionar un

segmento antes de seguir con la próxima unidad de marco. Al seleccionar un número aleatorio puede ser posible, salvo que se tomen precauciones especiales, ver el próximo número aleatorio en la lista. El conocimiento del próximo número aleatorio podría introducir sesgo en el trabajo de deslindar y numerar los segmentos en la próxima unidad de marco.

Otra ilustración de la posibilidad de introducir sesgo es el cambio de un lindero de segmento después de haber seleccionado aleatoriamente el segmento. Puede haber una fuerte tendencia a hacer esto cuando se encuentra que se necesita un mejor lindero para facilitar el trabajo del enumerador. Si se permiten cambios, los cambios deben ser mínimos y deben especificarse, para hacer los cambios, reglas estrictas que se crean insesgadas para fines prácticos. Tales prácticas siempre introducen la posibilidad de sesgo y un grado de incertidumbre acerca de la magnitud de cualquier sesgo que exista en los resultados. Por otro lado, algunos ajustes de los linderos podrían involucrar menos riesgo de sesgo que el permitir a los enumeradores que enumeren segmentos que tienen linderos ambiguos. Lo mejor es evitar esta situación al mayor grado factible. Tenga la mayor seguridad posible de que los linderos son satisfactorios antes de hacer las selecciones aleatorias. Esto da énfasis al punto mencionado anteriormente: los linderos de las unidades de marco deben coincidir con detalles topográficos permanentes y bien definidos.

A veces una diferencia en detalle parece carecer de importancia y se toma una decisión sobre la base de conveniencia. No corra riesgos innecesarios, usando detalles de procedimiento que pudieran introducir sesgo.

Ensaye ampliamente las alternativas factibles antes de fijar las especificaciones finales de un marco de muestreo. Los ensayos son necesarios para determinar costos, para evaluar alternativas y para depurar procedimientos.

8. CONSTRUCCION DE MARCO - ILUSTRACION NO. 1

Dos áreas que representan distintas situaciones topográficas y de uso de la tierra se han escogido para ilustrar marcos de muestreo por áreas y la selección de muestras. La primer área para la ilustración es una parte del condado de Mills, en el Estado de Iowa. Casi el 95% de la tierra en el condado de Mills consiste en explotaciones agrícolas. Aproximadamente el 85% de la tierra en las explotaciones es de cultivo y el tamaño medio de una explotación es de más de 300 acres (ó 121 hectáreas). Aproximadamente el 85% de los operadores viven en sus explotaciones. La densidad de explotaciones agrícolas es aproximadamente de dos por milla cuadrada.

En una gran parte de los Estados Unidos, incluyendo el condado de Mills, el Public Land Survey dividió la tierra en secciones (millas cuadradas). La sección estándar tiene 640 acres de tierra (casi 260 hectáreas). En el mapa de carreteras del condado (vea la figura 2a) se muestra cada sección como un cuadrado ($\frac{1}{2}$ por $\frac{1}{2}$ pulgadas) identificado por un número. La mayoría de las líneas seccionales siguen un detalle topográfico de algún tipo (camino, cerca o el lindero de un campo); pero a medida que han cambiado las prácticas agrícolas y los campos y las explotaciones han aumentado de tamaño, los detalles topográficos que siguen las líneas seccionales han desaparecido en parte. En el condado de Milla las secciones generalmente pueden identificarse por inspección visual de fotografías, pero las líneas seccionales no siempre son satisfactorias como linderos de unidades de marco o de segmentos.

El mapa de carreteras del condado, figura 2a, provee una base satisfactoria para la definición de unidades de marco. De hecho, en esta ilustración fue muy fácil deslindar las unidades de marco, como se ve en la figura 2b. Los linderos del condado se consideraron aceptables como linderos de unidades de marco. Con excepción de estos, no hubo necesidad de usar como linderos de unidades de marco otros detalles topográficos más que caminos permanentes. La figura 3a muestra una fotografía de la unidad de marco número 17. Para no tapar ningún detalle mostrado en la fotografía, el lindero de la unidad de marco número 17 se muestra en la figura 3b, que es la misma fotografía con la adición de linderos de la unidad de marco y de los segmentos. La figura 3b será comentada posteriormente. Algunos lectores podrán desear identificar los detalles topográficos en el mapa de carreteras, en la figura 2a o 2b, con los detalles en la fotografía, figura 3a.

Además de especificaciones para los linderos de unidades de marco, se necesita una especificación del tamaño mínimo de la unidad de marco. En

esta ilustración, se fijó en 4 millas cuadradas el mínimo preferido, con 3 millas cuadradas como mínimo absoluto. El tamaño máximo de la unidad de marco no es crítico. Fue de 6 a 7 millas cuadradas. La variación en el tamaño de las unidades de marco fue debida principalmente a la presencia de detalles topográficos adecuados para linderos de unidades de marco.

La agricultura y el uso dado a la tierra en el condado de Mills son tales que no es probable que se necesiten segmentos de más de 3 ó 4 millas cuadradas de extensión para una encuesta. Si la tierra hubiera de ser clasificada según su uso y las unidades de marco definidas dentro de estas clases de uso, se necesitarían especificaciones sobre: (1) las clases de uso dado a la tierra, (2) los detalles topográficos para los linderos de las clases y (3) el tamaño mínimo de una parcela para cada clase.

Se podría ahorrar tiempo en el deslinde de unidades de marco si se hicieran estas más grandes, pero tales ahorros no parecen tener importancia. Realmente, se requiere menos tiempo para seleccionar muestras cuando las unidades de marco son pequeñas. Si es necesario acomodar el uso de segmentos más grandes, las unidades de marco y los datos asociados pueden combinarse para formar unidades de marco más grandes. La cantidad de datos auxiliares que se podría necesitar para las unidades de marco no parecía ser factor importante en favor de unidades de marco más grandes (y por lo tanto, menos de ellas) en el caso de esta ilustración.

Las áreas de tierra en las unidades de marco podrían estimarse con el uso de un planímetro y el mapa del marco, figura 2b. Sin embargo, a ojo de buen cubero se pueden estimar las áreas de tierra con un error no mayor de una media milla cuadrada, lo que probablemente es suficientemente exacto para fines de muestreo. La columna (2) de la tabla 6 muestra el área aproximada de cada unidad de marco, tal como fue determinada por interpretación visual del mapa del marco. La necesidad de tener información auxiliar, aparte de áreas, sobre las unidades de marco será considerada a medida que progresa esta explicación. De paso, cada unidad de marco debe tener asignado por lo menos un segmento y tener oportunidad de ser seleccionada, salvo que haya evidencia positiva de que no contiene nada que pudiera contribuir a la población que se muestrea.

Para ilustrar cómo se puede usar el marco para diseñar y seleccionar muestras, se considerarán tres clases de encuesta: (1) una encuesta sobre áreas en cultivo, (2) una encuesta sobre datos económicos y (3) una encuesta sobre ganado de carne.

8.1 Una Encuesta sobre Areas en Cultivos

Suponga que se ha de realizar una encuesta por muestreo, después de la siembra, con el fin de estimar el área sembrada con cada cultivo. Para este propósito, el método de segmento cerrado es superior al de segmento abierto y al de segmento ponderado, suponiendo que los tracts son satisfactorios como unidades de información. Entre los aspectos importantes de un plan de muestreo se encuentran los criterios para la estratificación y el tamaño de la muestra, pero enfocaremos la atención principalmente en la especificación y deslinde de segmentos. Además, el problema de muestreo se considerará en el contexto de una muestra de propósito general para todos los cultivos en vez de una muestra diseñada para uno o dos cultivos específicos.

Con referencia a los propósitos y condiciones que se han esbozado, una meta apropiada al deslindar los segmentos cerrados es el igualamiento de los tamaños de los segmentos con respecto a la cantidad de tierra cultivada. El primer paso es asignar un número de segmentos a cada unidad de marco. Si los segmentos han de tener cantidades iguales de tierra cultivada, el número asignado de segmentos debe estar en proporción a las cantidades de tierra cultivada en las unidades de marco. En el condado de Mills, una gran proporción de la tierra es cultivada. De manera que el área de las unidades de marco, después de hacer cualquier deducción factible de tierras no agrícolas, es una muy buena medida de tamaño.

Puesto que existen fotografías para deslindar las unidades de marco, es factible fijar el tamaño medio de segmento como la mitad de una sección. Se podría considerar un promedio menor, como un cuarto de sección, pero esto no parece práctico y el error de cobertura tiende a aumentar a medida que disminuye el tamaño de los segmentos. La cuarta columna en la tabla 6 muestra el número de segmentos cerrados asignados a cada unidad de marco. Los números asignados son dos veces el número de millas cuadradas (columna (2), tabla 6), con excepción de la unidad de marco número 24. Las unidades de marco se revisaron rápidamente para identificar áreas aparentemente no agrícolas que fueran mayores que media milla cuadrada. La única área encontrada fue una población parcialmente en la unidad de marco número 24. Hay tres millas cuadradas en la unidad de marco número 24, pero se le asignaron cinco segmentos en vez de seis, porque tiene por lo menos media milla cuadrada de área residencial. Así que la idea era asignar números de segmentos proporcionales a las áreas de las unidades de marco después de haberles restado cualquier área no agrícola más grande que media milla cuadrada. Si la

Tabla 6. Unidades de marco y número de segmentos para la ilustración no. 1

Número de unidad de marco	Tamaño aproximado de las unidades de marco, en millas cuadradas	Número indicado de explotaciones	Segmentos cerrados o ponderados		Segmentos abiertos	
			No. asignado	No. acumulado	No. asignado	No. acumulado
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	7	25	14	14	16	16
2	4	15	8	22	10	26
3	4	10	8	30	6	32
4	5	11	10	40	7	39
5	5	17	10	50	11	50
6	3	8	6	56	5	55
7	8	14	16	72	9	64
8	6,5	20	13	85	13	77
9	6	17	12	97	11	88
10	4	14	8	105	9	97
11	5	11	10	115	7	104
12	4	12	8	123	8	112
13	4	11	8	131	7	119
14	3	17	6	137	11	130
15	5,5	19	11	148	12	142
16	4	13	8	156	8	150
17	4	6	8	164	4	154
18	4	12	8	172	8	162
19	6	14	12	184	9	171
20	6	18	12	196	12	183
21	6	25	12	208	16	199
22	6	24	12	220	16	215
23	7	15	14	234	10	225
24	3	8	5	239	5	230
25	4,5	17	9	248	11	241
Totales		373	248	248	241	241

unidad de marco número 24 resultara seleccionada para dividirse en segmentos, su área total sería incluida en los cinco segmentos, aunque la parte residencial no fue contada cuando se asignó el número de segmentos. Es decir, la parte residencial de la unidad de marco sería incluida en uno o más de los cinco segmentos. Además, uno procuraría definir los cinco segmentos de modo que contuvieran cantidades iguales de tierra cultivada.

La quinta columna de la tabla 6 muestra el número acumulado de segmentos para los métodos de segmento cerrado o de segmento ponderado. Los totales acumulados a menudo se generan como una manera conveniente de seleccionar unidades de marco con probabilidades proporcionales a los números de segmentos asignados. Una discusión de métodos alternativos de seleccionar una muestra de segmentos de los 248 asignados en la columna (4) de la tabla 6 involucra consideraciones técnicas que están fuera del alcance de esta publicación. Sin embargo, suponga que hay que seleccionar aleatoriamente un segmento. Se selecciona, en forma aleatoria, un número del 1 al 248 inclusive. Suponga que el número aleatorio es 157, que es, con referencia al total acumulado, mayor que 156 y menor que 165. Por lo tanto, se selecciona la unidad de marco número 17. Su probabilidad de selección es igual a $\frac{8}{248}$.

El próximo paso es dividir la unidad de marco número 17 en su número de segmentos asignados, que es 8. Esta unidad de marco se divide muy satisfactoriamente bajo el criterio de buenos linderos y uniformidad de tamaño con respecto a la cantidad de tierra cultivada (vea la figura 3b). Después de numerar los 8 segmentos de 1 a 8, se selecciona aleatoriamente uno de ellos. Suponga que se selecciona el número 7. Su probabilidad de ser seleccionado es igual a $\frac{8}{248} \left(\frac{1}{8}\right) = \frac{1}{248}$.

