

# SÍNTESIS METODOLÓGICA

## Introducción

Censar integralmente a todas las UPAs importantes y aplicar una muestra de áreas extensa para las demás, implicó el uso de un Marco Muestral de Lista (MML) para las primeras y un Marco Muestral de Áreas (MMA) para las segundas. El uso combinado de los dos marcos significó el empleo del método de investigación estadístico conocido con el nombre de Múltiples Marcos de Muestreo (MMM). La metodología de MMM es completamente objetiva y eficiente para alcanzar alta precisión de los datos o resultados minimizando el costo de investigación; permite efectuar estimaciones puntuales y construir intervalos de confianza alrededor del parámetro verdadero de la población. Ambos marcos fueron tratados de manera totalmente independiente, tanto en su elaboración como en la recolección de los datos; sin embargo, se complementaron a la hora de estimar resultados.

## La muestra de áreas (marco muestral de áreas -MMA-)

El muestreo de áreas consiste en la fragmentación o segmentación sin duplicación ni omisión del ámbito de interés o área total a ser investigada en “N” pedazos pequeños, denominados Segmentos de Muestreo (SMs), los cuales constituyen la población o universo. El conjunto total de SMs forman también el MMA, del cual selecciona una muestra probabilística y aleatoria de “n” SMs; y, con base en la investigación de los “n” SMs, se estima estadísticamente las cifras correspondientes a todos los “N” SMs. Para el la muestra de áreas del III Censo Nacional Agropecuario, el ámbito de interés constituyó todo el territorio nacional, excepto la Provincia de Galápagos.

## Diseño y selección de la muestra

Se diseñó una muestra de SMs estratificada; considerando al cantón como dominio mínimo de estudio; con afijación óptima; y, selección en dos etapas: en la primera, aleatoria y sistemática, con probabilidades proporcionales al tamaño de las Unidades Primarias de Muestreo [1], y en la segunda, selección aleatoria, con probabilidades iguales, de los SMs en que se dividieron las unidades de la primera etapa. Para discriminar territorios con características similares, en procura de una muestra más eficiente, en cada cantón de las provincias del país, excepto en los de Napo, Sucumbíos y Orellana, fueron conformados los siguientes estratos:

Tierras con predominio de pastos:	Estrato 10
Tierras con predominio de cultivos transitorios.	Estrato 20
Tierras con predominio de cultivos permanentes:	Estrato 30
Tierras con predominio de vegetación natural.	Estrato 40
Tierras con mínimo uso agropecuario:	Estrato 51
Tierras cubiertas por agua.	Estrato 52
Tierras bajo manglares sin uso agropecuario:	Estrato 53
Tierras de cumbres, nevados y selvas, sin uso agropecuario.	Estrato 54
Tierras con camaroneras:	Estrato 60
Centros poblados urbanos.	Estrato 71
Centros poblados rurales:	Estrato 72

[1] Las Unidades Primarias de Muestreo (UPM), son superficies que se aproximan a los 10 Km<sup>2</sup> en la mayoría de estratos. A las UPM se las dividió en SMs con áreas cercanas a los 2 Km<sup>2</sup>, y por cada UPM seleccionada se eligió un SM, en cuyo interior fueron censadas todas las Unidades de Producción Agropecuaria existentes.

La información disponible para las provincias nor-amazónicas de Napo, Sucumbíos y Orellana, permitió se definir los siguientes estratos:

Tierras con uso agropecuario menores de 2 km <sup>2</sup> .	Estrato 81
Tierras con uso agropecuario de 2 km <sup>2</sup> y más.	Estrato 82
Tierras con algún uso agropecuario menores de 2 km <sup>2</sup> .	Estrato 83
Tierras con algún uso agropecuario de 2 km <sup>2</sup> y más.	Estrato 84
Tierras cubiertas por agua.	Estrato 52
Tierras de cumbres, nevados y selvas, sin uso agropecuario.	Estrato 54
Centros poblados urbanos.	Estrato 71
Centros poblados rurales:	Estrato 72

Planteamientos iniciales de factibilidad respecto al tamaño nacional de la muestra [2], promovieron el uso de aproximadamente 10.000 SMs, asumiendo que este tamaño permitiría estimar resultados por cantón.

Con el tamaño de muestra establecido a priori, se efectuó su afijación o distribución entre provincias, cantones y estratos, utilizando, para el efecto, información sobre superficie sembrada con arroz, banano, café, maíz (de todas las variedades), palma africana, papas y pastos; además del total de tierra cultivable sembrada. Esta información fue tomada de la Encuesta Nacional de Superficie y Producción, levantada por el INEC en 1995.

