



Diseño de la Muestra 1999

Encuesta de Hogares de Propósitos Múltiples (EHPM)

San José, Costa Rica
1999

INDICE

PRESENTACIÓN

I. ENCUESTA DE HOGARES DE PROPÓSITOS MÚLTIPLES	1
Objetivos de la Encuesta	2
II. DESCRIPCIÓN DEL MARCO MUESTRAL DE VIVIENDAS	2
2.1 Estratificación del Marco Muestral de Viviendas	3
2.2 Información del Marco Muestral de Viviendas	5
III. DISEÑO DE LA MUESTRA	8
3.1 Población y Cobertura	8
3.2 Dominios de Estudio	8
3.3 Unidades de Muestreo	9
3.4 Definición de la Muestra	10
3.4.1 Determinación del Tamaño de Muestra	10
3.4.2 Distribución de la Muestra	13
3.5 Selección de la Muestra	17
3.5.1 Selección de Unidades Primarias de Muestreo (UPM)	17
3.5.2 Selección de Unidades Secundarias de Muestreo (USM)	18
3.6 Métodos de Estimación	19
3.6.1 Factor de Expansión Básico	19
3.6.2 Ajustes al Factor de Expansión	21
3.6.2.1 Ajuste por Actualización.	21
3.6.2.2 Ajuste por Tasa de Respuesta	24
3.6.3 Factor Final de Expansión.	24
3.6.4 Estimadores de la Encuesta	25
3.6.5 Ajuste por Proyecciones de Población	27
3.6.6 Estimación de la media	33
3.6.7 Estimación del Total	34
3.7 Cálculo de Errores Muestrales	36
3.8 Rotación de la Muestra	39
3.8.1 Esquema de Rotación	39
3.8.2 Grupos de Rotación	41
3.8.3 Procedimiento y Control de Rotación	42
4. Publicación de Errores Muestrales	46

PRESENTACIÓN

La anterior muestra de la Encuesta de Hogares de Propósitos Múltiples, (EHPM), diseñada en 1986, fue seleccionada del Marco Muestral de Viviendas (MMV) construido con base en los Censos Nacionales de Población y Vivienda de 1984. Debido a la edad de la muestra y a la falta de información censal reciente, se consideró necesario evaluar el actual diseño e incorporar los ajustes necesarios para responder a los nuevos requerimientos de confiabilidad y oportunidad de información que demanda el país.

En la revisión del diseño anterior y la reformulación del nuevo diseño muestral se contó con la colaboración de la Oficina Internacional del Trabajo (OIT) y de la Comisión Económica para América Latina y el Caribe (CEPAL) a través de dos de sus expertos, el Dr. Anis Maitra y el Dr. Fernando Medina respectivamente, quienes trabajaron en forma conjunta con el grupo técnico y profesional del Área de Estadística y Censos.

El presente documento presenta los procedimientos metodológicos seguidos por el Proceso de Muestreo para el rediseño de la muestra de la Encuesta de Hogares de Propósitos Múltiples (EHPM). Se incluyen aspectos relacionados con la construcción del Marco Muestral de Viviendas, al diseño, a la selección, a la actualización y al procedimiento de expansión de la muestra. Antes, durante y después de la participación de los asesores se llevaron a cabo una serie de pruebas y análisis los cuales no se incluyen en el presente informe pero que están debidamente documentados por el personal del Proceso de Muestreo.

I. ENCUESTA DE HOGARES DE PROPÓSITOS MÚLTIPLES

En Costa Rica la serie de encuestas de hogares tiene una larga trayectoria, esta se inició en 1966 con el programa de la Encuesta Centroamericana de Hogares. Este programa permaneció hasta 1971, año en que fue discontinuado por dificultades financieras. Posteriormente, de 1976 a 1986 se realizó la Encuesta Nacional de Hogares, Empleo y Desempleo, en un esfuerzo conjunto del Ministerio de Trabajo y de Seguridad Social y de la Dirección General de Estadística y Censos.

La EHPM se realiza en julio de cada año desde 1987. Este programa es desarrollado conjuntamente por el Ministerio de Economía, Industria y Comercio, a través de la Dirección General de Estadística y Censos, el Ministerio de Trabajo y Seguridad Social, a través de la Dirección General de Planificación del Trabajo, y la Caja Costarricense del Seguro Social. Cuenta además con el apoyo de diferentes instituciones que solicitan módulos especiales.

La EHPM constituye una valiosa fuente de información sobre diversos aspectos relativos a los hogares y a las personas que conforman esos hogares. Tiene especial importancia el Módulo de Empleo, en el cual se investiga sobre el empleo, el desempleo, el subempleo y los ingresos, en general sobre aspectos referentes a la fuerza de trabajo y sus características. La encuesta sirve también como medio para investigar otros temas complementarios acerca de las características demográficas y socioeconómicas de los hogares y de sus residentes.

La encuesta se denomina de propósitos múltiples porque cumple la función adicional de permitir que instituciones interesadas soliciten la inclusión de temas o módulos especiales de investigación. Esta información, es de gran utilidad para que los organismos oficiales elaboren planes y proyectos que respondan a los intereses y necesidades de la población en general, para dar seguimiento a las políticas de desarrollo económico y social del país, y en la evaluación de dichas políticas. También, aporta la información básica indispensable para investigaciones y estudios sobre temas relativos a la población y el empleo, siendo la principal fuente de consulta durante los años en que no se realiza el Censo Nacional de Población.

Objetivos de la Encuesta

De acuerdo a lo anterior, se establecieron los siguientes objetivos para la Encuesta de Hogares de Propósitos Múltiples:

- a) Mantener un flujo continuo de estadísticas relacionadas con la fuerza de trabajo, el empleo, el desempleo, el subempleo y los ingresos, así como de otras variables socioeconómicas necesarias para el establecimiento de políticas y la formulación de planes orientados al desarrollo económico y social del país, y para la evaluación de sus efectos.
- b) Proveer información periódica, sistemática y oportuna en los períodos intercensales, referente a las variables mencionadas.
- c) Servir de fuente de información a instituciones gubernamentales, universitarias, o de investigación, interesadas en temas relativos a la población y el empleo, y en otros temas que se introduzcan periódicamente en la encuesta.