Segmentos adicionales podrían seleccionarse del mismo modo. Sin embargo, a menudo se hace una selección sistemática, como sigue. Suponga que la fracción de muestreo es el 2 por ciento, ó 1 entre 50. Se seleccionaría aleatoriamente un número del grupo de 1 a 50 inclusive. Esta operación designa el primer número en una serie de números que tienen intervalo de 50. Suponga que el número seleccionado aleatoriamente es 12. La serie es 12, 62, 112, 162 y 212, lo que, con referencia a la tabla 6, designa las unidades de marco 1, 7, 11, 17 y 22, dentro de las cuales se han de deslindar segmentos y después seleccionar uno de los segmentos aleatoriamente. Puesto que hay dos pasos--la selección de unidades de marco y luego la selección de un segmento en cada unidad de marco--este procedimiento de selección a menudo se confunde con el muestreo en dos etapas. Pero en el caso que acabamos de describir los dos pasos son dos pasos de selección en un

plan de muestreo de una sola etapa.

La figura 3c es una fotografía del segmento número 7 en escala aumentada. Es ejemplo de una fotografía que un enumerador podría llevar consigo al segmento, excepto que las líneas que demarcan tracts y campos dentro del segmento no aparecerían. Al visitar el segmento y orientarse (es decir, identificar los linderos en la fotografía con la topografía del sitio), el enumerador divide el segmento en tracts. En el segmento 7 hay solamente tres tracts: A, B y C. Entoces, en una entrevista con el operador de un tract, el enumerador divide el tract en campos y obtiene la información deseada sobre los cultivos. Fijese que una fotografía del segmento es una ayuda importante para minimizar los errores de cobertura y de medición.

Si las fotografías tienen escala conocida, se pueden medir los campos con planímetro y usar los resultados como verificación de las áreas anotadas por los enumeradores. Aunque no esté presente un operador para la entrevista, el enumerador probablemente podrá obtener la información deseada sobre áreas en cultivo. Podría hablar con personas informadas o por observación visual podría deslindar campos y anotar, en cuanto fuera posible, el cultivo sembrado en cada campo. Las áreas de los cultivos podrían estimarse. Así es que el método de segmento cerrado provee un medio de obtener datos exactos y casi completos, en comparación con lo que sería posible o factible al usar otros métodos de encuesta.

Se dijo antes que, en este ejemplo, el área de una unidad de marco, menos la tierra no agrícola, era una buena medida de tamaño. Esto es verdad principalmente para cultivos comunes. Para cultivos menores (que ocupan áreas relativamente pequeñas) una variable auxiliar como área en cultivo o tierra laborable es generalmente de menor valor para la reducción de la variancia de muestreo.

Hay que poner atención especial a cualquier cultivo "menor" importante cuya variancia de muestreo tenga que ser pequeña. Un modo de hacerlo es seleccionar una muestra "general" diseñada para ser adecuada únicamente para los cultivos mayores, pero en la cual se recolectaría información sobre todos los cultivos. Además, se podría usar una o más muestras suplementarias diseñadas específicamente para los cultivos menores. Se combinarían los resultados de las muestras general y suplementaria, aplicando las ponderaciones apropiadas. Esto implica la existencia de una base para el diseño de muestras suplementarias para los cultivos menores dados. Si no hay tal base, no hay más alternativa que aumentar el tamaño de la muestra "general".

Puede ser muy útil en el diseño de muestras la información auxiliar sobre unidades de marco que dé alguna indicación de la cantidad (o la proporción) de tierra que probablemente se sembrará con cada uno de los cultivos menores.

Cuando esto es posible, el tamaño de las unidades de marco y el tamaño de los segmentos para una muestra suplementaria podrían ser muy diferentes de la medida de tamaño usada en la muestra general. Como ya se ha dicho, un asunto importante es ¿cuánto debe invertirse en la obtención de información auxiliar acerca de las unidades de marco? La pregunta análoga con respecto a marcos de lista (listas de operadores de explotaciones agrícolas para fines de muestreo) es ¿qué información debe acumularse y mantenerse acerca de las explotaciones individuales en la lista? De paso, la producción de algunos cultivos menores puede cambiar de año a año en fincas o localidades a tal grado que los datos auxiliares sobre el sitio donde se cultivaban antes ya no tienen valor.

Antes de proceder al próximo ejemplo, parece aconsejable mencionar el valor de las fotografías. En la ausencia de fotografías el requisito de que los linderos de los segmentos sean identificables con el uso de los mapas del condado habría significado segmentos más grandes y menor éxito en el igualamiento de tamaño de los segmentos. En otras palabras, por lo menos con respecto a la situación descrita arriba, alguna reducción de la variancia de muestreo puede atribuirse al uso de fotografías. Las fotografías también ayudan a reducir los errores de cobertura. La primera impresión podría ser que el costo de las fotografías es excesivo. Antes de llegar a tal conclusión, considere el costo de no usar fotografías. Es decir, considere el costo de lograr una reducción equivalente de la variancia de muestreo mediante un aumento en el tamaño de la muestra. También tome en cuenta la posibilidad de usar las mismas fotografías durante varias encuestas.

8.2 Una Encuesta sobre Datos Económicos

En una encuesta para recolectar datos sobre características económicas de todas las explotaciones, es posible usar el método de segmento abierto o de segmento ponderado. Puesto que el procedimiento ya descrito para el segmento cerrado es apropiado para una encuesta de explotaciones agrícolas que usa el segmento ponderado, la discusión que sigue se limitará al método de segmento abierto. Aunque una encuesta se considera de propósito general, puede existir la necesidad, para fines analíticos, de variar las fracciones de muestreo según características tales como tamaño o tipo de explotación. Este punto se examinará más tarde. Mientras tanto, se supone que todas las explotaciones deben tener la misma probabilidad de ser incluidas en la muestra.

La densidad de explotaciones agrícolas en el condado de Mills es

aproximadamente dos por milla cuadrada. La experiencia basada en análisis de variancias, costos y error de cobertura sugiere que el mejor tamaño medio de segmento abierto es probablemente menos de dos explotaciones, considerando la agricultura y la topografía asociadas con esta ilustración. Se supone un tamaño medio de una explotación por segmento, lo que quiere decir que queremos que el número de segmentos asignados a una unidad de marco sea igual al número de explotaciones que están "dentro" de la unidad de marco. No hay manera práctica de lograr esto con exactitud.

La base para asignar segmentos tiene que determinarse con respecto a la manera de definir la sede de la explotación. Si la residencia del operador es por definición la sede de la explotación, se necesita saber dónde vive el operador. En este caso, la meta es asignar números de segmentos a unidades de marco en proporción a los números de operadores que viven en las unidades de marco. Puede ser que no haya buena base para hacer esto. Por otro lado, suponga que un punto especificado dentro de los linderos de cada explotación es la sede de la explotación. Si no hay información sobre la ubicación de las sedes, se podrían asignar los segmentos en proporción a la cantidad de tierra agrícola o tierra laborable.

Con respecto al condado de Mills, aproximadamente el 85 por ciento de los operadores viven en sus explotaciones y algunos del otro 15 por ciento viven en campo abierto. Supongamos que la sede de la explotación es la residencia del operador si el operador vive en la explotación; si no vive allí, la sede es algún otro punto definido en la explotación. La información disponible y la discusión del párrafo anterior señalan dos alternativas. La primera es asignar segmentos a las unidades de marco en proporción a las áreas. La meta fue un promedio de una explotación por cada segmento abierto y, puesto que la densidad es de dos explotaciones por milla cuadrada, el tamaño medio de segmento sería media milla cuadrada. Por lo tanto, esta alternativa da una asignación de segmentos que resulta ser, en este caso, la misma que la asignación de segmentos cerrados en la columna (4) de la tabla 6. La división de unidades de marco en segmentos será diferente, sin embargo, porque el objetivo es igualar el número de explotaciones en los segmentos.

La segunda alternativa es derivar, como sigue, una indicación del número de explotaciones que están "dentro" de cada unidad de marco y entonces asignar segmentos en proporción al número indicado de explotaciones. En el campo abierto los mapas de carreteras tienen símbolos cuadrados, ■, que indican la posición de residencias rurales. Estos símbolos no están siempre correctos, pero son útiles. En algunas de estas localizaciones indicadas puede ser que la casa no esté ocupada por un operador de explotación agrícola. De hecho, puede no

haber ninguna casa en una de las localizaciones indicadas. Además, algunos operadores viven en sitios que no son identificados en los mapas. No obstante, se presenta en la tercera columna de la tabla 6 un recuento de las residencias rurales indicadas en el mapa del condado. Este recuento (373), a juzgar por el censo de agricultura, es como el 50 por ciento mayor que el número verdadero de explotaciones agrícolas. Probablemente se podría obtener, mediante fotografías, una indicación más exacta del número de residencias en las unidades de marco. En las fotografías se pueden identificar sitios edificados que probablemente son residencias de agricultores, pero eso tampoco da una identificación exacta y completa. Sin embargo, a menudo se han establecido estos números indicados de explotaciones agrícolas, los que han sido usados en la asignación de segmentos abiertos a las unidades de marco. Sin importar cómo se asignen segmentos abiertos a las unidades de marco, cuando se divide una unidad de marco, debe dividirse en el número asignado de segmentos de manera de tener el mismo número de explotaciones en cada segmento.

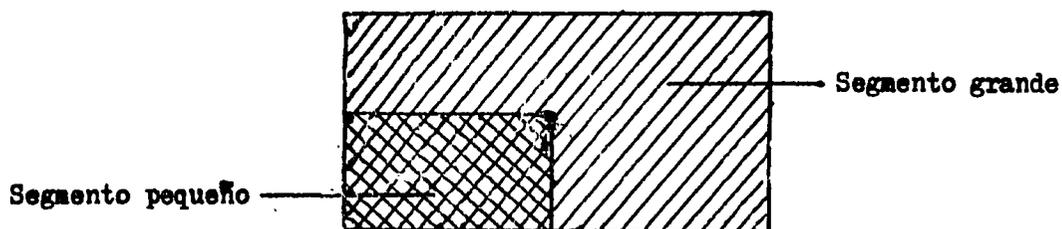
Para fines de ilustración, usaremos los números indicados de residencias rurales en la columna (3) de la tabla 6 para la asignación de segmentos. Recuerde que estos números indicados son como el 50 por ciento más grandes que el número verdadero de explotaciones. Buscamos un promedio de una explotación por segmento. Por lo tanto, los números asignados de segmentos abiertos en la columna (6) son como las dos terceras partes del número indicado de residencias rurales mostrado en la columna (3).

Como ejemplo, la unidad de marco número 17 será dividida en segmentos, porque fue usada anteriormente. El número de segmentos asignados fue 4 (vea la columna (6) en la tabla 6). La figura 3d muestra la unidad de marco número 17 dividida en 4 segmentos, suponiéndose el uso del método de segmento abierto. De paso, la fotografía (figura 3a) indica seis lugares donde probablemente resida un operador de explotación. Esto está de acuerdo con el mapa de carreteras.

Como caso especial, supongamos que los objetivos de una encuesta son tales que una fracción de muestreo uniforme es satisfactoria, salvo para explotaciones del tipo A, que forman una pequeña proporción de la población total. Es decir, se desean estimaciones separadas para un dominio compuesto de explotaciones del tipo A, pero las variancias de muestreo son demasiado grandes. ¿Cómo puede aumentarse el tamaño de la muestra de explotaciones del tipo A sin aumentar la muestra de todas las explotaciones? Si las explotaciones del tipo A están concentradas suficientemente, puede ser

posible definir el área de concentración y simplemente aumentar la fracción de muestreo para esa área únicamente. Si esta técnica no es apropiada, hay variaciones de por lo menos dos otros modos generales que podrían considerarse:

(1) El primero es más aplicable en situaciones en que las explotaciones del tipo A están distribuidas uniformemente entre todas las explotaciones. En este caso es apropiado hacer los segmentos que han de examinarse para encontrar las explotaciones del tipo A más grandes que los segmentos correspondientes a una muestra de todas las explotaciones. Esto sugiere la posibilidad de usar un segmento pequeño y uno grande, siendo el segmento pequeño también parte del segmento grande. Por ejemplo, suponga que se va a muestrear las explotaciones del tipo A con una fracción de muestreo que es cuatro veces el tamaño de la fracción de muestreo para todas las explotaciones. El primer paso es diseñar y seleccionar una muestra de segmentos grandes que se examinarán para encontrar las explotaciones del tipo A. Entonces se divide cada segmento grande en cuatro segmentos y se selecciona aleatoriamente uno de estos cuatro segmentos. El dibujo siguiente ilustra un par de segmentos grande y pequeño.



