Sistema Estadístico Agropecuario Nacional

SEAN

ENCUESTA DE SUPERFICIE Y PRODUCCION AGROPECUARIA CONTINUA

DISEÑO DE MUESTREO DE LA ESPAC

por
Guillermo Otáñez; Econ.; M.A.
CONSULTOR INEC/BID

Quito, Mayo del 2004

1 INTRODUCCION

El sector agropecuario continúa siendo de vital importancia para la economía del Ecuador, debido a que por un lado, su contribución al Producto Interno Bruto (PIB) en el último quinqueño ha fluctuado alrededor del 17 %; y por otro, sigue siendo un sector generador de empleo muy importante, puesto que la Población Económicamente Activa (PEA), tiene todavía una importancia relativa del orden del 40 % de la del área rural, según el Censo de Población y Vivienda del 2001.

También este sector continúa siendo una fuente crucial en la generación de divisas, a través de la exportación de los productos tradicionales del banano, café y cacao; pero, el Ecuador al momento trata de diversificar su cuota de participación en los mercados internacionales con nuevos productos de origen agrícola, tales como las flores, mangos, espárragos, brócoli, uvillas, tunas, especialmente desde el año 2000 en que se adoptó el modelo de la dolarización.

Además, el sector agropecuario desempeña un papel estratégico dentro del concepto global de seguridad alimentaria, ya que el mismo proporciona la mayor parte de alimentos a una población que crece a una tasa anual del 2.0 %. Para satisfacer esta demanda, la frontera agrícola se expandió a una tasa de crecimiento promedio anual del orden del 1.7 % y el ganado vacuno creció a un ritmo del 2.2 % en los años comprendidos entre 1974 al 2000, superando en el último año censal los 12 millones de hectáreas bajo alrededor de 843.000 Unidades de producción agropecuaria (UPAs), dedicadas a la agricultura, ganadería y otros usos, cifra que representa casi la mitad de la superficie física nacional; el hato de ganado vacuno fue de alrededor de 4'500.00 cabezas en el 2000 y su tasa de crecimiento anual en el período de 26 años (2.2 %) es un poco superior a la de la población ecuatoriana (2.0 %).[1]

En el mismo período inter-censal 1974-00, el número de UPAs (843.000) se ha incrementado a una velocidad un poco menor (1.3 %) que el ritmo de la frontera agrícola, fenómeno que contrasta con lo que sucede con la población rural del país, la cual mantiene una tasa casi nula de crecimiento promedio anual (0.3 %). Estas UPAs mantienen una población de 3'061.500 personas, que representa las 2/3 partes de la población rural del país, dando ocupación directa en las mismas a 2'522.800, de las cuales más de la cuarta parte son trabajadores remunerados permanentes y ocasionales.

En lo que se refiere al uso o aprovechamiento de la tierra, la superficie de labor (cultivos permanentes, transitorios y tierras en descanso) ocupa casi la cuarta parte del área bajo UPAs y los pastos (cultivados y naturales) abarcan alrededor de la tercera parte de la superficie total bajo las mismas, que sumados a la categoría de páramos (también utilizados para el pastoreo), llegan a representar el 42 %. Los montes y bosques tienen una importancia relativa del 31 % y el restante 3 % corresponde a otros usos (tierras improductivas, construcciones, caminos, canales de regadío, etc.).

[1] Ver 'Ecuador: Breve Análisis de los Resultados de las Principales Variables del Censo Nacional Aropecuario'; Guillermo Otáñez y Otros. www.sica.gov.ec

Los indicadores presentados en los párrafos anteriores actualizadas con los resultados del Tercer Censo Nacional Agropecuario (CNA) 2000[2], muestran la vital importancia estratégica que todavía tiene el sector agropecuario en los aspectos económicos y sociales del país.

Es de dominio de los especialistas sectoriales que existen restricciones socio-culturales, económicas, técnicas e institucionales para planificar el desarrollo sostenible del sector agropecuario. Paralelamente, durante la última década el país ha vivido los procesos de globalización de la economía, liberalización y apertura de mercados y al momento se apresta para entrar en las negociaciones del Tratado de Libre Comercio (TLC) de las Américas. Entonces, el desarrollo sostenible y la operación eficiente y apertura transparente de los mercados, requieren disponer de un sistema de información estadística que sea objetiva, confiable y, sobre todo, oportuna.

Sin embargo, el sistema de información estadística del sector agropecuario de Ecuador siempre se ha caracterizado por ser débil e incapaz de satisfacer la demanda de cifras de los tomadores de decisiones, planificadores y otros usuarios, tanto públicos como privados, incluyendo por supuesto, a los propios Productores Agropecuarios.

Para solucionar el problema permanente de oferta de información estadística que mida de manera permanente la dinámica del sector agropecuario, el Ecuador ha realizado esfuerzos aislados a través de los últimos 50 años. El Primer Censo Nacional Agropecuario se realizó en 1954, siendo el Ecuador el primer país que en Latino América utilizó la metodología del muestreo probabilístico en una operación censal (dos procesos separados: Directorio (censo) de UPAs con cuestionario reducido y muestra con cuestionario ampliado) con excelentes resultados cantonales, provinciales y nacionales[3], seguido por el de 1962 (un solo proceso: censo con cuestionario reducido y muestra con ampliado), que no llegó a producir cifras finales. Ante el fracaso del censo agropecuario de 1962, en 1968 se ejecutó la Encuesta Agropecuaria Nacional por Muestreo, que generó datos a nivel provincial y nacional; y, en 1974 se levantó el Segundo Censo Agropecuario Nacional, el cual ha sido la única operación censal por ENUMERACIÓN COMPLETA con cuestionario ampliado , que arrojó cifras a nivel sectorial, parroquial, cantonal, provincial y nacional.

El Programa Nacional del Trigo del Ministerio de Agricultura y Ganadería (MAG), con la asistencia técnica del actual Instituto Nacional de Estadística y Censos (INEC), emprendió a finales de la década de los sesenta la Encuesta Nacional del Trigo empleando la metodología del muestreo de conglomerados combinada con la del muestreo de lista; en 1975, se actualizó el marco de muestreo con los datos del Censo Agropecuario Nacional de 1975 en las provincias de Pichincha y Bolívar. Esta encuesta se interrumpió con la desaparición del Programa Nacional del Trigo. También el Programa Nacional del Café del mismo MAG condujo dos esfuerzos con metodologías objetivas con la realización de la Encuesta del Café a finales de la década del 60 y el Censo Cafetero de los años 80.

[2] Ver 'III Censo Nacional Agropecuario: Resultados Nacionales y provinciales', Volúmrn I. INEC, MAG, SICA

[3] La publicación de los Resultados Finales del Censo Agropecuario de 1954 indica que en el mismo no se aplicó la metodología del muestreo de áreas por la falta de una cartografía y fotografía aérea adecuadas.

A finales de 1979, el INEC y el MAG, en coordinación con el Centro de Levantamiento Integrado de Recursos Naturales por Sensores Remotos (CLIRSEN) iniciaron un esfuerzo común para producir información estadística permanente con metodologías científicas. En efecto, se estableció el Sistema de Estadísticas Agropecuarias Nacionales (SEAN), aplicando la metodología del muestreo de áreas con el uso de fotografía aérea, imágenes de satélite, mapas topográficos y cartografía censal, sistema que desde 1982 hasta 1995 llevó a cabo 18 rondas produciendo estimaciones a nivel provincial y nacional. Lamentablemente, a partir de 1996, la serie anual continua del SEAN, se interrumpió por falta de voluntad política de los encargados de proveer de los recursos financieros, puesto que los técnicos siempre estuvieron disponibles en el INEC.

En un nuevo esfuerzo para proveer de información estadística oportuna y confiable del sector agropecuario, el INEC y el MAG, con el apoyo financiero del Banco Mundial y la asistencia técnica del Departamento de Agricultura de los Estados Unidos/Servicio Nacional de Estadísticas Agropecuarias (USDA/NASS) y otros expertos nacionales conocedores del sistema de información estadística del sector, están terminando las actividades de análisis, publicación y difusión del denominado "III Censo Nacional Agropecuario (CNA) 2000", pero que en realidad se trata de una "Encuesta Grande por Muestreo", cuyos objetivos básicos fueron los de conocer la estructura actual del mismo y servir de base para restablecer el sistema permanente del SEAN.

En el CNA 2000 se utilizó la metodología del Muestreo de Marcos Múltiples (MMM), la cual combina el método del Muestreo del Marco de Areas (MMA) con el del Marco de Muestreo de Listas (MML).

El INEC inició en el 2002 el restablecimiento del SEAN interrumpido en 1996, a través de la planificación y realización de la primera Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC) empleando la metodología del MMM, para proveer información estadística confiable y oportuna de manera permanente, que mida los cambios coyunturales del sector agropecuario.

Este documento contiene el Diseño de Muestreo de la ESPAC, encuesta que constituye el primer MODULO PERMANENTE dentro del Plan para el Restablecimiento del SEAN.

2 OBJETIVOS DEL DISEÑO DE MUESTREO DE LA ESPAC

Presentar el diseño de muestreo de la ESPAC, para documentar y describir los procesos técnicos y operacionales adoptados en la ejecución del plan de muestreo de la misma.