II. DESCRIPCIÓN DEL MARCO MUESTRAL DE VIVIENDAS

El Marco Muestral de Viviendas (MMV) está constituido por un conjunto de unidades, llamadas Unidades de Marco, del cual se van a seleccionar las muestras. Este marco corresponde al tipo llamado Marco de Áreas, que se caracteriza porque sus unidades son superficies geográficas con límites claramente definidos. Estas áreas o Unidades de Marco pueden ser ciudades, centros poblados o áreas de enumeración censal (segmentos censales), en las cuales los hogares se encuentran asociados a las respectivas viviendas individuales, lo que hace posible tener una medida muy aproximada de las probabilidades de selección de los hogares.

En el caso del MMV construido por el INEC, las Unidades de Marco están definidas por los segmentos censales, los cuales son áreas geográficas en que se divide cada distrito del país y que contienen en promedio 60 viviendas en las zonas urbanas y 40 viviendas en promedio en las zonas

rurales.

Además del conjunto de segmentos censales, el MMV se complementa con información adicional que permite la ubicación, ordenamiento y estratificación de cada uno de estos segmentos, así como cualquier otra información que pueda ser útil para el diseño y ejecución de una encuesta de hogares por muestreo.

El MMV se construyó con base en información proveniente de la actualización cartográfica realizada durante el período 1993-1997, como parte de las actividades preparatorias de los Censos Nacionales de Población y Vivienda. Se utilizó una nueva delimitación de segmentos censales de zonas urbanas y rurales, e incluye el conteo del número de viviendas que se hizo en cada uno de estos segmentos.

2.1 Estratificación del Marco Muestral de Viviendas

En la construcción de un marco muestral es importante incorporar algunas variables de utilidad para la estratificación de las unidades. Estas variables son de gran importancia en el diseño y la selección de muestras ya que permiten asegurar una mayor precisión en las estimaciones y por consiguiente mejorar la calidad de las mismas. En el caso de los marcos muestrales para la selección de viviendas y hogares, es común que se utilice como criterio de estratificación un indicador del nivel socioeconómico, principalmente si existe interés de investigar sectores sociales específicos. Generalmente, estas variables de estratificación se obtienen de información censal o de encuestas diseñadas para tal propósito.

Como no se dispuso de información censal actualizada que permitiera la construcción de un indicador del nivel socioeconómico de los hogares, se decidió incorporar al nuevo MMV una variable de estratificación construida con base en una calificación visual de las viviendas y su

entorno en los segmentos de la zona urbana y la periferia urbana¹. Esta estratificación se realizó en cinco niveles:

1. Nivel Bajo
2. Nivel Medio-Bajo
3. Nivel Medio
4. Nivel Medio-Alto
5. Nivel Alto

Se incluyen en el marco otras variables que pueden utilizarse como criterios de estratificación estas son: las medidas de tamaño (número de viviendas por unidad) y el grado de urbanización (urbano, periferia urbana, rural concentrado y rural disperso). Los criterios utilizados para la clasificación por grado de urbanización son los siguientes:

Urbano: se define como área urbana a los centros administrativos de los cantones del país, parte o todo el distrito primero, además de otras áreas adyacentes o cuadrantes de otros distritos demarcados a priori con criterio físico y funcional, tomando en cuenta elementos tangibles tales como cuadrantes, calles, aceras, luz eléctrica, servicios urbanos, etc.

Periferia Urbana: son aquellas zonas que pertenecen a las áreas localizadas entre el límite del cuadrante urbano y la poligonal envolvente del área urbana. La poligonal envolvente consiste en una línea imaginaria que encierra tanto el cuadrante o casco urbano como las áreas adyacentes al mismo, que aunque no se encuentran dentro del cuadrante urbano, presentan características similares a las de estas zonas. Estas áreas se consideran en proceso de transición de rural a urbano.

Rural Concentrado: son aquellos centros poblados no ubicados en el área urbana

¹/ Ver documento "Asignación del Nivel Socioeconómico". Proceso de Muestreo. 1998.

que reúnen las siguientes características:

- a) El suelo está ocupado predominantemente por actividades no agropecuarias.
- b) Tienen 50 o más viviendas agrupadas o contiguas, en general, las distancias no son más de 20 metros entre sí.
- c) Disponen de algún servicio de infraestructura como electricidad domiciliaria, agua potable o teléfono.
- d) Cuentan con algunos servicios como escuela, iglesia, centro de salud, puestos de salud, guardia rural, etc.
- e) Disponen de pulpería y parque de esparcimiento.
- f) Se identifican por un nombre determinado, que los distingue de otras áreas pobladas.

Rural Disperso: estas comprenden las áreas no contempladas en el punto anterior.

Para efecto de análisis se une el grado de urbanización urbano con la periferia urbana para formar la zona urbana y el grado de urbanización rural concentrado con el rural disperso para formar la zona rural.

2.2 Información del Marco Muestral de Viviendas

El Marco Muestral de Viviendas quedó finalmente conformado por las siguientes variables:

- a) **Identificación Cartografía:** es la información necesaria que permite la ubicación geográfica de cada una de las Unidades de Marco por provincia, cantón, distrito, y número de segmento censal.
- b) **Medidas de Tamaño:** corresponde al número de viviendas individuales que hay en cada segmento censal de acuerdo con la última actualización cartográfica. Permiten asignar probabilidades de selección a cada unidad del marco.
- c) **Estratificación:** dependiendo del uso del marco, las características de estratificación son utilizadas

para la selección de muestras independientes y por consiguiente mejorar la calidad de las estimaciones. Se incluyen en el marco como variables de estratificación el grado de urbanización, la Región de Planificación, así como el indicador del nivel socioeconómico.