La muestra de segmentos pequeños da una muestra de todas las explotaciones y la muestra de segmentos grandes, que incluye los segmentos pequeños, es la muestra para las explotaciones del tipo A. Probablemente se instruiría al enumerador que enumerara primero el segmento pequeño y tratarlo como si no existiera el segmento grande. Entonces examinaría el resto del segmento grande en busca de explotaciones únicamente del tipo A.

Un ejemplo específico de un uso posible de un par de segmentos grande y pequeño es una encuesta de los costos, cantidades y clases de materiales correspondientes a la construcción de nuevos edificios y a la reparación y remodelación de edificios rurales viejos. En una gran proporción de todas las explotaciones se hacen cada año reparaciones, pero en cualquier año dado son pocos los edificios nuevos que se construyen. La nueva construcción es importante y la variancia de muestreo por explotación es relativamente grande, de manera que se necesita emplear una fracción de muestreo más grande para el caso de las construcciones nuevas que para el caso de las reparaciones y mantenimiento de edificios. Así, pues, si se emplea el método de segmentos pequeños y grandes, se

recolectaría información sobre toda reparación, mantenimiento y construcción nueva en los segmentos pequeños. El resto de cada segmento grande se examinaría en busca de edificios nuevos que se hubieran construido y se recolectarían datos sobre los edificios nuevos.

(2) El segundo modo es el diseño de dos muestras: una muestra de propósito general de todas las explotaciones y una muestra independiente diseñada específicamente para explotaciones del tipo A. En la sección 8.3 se presenta un ejemplo del muestreo de propósito especial. Pero primero interpondremos unas palabras de advertencia.

Aunque conceptualmente no debe haber ninguna diferencia, en la práctica es posible que las explotaciones identificadas como del tipo A en la muestra de segmentos pequeños difieran en promedio de las explotaciones identificadas como del tipo A en los segmentos grandes. Lo mismo se puede decir con respecto a las explotaciones identificadas como del tipo A en una muestra de propósitos generales y las explotaciones identificadas como del tipo A en una muestra suplementaria de propósito especial para explotaciones del tipo A. A menudo ocurren diferencias más grandes que las que pueden esperarse por error de muestreo, cuando se hacen cambios en los procedimientos de la encuesta, aunque los conceptos y las definiciones de los parámetros sean los mismos.

8.3 Una Encuesta sobre Ganado Bovino de Carne

Si más de un tercio o una mitad de los operadores agrícolas produjeran ganado de carne y si ninguno de ellos tuviera gran número de cabezas, un plan sencillo de muestreo por áreas que no utilizara datos auxiliares especializados sobre ganado de carne podría proveer una eficiencia de muestreo satisfactoria. Pero, a medida que la agricultura se vuelve más especializada y aparecen haciendas más grandes, se hace más necesario el tratamiento de cada producto (o grupo de productos) como problema especial de muestreo.

En el condado de Mills hay menos de 900 explotaciones agropecuarias. Los datos censales indican que casi el 40 por ciento de las explotaciones no tienen ganado y menos de 50 explotaciones tienen casi la mitad del ganado. Al considerar el muestreo por áreas y la posibilidad de usar secciones (áreas de una milla cuadrada) como unidades de muestreo, se ve que habría muchas secciones sin ganado de carne y un número muy pequeño de secciones con grandes potreros que pudieran tener más de 1,000 cabezas de ganado. El muestreo por áreas descrito en secciones 8.1 y 8.2 sería ineficiente. Es decir, se requerirían grandes fracciones de muestreo (quizá 0,5 ó más)

para lograr resultados satisfactorios. Una solución consiste en compilar una lista de empresas ganaderas grandes y utilizar el muestreo de marco múltiple, como se mencionó en la sección 2.5.2. Pero esta discusión se limita al muestreo por áreas.

Para tener una base para el muestreo eficiente por áreas en una encuesta sobre ganado, es esencial tener información sobre la ubicación del ganado. A menudo se pueden identificar en fotografías aéreas recientes los grandes potreros o construcciones para alimentación de gran número de ganado. En caso necesario, alguien podría visitar el área y hacer preguntas apropiadas para identificar y localizar por lo menos las empresas ganaderas grandes. ("Grande" en este contexto podría significar que las empresas son seleccionadas con probabilidad proporcional a tamaño.) Si es posible identificar las empresas grandes y medianas con un costo adicional moderado, probablemente valdría la pena hacerlo. De paso, el "tamaño" de una instalación dedicada al engorde de ganado debe medirse en términos de su capacidad en vez del número de ganado que se encuentre en ella en una fecha dada.

Como ilustración sencilla, suponga que se han localizado en los mapas 50 empresas grandes de producción de ganado de carne. Se definirían 50 segmentos, que incluirían las 50 empresas, un segmento correspondiente a cada empresa. El segmento debe ser suficientemente grande para incluir toda la empresa y debe observarse el requisito normal de tener linderos identificables. Estos 50 segmentos se tratarían como una subpoblación o estrato separado y se le aplicaría un plan de muestreo apropiado. Para muestrear el resto de la población, los 50 segmentos se quitarían de las unidades de marco en que se encuentran. Después de esta operación, el diseño y la selección de una muestra por áreas seguirían los principios que han sido mencionados ya. La subpoblación de 50 segmentos se muestrearía, usando una fracción de muestreo más grande que la usada para el resto.

Este procedimiento es aplicable en los métodos de segmento cerrado y segmento ponderado. Al usar el método de segmento abierto, hay que dar atención especial a la definición de "sede" de la explotación agrícola. Si la definición de "sede" resulta en la no inclusión de cualquiera de las 50 empresas en el estrato de 50 segmentos, podría ocurrir una pérdida grave de eficiencia de muestreo.

Se han esbozado brevemente tres ejemplos de muestreo por áreas para una región donde una proporción grande de la tierra fue cultivada y donde la topografía era relativamente favorable al muestreo por áreas. En la próxima ilustración son diferentes la topografía y el uso dado a la tierra.

9. CONSTRUCCION DE MARCO - ILUSTRACION NO. 2

Para la segunda ilustración se escogió una parte del condado de Johnson en el sur del Estado de Illinois. Las figuras 4a y 4b para el condado de Johnson corresponden a las figuras 2a y 2b para el condado de Mills. Se muestra todo el condado con excepción de una faja angosta a la orilla este, que se omitió para no tener que reducir la escala del mapa. A causa de la topografía, las unidades de marco son más grandes y más irregulares que las de la primera ilustración. La selección de detalles topográficos para linderos de unidades de marco es más difícil. Por ejemplo, los linderos de condados son descritos y mostrados en la documentación oficial de terrenos, pero los detalles visibles de la topografía no coinciden siempre con los linderos de los condados. Técnicamente, las unidades de marco podrían traslapar los linderos del condado. En ese caso, si los linderos del condado coinciden con los linderos de la población que ha de muestrearse, cada unidad de marco que traslapara el lindero del condado (lindero de la población) sería identificada antes del muestreo. Entonces la parte de cada una de tales unidades de marco que esté dentro del condado sería marcada y tratada como cualquier otra unidad de marco en la población. El permitir que las unidades de marco traslapen linderos del condado podría suministrar mejores linderos para las unidades de marco. Por otro lado, los mapas, fotografías y estadísticas generalmente se preparan por condados y hay cierta inconveniencia en permitir que las unidades de marco traslapen los linderos de condado. En esta ilustración se permitió que las unidades de marco traslaparan los linderos del condado (figura 4b).

Para ver la topografía y el uso dado a la tierra, vea las figuras 5 y 6. La figura 5 es un mosaico de fotografías aéreas de una parte del condado que incluye las unidades de marco 22, 23, 29 y la mayor parte de la 21. Este mosaico es parte de un índice de fotografías individuales que son identificadas por números en las esquinas derechas superiores; por ejemplo, BGS 1 MM-42. Al examinar el mosaico, no tome equivocadamente la orilla de la fotografía como detalle topográfico. Hay mucho traslape entre las fotografías y las fotos no se emparejan exactamente, a causa de diferencias de escala. La escala del mosaico de la figura 5 es aproximadamente 0,75 pulgada = 1 milla. Cada fotografía cubre un área de aproximadamente 2,25 x 2,25 pulgadas. La figura 6 es una fotografía de una parte de la unidad de marco número 23. La escala es mayor y muestra más detalles. Esta figura 6 se discutirá posteriormente.

Tres clases amplias de uso dado a la tierra pueden reconocerse en las fotografías: bosque, áreas residenciales o con muchas construcciones y el resto que es usado principalmente para la producción agropecuaria y que llamaremos "tierra agrícola". Se puede usar de distintas maneras esta información sobre el uso dado a la tierra. Una manera mencionada anteriormente es deslindar las clases según uso dado a la tierra y después deslindar unidades de marco dentro de cada clase. La topografía del condado de Johnson es tal que la proporción de tierra agrícola, por ejemplo, variaría mucho entre unidades de marco que pertenecen a la misma clase según uso de la tierra. Esto es inevitable, salvo que se afloje el requisito de que las unidades de marco deben tener linderos permanentes e identificables hasta un grado que permitiera linderos muy ambiguos para las unidades de marco. Las unidades de marco en la figura 4b se deslindaron sin tomar en cuenta el uso dado a la tierra. Es decir, la idea en esta ilustración es usar la información sobre uso de la tierra después de haber deslindado las unidades de marco. Se supone que las áreas de las unidades de marco ya han sido estimadas, probablemente con el uso del planímetro y los mapas del marco.

9.1 Una Encuesta sobre Areas Cultivadas

En el condado de Johnson la proporción de tierra agrícola varía del 35 al 75 por ciento entre las unidades de marco (vea la figura 5). Para una encuesta de áreas cultivadas, suponiendo el uso del método de segmento cerrado o de segmento ponderado, el área aproximada de tierra agrícola en cada unidad de marco parece ser una medida de tamaño mucho mejor que el área total. Existen por lo menos dos métodos factibles para aproximar la cantidad de tierra agrícola en las unidades de marco;

(1) Estimar la cantidad (o la proporción) de tierra agrícola en cada unidad de marco mediante el uso del planímetro o con la superposición de una rejilla transparente calibrada encima de las fotografías. Si se estiman las proporciones, las cantidades pueden estimarse multiplicando las proporciones por las áreas aproximadas de las unidades de marco. Este trabajo debe hacerse con cuidado, pero probablemente no valdría la pena gastar mucho tiempo en hacer las mediciones tan exactas como sea posible, en términos del efecto sobre la "variancia de muestreo". Un alto grado de exactitud, comparado con una aproximación cruda, podría hacer muy poca diferencia en el número de segmentos asignados a las unidades de marco. Además, al dividir las unidades de marco, el igualamiento de la cantidad de tierra agrícola en los segmentos es posible únicamente hasta cierto grado limitado, dependiendo de los detalles topográficos disponibles para linderos de los segmentos.

(2) El segundo método es menos exacto y consume menos tiempo. Por inspección visual de las fotografías se clasifican las unidades de marco como altas, medianas o bajas, con referencia a la proporción de terreno que es tierra agrícola. Por ejemplo, el objetivo podría ser clasificar visualmente como altas las unidades de marco con más del 60 por ciento de tierra agrícola, como medianas las con el 40 al 60 por ciento de tierra agrícola y como bajas las que contienen menos del 40 por ciento de tierra agrícola.