Definidos los tamaños de muestra para cada estrato cantonal con actividad agropecuaria, se seleccionó la misma de la siguiente forma.

**Listas de selección.-** Cada UPM del estrato cantonal fue incluida en una lista ordenada ascendentemente de acuerdo al número de identificación que le correspondió en el fotomosaico provincial y junto se colocó el número de SMs que contiene. Se conformó una columna adicional, acumulando el número de SMs de cada UPM; así, se “enumeró virtualmente” en forma consecutiva a los segmentos de muestreo. Entonces, la última línea de esta columna contuvo el número total de SMs existentes en el estrato de cantón.

**Réplicas de la muestra.-** El número de SMs asignados a la muestra en cada estrato de cantón fue un número par, para formar réplicas de al menos dos SMs. Dependiendo del número SMs considerados en la muestra, se estableció el número de réplicas no mayor de seis en cada estrato de cantón. El número de SMs a seleccionar en cada réplica se puede resumir por la siguiente relación:

$$n_{mh} = \frac{n_h}{r_h}; \text{ donde,}$$

$n_{mh}$  = número de SMs a seleccionar en la réplica m-ésima del estrato h-ésimo del cantón.  
 $n_h$  = número total de SMs a seleccionar en el estrato h-ésimo del cantón.  
 $r_h$  = número de réplicas en estrato h-ésimo del cantón, además  $1 \leq r_h \leq 6$ .

**Selección sistemática.-** Mediante el uso de software adecuado, en cada estrato de cantón se seleccionó la muestra sistemáticamente, de la siguiente forma. Primero se determinó el intervalo de selección (k), equivalente a:

[2] Ver página 14 del documento: Censo Agropecuario Nacional, Esquema General de Ejecución. INEC, octubre 1997

$$k = \frac{N_h}{n_h};$$

donde:  $N_h$  = número total de SMS en el estrato h-ésimo del cantón.

Luego se produjo automáticamente un arranque aleatorio ( $a$ ) de manera que:  $0 < a \leq k$ .

El valor de  $k$  identificó al primer SM incluido en la muestra y por ende a la primera UPM. Para seleccionar el segundo SM y los demás en la muestra se añadió sucesivamente a  $k$  el valor de  $a$ ; así:

Primer SM seleccionado	= $a$
segundo SM seleccionado	= $a+k$
tercer SM seleccionado	= $a+2k$
cuarto SM seleccionado	= $a+3k$ y se continúa para el resto de SMS.

Enseguida se muestra el número total de SMS seleccionados en la muestra. Los encuestadores, utilizando fotografía aéreas, imágenes satelitales y mapas topográficos, identificaron y encuestaron a los productores dentro de los SMS seleccionados del MMA. Se seleccionaron e investigaron 12,277 SMS del MMA, y dentro de ellos se efectuaron 145.219 entrevistas.

### **El Marco Muestral de Lista (MML)**

Se trata de un registro exhaustivo y ordenado de unidades de selección final o investigación que conforman una población. A más de ubicación geográfica o dirección, generalmente el MML incluye información respecto de las características más importantes de dichas unidades.

El MML complementó al MMA, mediante la investigación a todas y cada una de las personas productoras a cargo de *UPAs importantes*.

Para el III Censo Nacional Agropecuario, el MML estuvo conformado por unas 17,000 *UPAs importantes*, concepto bajo el cual se consideraron UPAs de las siguientes categorías:

- a) *Grandes*, con superficie iguales o mayores que 100 hectáreas;
- b) *Especiales o localizadas*, dedicadas principalmente al cultivo de productos cuya práctica no es extensiva y más bien está concentrada total o casi totalmente en zonas geográficamente determinadas; y
- c) *Sub-sectoriales*, dedicadas principalmente al cultivo de productos no tradicionales para exportación, como flores y camarones, entre otros.

Las UPAs clasificadas en dos o tres de los grupos descritos fueron registradas en uno solo de ellos, prevaleciendo, para el efecto, el criterio de extensión o tamaño sobre los otros.