La distribución de segmentos y viviendas del nuevo MMV por región de planificación y zona, y por región de planificación y nivel socioeconómico se presenta en los cuadros 1 y 2 respectivamente:

CUADRO 1
DISTRIBUCIÓN DE SEGMENTOS Y VIVIENDAS DEL NUEVO MARCO MUESTRAL DE VIVIENDAS
POR ZONA URBANA Y RURAL SEGÚN REGIÓN DE PLANIFICACIÓN

REGIÓN	TOTAL		ZONA			
			URBANA		RURAL	
	Segmentos	Viviendas	Segmentos	Viviendas	Segmentos	Viviendas
COSTA RICA	16.107	826.596	7.781	463.594	8.326	363.002
Central	9.669	525.076	6.105	363.854	3.564	161.222
Chorotega	1.419	67.722	387	23.593	1.032	44.129
Pacífico Central	992	50.135	461	26.359	531	23.776
Brunca	1.453	64.716	273	16.653	1.180	48.063
Huetar Atlántica	1.669	78.349	439	25.936	1.230	52.413
Huetar Norte	905	40.598	116	7.199	789	33.399

CUADRO 2
DISTRIBUCIÓN DE SEGMENTOS Y VIVIENDAS DEL NUEVO MARCO MUESTRAL DE VIVIENDAS
POR NIVEL SOCIOECONÓMICO SEGÚN REGIÓN DE PLANIFICACIÓN

REGIÓN	TOTAL	BAJO	MEDIO-BAJO	MEDIO	MEDIO-ALTO	ALTO	RURAL
COSTA RICA							
Segmentos	16.107	978	2.325	3.879	389	210	8.326
Viviendas	826.596	57.686	138.734	231.069	23.430	12.675	363.002
CENTRAL							
Segmentos	9.669	663	1.505	3.345	383	209	3564
Viviendas	525.076	39.156	89.553	199.482	23.049	12.614	161.222
CHOROTEGA							
Segmentos	1.419	4	263	120	0	0	1.032
Viviendas	67.722	255	15.890	7.448	0	0	44.129
PAC. CENTRAL							
Segmentos	992	14	188	256	3	0	531
Viviendas	50.135	783	10.958	14.425	193	0	23.776
BRUNCA							
Segmentos	1.453	15	196	61	0	1	1.180
Viviendas	64.716	898	11.919	3.775	0	61	48.063
HUETAR ATLÁNTICA							
Segmentos	1.669	273	124	42	0	0	1.230
Viviendas	78.349	16.095	7.385	2.456	0	0	52.413
HUETAR NORTE							
Segmentos	905	9	49	55	3	0	789
Viviendas	40.598	499	3.029	3.483	188	0	33.399

III. DISEÑO DE LA MUESTRA

En la definición del Diseño Muestral de la Encuesta de Hogares de Propósitos Múltiples (EHPM) se contó con la asesoría del consultor Dr. Anis Kumar Maitra para evaluar el actual diseño muestral y proponer ajustes de acuerdo a los requerimientos actuales de la encuesta. Uno de estos requerimientos es la inclusión de módulos sobre temas relacionados con el trabajo infante-juvenil. En esta etapa se contó además con la participación del Dr. Fernando Medina consultor regional de la CEPAL.

3.1 Población y Cobertura

La población que investiga la EHPM está definida como el conjunto de todas las viviendas individuales y sus ocupantes, **residentes permanentes** en esas viviendas, tanto del área urbana como del área rural de Costa Rica. Se excluye del estudio a la población residente en las viviendas colectivas (hoteles, hospitales, asilos, cárceles, conventos, etc.).

El ámbito geográfico del MMV en principio se extiende a todo el territorio nacional, sin embargo, por razones de costo se excluyó para la encuesta un grupo de segmentos censales por considerarse de baja densidad poblacional: segmentos de menos de 15 viviendas. Esto significa que la población muestreada es un subconjunto de la población objetivo lo que podría tener como resultado un sesgo en las estimaciones. Como los segmentos excluidos representan un 0.28% del total de segmentos del marco se consideró que el sesgo es poco significativo y por otro lado representa un ahorro importante de recursos para la encuesta.

3.2 Dominios de Estudio

Desde el punto de vista de terminología es útil entender los términos estrato y dominio, la distinción reside principalmente en el objetivo de cada uno. El estrato es una parte del marco y representa un subconjunto de elementos de la población agrupados de acuerdo a alguna característica en común; el considerar esta estratificación en el diseño y selección de la muestra garantiza un

aumento en la precisión de las estimaciones que brinda la encuesta. El dominio también es una subpoblación pero enfocada más al proceso de análisis y para los cuales se requiere que las estimaciones tengan una precisión conocida, generalmente se relacionan los dominios de estudio a áreas geográficas. Los estratos pueden coincidir con los dominios pero no necesariamente.

Para el rediseño de la muestra de la encuesta, se definen como subpoblaciones de estudio los dominios de estudio definidos para el programa de encuestas de 1987-1998, esto con el fin de mantener la comparabilidad de la serie de datos. Los dominios de estudio geográficos propuestos para la encuesta de 1999 son:

1. Nacional
2. Nacional Urbano y Nacional Rural
3. Región Central: Urbano y Rural
4. Resto de Regiones: Urbano y Rural

Además, de acuerdo al diseño muestral se podrán producir resultados para otras subpoblaciones de interés: hombres y mujeres, grupos de edad, menor trabajador, personas de la tercera edad, etc., y su nivel de desagregación dependerá fundamentalmente de la precisión con que se estime el dato y del tamaño de muestra resultante en cada caso.

3.3 Unidades de Muestreo

Las unidades de muestreo corresponden a las unidades que serán objeto de selección para la muestra. Es factible combinar o dividir las Unidades de Marco (segmentos censales) para crear unidades más apropiadas para ser utilizadas en diferentes encuestas. Además, la definición de las unidades de muestreo es diferente según las etapas de selección de la muestra, en el caso de la EHPM el diseño muestral incorpora dos etapas de selección y por tanto dos diferentes unidades de muestreo.

Para propósitos muestrales es importante distinguir entre la unidad de análisis y la unidad de muestreo.

- a) **Unidad de Análisis:** Es el hogar particular, entendiéndose por hogar a la persona sola o grupo de personas, con o sin vínculos familiares, que residen habitualmente en una vivienda individual, y que atienden sus necesidades vitales sujetas a un presupuesto común.
- b) **Unidad Última de Selección:** Es la vivienda individual definida como la estructura separada e independiente, ocupada o destinada a ser ocupada por uno o más hogares.
- c) **Unidades de Muestreo:** Las unidades de muestreo están definidas según la etapa de selección:
- **Unidad Primaria de Muestreo (UPM):** en la primera etapa de selección la unidad de muestreo corresponde al segmento censal del MMV.
 - **Unidad Secundaria de Muestreo (USM):** en la segunda etapa de selección la unidad de muestreo corresponde a la vivienda y esta es además la unidad última de selección.