Servir de documento de capacitación y consulta para los técnicos y funcionarios especializados del INEC y otras instituciones interesadas, para cumplir con el proceso de selección de las muestras y el proceso de estimación de los resultados y sus errores de muestreo, tanto de la ESPAC como de otros módulos que se deban diseñar y desarrollar en el futuro inmediato (pronósticos de la producción de cultivos, por ejemplo).

3 EL CNA 2000

3.1 METODOLOGIA DEL CNA 2000

El CNA 2000 utilizó la metodología del Muestreo de Marcos Múltiples (MMM), la cual combina el método del Muestreo del Marco de Areas (MMA) con el del Muestreo del Marco de Lista (MML), con la finalidad de hacer más eficiente el diseño de muestro minimizando las varianzas para un costo dado. La disponibilidad de recursos fue la razón fundamental para utilizar el muestreo en el III CNA 2000. La recolección de los datos se realizó mediante ENTREVISTA DIRECTA a Informantes calificados de las UPAs incluidas en la muestra.

3.2 PLAN DE MUESTREO DEL CNA 2000: PROCESO DE SELECCION

3.2.1 Bosquejo del diseño del MMA

El diseño del MMA empleado en el CNA 2000 fue probabilístico, estratificado y replicado de áreas (conglomerados) con dos etapas de selección, mediante el cual en la primera se seleccionaron sistemáticamente las Unidades Primarias de Muestreo (UPMs) y en la segunda, se escogió aleatoriamente un Segmento de Muestreo (SM) en cada UPM incluida en la muestra de primera etapa de selección.

3.2.2 Nivel de las estimaciones del MMA

El diseño del MMA produjo estimaciones (agregados) a nivel nacional, provincial (21) y cantonal (215) y cuatro zonas con problemas de límites. En la provincia de Galápagos se realizó el CNA por enumeración completa.

3.2.3 Construcción del Marco de Areas (MA)

La superficie total del país continental (249.585 Km^2), con excepción de Galápagos, fue dividida sobre mapas topográficos escala 1:50.000, en pequeños pedazos (26.279 UPMs) de 10 Km^2 en promedio, mediante límites físicos y/o culturales fácilmente identificables en el terreno. El MA fue construido a nivel cantonal y el cuadro 1 muestra un resumen del mismo a nivel nacional.

Del cuadro siguiente se desprende que solamente alrededor del 57 % de la superficie total del país continental (sin la provincia insular de Galápagos), fue incluida en el CNA, con las UPMs y SMs representando el 69 % y 60 % de los pequeños pedazos construidos en el MA respectivamente, con un promedio de 2.03 Km^2 por SM.

CUADRO 1

CNA: RESUMEN DEL MA A NIVEL NACIONAL

ESTRATO	Area (Km^2)	UPMs	SMs	Km^2/SM	Muestra
10 Predominio de pastos	57.074	6.578	28.520	2,00	3.841 (13,5 %)
20 Predominio de cultivos transitorios	16.569	2.078	8.294	2,00	2.320 (28,0 %)
30 Predominio de cultivos permanentes	15.654	1.926	7.830	2,00	1.768 (22,6 %)
40 Predominio de vegetación natural	14.956	1.631	7.461	2,00	1.284 (17,2 %)
51 Tierras sin uso agropecuario	25.849	399	10.047	2,57	1.198 (11,9 %)
52 Tierras cubiertas por agua	2.952	973	973	3,03	*
53 Tierras de manglares sin uso agropecuario	1.523	249	249	6,12	*
54 Cumbres, nevados y selva sin uso agropecuario	100.656	5.420	40.311	2,50	*
60 Tierras con camaroneras	1.741	541	541	3,22	*
71 Centros poblados urbanos	1.361	230	5.427	0,25	*
72 Centros poblados rurales	667	1.057	2.522	0,26	877 (34,8 %)
81 Tierras con uso agropecuario menores a 2 Km^2 **	868	1.281	1.276	0,68	266 (20,8 %)
82 Tierras con uso agropecuario de 2 Km^2					

$$ {}_p y_i = \sum_{j=1}^{v_i} w_{ij} y_{ij} $$

donde:

${}_p y_i$ = *Total de una* variable *en el* i *– ésimo SM ponderado*;

w_{ij} = Proporción *de la superficie de la* j *– ésima UPA dentro del* i *– ésimo SM dividido por la superficie total de la UPA* (*dentro y fuera del SM*);

y_{ij} = *Valor de una variable en la* j *– ésima UPA dentro del* i *– ésimo SM* ($j = 1, 2, ..., v_i$);

v_i = *Número de UPAs en el* i *– ésimo SM*.

- ***PASO 2.- Estimación del TOTAL de una variable en un Estrato***: para la estimación con MMA del ***TOTAL*** de una variable en un estrato, con replicaciones independientes seleccionadas sistemáticamente y cuando se tiene igual el número de SMs por replicación, se utiliza la siguiente fórmula:

$$ \hat{Y}_h = N_h \bar{y}_h = N_h \bar{\bar{y}}_r = N_h \frac{\sum_{m=1}^{r_h} \bar{y}_{hm}}{r_h} = N_h \frac{\sum_{m=1}^{r_h} \sum_{i=1}^{n_m} {}_p y_{hmi}}{n_{hm}} = $$

$$ = N_h \frac{\sum_{m=1}^{r_h} \sum_{i=1}^{n_m} {}_p y_{hmi}}{n_h} = F_h \sum_{m=1}^{r_h} y_{hm} = F_h y_h = \sum_{i=1}^{n_h} F'_{hi} \, {}_p y_{hi}, $$

donde:

$\hat{Y}_h$ = *Estimación del agregado de una variable en el* h *– ésimo estrato*;

N_h = *Número de SMs en la población o universo del* h *– ésimo estrato*, sin *icluir las UPAs del MML*;

- ***PASO 4.- Estimación del TOTAL de una Variable de un Cantón, una Provincia y del Ecuador***: simplemente se agregan sucesivamente las estimaciones de los ***TOTALES*** de cada nivel geográfico. La fórmula es:

$$\hat{Y}_{Ae} = \sum_{p=1}^{P} \sum_{c=1}^{C} \sum_{h=1}^{H} F_h^{'} y_h,$$

donde:

$\hat{Y}_{Ae}$ = *Estimación con MMA del TOTAL de una* var*iable en el Ecuador*;

P = *Número total de provincias en el Ecuador* ($p = 1, 2, 3, ..., P$);

C = *Número total de cantones en la provincia* p – *ésima* ($c = 1, 2, 3, ..., C$);

H = *Número total de estratos en el cantón* c – *ésimo* ($h = 1, 2, 3, ..., H$).

- ***PASO 5.- Estimación de la Varianza del TOTAL de una Variable de un Cantón, una Provincia y del Ecuador***: simplemente se agregan sucesivamente las estimaciones de las ***VARIANZAS*** de los ***TOTALES*** de cada nivel geográfico. La fórmula es:

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

donde:

$\text{var}(\hat{Y}_{Ae})$ = *Estimación con MMA de la* var*ianza del TOTAL de una* var*iable en el Ecuador*;

P = *Número total de provincias en el Ecuador* ($p = 1, 2, 3, ..., P$);

C = *Número total de cantones en la provincia* p – *ésima* ($c = 1, 2, 3, ..., C$);

H = *Número total de estratos en el cantón* c – *ésimo* ($h = 1, 2, 3, ..., H$).

3.3.2 Proceso de Estimación con el MML

El TOTAL de una variable con el MML, es simplemente la suma de la misma a nivel de Cantón multiplicada por un FACTOR DE AJUSTE para la falta de respuesta. Luego, se suman todos los Cantones y Provincias, para obtener el TOTAL de una variable a nivel del País. La varianza es igual a cero (0) dado que se incluyó el 100 % de las UPAs registradas en el Directorio del ML del CNA 2000 La fórmula del TOTAL de una variable es:

$$Y_{Le} = \sum_{p=1}^{P} \sum_{c=1}^{C} FA_c Y_c ,$$

donde :

$Y_{Le} = TOTAL\ de\ una\ \text{var}iable\ con\ MML\ en\ el\ Ecuador$;

$FA_c = Factor\ de\ Ajuste\ por\ falta\ de\ respuesta\ en\ el\ Cantoón\ c-ésimo$;

$Y_c = TOTAL\ de\ una\ \text{var}iable\ con\ MML\ en\ el\ Cantón\ c-ésimo$;

$P = Número\ total\ de\ provincias\ en\ el\ Ecuador\ (p = 1, 2, 3, ..., P)$;

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

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

3.3.3 Proceso de Estimación con el MMM

La estimación del TOTAL de una variable con el MMM, es la unión de la estimación del TOTAL producida por el MMA con el agregado generado por el MML. La varianza de la estimación de una variable del MMM, es igual solamente a la varianza del MMA, puesto que el 100 % de las UPAs del ML fueron incluidas en el CNA 2000. Las fórmulas empleadas en este proceso de estimación del TOTAL de una variable y su varianza son:

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

donde :

$\hat{Y}_{Me} = Estimación\ con\ MMM\ del\ TOTAL\ de\ una\ \text{var}iable\ en\ el\ Ecuador$;

$\hat{Y}_{Ae} = Estimación\ con\ MMA\ del\ TOTAL\ de\ una\ \text{var}iable\ en\ el\ Ecuador$;

$Y_{Le} = TOTAL\ de\ una\ \text{var}iable\ con\ MML\ en\ el\ Ecuador$.