El censo de agricultura indica que aproximadamente la mitad del área total del condado de Johnson consiste en explotaciones agrícolas. Del terreno de las explotaciones agrícolas, una cuarta parte es tierra cultivada, casi la cuarta parte es bosque y una proporción importante sirve de pastoreo. El tamaño medio de las explotaciones es aproximadamente de 200 acres y hay aproximadamente 1,8 explotaciones agrícolas por milla cuadrada. La cantidad de tierra juzgada como agrícola por inspección visual de fotografías (es decir, terreno no cubierto de árboles y que no parece ser usado para fines residenciales o industriales) puede ser muy diferente de la cantidad de tierra agrícola establecida por el censo. No obstante, para una encuesta sobre cultivos, usando el método de segmento cerrado o de segmento ponderado, la cantidad de tierra agrícola deducida de las fotografías es una medida útil y factible de tamaño para la asignación de segmentos a las unidades de marco.

La topografía del condado de Johnson es tal que el tamaño medio de los segmentos no debe ser menor de 500 ó 600 acres de tierra agrícola. Por lo tanto, se especifica para esta ilustración un tamaño medio de segmento de una milla cuadrada (640 acres). Como promedio, un segmento tendrá como 160 acres ($\frac{1}{4}$ de milla cuadrada) de tierra que produce cosechas. Si una estimación de la cantidad de tierra agrícola, expresada en millas cuadradas, está disponible para cada unidad de marco, el número de segmentos asignados a las unidades de marco sería el número de millas cuadradas de tierra agrícola, redondeado al número entero más cercano. Cada unidad de marco debe tener por lo menos un segmento, con la excepción de cualquier unidad de marco que haya sido omitida intencionalmente de la población que se va a muestrear.

Suponga que cada unidad de marco ha sido clasificada como alta, mediana o baja con respecto a la proporción de terreno total que es tierra agrícola. Suponga que las proporciones medias de tierra agrícola correspondientes a estas tres clases son 0,7, 0,5 y 0,3. Las áreas de terreno en millas cuadradas en las unidades de marco serían multiplicadas respectivamente por 0,7, 0,5 y 0,3 para determinar los números de segmentos asignados. Es

posible hacer con más exactitud la asignación de segmentos. Hay que decidir si un método más exacto valdría la pena. Fíjese que la clasificación de unidades de marco según el uso dado a la tierra se discutió como un mecanismo para asignar segmentos y no como criterio para estratificación en el sentido de muestreo aleatorio estratificado. Las unidades de marco pueden estratificarse de cualquier modo que sea apropiado para la encuesta.

Se ha seleccionado la unidad de marco número 23 como ilustración. Por el uso del planímetro sobre el mapa del marco, que está a escala conocida, se estimó un área de 8,6 millas cuadradas en dicha unidad. De fotografías de escala mayor que la de la figura 5 se estimó, con la ayuda de una rejilla transparente, que aproximadamente el 60 por ciento del terreno en la unidad de marco 23 era tierra agrícola. Esto da 5,2 millas cuadradas ($8,6 \times 0,6$) como estimación de la cantidad de tierra agrícola. Así, según las especificaciones para los segmentos, que se trataron anteriormente, la unidad de marco 23 contendrá 5 segmentos. Suponga que ha sido asignado un número de segmentos a cada unidad de marco de modo similar. Suponga además que, para una encuesta sobre áreas cultivadas, ha sido seleccionada la unidad de marco 23 y hay que dividirla en 5 segmentos.

Un estudio de fotografías con detalle comparable al de la figura 7 indicó que la unidad de marco 23 no se podía dividir fácilmente en 5 segmentos con casi la misma cantidad de tierra agrícola en cada uno. La situación presenta el problema típico de ajuste entre claridad de linderos de segmentos e igualamiento de tamaños de segmentos. Sin embargo, la unidad de marco 23 se podría dividir en cuatro partes bien definidas por los caminos mostrados en la figura 4b. Las cuatro partes se muestran en la figura 6.

Las fotografías indican que la parte número 1 tiene más tierra agrícola y puede dividirse satisfactoriamente en dos partes. De modo que se presentan dos alternativas: (1) aceptar las partes 2, 3 y 4 como segmentos y dividir la parte 1 en dos segmentos, dando un total de 5, 6 (2) permitir ambigüedad de los linderos para igualar la cantidad de tierra agrícola. La primera alternativa no aprovecha al máximo la información sobre uso dado a la tierra. La segunda alternativa reduce la variancia de muestreo, pero aumenta la posibilidad de sesgo. En estas circunstancias, el autor prefiere la primera alternativa, salvo que ensayos en las condiciones reales de operación demuestren que la segunda alternativa es operacionalmente factible y que se puede evitar el sesgo.

La figura 7 muestra la parte 1 de la unidad de marco 23 dividida en dos segmentos. Se usó como lindero una quebrada pequeña pero bien definida. Observe la pequeña población y cómo los linderos de segmento siguen caminos o calles hasta el centro de la población. Con los métodos de segmento cerrado

o de segmento ponderado, la existencia de un área residencial dentro de un segmento no debe presentar, en la mayoría de los casos, dificultades para el enumerador. Desde el punto de vista de muestreo, la parte importante de su tarea es deslindar exactamente los tracts dentro del segmento. Al usar el método de segmento abierto, las áreas residenciales presentan problemas especiales.

9.2 Una Encuesta sobre Todas las Explotaciones Agrícolas

Para una encuesta de todas las explotaciones agrícolas, usando el método de segmento ponderado, probablemente se definirían y se deslindarían los segmentos tal como se explicó en la sección anterior. Como se ha dicho anteriormente, el método de segmento abierto ha sido usado muchas veces y se han ensayado y estudiado muchas maneras alternativas de aplicarlo. Ninguna manera de aplicar el método de segmento abierto puede recomendarse como generalmente superior.

Con respecto a la aplicación del método de segmento abierto para obtener una muestra de todas las explotaciones en el condado de Johnson, no hay puntos nuevos que tratar. Para repetir, el objetivo general es (1) asignar números de segmentos a las unidades de marco en proporción a los números de explotaciones con sede dentro de las unidades y (2) dividir las unidades de marco en segmentos de tal manera que haya número igual de explotaciones con sede en cada segmento. Los medios limitados para lograr este objetivo dejan mucho que desear. Pero el problema del error de cobertura es más grave, debido a la falta de una definición conceptualmente válida y práctica de "sede" de una explotación agrícola. Recuerde que "sede" es el nombre de un punto único que determina si la explotación agrícola está dentro de la muestra. Un marco de muestreo que se construye solamente para la aplicación de los métodos de segmento cerrado o de segmento ponderado representa una labor más sencilla, porque no involucra consideraciones de la definición de "sede" y su localización. Se vuelve urgente la necesidad de explorar ampliamente el método de segmento ponderado como alternativa al método de segmento abierto.

10. RESUMEN Y BREVE MIRADA AL FUTURO

Los marcos de muestreo deben construirse con reconocimiento del hecho de que la agricultura se compone de varias subpoblaciones que tienen que muestrearse. Una muestra diseñada eficientemente para una subpoblación puede ser de poco valor para otra. Así es que pueden ser necesarios varios marcos de muestreo; o, si ha de construirse un solo marco de muestreo, probablemente debe ser de propósitos múltiples.

En general, a medida que las empresas agrícolas se vuelven más especializadas y más grandes, más necesario es el desarrollo de marcos de muestreo más flexibles para la selección de muestras para muchos propósitos. Por ejemplo, hace 30 años en algunas regiones de los Estados Unidos la misma muestra podría haber sido útil tanto para cultivos como para ganado. Pero este ya no es el caso. Para muestrear eficientemente un ítem como el ganado bovino de carne, es necesario (1) tener una lista de productores de ganado adecuada para fines de muestreo, (2) usar el muestreo de marco múltiple incluyendo el muestreo por áreas y una lista de, por lo menos, los productores más grandes, o (3) desarrollar el muestreo por áreas sobre una base eficiente para propósitos especiales, como en la sección 8.3. El desarrollo de marcos de muestreo mejorados es requerido por (1) la tendencia hacia explotaciones agrícolas más grandes y más especializadas, (2) la demanda general para estadísticas más exactas y (3) la necesidad de mantener tan bajos como sea posible los tamaños de muestras y los costos. También, hasta cierto grado, el tamaño de la muestra se relaciona inversamente con la capacidad de controlar el error no muestral, lo que es otro punto a favor del muestreo eficiente para mantener el tamaño de la muestra tan pequeño como sea posible. El problema de la carga que se impone sobre los entrevistados que contestan las preguntas en una encuesta también apoya el uso de muestras más pequeñas y más eficientes. Estos factores parece estar requiriendo la dedicación de más recursos a la construcción y mantenimiento de marcos de muestreo que den más eficiencia en el diseño de muestras

Hay muchas fuentes de error y maneras de reducir el error. Los planes para una encuesta deben incluir provisiones para el estudio de la variancia de muestreo, la variancia de respuesta, el error de cobertura y los costos. Tales estudios deben proveer una base permanente para ajustar la distribución de recursos en un esfuerzo por lograr el máximo de exactitud para un costo dado.

No es probable que el muestreo por áreas reemplace el muestreo de listas de operadores de explotaciones agrícolas, ni viceversa. Uno de los

problemas más importantes en encuestas sobre empresas agrícolas estriba en el enlace obscuro entre operadores y explotaciones agrícolas, lo que resulta en errores de cobertura. Los problemas de enlace son comunes cuando se muestra de listas y en el muestreo por áreas, especialmente cuando se usa el método de segmento abierto. El uso del método de segmento cerrado evita la mayor parte del error de cobertura causado por el enlace obscuro entre operadores y explotaciones. Este es un punto de mayor importancia a favor del uso del método de segmento cerrado. En las encuestas en las que los tracts son adecuados como unidades de información, el método de segmento cerrado probablemente seguirá usándose como método efectivo.

En encuestas en las que las explotaciones agrícolas son las unidades de información, el autor cree que debe explorarse ampliamente el segmento ponderado como una alternativa al segmento abierto. La variancia de muestreo, por segmento, es menor en el caso del segmento ponderado que en el caso del segmento abierto. Tenemos que saber más acerca de costos comparados y el error de cobertura para poder entender mejor las circunstancias en que un método podría ser mejor que el otro.

En una situación donde es aplicable el segmento cerrado a solamente una parte de las preguntas, podría usarse el segmento cerrado en combinación con el abierto o el ponderado para aprovechar al máximo el segmento cerrado. En el juicio del autor, la experiencia demostrará que es mejor la combinación de cerrado con ponderado. Si resulta que la experiencia de muestreo que el segmento ponderado tiene error pequeño de cobertura, puede ser necesario decidir entre usar la combinación de cerrado y ponderado, o usar solamente el ponderado, porque este último tiene la ventaja de usar una sola definición de "segmento" en la misma encuesta. De paso, con el equipo moderno de computación, la ponderación de datos ya no debe considerarse un obstáculo mayor al uso del segmento ponderado.

En años recientes, muchas personas se han interesado en la detección remota, incluyendo el efecto que podría tener sobre el muestreo por áreas y sobre los procedimientos usados en general para estimaciones agropecuarias. Este es un tema mayor que involucra mucha conjetura. Sin embargo, tal vez vale la pena que el autor exprese algunas de sus opiniones.

Un efecto inmediato de la detección remota será un aumento en la demanda de datos "terrestres" de muestras por áreas que puedan correlacionarse con los datos obtenidos por detección remota. Esa demanda ya empezó. En el futuro, a medida que se desarrolla la tecnología de la detección remota, probablemente se dispondrá de que pueda ser utilizada para mejorar

substancialmente los marcos de muestreo por áreas y la eficiencia de muestreo por áreas. Esto podría conducir a importantes reducciones del tamaño de muestras por áreas para ciertos propósitos, especialmente para características estrechamente ligadas con el uso dado a la tierra y el ambiente físico.

Una proporción grande de las estadísticas agropecuarias involucran cantidades o actividades que no se prestan a medición por detección remota. Pero considere las áreas cultivadas y los rendimientos. ¿Será posible que la tecnología de detección remota elimine por completo la necesidad de recolectar datos sobre áreas y rendimientos mediante los métodos actuales?

El desarrollo de modelos para estimar o pronosticar rendimientos de cultivos mediante datos de detección remota requiere datos exactos sobre rendimientos obtenidos de una fuente independiente; es decir, de mediciones hechas en el terreno. Suponiendo que se desarrollan modelos operacionales prácticos, se espera que habrá que mejorar continuamente los modelos y esto requerirá, en cierto grado, la continua recolección de datos sobre rendimientos por los métodos actualmente usados. Además, los cambios en el rendimiento asociados con avances tecnológicos cambiarán los parámetros de los modelos y harán necesarios esfuerzos continuos para actualizar los modelos. Esto significa observaciones terrestres en una muestra de campos cultivados que representen la gama de condiciones involucradas.