3.4 Definición de la Muestra

La muestra de la EHPM corresponde a una muestra probabilística de áreas, estratificada y bietápica. Es probabilística porque en cada una de las etapas se utilizan métodos aleatorios de selección que permiten hacer inferencias a la población con base en la teoría de probabilidades. Es de áreas porque la probabilidad de selección está asociada a áreas geográficas dentro de los dominios de estudio. Es estratificada porque previamente a la selección la población se dividió en doce estratos -cada región de planificación dividida por zona urbana y zona rural- con la finalidad de mejorar su representatividad y aumentar así la eficiencia relativa del diseño muestral. Es bietápica porque en una primera etapa se selecciona segmentos censales (UPM) y en una segunda etapa viviendas (USM) dentro de las UPM seleccionadas en primera etapa.

3.4.1 Determinación del Tamaño de Muestra

Antes de determinar el tamaño de muestra del diseño revisado, el asesor tomó en consideración las recomendaciones que sobre el tópico de tamaños de muestras se dan en el manual

sobre Encuestas de Población Económicamente Activa de la OIT ².

Para determinar el tamaño de muestra se utiliza la fórmula del muestreo aleatorio simple ajustada por el efecto del diseño, a saber:

$$n = \frac{Z_{(\alpha)}^2 \sigma^2}{E^2} * D_{eff}$$

donde:

$Z_{(\alpha)}$: es el valor crítico de la curva normal estándar para un intervalo de confianza del $(1-\alpha)$ por ciento.

σ^2 : es la variancia de la variable de interés. En el caso de una proporción esta variancia es igual a $P(1-P)$.

E : es el margen de error propuesto como aceptable para estimar la variable de interés.

D_{eff} : es el valor del efecto del diseño obtenido de encuestas anteriores y representa el aumento en la variancia al utilizar diseños muestrales complejos en lugar de un muestreo aleatorio simple.

La encuesta más reciente (1997), da una estimación de la tasa de desempleo de alrededor del 5% para la mayoría de las regiones: $P=0.05$ y $\sigma^2 = P(1-P) = 0.05 * 0.95 = 0.0475$.

El valor de α es 0.05 y el valor para este nivel de significancia en la tabla de la distribución normal estándar es de $Z_{(\alpha)} = 1.96$. El margen de error propuesto es $E=0.01$, esto significa que si la tasa observada de desempleo es $p=0.05$, el verdadero valor de la tasa de desempleo será estimado con un 95% de confianza en el intervalo (0.04, 0.06).

La región Central tiene un mayor peso sobre las estimaciones totales y es para esta región que se dan estimaciones por zona urbana y rural, en el resto de las regiones se agregan los resultados de zona urbana y rural. Siguiendo el criterio del asesor de utilizar el efecto del diseño más alto para la

² / OIT. *Manual sobre Encuestas de la Población Económicamente Activa. Página 221.*

tasa de desempleo se tiene que el mayor valor del efecto del diseño se da en la zona urbana de la región Central cercano a 2.45.

De los datos de la reciente encuesta se ha encontrado que hay un promedio de 1.5 personas económicamente activas por vivienda y cerca de 1.1 hogares por vivienda en el país.

Con la información anterior se tiene:

$$\begin{aligned}n &= \frac{(1.96)^2 (0.0475)}{(0.01)^2} * 2.45 \\&= 1,824.75 * 2.45 \\&= 4,470.662 \\&= 4,471 \quad \text{personas económicamente activas (y dividido por 1.5)} \\&= 2,980.667 \quad \text{hogares (y dividido por 1.1).} \\&= 2,709.697 \cong 2710 \quad \text{viviendas.}\end{aligned}$$

El diseño muestral considera cuatro estratos importantes de estimación: Central Urbano, Central Rural, Resto del País Urbano y Resto del País Rural. El primer supuesto es que se requiere aproximadamente el mismo grado de precisión para las estimaciones de cada uno de estos cuatro subniveles nacionales. Sobre la base de este supuesto, el tamaño de muestra total requerido para el total país es de:

$$n = 4 * 2,710 = 10,840 \text{ viviendas.}$$

La tasa de respuesta de la EHPM se ha mantenido alrededor del 90%, así el tamaño de muestra efectivo que se espera obtener es de 9,756 viviendas, o 10,732 hogares. El asesor recomendó que el tamaño de muestra no debería ser incrementado en esta etapa para compensar la no respuesta y hacer los esfuerzos necesarios para que se mantenga a estos niveles.

3.4.2 Distribución de la Muestra

Costa Rica y otros países de la región se han unido al programa de erradicación del trabajo infantil, pasando a ser esta otra área de gran interés en la encuesta. Por este motivo en el diseño muestral se ha tomado en consideración esta variable para la afijación o distribución de la muestra de modo que se garantice para esta variable una buena estimación a nivel nacional y por zona urbana y rural.

Se ha observado que existen algunas diferencias apreciables en el error estándar y en las tasas de participación económica del menor trabajador entre la zona urbana y rural, por ello, la distribución de la muestra toma en consideración esta variación. En el cuadro 3 se presenta la afijación que se realizó de la muestra a partir de la distribución del número de viviendas en la zona urbana y rural obtenida de la actualización cartográfica y de los datos sobre trabajo infantil de la encuesta de julio de 1995.

CUADRO 3

AFIJACIÓN DE LA MUESTRA POR ZONA URBANA Y RURAL A PARTIR DEL ERROR ESTÁNDAR DE LA PARTICIPACIÓN ECONÓMICA DEL MENOR TRABAJADOR

Zona	Viviendas N_i	Error Estándar (σ)	$N_i * \sigma_i$	$\frac{N_i * \sigma_i}{\sum N_i * \sigma_i}$	n_i (5) * n
(1)	(2)	(3)	(4)	(5)	(6)
Total	826.596	-	5.221,986	1,000	10.840
Urbana	463.594	0,005	2.317,970	0,444	4.812
Rural	363.002	0,008	2.904,016	0,556	6.028

donde:

N_i : número de viviendas en el MMV en la zona i , para $i=1, 2$

σ : error estándar de la participación económica del menor trabajador.

n : muestra total de viviendas a seleccionar.

n_i : muestra de viviendas a seleccionar por zona urbana y rural.

Lo anterior indica que se necesitan 4,812 viviendas en la zona urbana y 6,028 viviendas en la zona rural para dar estimaciones bastantes confiables del trabajo infantil a nivel urbano y rural.