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

donde :

$\text{var}(\hat{Y}_{Me}) = Estimación\ con\ MMM\ de\ la\ \text{var}ianza\ del\ TOTAL\ de\ una\ \text{var}iable\ en\ el\ Ecuador$;

$\text{var}(\hat{Y}_{Ae}) = Estimación\ con\ MMA\ de\ la\ \text{var}ianza\ del\ TOTAL\ de\ una\ \text{var}iable\ en\ el\ Ecuador$;

$\text{var}(Y_{Le}) = Varianza\ del\ TOTAL\ de\ una\ \text{var}iable\ con\ MML\ en\ el\ Ecuador$

$= 0.$

3.3.4 Otras Fórmulas para el Proceso de Estimación con MMA

Es importante señalar que las fórmulas de estimación del TOTAL de una variable y de su varianza presentadas en el numeral 3.3.1, consideran que las replicaciones tienen igual número de SMs. Sin embargo, en algunas ocasiones esta igualdad no se mantiene

debido a la FALTA DE RESPUESTA dado que algunos SMs no son investigados en el campo por diferentes razones (rechazos, dificultades de transporte y accesibilidad, etc.). Entonces, en estos casos es necesario introducir factores de ponderación adecuados en el FACTOR DE EXPANSION para suplir por la falta de respuesta. Sin embargo, no se debe confundir la FALTA DE RESPUESTA con el caso en que un SM completo es SIN USO AGROPECUARIO y por lo tanto, no existe ninguna UPA; en el primer caso, el SM no entra en el cálculo de la varianza porque su efecto debe considerarse en el ajuste del FACTOR DE EXPANSION; en el segundo, el SM tiene un valor de cero (0), por lo que no se necesita ajustar el FACTOR DE EXPANSION, dado que se trata de un SM que no es FALTA DE RESPUESTA.

A continuación se presentan otras fórmulas alternativas para la estimación del TOTAL de una variable y su varianza en un estrato, que producen los mismos resultados de las presentadas en los numerales anteriores.

FORMULA´SAS´ POR ESTRATO EMPLEADA EN EL CNA

$$_{SAS}\hat{Y}_h = \sum_{m=1}^{r_h} {}_{SAS}\hat{Y}_{hm} = \sum_{m=1}^{r_h}\sum_{i=1}^{n_m} F'_{hmi}\ {}_p y_{hmi},$$

donde:

$_{SAS}\hat{Y}_h$ = *Estimación SAS del TOTAL de una* var*iable en el estrato h – ésimo*;

$_{SAS}\hat{Y}_{hm}$ = *Estimación del TOTAL de una* var*iable en la réplica m – ésima del estrato h – ésimo*;

F'_{hmi} = *Factor de Expansión ajustado por falta de respuesta en el i – ésimo SM de la réplica m – ésima del estrato h – ésimo*;

$_p y_{hmi}$ = *Total de una* var*iable ponderada en la muestra del i – ésimo SM de la réplica m – ésima del estrato h – ésimo.*

$$\text{var}(_{SAS}\hat{Y}_h) = (1 - \frac{r_h}{(N_h / n_m)})(r_h)\frac{\sum_{m=1}^{r_h}(\hat{Y}_{hm} - \hat{\bar{Y}}_h)^2}{r_h - 1},$$

donde:

$$\hat{\bar{Y}}_h = \frac{\sum_{m=1}^{r_h}\hat{Y}_{hm}}{r_h} = \textit{Media por réplica de la estimación del TOTAL en el h – ésimo estrato.}$$

FORMULA´FAO´ PARAUN ESTRATO

$$_{FAO}\hat{Y}_h = \frac{\sum_{m=1}^{r_h} {}_{FAO}\hat{Y}^*_{hm}}{r_h} = \frac{\sum_{m=1}^{r_h}\sum_{i=1}^{n_m} {}_{FAO}Y^*_{hmi}}{r_h} = \sum_{m=1}^{r_h}\sum_{i=1}^{n_m} F_{h\ p}\, y_{hmi},$$

donde:

$_{FAO}\hat{Y}_h$ = *Estimación FAO del TOTAL de una* var*iable en el estrato h – ésimo*;

$$_{FAO}\hat{Y}^*_{hm} = \sum_{i=1}^{n_m} {}_{FAO}\hat{Y}^*_{hmi};$$

$$_{FAO}\hat{Y}^*_{hmi} = r_h F_{h\ p}\, y_{hmi};$$

$$F_h = \frac{N_h}{n_h}, \text{cuando el número de SMs por réplica es igual};$$

$$F_h = \frac{N_h}{n_h^{'}}, \text{cuando el número de SMs por réplica no es igual por falta de respuesta.}$$

$$\text{var}(_{FAO}\hat{Y}_h) = (1 - f_h)\frac{\sum_{m=1}^{r_h}(_{FAO}\hat{Y}^*_{hm} - _{FAO}\hat{\bar{Y}}^*_h)^2}{r_h(r_h - 1)} = (1 - f_h)\frac{\sum_{m=1}^{r_h}(_{FAO}\hat{Y}^*_{hm} - _{FAO}\hat{Y}_h)^2}{r_h(r_h - 1)},$$

donde:

$$_{FAO}\hat{\bar{Y}}^*_h = \frac{\sum_{m=1}^{r_h} {}_{FAO}\hat{Y}^*_{hm}}{r_h} = {}_{FAO}\hat{Y}_h.$$

FORMULA´KISH´ PARAUN ESTRATO PARA REPLICAS DESIGUALES

$${}_{KISH}\hat{Y}_h = N_h\, {}_{KISH}\bar{y}_h,$$

donde:

$${}_{KISH}\bar{y}_h = \frac{1}{r_h}\sum_{m=1}^{r_h}\frac{n_{hm}}{n_h / r_h}\bar{y}_{hm},$$

donde:

$$\bar{y}_{hm} = \frac{\sum_{i=1}^{n_m}{}_p\, y_{hmi}}{n_{hm}} = \textit{media de la réplica } m-\textit{ésima en el estrato } h-\textit{ésimo.}$$

$$\text{var}({}_{KISH}\hat{Y}_h) = N_h^2\ \text{var}({}_{KISH}\bar{y}_h),$$

donde:

$$\text{var}({}_{KISH}\bar{y}_h) = \frac{(1-f_h)}{r_h(r_h-1)}\sum_{m=1}^{r_h}\left[\frac{n_{hm}}{n_h / r_h}(\bar{y}_{hm}-\bar{y}_h)\right]^2.$$

3.3.5 Intervalo de Confianza

La fórmula para construir el intervalo de confianza para un nivel de confiabilidad dado.

$$\Pr\left[\hat{Y} - z_{\alpha/2}ee(\hat{Y}) \le Y \le \hat{Y} + z_{\alpha/2}ee(\hat{Y})\right] = (1-\alpha),$$

donde:

Pr = Pr *obabilidad o área bajo la curva de la distribución normal*;

$\hat{Y}$ = *Valor de la estimación de una* var *iable en estudio y*;

$z_{\alpha/2}$ = *Valor de la abscisa de la distribución normal estándar para un nivel de confiabilidad (probabilidad) dado* (95%, *por ejemplo*);

$ee(\hat{Y})$ = *Error estándar de la estimación* $\hat{Y}$;

Y = *Parámetro verdadero de la población o universo*;

α = Pr *obabilidad o área bajo la curva normal de que el* int *ervalo a construirse no encierre al parámetro verdadero de la población o universo* ($\alpha = 0.05 = 5\%$ *con un* 95 % *de confiabilidad*, *por ejemplo*).

- **Ejemplo de Intervalo de Confianza:** a continuación se presenta un ejemplo de la construcción del intervalo de confianza con el 95 % de confiabilidad en la provincia de Bolívar, en la que el CNA 2000 estima 3.901 Ha con un error relativo del 10.8 %. La varianza de la estimación es 178.359 (ver Anexo I).

$$\Pr\left[\hat{Y} - z_{\alpha/2} ee(\hat{Y}) \le Y \le \hat{Y} + z_{\alpha/2} ee(\hat{Y})\right] = (1-\alpha),$$

Entonces:

$$\Pr\left[3.900 - 1{,}96 * 422 \le Y \le 3.900 + 1{,}96 * 422\right] = (1 - 0{,}05),$$

$$\Pr\left[3.900 - 844 \le Y \le 3.900 + 844\right] = 0{,}95,$$

$$\Pr\left[3.056 \le Y \le 4.744\right] = 0{,}95 = 95\,\%.$$

- **Interpretación del Intervalo de Confianza:** la superficie verdadera de cultivos permanentes existente en el Cantón 1 de la Provincia de Bolívar en el 2000, puede oscilar entre 3.056 y 4.744 Ha con un 95 % de confiabilidad (seguridad) de no equivocarse. Pero, existe un 5 % de probabilidad de que la superficie real bajo cultivos permanentes esté por fuera del intervalo construido.

3.3.6 Ejemplos del Proceso de Estimación del CNA

Los Anexos I y II presentan dos ejemplos del proceso de estimación del TOTAL de una variable, su varianza, error estándar (error de muestreo), error estándar relativo (coeficiente de variación) e intervalo de confianza. Estos ejemplos han sido desarrollados en EXCEL utilizando los cuatro tipos de fórmulas presentados anteriormente.