Las cifras oficiales del Censo se generaron de los dos componentes, uno de área (MMA) y otro de lista (MML). En el componente de lista, todas las UPAs incluidas en la lista fueron censadas. En cambio, como se describió en "Diseño y selección de la Muestra", en el componente de áreas se seleccionó una muestra de SMS para obtener estimaciones

A cada UPA seleccionada en la muestra del MMA se le asignó un factor de expansión para propósitos de generar estimaciones para el componente MMA. Los factores de

expansión varían por estrato y cantón, de acuerdo al número de SMs seleccionados en cada uno de ellos (el número de SMs en la muestra de un estrato - cantón dependió de la intensidad y homogeneidad con que se realiza la actividad agropecuaria en su interior). Los datos recolectados en los cuestionarios del MMA fueron multiplicados por los factores de expansión para representar a las UPAs no incluidas en la muestra. Las cifras oficiales del III Censo Nacional Agropecuario representan, entonces, la suma de las cifras expandidas del componente de áreas, más las cifras –no expandidas- del componente de lista; esto podría resumirse en la siguiente expresión:

$$\hat{Y}_{Me} = \hat{Y}_{Ae} + Y_{Le}.$$

donde :

$\hat{Y}_{Me}$  = Estimación con MMM del total de la variable y en el Ecuador;

$\hat{Y}_{Ae}$  = Estimación con MMA del total de la variable y en el Ecuador;

$Y_{Le}$  = Valor con MML del total de la variable y en el Ecuador;

A su vez :

$$\hat{Y}_{Ae} = \sum_{p=1}^P \sum_{c=1}^C \sum_{h=1}^H R_h Y_h$$

donde :

$P$  = Total de provincias en Ecuador ( $p= 1, 2, 3, \dots P$ );

$C$  = Total de cantones en la provincia  $p$ -ésima ( $c= 1, 2, 3, \dots C$ );

$H$  = Total de estratos en el cantón  $c$ -ésimo ( $h= 1, 2, 3, \dots H$ ).

Las cifras para el III Censo Nacional Agropecuario en Ecuador están sujetas a errores de muestreo. Este tipo de error se introduce porque se seleccionó una muestra de SMs para obtener resultados del componente del MMA. Como las cifras fueron obtenidas sobre la base a una muestra de SMs, estas cifras no son necesariamente iguales a los resultados que se hubiesen obtenido mediante una enumeración completa de todos los SMs del MMA.

El error de muestreo, es decir el error estándar, provee al usuario con una medida de la precisión de la cifra.

El error relativo, expresado en porcentajes, es igual al error estándar de la estimación, dividido por el estimador.

La siguiente expresión puede resumir la forma de cálculo del error muestral:

$$\text{var}(\hat{Y}_{Me}) = \text{var}(\hat{Y}_{Ae}) + \text{var}(Y_{Le})$$

Donde:

$\text{var}(\hat{Y}_{Me})$  = Varianza del total nacional de una variable para el MMM;

$\text{var}(\hat{Y}_{Ae})$  = Varianza del total nacional de una variable y para el MMA

$$= \sum_{p=1}^P \sum_{c=1}^C \sum_{h=1}^H \text{var}(\hat{Y}_h)$$

a su vez:  $\text{var}(\hat{Y}_h)$  = Varianza del total de una variable en el estrato  $h$ -ésimo.

Se sabe que el ML fue investigado totalmente, sin seleccionar muestra; entonces, por definición el término  $\text{var}(\hat{Y}_{Le}) = 0$ .

Para los cultivos asociados se adopta una metodología que permite observar la real estructura de los sistemas de producción del sector agrícola ecuatoriano. Este tipo de práctica la realiza un significativo número de productores y refleja una cultura agrícola tradicional.

Los productos agrícolas componentes de una asociación están relacionados únicamente por el área física donde se encuentran sembrados o plantados, de tal forma que no es posible identificar áreas individuales para cada uno de ellos y, por lo mismo, requieren ser tratados en forma conjunta. Esto da lugar a características productivas propias para estas asociaciones, en dependencia de la forma en que están constituidas. En el III Censo Nacional Agropecuario se han encontrado 4.328 combinaciones de productos agrícolas (o distintos tipos de asociación).

Entonces, el objetivo de la metodología es recoger la mayor cantidad de información de los cultivos asociados, tanto de su variedad como de sus características. En este sentido múltiples análisis evidenciaron que alrededor de 25 asociaciones o combinaciones describen más del 70 % de la información sobre este tipo de cultivos.