Para determinar el número de UPM a seleccionar en la muestra es necesario fijar el número de viviendas, unidades secundarias de muestreo (USM), que se van a seleccionar por UPM. Este número de USM depende del grado de homogeneidad que existe dentro de las UPM, entre más homogéneas menos USM se deben seleccionar y por tanto habrá más UPM con lo cual la muestra queda más dispersa en la población, en caso contrario, más heterogeneidad, se necesita más USM por UPM y la muestra queda concentrada en menos UPM en la población.

El grado de homogeneidad se refleja además en el efecto del diseño ($Deff$), mientras menor sea el grado de homogeneidad menor es el efecto del diseño bietápico y menor será el error de muestreo. En la EHPM se ha observado para la tasa de desempleo coeficientes de homogeneidad bastante bajos. Además, debido a los costos que implica un mayor número de UPM en la muestra (actualización, trabajo de campo, proceso, etc.), se recomendó, para determinar la muestra de UPM, mantener el promedio de 15 entrevistas por UPM que actualmente se realizan en la encuesta:

CUADRO 4
NÚMERO DE UPM A SELECCIONAR POR ZONA URBANA Y RURAL

Zona	n_i	USM por UPM	Número de UPM
(1)	(6)	(7)	(6) / (7)
Total	10.840		724
Urbana	4.812	15	322
Rural	6.028	15	402

Como las seis regiones de planificación son subpoblaciones importantes de estudio y se requieren estimaciones confiables a este nivel, se consideró importante distribuir la muestra de segmentos por zona urbana y rural entre las seis regiones de planificación. En este sentido, la distribución de la muestra de segmentos se realizó tomando en cuenta la variación relativa, coeficiente de variación, en la estimación del número de desempleados. Con el fin de suavizar las variaciones de un año a otro se tomó la variación relativa promedio de esta variable entre los años 1995, 1996 y 1997. La distribución final de la muestra de segmentos en cada una de las regiones por zona urbana y rural se presenta en los cuadros 5 y 6 respectivamente:

CUADRO 5
DISTRIBUCIÓN DE LA MUESTRA DE UPM URBANAS POR REGIÓN
SEGÚN LA VARIABILIDAD RELATIVA DEL NÚMERO DE DESEMPLEADOS

REGIÓN	UPM (N_i)	Variación Relativa (V_i)	N_i*V_i	$\frac{N_i * V_i}{\sum N_i * V_i}$	n_i
Total	7.786		1.116,30		322
Central	6.110	0,1149	702,04	0,6289	202
Chorotega	387	0,2478	95,90	0,0859	28
Pac. Central	461	0,2324	107,14	0,0960	31
Brunca	273	0,3039	82,96	0,0743	24
H. Atlántica	439	0,2024	88,85	0,0796	26
H. Norte	116	0,3397	39,41	0,0353	11

CUADRO 6
DISTRIBUCIÓN DE LA MUESTRA DE UPM RURALES POR REGIÓN
SEGÚN LA VARIABILIDAD RELATIVA DEL NÚMERO DE DESEMPLEADOS

REGIÓN	UPM (N_{hi})	Variación Relativa (V_{hi})	$N_{hi} * V_{hi}$	$\frac{N_{hi} * V_{hi}}{\sum N_{hi} * V_{hi}}$	n_{hi}
Total	8.321		1.325,53		402
Central	3.559	0,1249	444,52	0,3354	135
Chorotega	1.032	0,1732	178,74	0,1348	54
Pac. Central	531	0,2592	137,64	0,1038	42
Brunca	1.180	0,1860	219,48	0,1656	67
H. Atlántica	1.230	0,1749	215,13	0,1623	65
H. Norte	789	0,1648	130,03	0,0981	39

donde:

N_{hi} : número de UPM en el MMV en la i-ésima zona de la h-ésima región.

V_{hi} : es el coeficiente de variación promedio en la i-ésima zona de la h-ésima región.

n_{hi} : muestra de UPM a seleccionar en la i-ésima zona de la h-ésima región. .

Finalmente se realizaron pequeños ajustes de redondeo que dio como resultado la distribución final de la muestra de segmentos y el número esperado de viviendas para los cuatro estratos definidos anteriormente:

CUADRO 7
DISTRIBUCIÓN DE LA MUESTRA DE UPM POR ZONA SEGÚN REGIÓN

REGIÓN	Segmentos			Viviendas		
	Total	Urbano	Rural	Total	Urbano	Rural
Total	726	322	404	10.890	4.830	6.060
Central	334	200	134	5.010	3.000	2.010
Resto	392	122	270	5.880	1.830	4.050
Chorotega	82	28	54	1.230	420	810
Pac. Central	74	32	42	1.110	480	630
Brunca	92	24	68	1.380	360	1.020
H. Atlántica	92	26	66	1.380	390	990
H. Norte	52	12	40	780	180	600

3.5 Selección de la Muestra

Para la selección de la muestra, se dividió el MMV en doce submarcos, correspondientes a las seis regiones de planificación divididas por zona urbana y rural. El procedimiento de selección dentro de cada submarco se realizó dando un ordenamiento de las UPM por tamaño (número de viviendas) de menor a mayor.

3.5.1 Selección de Unidades Primarias de Muestreo (UPM)

La selección de la muestra de UPM sobre cada submarco se efectuó en forma sistemática después de un ordenamiento según el número de viviendas por UPM. El criterio de hacer una selección sistemática de una lista ordenada por tamaño, se debe a que el ordenamiento por tamaño de las unidades produce una estratificación natural, lo cual garantiza una mejor representatividad de la muestra bajo el supuesto de que las UPM con tamaño similar presentan características bastante homogéneas.

La selección sistemática da igual probabilidad de selección para cada Unidad del Marco, por tanto la probabilidad de selección de cada UPM está dada por $P_{hi} = 1 / N_{hi}$, donde P_{hi} es la probabilidad de selección y N_{hi} es el número de UPM en la i -ésima zona de la h -ésima región (para $i=1, 2$ y $h=1, 2, 3, 4, 5, 6$).

3.5.2 Selección de Unidades Secundarias de Muestreo (USM)

Después de la selección de la muestra de UPM en la primera etapa, se procedió al levantamiento del marco de segunda etapa. Este marco corresponde a un Marco de Lista llamado Registro de Edificios y Viviendas (REV) que tiene como objetivo identificar, registrar y describir el uso o destino de los diferentes tipos de edificios y viviendas, así como levantar el listado de jefes de hogar para aquellas estructuras dedicadas a viviendas dentro de cada una de las UPM seleccionadas en la primera etapa. Una vez digitada y procesada esta información se pasó a la etapa de selección de las viviendas o USM.