ANEXO I: con igual número de SMs en cada replicación y un SM sin uso agropecuario

- **Provincia**: Bolívar;
- **Cantón**: 1;
- **Variable**: superficie bajo cultivos permanentes;
- **Archivo EXCEL**: *SICA BOLIVAR CANTON 1*.

ANEXO II: con un estrato que no tiene igual número de SMs en cada replicación (falta de respuesta) y otro que tiene un SM sin uso agropecuario

- **Provincia**: Guayas;
- **Cantón**: 7;
- **Variable**: superficie sembrada de arroz;
- **Archivo EXCEL**: *SICA GUAYAS CANTON 7*.

4 LA ESPAC

4.1 METODOLOGIA DE LA ESPAC

La ESPAC usa la misma metodología del MMM del CNA, la cual combina el método del MMA con el del MML, bajo el enfoque o paraguas de los principios del MUESTREO EN DOS FASES o MUESTREO DOBLE La recolección de los datos se realiza mediante ENTREVISTA DIRECTA a Informantes calificados de las UPAs incluidas en:

- La sub-muestra de la muestra del MA;
- La muestra del ML.

4.2 USO DEL MUESTREO EN DOS FASES EN LA ESPAC

El diseño de muestreo de la ESPAC utiliza los principios y ventajas del MUESTREO DOBLE O EN DOS FASES, el cual eficientemente combinado con el método del MMM estratificado aplicado en el CNA, tiene la potencialidad de mejorar las estimaciones de las variables en estudio reduciendo sus errores de muestreo para un costo dado. Entonces, mediante el MUESTREO EN DOS FASES (DOBLE):

- **LA PRIMERA FASE DEL MMA**: está conformada por la muestra GRANDE de 12.277 SMs seleccionada de los 70.000 SMs del MA en el CNA;
- **LA SEGUNDA FASE DEL MMA**: está integrada por la sub-muestra PEQUEÑA de 2.000 SMs escogida en la ESPAC de la muestra GRANDE de 12.277 SMs del CNA;
- **LA PRIMERA FASE DEL MML**: está compuesta por las por el 100 % de las 12.341 UPAs que perteneciendo al ML en el CNA, fueron incluidas en el ESPAC (en esta encuesta no se toman en cuenta los estratos 11 y 15 de camaroneras);
- **LA SEGUNDA FASE DEL MML**: está conformada por la muestra PEQUEÑA de 4.000 UPAs seleccionada en la ESPAC desde el 100 % de las 12.341 UPAs del ML del CNA, a la que se considera como la muestra GRANDE dentro del concepto del MUESTREO EN DOS FASES (DOBLE).

Es necesario aclarar que, si bien no se conocía el hecho de que cuando se diseñó la muestra del CNA, no se tuvo en cuenta el empleo del MUESTREO EN DOS FASES (DOBLE) para el sistema de encuestas continuas posteriores al mismo, en el diseño de muestreo de la ESPAC se consideró que este tipo de muestreo ofrecía técnicas de selección y análisis que podrían ser aplicadas, especialmente en el proceso de estimación. En otras palabras, en la ESPAC el diseño de la sub-muestra de 2.000 SMs deL MA y de la muestra de 4.000 UPAs del ML (que de aquí en adelante se mencionarán como “muestra PEQUEÑA de segunda fase del MA” y “muestra PEQUEÑA de segunda fase del ML”), toman a la sub-muestra de 12.277 SMs como la muestra GRANDE de primera fase y al 100 % de las 12.341 UPAs del ML como la muestra GRANDE de primera fase.

Por otra parte, los tamaños de la sub-muestra del MA de 2.000 SMs y de la muestra del ML de 4.000 UPAs de la ESPAC, fueron fijados con base en el presupuesto disponible sin consideración de grados de precisión y confiabilidad y costos, que son factores a considerarse bajo otras circunstancias cuando se usa el MUESTREO EN DOS FASES (DOBLE).
Por lo tanto, la eficiencia estadística de las estimaciones generadas por el MUESTREO EN DOS FASES (DOBLE) se debe a la posibilidad de correlacionar la información obtenida de la muestra de PRIMERA FASE con la información recolectada en la muestra de SEGUNDA FASE.

4.3 MUESTRA PEQUEÑA DE SEGUNDA FASE DEL MA: PROCESO DE SELECCION

4.3.1 Objetivos de la Muestra PEQUEÑA DE SEGUNDA FASE del MA

Con una restricción muy fuerte en el tamaño de muestra predeterminado demasiado pequeño de 2.000 SMs, el diseño de la muestra PEQUEÑA DE SEGUNDA FASE del MA en la ESPAC persiguió los siguientes objetivos fundamentales:

- Mantener el diseño de la muestra PEQUEÑA DE SEGUNDA FASE del MA lo más ajustado posible al diseño del CNA (replicaciones independientes dentro de un estrato en cada Cantón), por múltiples razones estadísticas y técnicas;
- Producir las estimaciones a nivel provincial con grados de precisión y niveles de confiabilidad aceptables, que no sean muy diferentes a los del CNA;
- Diseñar la muestra PEQUEÑA DE SEGUNDA FASE del MA de tal manera que en el futuro permita el incremento del tamaño de la misma, cuando se consigan mayores recursos.

4.3.2 Tamaño de la Muestra PEQUEÑA DE SEGUNDA FASE del MA

Como se indicó anteriormente, en el CNA se utilizó una muestra de 12.277 SMs que fueron seleccionados dentro de cada Cantón en estratos de usos del suelo con réplicas independientes, para producir estimaciones a partir del nivel cantonal.

El tamaño de la muestra PEQUEÑA DE SEGUNDA FASE del MA fue fijado en 2.000 SMs con base al presupuesto asignado al INEC con anterioridad. En realidad, en el 2001 se habían presentado a las autoridades del INEC otras alternativas de tamaños de muestra: una ideal de 6.000 SMs; una segunda deseable de 3.700 SMs, que era el tamaño de muestra de la última encuesta del SEAN de 1995. Finalmente se consiguió financiamiento para 2.000 SMs.

4.3.3 Nivel de las Estimaciones del MMA

Con la finalidad de lograr los objetivos del numeral anterior bajo la restricción de un tamaño de muestra muy pequeño, se adoptó producir las estimaciones para los siguientes niveles:

- Provincias auto-representadas (importancia alta en la producción agropecuaria, especialmente de exportación), que se espera producirán estimaciones con buenos grados de precisión y niveles de confiabilidad y que mantienen al máximo posible el diseño de muestreo del CNA: Guayas, Manabí, Los Ríos, Pichincha. A este grupo se añadió Chimborazo por razones técnicas de facilidad de mantener el diseño del CNA, aun cuando no mantiene las mismas características de las otras cuatro anteriores;
- Provincias no-auto-representadas, que se espera generarán estimaciones provinciales con regulares grados de precisión y niveles de confiabilidad y que no mantienen al máximo posible el diseño de muestreo del CNA: Azuay, Bolívar, Cañar, Carchi, Cotopaxi, El Oro, Esmeraldas, Imbabura, Morona Santiago, Loja y Tungurahua;
- Grupos de Provincias: Nororiente (Napo, Orellana y Sucumbíos), Suroriente (Pastaza y Zamora Chinchipe);
- Areas no incluidas: Galápagos y Zonas en Discusión.

4.3.4 Reparto del Tamaño de la Muestra PEQUEÑA DE SEGUNDA FASE del MA entre Provincias

El reparto entre provincias del tamaño de la muestra PEQUEÑA DE SEGUNDA FASE del MA de 2.000 SMs, se realizó mediante el método de ´COMPROMISO´ entre las afijaciones proporcional, tamaños iguales y la de afijación óptima. Para esta última, se usaron las variables correspondientes a las categorías de uso del suelo y combinaciones de las mismas, superficies sembradas de 14 de los principales cultivos, existencias de ganado vacuno, porcino, ovino y aves de corral, con sus respectivas varianzas obtenidas del CNA para el MA.[4] En el cuadro 3 se puede observar el reparto final del tamaño de la muestra PEQUEÑA DE SEGUNDA FASE del MA entre las provincias.

4.3.5 Diseño de la Muestra PEQUEÑA DE SEGUNDA FASE de las Provincias Auto-Representadas

El bosquejo del diseño de la muestra PEQUEÑA DE SEGUNDA FASE del MA de las provincias auto-representadas es el siguiente:

- Clasificación mediante la técnica del ´ANALISIS DE CONGLOMERADOS (CLUSTER ANÁLISIS)´ de los cantones de cada provincia en dos Dominios de Estudio (DEs) de acuerdo a variables compuestas de cultivos (permanentes, transitorios y barbecho), pastos (cultivados, naturales y páramos) y superficie total bajo la Unidad de Producción Agropecuaria (UPA). Estos DEs son:
  - ´I´: cantones más importantes;
  - ´II´: cantones menos importantes.

  El cuadro 4 indica los dos DEs formados con el análisis de conglomerados.
- Sub-clasificación de cada DE en dos Sub-Dominios de Estudio (SDEs) de acuerdo a la estratificación por uso del suelo del CNA. Estos SDEs son:
  - ´A´: conformado por los cantones, replicaciones y estratos 10, 20 y 30 del CNA 2000 dentro del DE ´I´;

[4] Para mayores detalles del reparto del tamaño de la sub-muestra del MA entre provincias, consultar en el Departamento de Muestreo de la Dirección de Planificación del INEC.