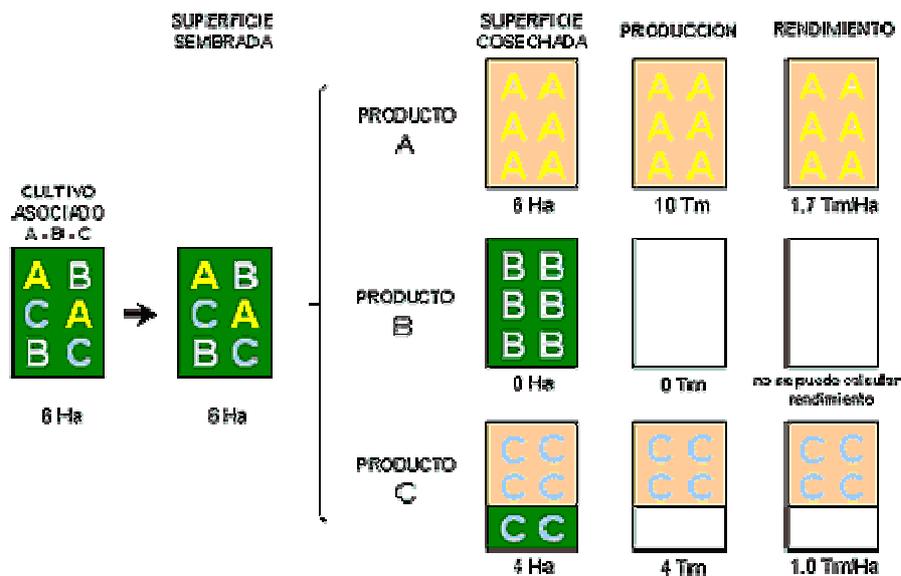
Por ejemplo eligiendo la variable *superficie sembrada* para el análisis se encontró que casi el 75 % de la información corresponde a las 25 asociaciones (ver el siguiente gráfico). Algo similar ocurre si se considera el *número de UPAs* que tienen el cultivo asociado o si se toma en cuenta un indicador sintético de las dos variables en mención. Adicionalmente, para estos 25 cultivos asociados principales, se puede mencionar las características particulares de los productos agrícolas que componen las asociaciones, así la superficie cosechada y la producción, entre otras.

Si se realiza el mismo procedimiento para cada provincia entonces resulta que, por la especialización agrícola de las mismas, los porcentajes de información descrita superan el 85 % y en algunos casos más del 90 %.

Más aún, la información restante, es decir la proveniente de los cultivos asociados no incluidos en aquellos 25 principales, fue nuevamente analizada, pero esta vez bajo el sistema de clasificación agrícola de la FAO[3]. A partir de ello ya no corresponde analizar más de 4.300 asociados distintos sino únicamente 14. Para alcanzar tal simplificación, de cada cultivo asociado se seleccionó el producto componente de mayor rendimiento relativo al rendimiento promedio en cultivos solos, y luego este cultivo seleccionado fue clasificado de acuerdo a la nomenclatura de la FAO. Sin embargo, debido a este último procedimiento adoptado ya no es posible describir las características particulares de cada producto componente (superficie sembrada, superficie cosechada, producción, etc.), únicamente se puede mencionar características generales como la superficie sembrada y el número de UPAs que tienen el cultivo clasificado.

Como se mencionó, la metodología adoptada permite observar el sistema de producción tal cual es practicado por un alto porcentaje de pequeños y medianos productores agropecuarios y en algunos casos inclusive por los grandes (cacao, café y plátano, por ejemplo), lo cual es una ventaja comparativa muy apreciable, frente a otras metodologías que asignan superficies a cada producto agrícola componente de la asociación, imposibilitando así, identificar y analizar un comportamiento productivo agrícola con arraigo en el país.

A continuación se muestra un ejemplo gráfico de esta aplicación metodológica:



En este ejemplo se tiene un cultivo asociado cuyos productos componentes son A, B y C, todos ellos sembrados en forma entrecruzada sobre la misma superficie. Además para cada uno de los productos componentes se indica la superficie cosechada y la producción, lo que en particular permite el cálculo del rendimiento del producto componente en el cultivo asociado en consideración. Acerca de la información sobre superficie cosechada y producción se han propuesto tres casos: la cosecha de toda la superficie sembrada, el caso del producto A; ninguna cosecha, correspondiente al producto B; y, la cosecha de una parte de la superficie sembrada, el caso del producto C. De ello se tiene que es factible el cálculo de los rendimientos para los productos A y C (a partir de la producción dividida para la superficie cosechada), pero no para el producto B ya que, como se indicó, no tiene superficie cosechada.