Antes de seleccionar las USM se tomó la decisión de que la muestra fuera autoponderada, esto significa que todas las viviendas dentro de una misma región de planificación y zona tienen la misma fracción de muestreo. En este sentido, para mantener la autoponderación, se fijó en un cuarto la fracción de muestreo en segunda etapa para los segmentos urbanos y una fracción de muestreo de un tercio en la segunda etapa para los segmentos rurales.

La selección de USM se efectúa de forma sistemática en conglomerados de tres viviendas con probabilidades iguales. Aunque la selección de viviendas en conglomerados compactos tiende a aumentar el grado de homogeneidad dentro de las UPM, se espera que este aumento no sea muy importante y se prefiere a una selección aleatoria o sistemática de viviendas individuales por las ventajas que implica en cuanto al recorrido dentro de las UPM y a una mejor labor de supervisión en el trabajo de campo. Debido a que se selecciona una fracción fija en segunda etapa, el número final de USM dependerá del tamaño de las UPM.

3.6 Métodos de Estimación

Antes de proceder al paso de estimación, se asume que para una zona y región dada, el número de viviendas listadas ha sido ajustado con la información muestral de los cambios encontrados en el campo en las UPM seleccionadas. Por ser hecho con datos muestrales, este ajuste tiene un componente aleatorio y se espera sea muy pequeño año tras año comparado con el número inicial de viviendas listadas. Para propósitos prácticos, se asume que este ajuste no está sujeto a variación muestral.

En el proceso de estimación hay que considerar que la selección se realizó independientemente para cada submarco muestral, uno por cada región y zona, doce en total, por tanto se tienen doce ponderaciones o factores de expansión con los cuales expandir los resultados de la EHPM.

3.6.1 Factor de Expansión Básico

El factor de expansión básico para cada vivienda seleccionada es determinado por el diseño de la muestra y en un diseño bietápico, como es el caso de la EHPM, este factor es calculado al multiplicar el inverso de las probabilidades de selección de cada etapa:

$$F_{hij} = 1 / f_{hi1} * 1 / f_{hij2}$$

donde

F_{hij} : es el factor de expansión básico para la h-ésima región, la i-ésima zona en la j-ésima UPM.

f_{hi1} : fracción de muestreo en la primera etapa de selección para la h-ésima región y la i-ésima zona.

f_{hij2} : fracción de muestreo en la segunda etapa de selección para la h-ésima región, la i-ésima zona en la j-ésima UPM.

En primera etapa las UPM se seleccionaron sin reposición y con probabilidades iguales por lo que la fracción de muestreo es:

$$f_{hi\ 1} = n_{hi} / N_{hi}$$

donde :

n_{hi} : es el número de UPM seleccionadas en la h-ésima región y la i-ésima zona.

N_{hi} : es el número total de UPM que hay en la h-ésima región y la i-ésima zona.

En segunda etapa la fracción de muestreo dentro de cada UPM se supone siempre que se realiza sin reposición y con probabilidades iguales:

$$f_{hi\ 2} = m_{hij} / M_{hij}$$

donde :

m_{hij} : es el número de USM seleccionadas en la h-ésima región, en la i-ésima zona y la j-ésima UPM.

M_{hij} : es el número total de USM que hay en la h-ésima región y la i-ésima zona y la j-ésima UPM.

Como la muestra se propone autoponderada la fracción final de muestreo debe ser constante:

$$f_{hi} = f_{hi\ 1} * f_{hi\ 2} = \text{constante}$$

Para que esta condición se cumpla se propone que la fracción de muestreo de segunda etapa sea una fracción fija:

$f_{h\ 12} = 1/4$ para la zona urbana (1) en la h-ésima región.

$f_{h\ 22} = 1/3$ para la zona rural (2) en la h-ésima región.

El factor de expansión está dado por:

$$F_{h1} = \frac{N_{h1}}{n_{h1}} * \frac{M_{h1j}}{(1/4) M_{h1j}} = \frac{M_{h1}}{m_{h1}} \quad \text{urbano}$$

$$F_{h2} = \frac{N_{h2}}{n_{h2}} * \frac{M_{h2j}}{(1/3) M_{h2j}} = \frac{M_{h2}}{m_{h2}} \quad \text{rural}$$

3.6.2 Ajustes al Factor de Expansión

Es común que se presenten diferencias entre lo planificado en oficina y lo observado en el campo, debido a esto el factor de expansión básico debe ser ajustado para mantener las probabilidades iniciales de selección. Estas diferencias que afectan las probabilidades de selección se deben principalmente a dos aspectos: a diferencias en el tamaño de las UPM en cuanto al número de viviendas (actualización), y a diferencias en cuanto al número de viviendas seleccionadas y viviendas realmente encuestadas (tasa de respuesta).

3.6.2.1 Ajuste por Actualización.

La población del país no permanece estática, experimenta cambios tanto en su estructura demográfica como en sus características socioeconómicas, pero la población a la que se refiere el MMV se refiere a la población en un momento específico, no refleja esos cambios que se dan año a año. Esto implica, otro que las probabilidades iniciales de selección de la muestra se van desactualizando y se va perdiendo confiabilidad en las estimaciones que brinda la encuesta. Para solventar este problema se hace necesario actualizar el dato poblacional del marco de forma tal que se vaya incorporando ese dinamismo poblacional.

La actualización del dato poblacional del MMV incluye, por una parte, una continua revisión de los

tamaños de las UPM con la consiguiente alteración de los factores de expansión, por otra, una nueva selección de las USM incorporando así los cambios sociodemográficos que se dan en la UPM. En este sentido el asesor de la OIT planteó dos formas de actualización de las UPM para incorporar estos cambios:

- **Total de la UPM:** el supervisor debe registrar el número de viviendas no listadas, ya sea por resultado de una mala actualización cartográfica o por haber sido construidas después de la actualización.
- **Dentro de las USM:** el supervisor debe registrar el número de viviendas no listadas en los conglomerados seleccionados dentro de la UPM. En este caso, una estimación del número de nuevas viviendas en la totalidad de la UPM se obtiene multiplicando el número de nuevas viviendas por el inverso de la fracción de muestreo dentro de las UPM.