- ü ´B´: constituido por los estratos 40 y 51, que previamente fueron colapsados entre cantones dentro del DE ´I´;
- ü ´C´: integrado por los estratos 10, 20 y 30, que previamente fueron colapsados entre cantones dentro del DE ´II´;
- ü ´D´: compuesto por los estratos 40 y 51, que previamente fueron colapsados entre cantones dentro del DE ´II´;
- ü El estrato 72 de cabeceras parroquiales y centros poblados de otras áreas rurales no fue incluido en la ESPAC, por su escasa representatividad con respecto a la producción agropecuaria y otras variables correlacionadas.

- Selección sistemática de dos SMs por cada replicación dentro de cada estrato en un cantón, con el método de Brewer[5] con Probabilidad Proporcional al Tamaño (PPT) de la superficie bajo ´cultivos y pastos´ o superficie total bajo UPAs en el SDE ´A´. En este SDE A se mantiene el diseño de muestreo del CNA;
- Selección sistemática controlada de la sub-muestra asignada (generalmente con afijación proporcional o igual) de SMs a cada estrato colapsado entre cantones en los SDEs ´B´, ´C´ y ´D´, en los cuales no se mantiene el diseño de muestreo del CNA. Las variables compuestas de control fueron cultivos y pastos o superficie total bajo UPAs y la importancia de la producción agropecuaria en estos SDEs, es generalmente baja;
- Además, se introdujo una nueva variable de control para los cuatro SDEs (´A´, ´B´, ´C´ y ´D´) consistente en que los SMs cuyas variables compuestas de cultivos y pastos o superficie total bajo UPAs, eran igual a cero (0), fueron excluidas de la selección sistemática, circunstancia que hay que tener en cuenta en el proceso de estimación.

  El cuadro 5 contiene un resumen de la selección de la muestra de la provincia de Guayas.

4.3.6 Diseño de la Muestra PEQUEÑA DE SEGUNDA FASE del MA en las Provincias No-Auto-Representadas

El diseño de la muestra PEQUEÑA DE SEGUNDA FASE del MA en las provincias no-auto-representadas, es similar al de las auto-representadas, salvo que en el SDE ´A´ el proceso de selección es el mismo que el de los SDEs ´B´, ´C´ y ´D´. Es importante indicar que en estas provincias no se mantiene el diseño de replicaciones y estratos a nivel cantonal del CNA, pues los estratos fueron colapsados entre cantones, como se indicó en el numeral anterior. Algunas de estas provincias no tienen los cuatro SDEs.

4.3.7 Diseño de la Muestra PEQUEÑA DE SEGUNDA FASE del MA en las Provincias Agrupadas

El diseño de la muestra PEQUEÑA DE SEGUNDA FASE del MA en las provincias agrupadas, es similar al de las no-auto-representadas, pero los estratos 10, 20, 30, 40 y 51 fueron colapsados entre cantones y provincias.

[5] Ver ´Técnicas de Muestreo´; William G. Cochran. John Wiley & Sons, New York

transitorios y barbecho para el estrato 3 de flores; la superficie de permanentes en el estrato 5 de mango; la superficie de transitorios y barbecho en el caso del estrato 8 de tabaco; el número de porcinos para el estrato 12;

- Además, cuando las variables compuestas de control de las UPAs, eran igual a cero (0), fueron excluidas de la selección sistemática, circunstancia que hay que tener en cuenta en el proceso de estimación.

4.5 RESUMEN DE LOS TAMAÑOS DE MUESTRA

El cuadro 7 presenta un resumen por provincias de las muestras PEQUEÑAS DE SEGUNDA FASE seleccionadas del MA y ML, para la realización de la ESPAC.

CUADRO 7

ESPAC: RESUMEN DE LOS TAMAÑOS DE MUESTRAS PEQUEÑAS DE SEGUNDA FASE DEL MA Y ML, SEGUN PROVINCIAS.

Provincia	MA					ML		
	Universo	Tamaño muestra		% submuestra del		Universo	muestra	% de la muestra
		CNA	ESPAC	Universo	CNA			
Azuay	3117	666	65	2.1	9.8	607	117	19.3
Bolívar	1864	409	80	4.3	19.6	255	37	14.5
Cañar	1290	303	40	3.1	13.2	170	130	76.5
Carchi	1113	285	30	2.7	10.5	238	48	20.2
Cotopaxi	2351	479	100	4.3	20.9	401	156	38.9
Chimborazo	2174	500	130	6.0	26.0	406	128	31.5
El Oro	2320	523	100	4.3	19.1	683	142	20.8
Esmeraldas	4537	576	80	1.8	13.9	881	240	27.2
Guayas	7400	1453	400	5.4	27.5	1461	859	58.8
Imbabura	1752	375	50	2.9	13.3	293	103	35.2
Loja	5126	893	121	2.3	13.5	542	202	37.3
Los Ríos	3563	749	196	5.5	26.2	687	317	46.1
Manabí	8603	1258	298	3.5	23.7	1775	534	30.1
Pichincha	5755	774	170	3.0	22.0	1220	486	39.8
Tungurahua	1020	304	40	3.9	13.2	573	98	17.0
Morona S.	5866	669	35	0.6	5.2	1065	125	11.7
Nor-oriente	5426	1036	39	0.7	3.8	479	163	34.0
Sur-oriente	5995	826	26	0.4	3.1	605	115	19.0
TOTAL	**69272**	**12078**	**2000**	**2.9**	**16.6**	**12341**	**4000**	**32.4**

5 PROCESO DE ESTIMACIÓN DE LA ESPAC MEDIANTE EXPANSION DIRECTA

5.1 SIMULACION DEL PROCESO DE ESTIMACIÓN DE LA ESPAC MEDIANTE EXPANSION DIRECTA CON LOS DATOS DEL CNA

El cuadro 8 contiene un resumen comparativo a nivel nacional entre las estimaciones SIMULADAS desde la ESPAC para el uso del suelo y los resultados generados por el CNA con sus respectivos errores relativos de muestreo o coeficientes de variación (cv). Estas estimaciones y sus errores de muestreo han sido producidas a través del método de EXPANSION DIRECTA utilizando las mismas fórmulas del CNA presentadas en los capítulos anteriores. El Anexo III (archivo EXCEL 'ESPAC ERRORES USO 1') presenta estos mismos resultados, pero a nivel regional y provincial.

CUADRO 8

COMPARACIÓN DE LAS ESTIMACIONES SIMULADAS DEL USO DEL SUELO DE LA ESPAC CON EL CNA

Uso del Suelo	CNA*		ESPAC (simulación)		Razón=ESPAC/CNA	
	Ha (000)	cv (%)	Ha (000)	Cv (%)	Ha	Cv
Permanentes	1305.1	0.9	1334.1	2.4	+ 1.02 veces	+ 2.7 veces
Transitorios*	1223.4	0.8	1250.5	2.3	+ 1.02 veces	+ 2.9 veces
Descanso	376.9	1.7	362.4	4.6	- 1.04 veces	+ 2.7 veces
Pastos Cultiv	3302.7	0.7	3485.7	2.2	+ 1.05 veces	+ 3.1 veces
Pastos Natur	1126.5	1.3	1268.6	3.3	+ 1.13 veces	+ 2.5 veces
Páramos	600.3	2.1	619.5	4.1	+ 1.03 veces	+ 2.0 veces
Bosques	3861.9	0.9	3506.9	2.3	- 1.10 veces	+ 2.6 veces
Otros**	173.8	1.6	213.9	4.1	+ 1.23 veces	+ 2.6 veces
TOTAL	**11970.6**	**0.4**	**12041.8**	**0.9**	**+ 1.01 veces**	**+ 2.3 veces**

* Se excluyen Galápagos y "Zonas No Asignadas" las cuales no están tomadas en cuenta en la ESPAC.

** Se excluyen los estratos 11 y 15 de camaroneras no consideradas en la ESPAC.

De un somero análisis de los resultados que constan en el cuadro 8 y Anexo III, se pueden establecer los siguientes aspectos importantes:

- A nivel nacional (cuadro 8), las diferencias en las superficies estimadas de cultivos es de alrededor del 2 % y de pastos considerados en conjunto (cultivados+naturales +páramos) es casi del 7 %; la diferencia mayor se observa en la categoría de otros usos en la que la ESPAC estima casi ¼ más del área del CNA. La superficie total bajo UPAs mantiene una pequeña discrepancia del 1 %;
- A nivel nacional (cuadro 8), los errores relativos estándar de muestreo de la ESPAC que son menores que el 5 %, naturalmente aumentan con respecto a los obtenidos en el CNA, pero no en la magnitud prevista, debido a que si el tamaño de la muestra disminuye en un poco más de seis veces, era de esperarse que los mismos aumenten por un factor de 36. En realidad, los errores relativos de muestreo o coeficientes de variación (cv) se incrementan entre alrededor del 2 y 3 veces. Esto se debe a la eficiencia ganada en el diseño de las dos muestras PEQUEÑAS DE SEGUNDA FASE, tanto del MA como del ML, gracias a la abundante información generada por el CNA;

- A nivel provincial (Anexo III) naturalmente los coeficientes de variación son superiores a los nacionales y regionales. En el grupo de las cuatro provincias auto-representadas (Guayas, Manabí, Los Ríos y Pichincha), cultivos y pastos son estimados en la ESPAC 2000 con errores relativos de muestreo inferiores al 10 % (con varios casos que son inferiores al 5 %), excepto pastos naturales en Los Ríos y Manabí, cuyos coeficientes de variación son del orden del 16 y 19 % respectivamente. En las otras provincias se observan errores relativos de muestreo menores del 10 % y la gran mayoría de los más altos oscilan en un rango del 25 % al 40 %;
- En general a nivel provincial (Anexo III), el coeficiente entre los errores relativos de muestreo de la ESPAC y los del CNA, oscila entre 3 y 5 veces.