Algo similar ocurre en la estimación de áreas cultivadas. Se necesitan modelos para la interpretación de datos obtenidos por detección remota. Probablemente los modelos, o los parámetros de los modelos, estarán sujetos a cambios, que en el mejor de los casos requerirá algo de muestreo por áreas en el terreno, concurrentemente con la recolección de datos por detección remota. Además, para servir los propósitos analíticos de algunas encuestas agrícolas, es necesario obtener datos sobre áreas cultivadas y rendimientos por explotación agrícola. La única fuente de tal información consiste en los operadores mismos.

Es posible prever el valor potencial de la detección remota en el mejoramiento de marcos de muestreo por áreas, que resulta en una selección entre estimaciones más exactas o muestras más pequeñas para lograr los niveles actuales de exactitud. Esto, por supuesto, se refiere únicamente a los datos agropecuarios que estén correlacionados por lo menos moderadamente con la información obtenida de la detección remota. Correlaciones menores que 0,6 ó 0,7 generalmente no son suficientemente altas para considerarse seriamente. Un valor potencial también previsible es que podría servir como base para mejorar ciertas clases de estadísticas para áreas pequeñas, como condados o partes de condados.

Suponiendo que se vuelve operacional un sistema adecuado de coordenadas para representar los linderos de unidades de marco o de segmentos en una cinta magnética, se tendría a la mano una gran cantidad de datos derivados de la detección remota para las unidades de marco. Por esto, hay una posibilidad de que en el futuro se mantengan marcos de muestreo por áreas en una cinta magnética. Para algunos propósitos, tal marco de muestreo podría ser muy eficiente con respecto a la variancia de muestreo. Se podría programar la computadora para que produjera muestras bien diseñadas para propósitos específicos.

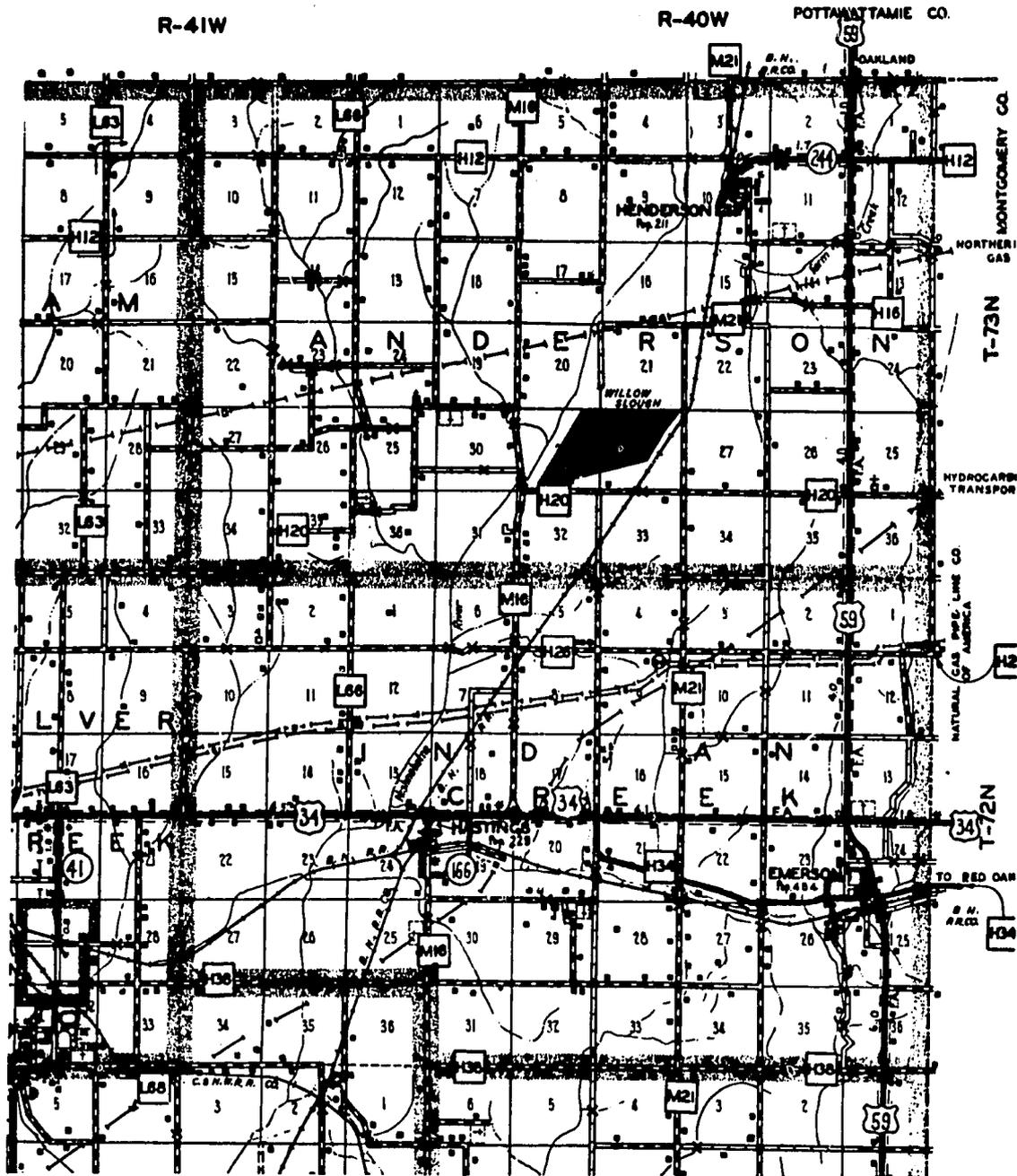


Figura 2a. Mapa de Caminos del Condado de Mills (parte nordeste del condado)
Escala: $\frac{1}{2}$ pulgada = 1 milla

Leyenda (incompleta)

<u>Caminos</u>		<u>Otros</u>	
No mejorados	=====	Casa ocupada	■
Nivelados y drenados	===== ===== =====	Casa desocupada	□
De tierra	-----	Negocio	+
Asfaltados	===== ===== =====	Grupo residencial	■
Pavimentados	=====	Lindero de sección	—