Ambas formas de actualización pueden utilizarse en diferentes UPM, pero si es importante indicar cuál procedimiento se utilizó para el ajuste final del número de nuevas viviendas. Por ejemplo, si para un estrato se tienen N_{hi} UPM en el marco y n_{hi} UPM en la muestra, y de las n_{hi} UPM se actualizan n_{hi1} completamente y las restantes n_{hi2} se actualizan sólo en los conglomerados seleccionados, el número estimado de nuevas viviendas (m'_{hi}) en las n_{hi} UPM de la muestra es:

$$m'_{hi} = m'_{hi1} + (m'_{hi2} * F_{hi2})$$

donde:

m'_{hi} : número de nuevas viviendas en la muestra de la h-ésima región y de la i-ésima zona.

m'_{hi1} : número de nuevas viviendas encontradas en las n_1 UPM actualizadas completamente en la h-ésima región y en la i-ésima zona. .

m'_{hi2} : número de nuevas viviendas encontradas en las n_2 UPM actualizadas solo en los conglomerados seleccionados de la h-ésima región y de la i-ésima zona.

F_{hi2} : inverso de la fracción de muestreo de segunda etapa para la h-ésima región y la i-ésima zona.

El número de nuevas viviendas estimado en el marco está dado por:

$$M'_{hi} = \frac{N_{hi}}{n_{hi}} * m'_{hi}$$

con lo que el total actualizado de viviendas en el marco es:

$$M_{hi}^* = M_{hi} + M'_{hi}$$

donde

M'_{hi} : es el número estimado de nuevas viviendas en el marco para la h-ésima región y la i-ésima zona.

N_{hi} : es el número de segmentos en el marco en la h-ésima región y la i-ésima zona.

n_{hi} : es el número de segmentos en la muestra en la h-ésima región y la i-ésima zona.

m'_{hi} : es el número estimado de nuevas viviendas en la muestra en la h-ésima región y la i-ésima zona.

M_{hi} : es el número actual de viviendas registradas en el marco para la h-ésima región y la i-ésima zona.

M_{hi}^* : es el número actualizado de viviendas en el marco para la h-ésima región y la i-ésima zona.

M_{hi}^*/M_{hi} : es el factor de ajuste por el cual es multiplicado el factor de expansión inicial.

El factor básico de expansión es ajustado según los resultados anteriores. Es importante recordar que el factor de expansión puede ser determinado solo después del trabajo de campo, porque tanto M_{hi}^* , M'_{hi} y m'_{hi} dependen del número de viviendas no listadas estimadas en la muestra.

3.6.2.2 Ajuste por Tasa de Respuesta

Es necesario ajustar los factores básicos de expansión para tomar en consideración las viviendas seleccionadas pero no encuestadas con la finalidad de eliminar el sesgo debido a la no entrevista. Este ajuste a los factores básicos de expansión se efectúa a nivel de zona dentro de cada región de planificación:

$$\mathbf{TRV}_{hi} = \frac{\mathbf{V}_{hi(e)}}{\mathbf{V}_{hi(e)} + \mathbf{V}_{hi(ne)}} = \frac{\mathbf{V}_{hi(e)}}{\mathbf{V}_{hi}}$$

donde

$\mathbf{V}_{hi(e)}$: total de viviendas seleccionadas y entrevistadas en la h-ésima región y la i-ésima zona.

$\mathbf{V}_{hi(ne)}$: total de viviendas seleccionadas y no encuestadas en la h-ésima región y la i-ésima zona.

\mathbf{V}_{hi} : total de viviendas seleccionadas en la h-ésima región y la i-ésima zona.

Los factores que influyen en la no-respuesta y que impiden la realización de las entrevistas merecen un análisis adicional buscando alternativas para su tratamiento y posibles ajustes.

3.6.3 Factor Final de Expansión.

El factor final de expansión para cada vivienda se obtiene al multiplicar el factor de expansión básico (F) por el crecimiento del marco y por el inverso del factor de ajuste por vivienda no encuestada (1/TRV)

$$\mathbf{F}_{hi}^* = \frac{1}{\mathbf{f}_{hi1}} * \frac{1}{\mathbf{f}_{hi2}} * \frac{\mathbf{M}_{hi}^*}{\mathbf{M}_{hi}} * \frac{1}{\mathbf{TRV}_{hi}} = \mathbf{F}_{hi} * \frac{\mathbf{M}_{hi}^*}{\mathbf{M}_{hi}} * \frac{1}{\mathbf{TRV}_{hi}} = \frac{\mathbf{M}_{hi}^*}{\mathbf{m}_{hi}}$$

donde:

- $f_{hi1} : n_{hi} / N_{hi}$ es la fracción de muestreo en primera etapa.
- $f_{hi2} : 1/4$ es la fracción de muestreo en segunda etapa en la zona urbana.
- $f_{hi2} : 1/3$ es la fracción de muestreo en segunda etapa en la zona rural.
- $TRV_{hi} :$ es la tasa de respuesta de viviendas.
- $M_{hi}^* :$ es el total ajustado de viviendas en la h-ésima región y la i-ésima zona.
- $M_{hi} :$ es el número de viviendas en el marco en la h-ésima región y la i-ésima zona.
- $m_{hi} :$ es el número de viviendas seleccionadas en la h-ésima región y la i-ésima zona.

Las m_{hi} es el número actual de viviendas que se espera proporcionen los datos. Este número debe ser conocido tan rápido como las UPM sean seleccionadas y no se modifica en forma alguna por el número de nuevas viviendas ni por las desaparecidas recientemente. Es importante recordar que este número está sujeto a variación muestral porque depende del tamaño de las UPM.

El factor de expansión es determinado solo después del trabajo de campo, porque M_{hi}^* depende del número de viviendas no listadas estimadas de la muestra, por efecto práctico M_{hi}^* se supone constante. Hay un factor de expansión para cada estrato, doce factores en total.

3.6.4 Estimadores de la Encuesta

La metodología para la estimación de las características investigadas en la EHPM involucra el uso del peso o factor de expansión para cada vivienda de la muestra, este factor es multiplicado a todos los datos que provienen de cada una de estas viviendas. Los estimadores así propuestos se denominan estimadores naturales porque los resultados de la muestra se expanden con información proveniente de la misma muestra y del diseño muestral utilizado.