El análisis anterior indica, que para los tamaños de muestras establecidos con la restricción del presupuesto escaso disponible, que para las principales variables estimadas de la ESPAC se puede esperar lo siguiente:

- Estimaciones muy buenas a nivel nacional y regional;
- Estimaciones buenas a nivel de las cuatro provincias auto-representadas (Guayas, Manabí, Los Ríos, Pichincha y Chimborazo);
- Estimaciones regulares para el resto de provincias.

La expectativa de la última viñeta anterior, debe llevar a estudiar detenidamente diferentes alternativas para mejorar la eficiencia de las estimaciones:

- Colapsar provincias: las seis de la amazonía pueden agruparse en una sola, por ejemplo;
- Utilizar los métodos de razón y regresión aprovechando de que se ha empleado el MUESTREO EN DOS FASES (DOBLE) en el proceso de selección de las muestras PEQUEÑAS DE SEGUNDA FASE, tanto del MA como del ML;
- Experimentar nuevos métodos de estimación, como el de razón y regresión, sin usar el MUESTREO EN DOS FASES (DOBLE). En las estimaciones del MMM presentadas en el cuadro 8 y anexo III, solamente se han sumado las estimaciones del MA con la del ML. Sin embargo, en la literatura de la metodología del MMM existen muchos otros estimadores que se pueden aplicar, como los producidos por regresión múltiple, por ejemplo.

5.2 FACTOR ORIGINAL Y FINAL DE EXPANSION DIRECTA DE UN ESTRATO EN LA MUESTRA PEQUEÑA DE SEGUNDA FASE DEL MA

En general, en un estrato de la muestra PEQUEÑA DE SEGUNDA FASE del MA, la fracción de muestreo o probabilidad de selección de un SM y su correspondiente Factor Original de Expansión Directa ($_{MA}FOED_{ESPAC}$) y Factor Final de Expansión Directa ($_{MA}FFED_{ESPAC}$) dentro de un estrato, responde a la fórmula siguiente:

$$_{MA}f_{ESPAC(f)} = {}_{MA}f_{CNA(G)} * {}_{MA}f_{ESPAC(p)} = \frac{{}_{MA}n_{CNA(G)}}{{}_{MA}N_{CNA}} * \frac{{}_{MA}n_{ESPAC(p)}}{{}_{MA}n_{CNA(G)}} = \frac{{}_{MA}n_{ESPAC(p)}}{{}_{MA}N_{CNA}},$$

donde:

$_{MA}f_{ESPAC(f)}$ = *Fracción de muestreo final de la ESPAC en el MA*;

$_{MA}f_{CNA(G)}$ = *Fracción de la muestra GRANDE DE PRIMERA FASE del CNA en el MA*;

$_{MA}f_{ESPAC(p)}$ = *Fracción de la muestra PEQUEÑA DE SEGUNDA FASE de la ESPAC en el MA*;

$_{MA}n_{CNA(G)}$ = *No. de SMs de la muestra GRANDE DE PRIMERA FASE del CNA* 2000 *en el MA*;

$_{MA}N_{CNA}$ = *No. de SMs en la población del CNA en el MA*;

$_{MA}n_{ESPAC(p)}$ = *No. de SMs de la muestra PEQUEÑA DE SEGUNDA FASE de la ESPAC en el MA.*

Por lo tan *to, el Factor Original de Expansión Directa en el MA es*:

$$_{MA}FOED_{ESPAC} = \frac{1}{{}_{MA}f_{ESPAC(f)}} = \frac{{}_{MA}N_{CNA}}{{}_{MA}n_{ESPAC(p)}},$$

donde:

$_{MA}FOED_{ESPAC}$ = *Factor Original de Expansión Directa de la ESPAC en el MA.*

Luego, el $_{MA}FOED_{ESPAC}$ *tiene que ser ajustado de la siguiente forma para obtener el Factor Final de Expansión Directa en el MA*:

$$_{MA}FFED_{ESPAC} = {}_{MA}FOED_{ESPAC} * {}_{MA}FA_{sub} * {}_{MA}FA_{dis} * {}_{MA}FA_{cob} * {}_{MA}FA_{res},$$

donde:

$_{MA}FFED_{ESPAC}$ = *Factor Final de Expansión Directa de la ESPAC en el MA*;

$_{MA}FA_{sub}$ = *Factor de Ajuste por el número de subdivisiones del SM del CNA en el MA*;

$_{MA}FA_{dis} = \frac{{}_{MA}dis_{G2}}{{}_{MA}dis_{G1}}$ = *Factor Ajuste por diseño de la muestra PEQUEÑA de la ESPAC en el MA, pero calculado en la muestra GRANDE DE PRIMERA FASE del CNA,*

donde:

(*i*) *En el do* min *io A de las provincias autorepresentadas, el* $_{MA}FA_{dis}$ *es*:

$$_{MA}FA_{dis} = \frac{{}_{MA}dis_{G2} = \text{No de SMs por réplica con la var iable de control} > 0 \text{ en un estrato del CNA}}{{}_{MA}dis_{G1} = \text{No de SMs por réplica de un estrato en la muestra del CNA en el MA}};$$

(*ii*) *En los do* min *ios B, C y D de las provincias autorepresentadas y en los do* min *ios A, B, C y D de las provincias no autorepresentadas, el* $_{MA}FA_{dis}$ *es*:

$$_{MA}FA_{cob} = \frac{{}_{MA}dis_{G2} = \text{No de SMs por estrato con la var iable de control} > 0 \text{ en el CNA}}{{}_{MA}dis_{G1} = \text{No de SMs por estrato en la muestra del CNA en el MA}},$$

donde:

'*Variable de Control*' = *Cultivos*(*permanentes* + *transitorios* + *barbecho*) + *Pastos* (*cultivados* + *naturales* + *páramos*) *ó* '*Superficie Total*' *de la UPA*;

$_{MA}FA_{cob}$ = *Factor de Ajuste por la falta de cobertura de la ESPAC en el MA*;

$_{MA}FA_{res}$ = *Factor de Ajuste por la falta de respuesta de la ESPAC en el MA.*

Lógicamente, un $_{MA}FOED_{ESPAC}$ y un $_{MA}FFED_{ESPAC}$ de un estrato se calculan de acuerdo a las características peculiares del diseño de la muestra PEQUEÑA DE SEGUNDA FASE del MA de la ESPAC en cada provincia. Así por ejemplo, en Guayas (una provincia auto-representada) el diseño de la muestra PEQUEÑA DE SEGUNDA FASE del MA contempla cuatro DEs (A, B, C, D): en el A la selección de los SMs se realizó independientemente por replicaciones, estratos y cantones, manteniendo el proceso de selección del CNA; en los otros tres DEs (B, C, D) se escogieron directamente los SMs en los estratos considerados para la ESPAC, sin tomar en cuenta las replicaciones ni los cantones.

En cambio, Cañar es una provincia no auto-representada con tres DEs (A, B y C), cuyas muestras de SMs del MA se seleccionaron solamente por estratos como en los DEs B, C y D de Guayas o sea sin respetar las replicaciones ni los cantones. Es necesario señalar que solamente Imbabura y Cotopaxi de las provincias no auto-representadas y Guayas y Manabí de las auto-representadas, tienen el DE ´D´.

Los cuadros 9 y 10 contiene una ilustración por DEs y estratos del cálculo de los $_{MA}FFED_{ESPAC}$ para el MA de las provincias de Guayas y Cañar, cuando los $_{MA}FA_{cob}$ y $_{MA}FA_{res}$ son iguales a la unidad o sea no se tienen que realizar ajuste por falta de cobertura y respuesta.