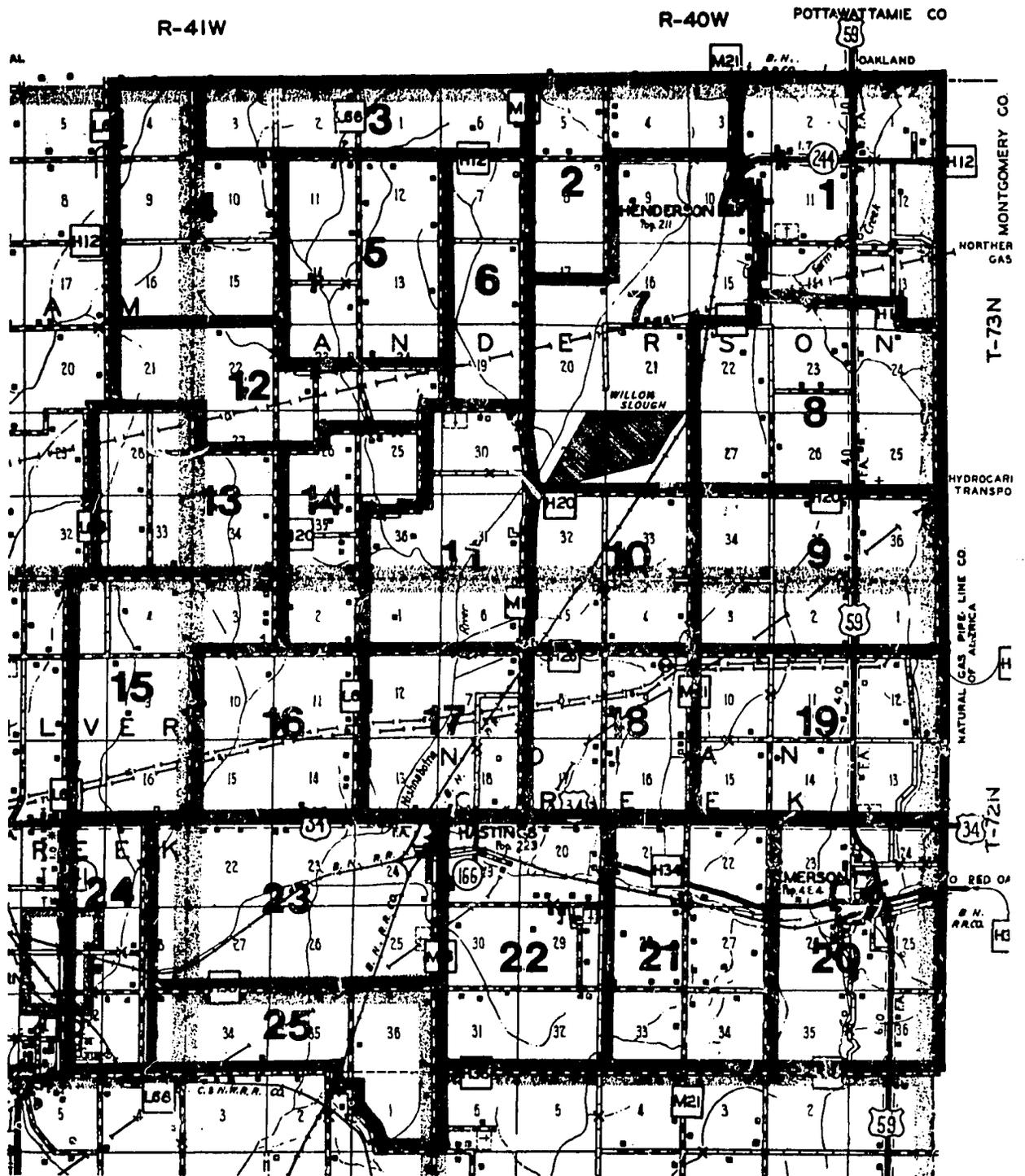


Figura 2b. Unidades de Marco para Parte del Condado de Mills

Legenda

- Lindero de unidad de marco
- 12** Número de unidad de marco

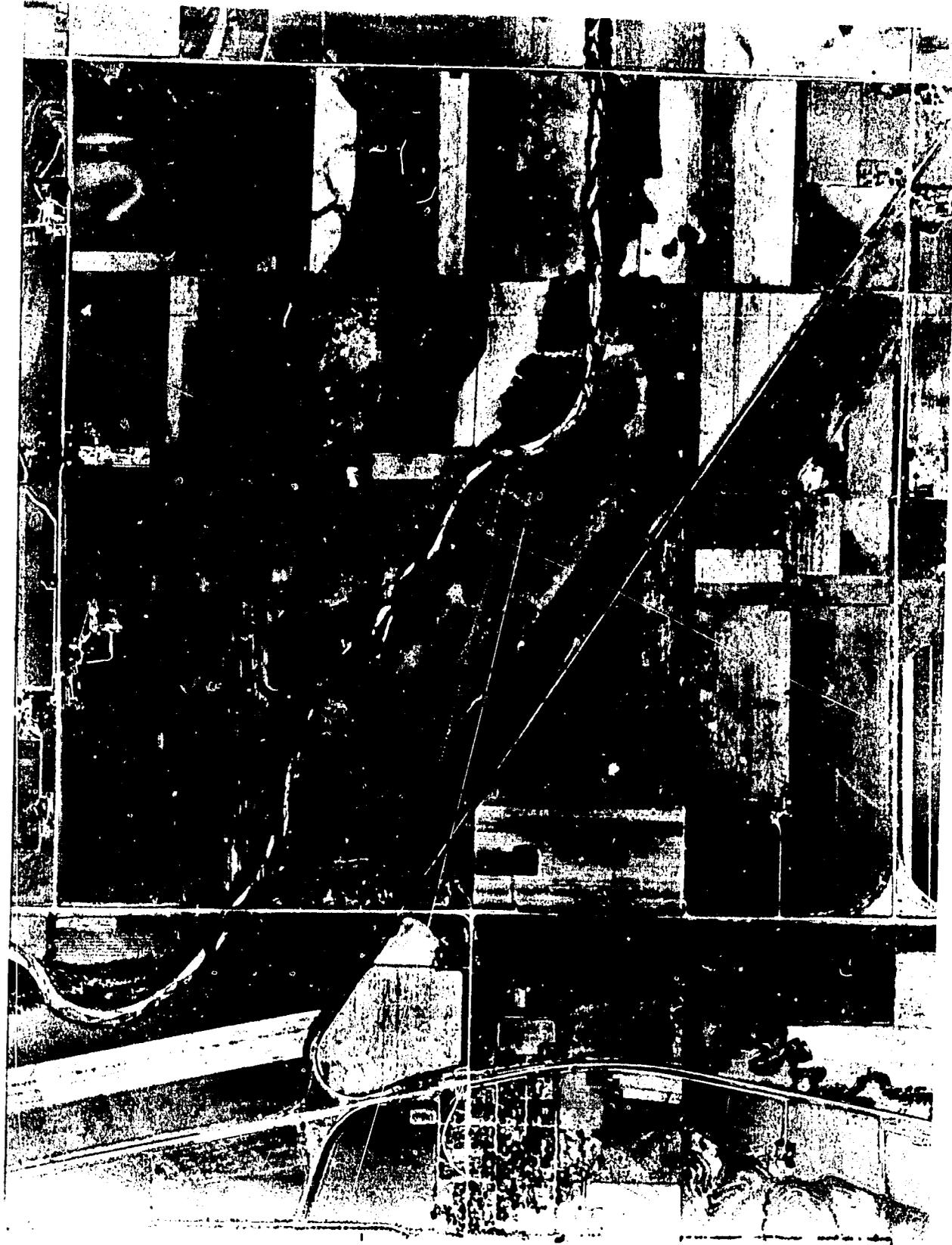


Figura 3a. Fotografía de la Unidad de Marco No. 17

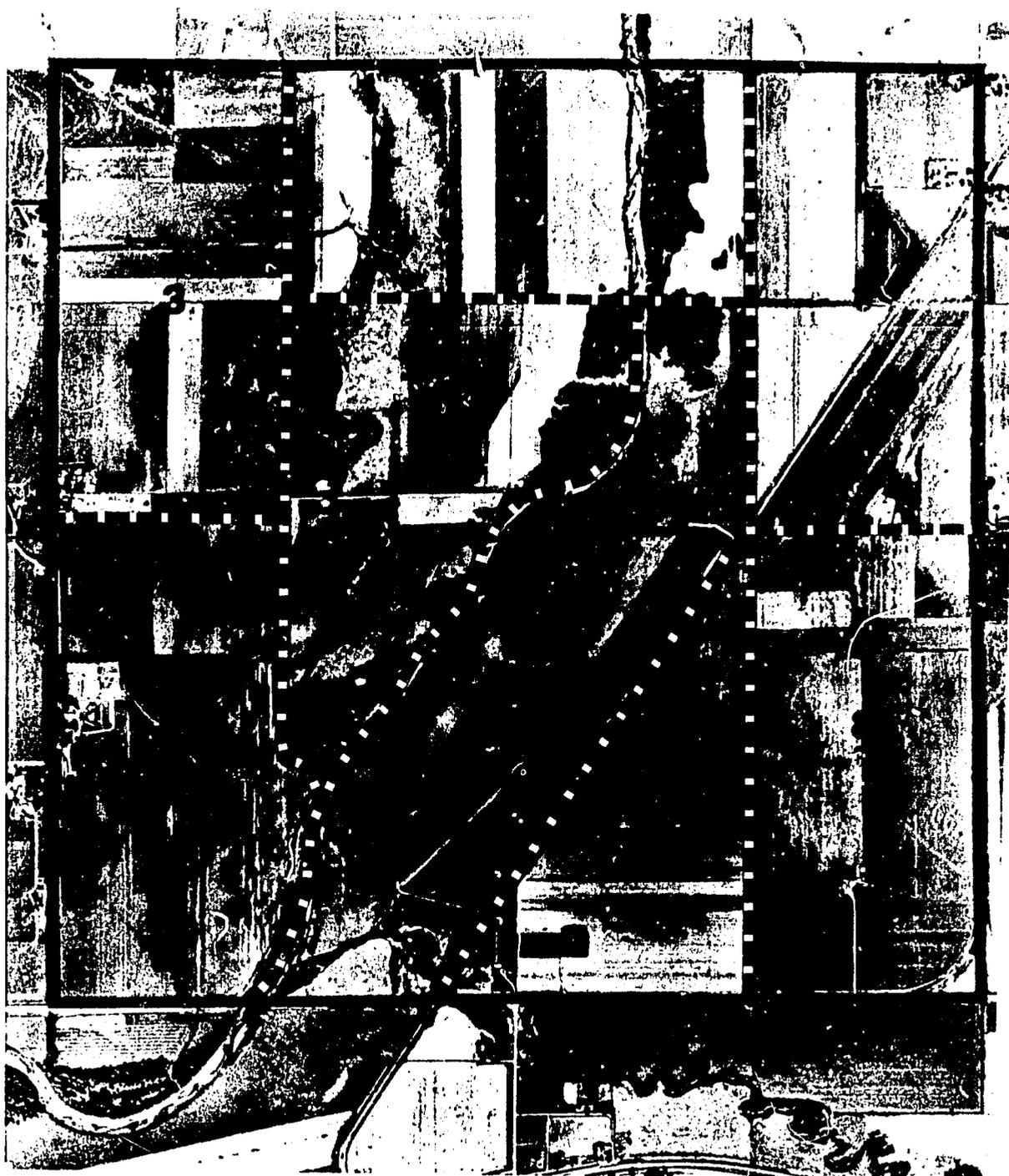


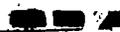
Figura 3b. Unidad de Marco No. 17 Dividida en Ocho Segmentos Cerrados

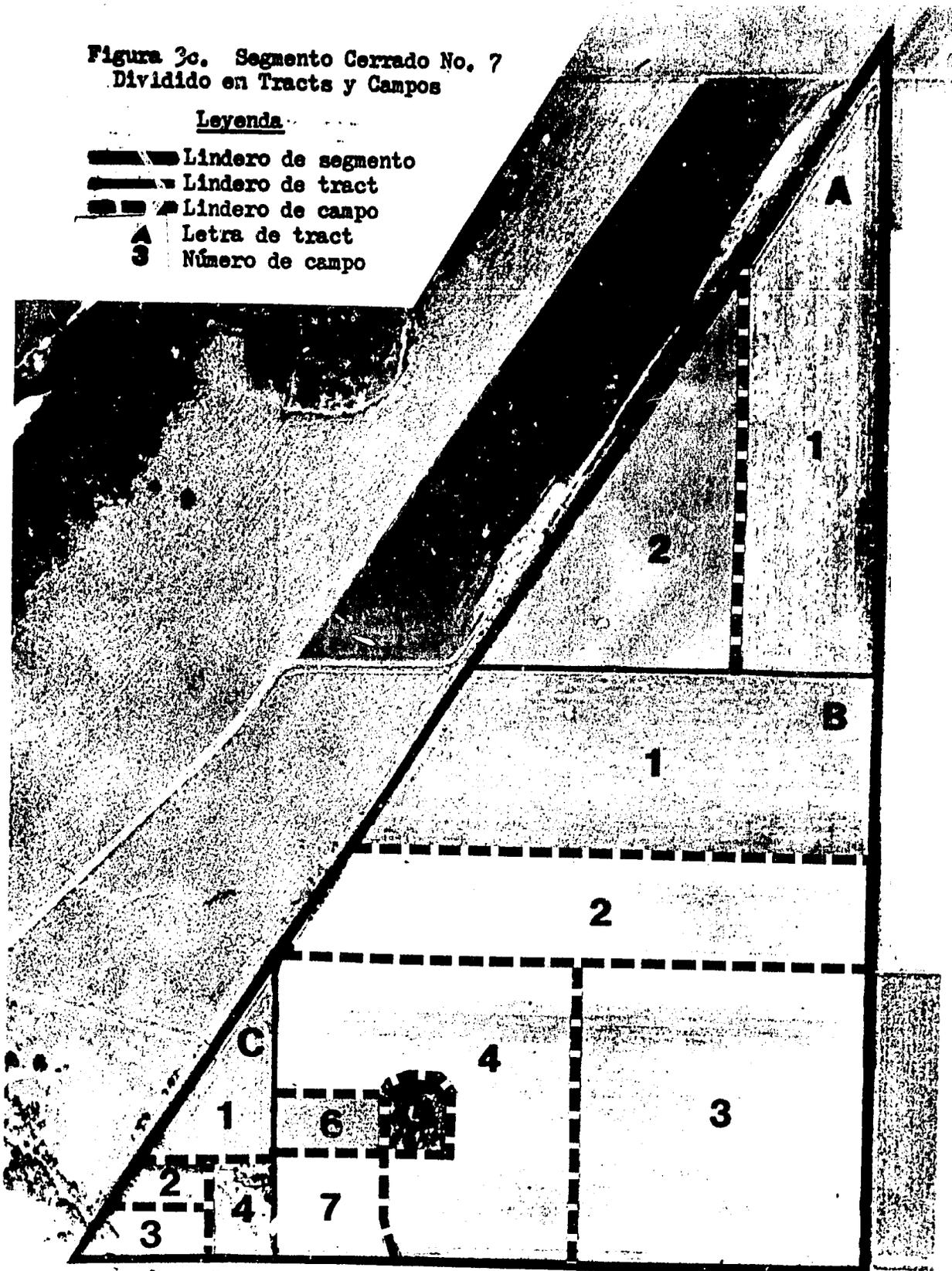
Leyenda

- Lindero de unidad de marco
- ■ ■ Lindero de segmento
- 3 Número de segmento

Figura 3c. Segmento Cerrado No. 7
Dividido en Tracts y Campos

Leyenda

-  Lindero de segmento
-  Lindero de tract
-  Lindero de campo
-  Letra de tract
-  Número de campo