CUADRO 9

ESPAC: EJEMPLOS DE FACTORES FINALES DE EXPANSIÓN DIRECTA DEL MA DE GUAYAS, POR DOMINIOS Y ESTRATOS

UPM	SM	$_{MA}N_{CNA}$	$_{MA}n_{ESPAC(p)}$	$_{MA}FA_{sub}$			$_{MA}FA_{dis}$		$_{MA}FA_{cob}$	$_{MA}FA_{res}$	$_{MA}FFED_{ESPAC}$
				sub_1	sub_2	sub_3	$_{MA}dis_{G1}$	$_{MA}dis_{G1}$			
DOMINIO A: CANTON 4, ESTRATO 10											
116	1	479	12	1	1	1	8	7	1	1	34.9271
130	2	479	12	1	1	1	8	8	1	1	39.9167
DOMINIO A: CANTON 11, ESTRATO 20											
1731	1	47	4	1	1	1	3	2	1	1	7.8333
1868	5	47	4	2	1	1	3	2	1	1	15.6667
DOMINIO B: ESTRATO 40											
708	5	281	5	1	1	1	72	52	1	1	40.5889
940	3	281	5	1	1	1	72	52	1	1	40.5889
DOMINIO C: ESTRATO 10											
338	2	880	25	1	1	1	154	103	1	1	23.5429
508	2	880	25	2	1	1	154	103	1	1	47.0857
DOMINIO D: ESTRATO 51											
1014	1	405	5	1	1	1	59	35	1	1	48.0508
1417	3	405	5	2	1	1	59	35	1	1	96.1017

CUADRO 10

ESPAC: EJEMPLOS DE FACTORES FINALES DE EXPANSIÓN DIRECTA DEL MA DEL CAÑAR, POR DOMINIOS Y ESTRATOS

UPM	SM	$_{MA}N_{CNA}$	$_{MA}n_{ESPAC(p)}$	$_{MA}FA_{sub}$			$_{MA}FA_{dis}$		$_{MA}FA_{cob}$	$_{MA}FA_{res}$	$_{MA}FFED_{ESPAC}$
				sub_1	sub_2	Sub_3	$_{MA}dis_{G1}$	$_{MA}dis_{G1}$			
DOMINIO A: ESTRATO 10											
213	2	464	15	1	1	1	76	72	1	1	29.3053
255	2	464	15	2	1	1	76	72	1	1	58.6105
DOMINIO A: ESTRATO 20											
127	3	182	6	1	1	1	68	67	1	1	29.8873
224	4	182	6	2	2	1	68	67	1	1	119.5490
DOMINIO B: ESTRATO 40											
2	3	191	6	1	1	1	40	40	1	1	31.8383
163	1	191	6	1	1	1	40	40	1	1	31.8383
DOMINIO C: ESTRATO 10											
144	1	24	2	2	2	1	12	12	1	1	48.0000
144	8	24	2	2	4	1	12	12	1	1	96.0000
DOMINIO C: ESTRATO 30											
26	1	109	2	1	1	1	30	25	1	1	45.4167
178	4	109	2	1	1	1	30	25	1	1	45.4167

5.3 FACTOR ORIGINAL Y FINAL DE EXPANSION DIRECTA DE UN ESTRATO DE LA MUESTRA PEQUEÑA DE SEGUNDA FASE DEL ML

En general, en un estrato de la muestra PEQUEÑA DE SEGUNDA FASE del ML, la fracción de muestreo o probabilidad de selección de una UPA y su correspondiente Factor Original de Expansión Directa ($_{ML}FOED_{ESPAC}$) y Factor Final de Expansión Directa ($_{ML}FFED_{ESPAC}$) dentro de un estrato, responde a la fórmula siguiente:

$$_{ML}f_{ESPAC(f)} = {}_{ML}f_{CNA(G)} * {}_{ML}f_{ESPAC(p)} = \frac{{}_{ML}n_{CNA(G)}}{{}_{ML}N_{CNA}} * \frac{{}_{ML}n_{ESPAC(p)}}{{}_{ML}n_{CNA(G)}} = \frac{{}_{ML}N_{CNA}}{{}_{ML}N_{CNA}} * \frac{{}_{ML}n_{ESPAC(p)}}{{}_{ML}N_{CNA}} = \frac{{}_{ML}n_{ESPAC(p)}}{{}_{ML}N_{CNA}},$$

donde:

$_{ML}f_{ESPAC(f)}$ = *Fracción de muestreo final de la ESPAC en el ML*;

$_{ML}f_{CNA(G)} = 1$ = *Fracción de la muestra GRANDE DE PRIMERA FASE del CNA en el ML*;

$_{ML}f_{ESPAC(p)}$ = *Fracción de la muestra PEQUEÑA DE SEGUNDA FASE de la ESPAC en el ML*;

$_{ML}n_{CNA(G)} = {}_{ML}N_{CNA}$ = *No. de UPAs de la muestra GRANDE DE PRIMERA FASE del CNA en el ML*;

$_{ML}N_{CNA}$ = *No. de UPAs en la población del CNA en el ML*;

$_{ML}n_{ESPAC(p)}$ = *No. de UPAs de la muestra PEQUEÑA DE SEGUNDA FASE de la ESPAC en el ML.*

Por lo tan *to, el Factor Original de Expansión Directa en el ML es*:

$$_{ML}FOED_{ESPAC} = \frac{1}{{}_{ML}f_{ESPAC(f)}} = \frac{{}_{ML}N_{CNA}}{{}_{ML}n_{ESPAC(p)}},$$

donde:

$_{MA}FOED_{ESPAC}$ = *Factor Original de Expansión Directa de la ESPAC en el ML.*

Luego, el $_{ML}FOED_{ESPAC}$ *tiene que ser ajustado de la siguiente forma para obtener el Factor Final de Expansión Directa en el ML*:

$$_{ML}FFED_{ESPAC} = {}_{ML}FOED_{ESPAC} * {}_{ML}FA_{dis} * {}_{ML}FA_{cob} * {}_{ML}FA_{res},$$

donde:

$_{ML}FFED_{ESPAC}$ = *Factor Final de Expansión Directa de la ESPAC en el ML*;

$_{ML}FA_{dis} = \frac{{}_{ML}dis_{G2}}{{}_{ML}dis_{G1}}$ = *Factor Ajuste por diseño de la muestra PEQUEÑA de la ESPAC en el ML, pero calculado en la muestra GRANDE DE PRIMERA FASE del CNA,*

donde:

(*i*) *En los estratos* 1 *y* 2, *subestratos* (*grupos*) 1, 2 *y* 3 (*no incluye el* 4), *el* $_{ML}FA_{dis}$ *es*:

$$_{ML}FA_{dis} = \frac{{}_{ML}dis_{G2} = \text{No de UPAs por estrato con la variable de control} > 0 \text{ en el CNA}}{{}_{ML}dis_{G1} = \text{No de UPAs por subestrato (grupo) del CNA en el ML}},$$

con la 'Variable de Control' = *Cultivos* (*permanentes* + *transitorios* + *barbecho*) + *Pastos* (*cultivados* + *naturales* + *páramos*);

(*ii*) *En los estratos* 4 *al* 14 (sin *incluir* 11 *y* 15 *de camaroneras*), *el* $_{ML}FA_{dis}$ *es*:

$$_{ML}FA_{cob} = \frac{{}_{ML}dis_{G2} = \text{No de UPAs por estrato con la variable de control} > 0 \text{ en el CNA}}{{}_{ML}dis_{G1} = \text{No de UPAs por estrato del CNA en el ML}},$$

con la 'Variable de Control' = *Característica que define el estrato* (*flores, maracuyá, ..., mango, brócoli, palmito, aves, porcinos, etc.*);

$_{ML}FA_{cob}$ = *Factor de Ajuste por la falta de cobertura de la ESPAC en el ML*;

$_{ML}FA_{res}$ = *Factor de Ajuste por la falta de respuesta de la ESPAC en el ML.*

Los cuadros 11 y 12 contienen ejemplos por estratos del cálculo de los $_{ML}FFED_{ESPAC}$ para el MA de las provincias de Guayas y Cañar.

CUADRO 11

ESPAC: EJEMPLOS DE FACTORES FINALES DE EXPANSIÓN DIRECTA DEL ML DE GUAYAS, POR ESTRATOS

(Cantón) Estrato	Sub-Estrato (Grupo)	UPA No	$_{ML}N_{CNA}$	$_{ML}n_{ESPAC(p)}$	$_{ML}FA_{dis}$		$_{ML}FA_{cob}$	$_{ML}FA_{res}$	$_{ML}FFED_{ESPAC}$
					$_{ML}dis_{G1}$	$_{ML}dis_{G2}$			
(1) 1	1	9	389	268	389	380	1	1	1.4179
(1) 1	2	2	282	264	282	270	1	1	1.0227
(1) 1	3	3	99	97	99	97	1	1	1.0000
(1) 1	4	18	66	66	66	66	1	1	1.0000
(1) 2	1	5	162	15	162	45	1	1	9.6667
(1) 2	2	2	13	4	13	12	1	1	3.0000
(1) 2	3	9	4	2	4	4	1	1	2.0000
(17) 2	4	4	1	1	1	1	1	1	1.000
(1) 5	-	11	94	47	94	86	1	1	1.8298
(1) 10	-	31	15	6	15	10	1	1	1.6667

CUADRO 11

ESPAC: EJEMPLOS DE FACTORES FINALES DE EXPANSIÓN DIRECTA DEL ML DE CAÑAR, POR ESTRATOS

(Cantón) Estrato	Sub-Estrato (Grupo)	UPA No	$_{ML}N_{CNA}$	$_{ML}n_{ESPAC(p)}$	$_{ML}FA_{dis}$		$_{ML}FA_{cob}$	$_{ML}FA_{res}$	$_{ML}FFED_{ESPAC}$
					$_{ML}dis_{G1}$	$_{ML}dis_{G2}$			
(1) 1	1	7	88	72	88	88	1	1	1.2222
(1) 1	2	4	32	30	32	30	1	1	0.9999
(3) 1	3	2	8	8	8	8	1	1	1.0000
(4) 1	4	8	15	15	15	15	1	1	1.0000
(-) 2	1	-	-	-	-	-	-	-	-
(-) 2	2	-	-	-	-	-	-	-	-
(-) 2	3	-	-	-	-	-	-	-	-
(-) 2	4	-	-	-	-	-	-	-	-
(2) 3	-	1	4	4	4	4	1	1	1.0000
(4) 9	-	1	1	1	1	1	1	1	1.0000