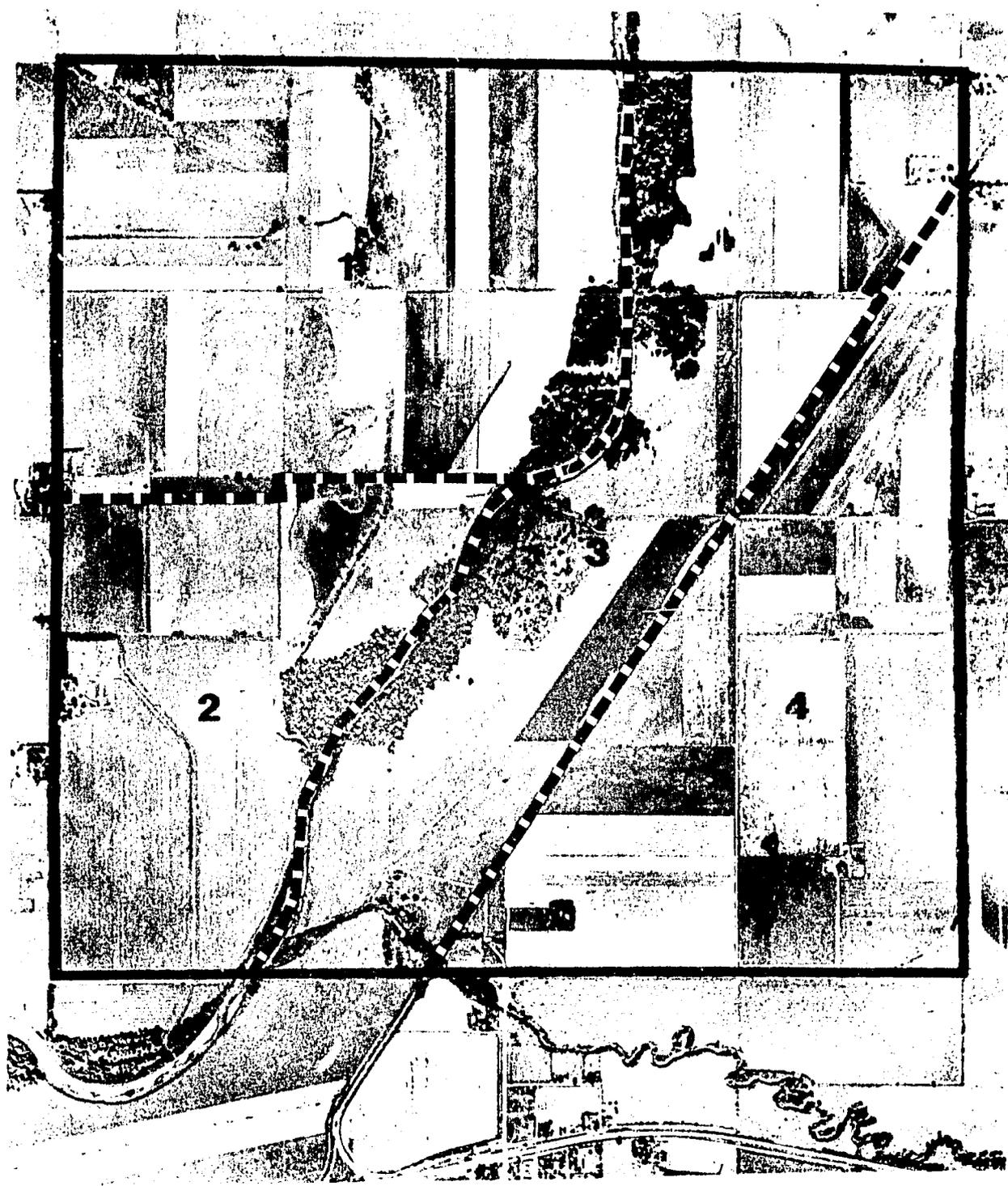


Figura 3d. Unidad de Marco No. 17 Dividida en Cuatro Segmentos Abiertos

Leyenda

- Lindero de unidad de marco
- - - - - Lindero de segmento
- 2 Número de segmento

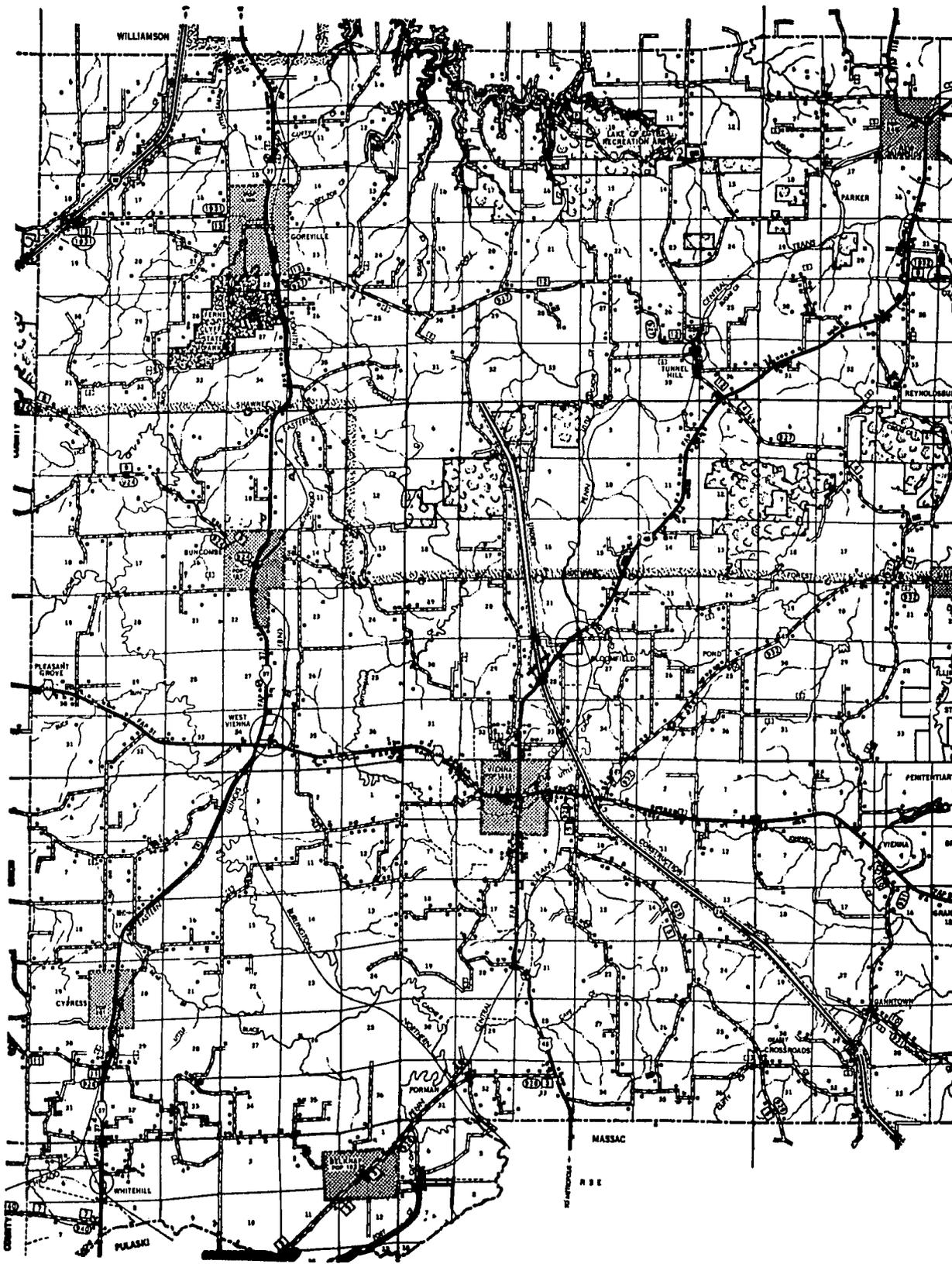


Figura 4a. Mapa de Caminos del Condado de Johnson

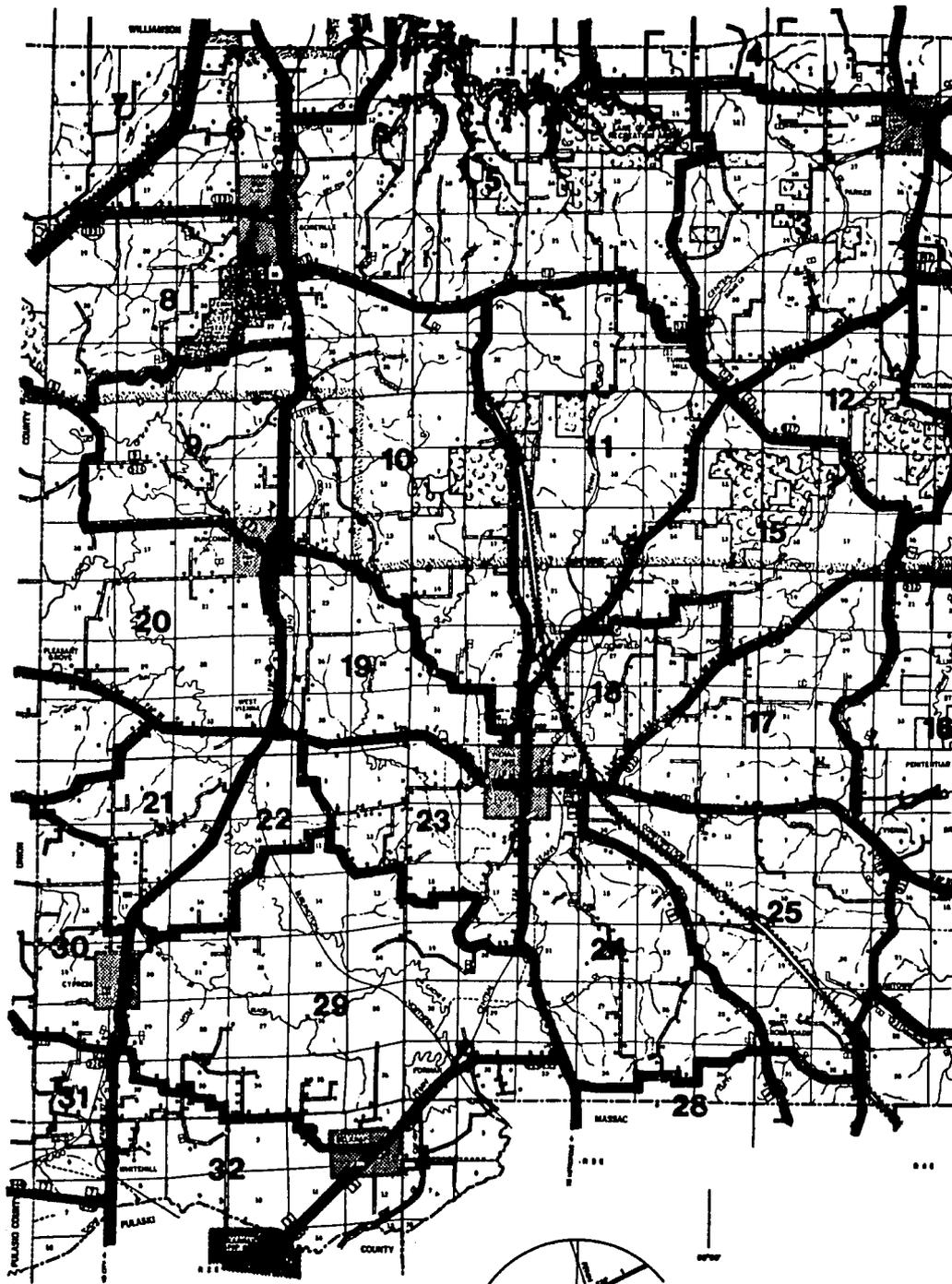


Figura 4b. Unidades de Marco para el Condado de Johnson

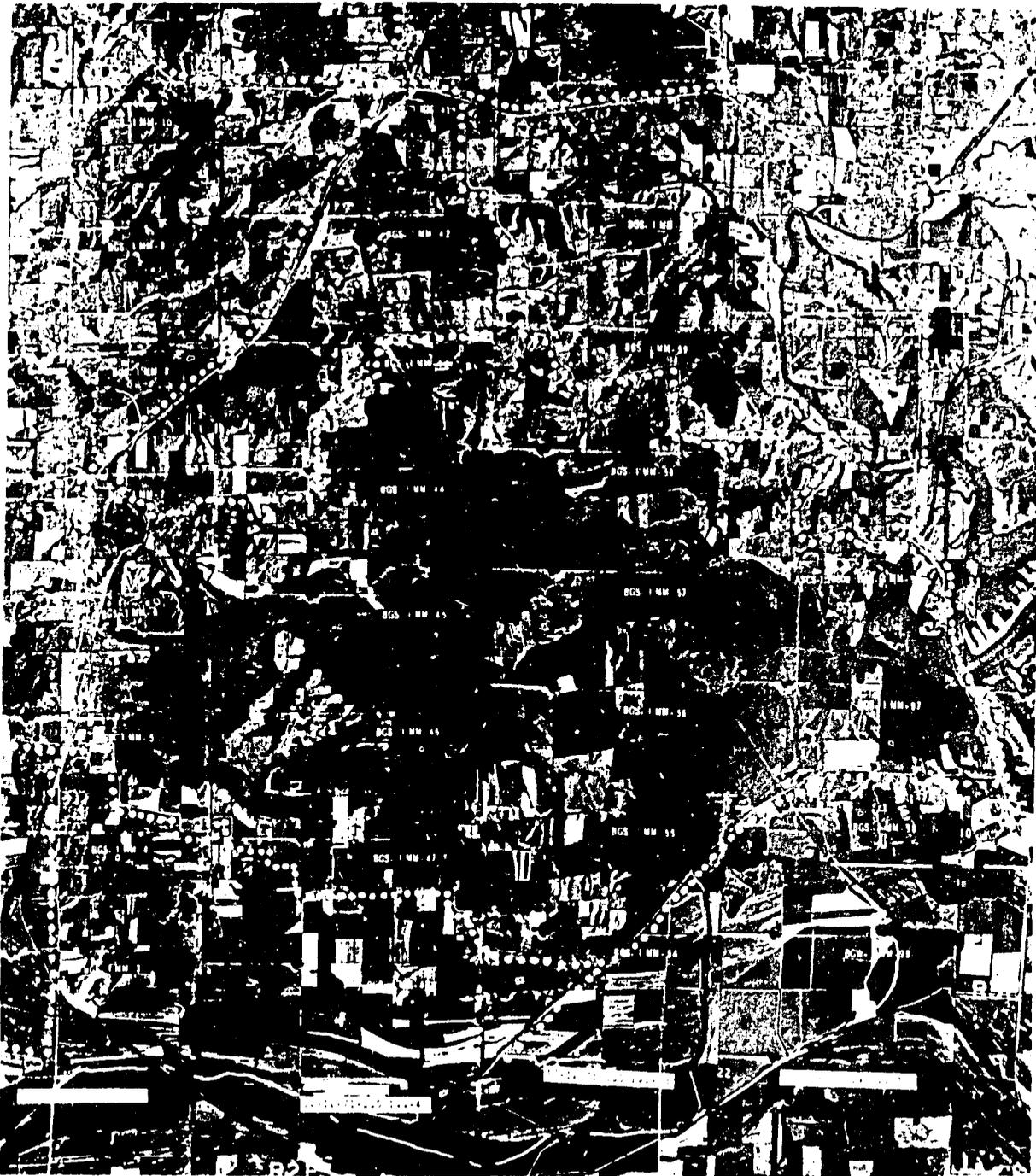


Figura 5. Fotoíndice para una Parte del Condado de Johnson

Leyenda

- | | |
|---|----------------------------|
|  | Lindero de unidad de marco |
| 22 | Número de unidad de marco |
| BGS-1-MM-57 | Número de fotografía |