5.4 EQUIVALENCIA DE LA SIMBOLOGIA USADA EN LOS FACTORES FINALES DE EXPANSION DIRECTA

El cuadro 13 indica las equivalencias de los símbolos utilizados en las fórmulas de los MA

´ ´ ,

^ ^ ,

1 1

:

^ ´ ´ var ;

^ ´ ´ var

;

;

var

;

.

$$\text{var}(\hat{\quad}) \quad (1 \quad \text{—}) \frac{1 \quad (\hat{\quad} \quad \hat{\bar{\quad}})^2}{1},$$

:

var(^) var ´ ´ var

;

^

$\hat{\bar{\quad}}$ $\frac{1}{\quad}$ *Media de la estimación´SAS´del TOTAL de una* var*iable*

en el estrato h ésimo del MA;

N Número de SMs en la población del estrato h ésimo del MA.

$$\text{var}({}_{ESPAC}\hat{Y}_r) = \frac{1}{n_p}(s^2_{\hat{Y}_p} + R_d^2 s^2_{\hat{Y}_p} - 2R_d s_{\hat{Y}_p\hat{X}_p}) + \frac{2R_d s_{\hat{Y}_p\hat{X}_p} - R_d^2 s^2_{\hat{X}_p}}{n_G},$$

donde:

$\text{var}({}_{ESPAC}\hat{Y}_r)$ = var *ianza de la estimación de RAZON del TOTAL de una* var *iable en un estrato de la ESPAC con MUESTREO EN DOS FASES* (*DOBLE*);

n_G = *Tamaño de la muestra* (*SMs o UPAs*) *GRANDE DE PRIMERA FASE en el CNA*;

n_p = *Tamaño de la muestra* (*SMs o UPAs*) *PEQUEÑA DE SEGUNDA FASE en la ESPAC*;

$s^2_{\hat{Y}_p}$ = var *ianza de* $({}_{ESPAC}\hat{Y}_p)$;

$s^2_{\hat{Y}_p}$ = var *ianza de* $({}_{CNA}\hat{X}_p)$;

$s_{\hat{Y}_p\hat{X}_p}$ = cov *arianza de* $({}_{ESPAC}\hat{Y}_p, {}_{CNA}\hat{X}_p)$.

6.2 PROCESO DE ESTIMACIÓN DE LA ESPAC MEDIANTE REGRESION

Este método de estimación del TOTAL de una variable y de su varianza en la ESPAC, tanto en un estrato del MA como del ML, responde a las siguientes fórmulas:

$${}_{ESPAC}\hat{Y}_{rgd} = {}_{ESPAC}\hat{Y}_p + B_d({}_{CNA}\hat{X}_G - {}_{CNA}\hat{X}_p),$$

donde:

${}_{ESPAC}\hat{Y}_{rgd}$ = *Estimación de REGRESION del TOTAL de una* var *iable en un estrato de la ESPAC con MUESTREO EN DOS FASES* (*DOBLE*);

$$B_d = \frac{\sum_{i=1}^{n_p}({}_{ESPAC}\hat{Y}_{ip} - {}_{ESPAC}\hat{\bar{Y}}_p)({}_{CNA}\hat{X}_{ip} - {}_{CNA}\hat{\bar{X}}_p)}{\sum_{i=1}^{n_p}({}_{CNA}\hat{X}_{ip} - {}_{CNA}\hat{\bar{X}}_p)^2}$$

= *Coeficiente de REGRESION de la* var *iable* $({}_{ESPAC}\hat{Y}_p)$ *con la* $({}_{CNA}\hat{X}_p)$ *en un estrato de la muestra PEQUEÑA DE SEGUNDA FASE en la ESPAC*,

donde:

${}_{ESPAC}\hat{\bar{Y}}_p$ = *Media de la estimación del TOTAL de la* var *iable* $({}_{ESPAC}\hat{Y}_p)$ *en un estrato de la muestra PEQUEÑA DE SEGUNDA FASE en la ESPAC*;

${}_{CNA}\hat{\bar{X}}_p$ = *Media de la estimación del TOTAL de la* var *iable* $({}_{CNA}\hat{X}_p)$ *en un estrato de la muestra PEQUEÑA DE SEGUNDA FASE*, *pero obtenida con datos del CNA.*

FORMULAS APROXIMADAS DE LA VARIANZA DE LA ESTIMACION POR REGRESION DEL TOTAL DE UNA VARIABLE EN UN ESTRATO DEL MA o ML EN LA ESPAC

PRIMERA FORMULA DE W. G. COCHRAN

$$\text{var}({}_{ESPAC}\hat{Y}^{C_1}_{rgd}) = \frac{s^2_{\hat{Y}_p\hat{X}_p}}{n_p} + \frac{s^2_{\hat{Y}_p} - s^2_{\hat{Y}_p\hat{X}_p}}{n_g} - \frac{s^2_{\hat{Y}_p}}{N},$$

donde:

$\text{var}({}_{ESPAC}\hat{Y}^{C_2}_{rgd\ rgd}) = \text{var}$ *ianza de la estimación por REGRESION del TOTAL de una* var *iable en un estrato de la ESPAC con MUESTREO EN DOS FASES*;

$$s^2_{\hat{Y}_p\hat{X}_p} = \frac{1}{n_p - 2}\left[\sum_{i=1}^{n_p}({}_{ESPAC}\hat{Y}_{IP} - {}_{ESPAC}\hat{\bar{Y}}_P)^2 - B_d^2\sum_{i=1}^{n_p}({}_{CNA}\hat{X}_{IP} - {}_{CNA}\hat{\bar{X}}_p)^2\right];$$

N = Número de SMs en la población del MA o número de UPAs en la población del ML.

SEGUNDA FORMULA DE W. G. COCHRAN

$$\text{var}({}_{ESPAC}\hat{Y}_{rgd}) = s^2_{\hat{Y}_p\hat{X}_p}\left[\frac{1}{n_p} + \frac{({}_{CNA}\hat{X}_G - {}_{CNA}\hat{X}_p)^2}{\sum_{i=1}^{n_p}({}_{CNA}\hat{X}_{IP} - {}_{CNA}\hat{\bar{X}}_p)^2}\right] + \frac{s^2_{\hat{Y}_p} - s^2_{\hat{Y}_p\hat{X}_p}}{n_g} - \frac{s^2_{\hat{Y}_p}}{N}.$$

FORMULA DE L. KISH

$$\text{var}({}_{ESPAC}\hat{Y}^{K}_{rgd}) = \frac{(s^2_{\hat{Y}_p} - \rho^2_{\hat{Y}_p\hat{X}_p})}{n_p - 2} + \frac{\rho^2_{\hat{Y}_p\hat{X}_p}s^2_{\hat{Y}_p}}{n_G},$$

donde:

$$\rho_{\hat{Y}_p\hat{X}_p} = \frac{\sum_{i=1}^{n_p}({}_{ESPAC}\hat{Y}_{ip} - {}_{ESPAC}\hat{\bar{Y}}_p)({}_{CNA}\hat{X}_{ip} - {}_{CNA}\hat{\bar{X}}_p)}{\sqrt{\sum_{i=e}^{n_p}({}_{ESPAC}\hat{Y}_{ip} - {}_{ESPAC}\hat{\bar{Y}}_p)^2\sum_{i=1}^{n_p}({}_{CNA}\hat{X}_{ip} - \hat{\bar{X}}_p)^2}} = \textit{coeficiente de correlación entre las}$$

las var *iables* $({}_{ESPAC}\hat{Y}_{ip}, {}_{CNA}\hat{X}_{ip})$;

$\rho^2_{\hat{Y}_p\hat{X}_p} =$ *coeficiente de* det *er* min *ación entre las las* var *iables* $({}_{ESPAC}\hat{Y}_{ip}, {}_{CNA}\hat{X}_{ip})$.

7 SOFTWARE UTILIZADO

En los procesos de selección de la muestra y estimación de los resultados, se han utilizado básicamente el "software" desarrollado con el Statistical Analysis System (SAS)", el cual se encuentra en el Departamento del SEAN/INEC y que se detalla a continuación:

- Programas ad-hoc desarrollados en SAS para el manejo de bases de datos del CNA/00 y las ESPAC/02-03 tanto para el MA como el ML.
- Rutina SAS "PROC SURVEY SELECT" para la selección de las muestras PEQUEÑAS DE SEGUNDA FASE del MA y ML.
- Rutina SAS "PROC TABULATE" para la producción de tablas con los resultados de las ESPAC/02-03.
- Rutina SAS "PROC SURVEYMEANS" para el proceso de estimación por Expansión Directa de los agregados de las variables y sus errores de muestreo.
- Programas ad-hoc desarrollados en SAS para el proceso de estimación por Expansión Directa de los agregados de las variables y sus errores de muestreo, a nivel provincial, regional y nacional.
- Programas ad-hoc desarrollados en SAS para el proceso de estimación por Regresión de los agregados de las variables y sus errores de muestreo.
- Rutina SAS "PROC SURVEYREG" para el proceso de estimación por Regresión de los agregados de las variables y sus errores de muestreo.
- Variedad de programas ad-hoc desarrollados en SAS para la adecuación de las bases de datos del CNA/00 y las ESPAC/02-03 que sirvan de insumos en la aplicación de las rutinas y programas SAS mencionados en las viñetas anteriores.
- Variedad de programas ad-hoc desarrollados en SAS para resolver los problemas diarios encontrados en el análisis de los resultados de las ESPAC/02-03.