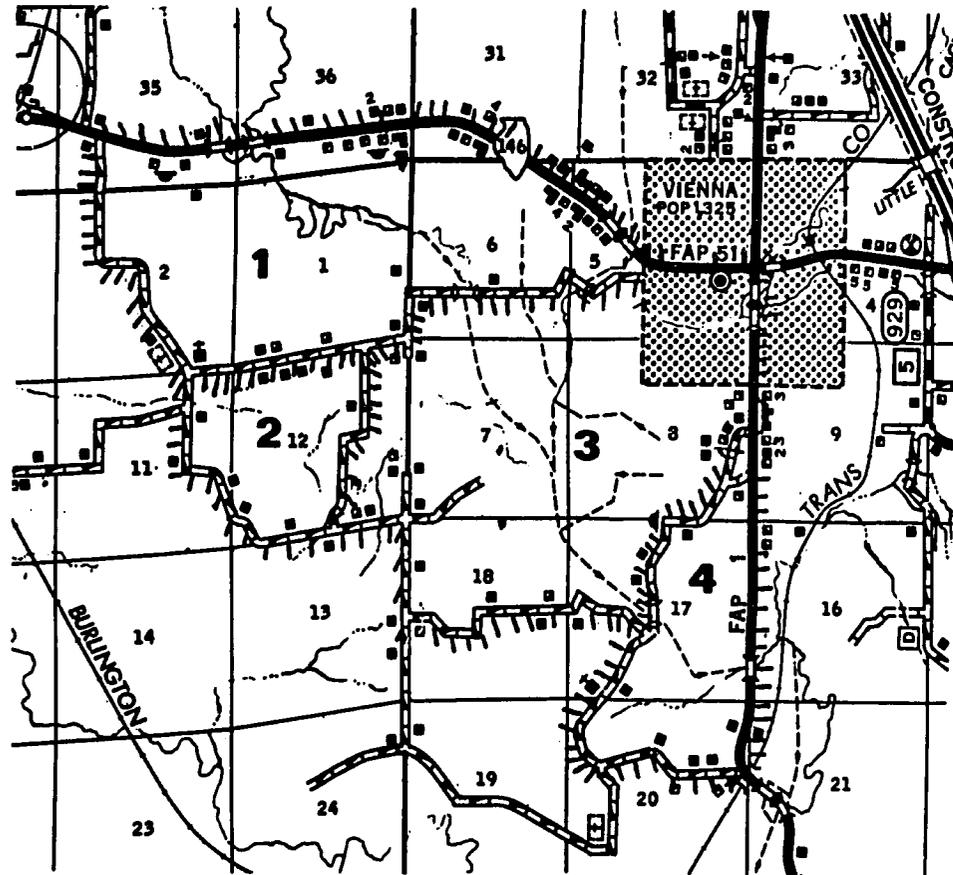


Figura 6. Unidad de Marco No. 23 Dividida en Cuatro Partes

Leyenda

 Lindero de parte

3 Número de parte

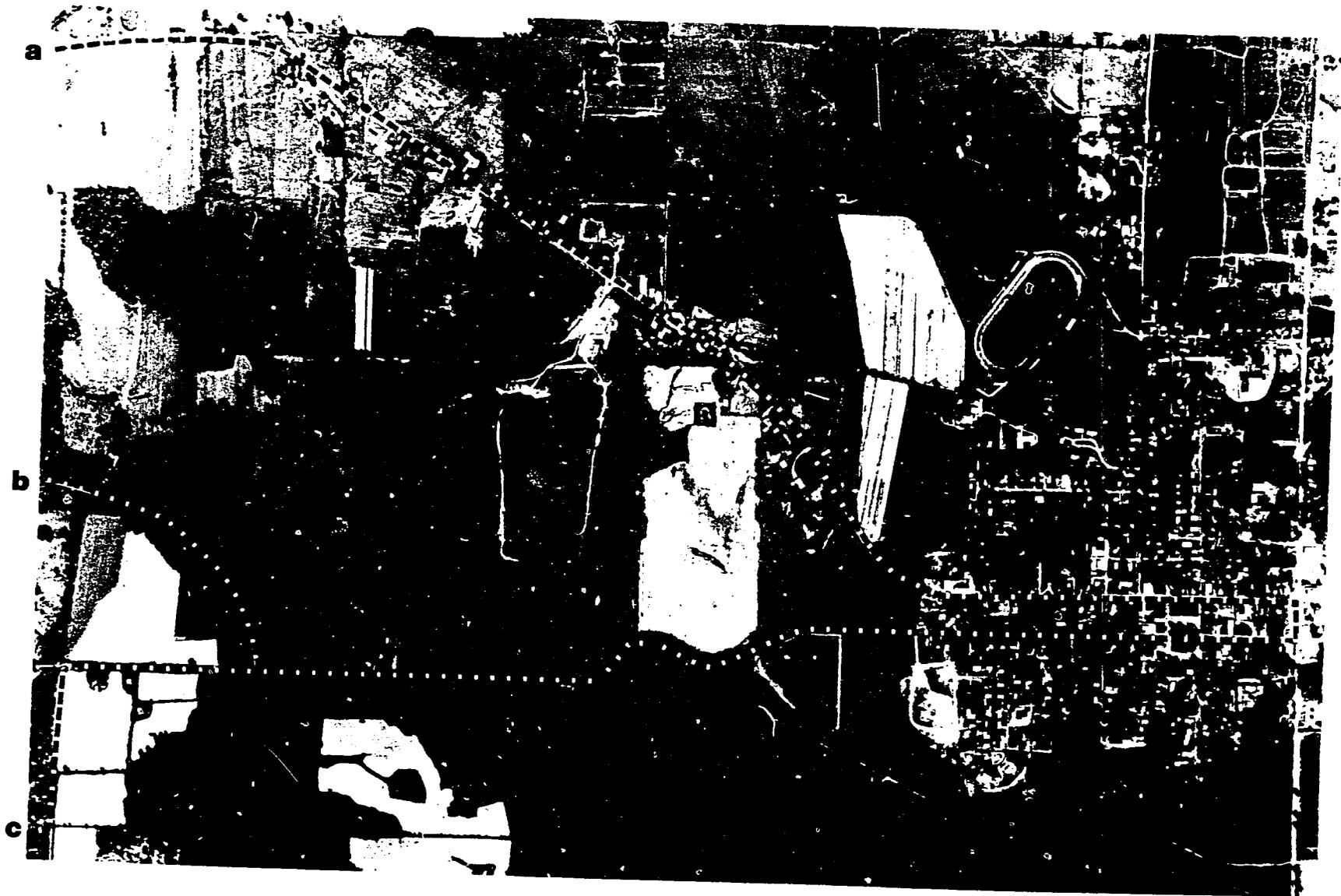


Figura 7. Parte 1 de la Unidad de Marco 23 Dividida en Dos Segmentos
----- Lindero de segmento

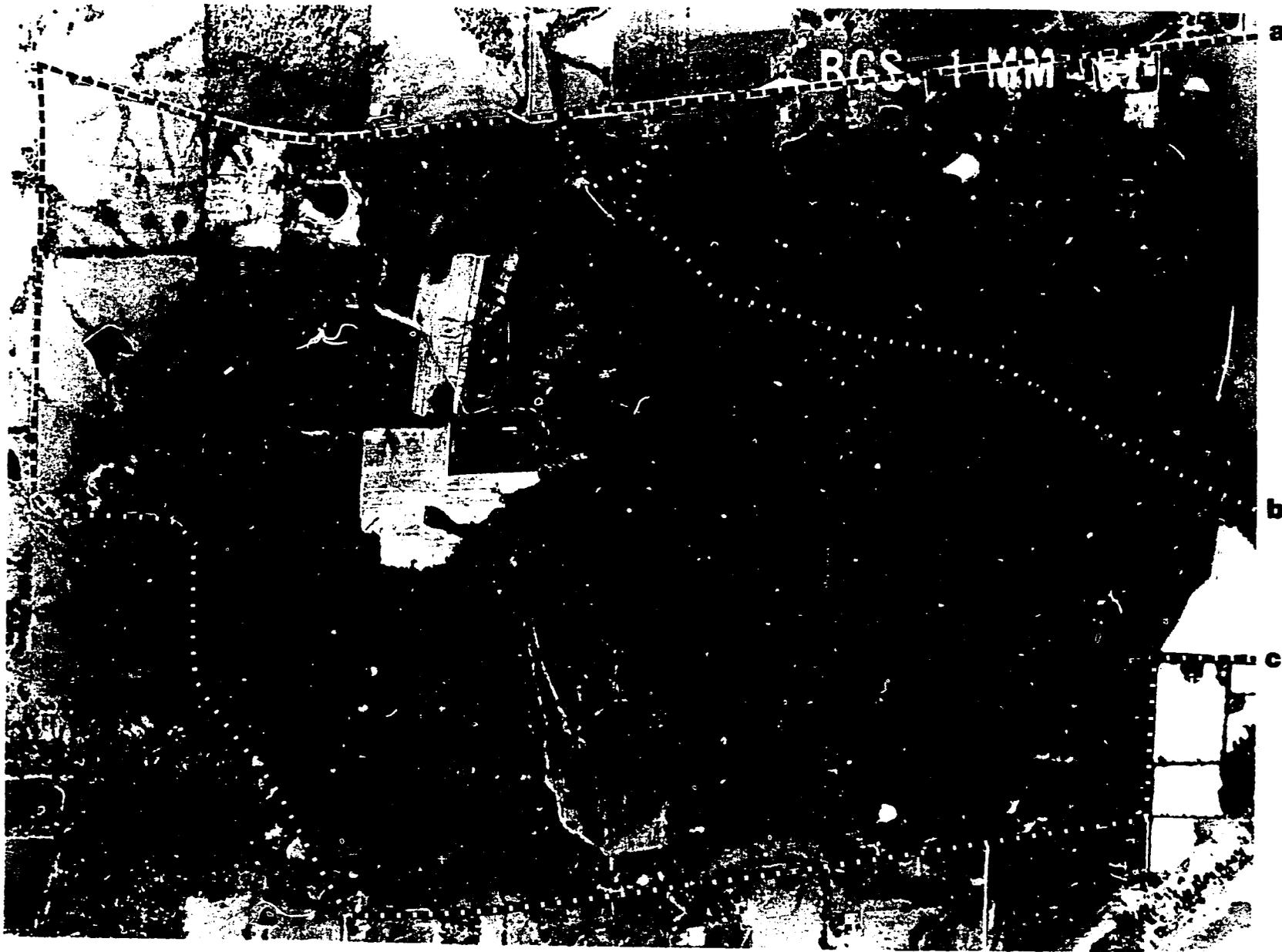


Figura 7. (continuación)