La estimación del total para un estrato dado se obtiene al multiplicar el factor de expansión

propuesto en el punto anterior por el total muestral de la característica investigada en ese mismo estrato:

$$\hat{Y}_{hi} = \frac{N_{hi}}{n_{hi}} \sum_{i=1}^{n_{hi}} \hat{Y}_{hij} = \frac{N_{hi}}{n_{hi}} \sum_{i=1}^{n_{hi}} \frac{1}{f_{hi}} y_{hij} = \frac{M_{hi}^*}{m_{hi}} y_{hi} = F_{hi} Y_{hi}$$

donde

\hat{Y}_{hi} : es el total estimado de la variable "Y" en la i-ésima zona de la h-ésima región.

\hat{Y}_{hij} : es el total estimado de la variable "Y" en la j-ésima UPM.

$y_{hij} = \sum y_{hijk}$: es la suma del valor de la variable "Y" para la k-ésima vivienda.

y_{hi} : es el total de la variable "Y" en la i-ésima zona de la h-ésima región.

En una región el total estimado se obtiene al agregar las estimaciones obtenidas por zona urbana y rural:

$$\begin{aligned} \hat{Y}_h &= \frac{M_{h1}^*}{m_{h1}} y_{h1} + \frac{M_{h2}^*}{m_{h2}} y_{h2} \\ &= F_{h1} y_{h1} + F_{h2} y_{h2} \\ &= \sum_{i=1}^2 \frac{M_{hi}^*}{m_{hi}} y_{hi} = \sum_{i=1}^2 F_{hi} y_{hi} \end{aligned}$$

El total nacional se obtiene agregando las estimaciones individuales por región de planificación:

$$\hat{Y} = \sum_{h=1}^6 \hat{Y}_h = \sum_{h=1}^6 \sum_{i=1}^2 \frac{M_{hi}^*}{m_{hi}} y_{hi} = \sum_{h=1}^6 \sum_{i=1}^2 F_{hi} y_{hi}$$

Todas estos estimadores son de razón porque tanto el numerador (y_{hi}), como el denominador (m_{hi}) están sujetos a variación muestral, son variables aleatorias. Solo por un supuesto a priori M_{hi}^* se considera constante. El estimador de la razón entre dos variables Y y X presenta un sesgo que es despreciable en muestras grandes, cuando el coeficiente de variación de Y y X tome valores menores

al 10%. En general, para cualquier muestra aleatoria o no, el estimador de razón es el mejor estimador lineal insesgado si se cumplen las condiciones siguientes³:

- a) La línea de regresión de y_i sobre x_i es una recta que pasa por el origen.
- b) La variancia de y_i alrededor de esta línea es proporcional a x_i .
- c) Es importante destacar que esta segunda opción exige que la variancia de y_i para un x_i fijo aumente proporcionalmente a medida que lo hace y_i .

La estimación de parámetros tales como la tasa de ocupación, la tasa de desocupación, la tasa de participación u otras, en los distintos estratos y niveles de estimación, se calculan como cocientes de estimaciones separadas para el total de ocupados, desocupados, integrantes de la fuerza de trabajo, o la variable que corresponda usar. La estimación de estas tasas se obtienen también por el método de razón:

$$\hat{R}_{hi} = \frac{\hat{Y}_{hi}}{\hat{X}_{hi}} = \frac{\frac{M_{hi}^*}{m_{hi}} y_{hi}}{\frac{M_{hi}^*}{m_{hi}} x_{hi}} = \frac{y_{hi}}{x_{hi}}$$

3.6.5 Ajuste por Proyecciones de Población

La desactualización del marco muestral afecta las probabilidades iniciales de selección de la muestra y aunque los estimadores propuestos continúan siendo insesgados sus variancias aumentan considerablemente. Una solución al problema anterior ha sido el uso del marco de nuevas construcciones que va incorporando las nuevas viviendas encontradas en el campo a este marco. Con este método, la distribución de las UPM que permanecen en el marco original no es muy diferente a la utilizada inicialmente en la selección, lo cual permite mantener la precisión del marco.

³/ Cochran, William. *Técnicas de Muestreo*. Páginas 204,205,206.

Otra solución, es efectuar una actualización completa de las UPM seleccionadas que permita comparar el tamaño actual con el usado en la selección, en caso de que se detecten discrepancias bastante acentuadas se debe de tomar con cautela la información proveniente de la encuesta pues se podría estar en niveles de estimación bastante alejados del marco original.

Los estimadores naturales se ven afectados por los problemas identificados en el marco, otra forma de solucionar estos problemas es el uso de estimadores que incorporan fuentes externas a la muestra como una compensación a los problemas detectados: envejecimiento del marco muestral, desconocimiento del total de viviendas en la población, falta de capacidad de la muestra para captar una variación en el tamaño de las UPM o desconocimiento de la precisión de los listados de viviendas.

Los estimadores naturales incorporan en el factor de expansión las características propias del diseño muestral, pero si además se dispone de información exógena a la muestra esta puede incorporarse para mejorar la precisión de las estimaciones. Esta información se refiere a proyecciones intercensales de población, tanto para la población total como para subpoblaciones de acuerdo a características como sexo, grupos de edad y otros criterios, en los distintos estratos y niveles de estimación.

La incorporación de información exógena presenta la ventaja de proporcionar estimaciones confiables basadas en una evolución consistente y suavizada de la población a través del tiempo, y adicionalmente, mejoran la precisión de las estimaciones muestrales a partir de estimadores de razón, que pueden presentar importantes ganancias en términos de una menor variabilidad en el proceso de estimación.

El uso de información proveniente de proyecciones de población no debe ser visto como obligatorio si se cree que la calidad de los datos muestrales obtenidos a través de estimaciones naturales pueden ofrecer una mayor confiabilidad que los proyectados, en cuyo caso el ajuste por sexo y edad no debería efectuarse. Si se hace necesario, la incorporación de proyecciones de

población debería apenas ajustar las desviaciones en las estimaciones por sexo y edad, y este ajuste debe ser equilibrado a fin de no exagerar en una innecesaria artificialidad que anulen las características propias de expansión de la muestra.

Por otro lado, a medida que los supuestos en que se basan las proyecciones se alejan de la realidad, los valores proyectados pueden estar bastante alejados de los valores poblacionales reales, especialmente para niveles geográficos menores y entre más se distancien del último censo disponible.

Los ajustes sobre proyecciones por sexo y edad pueden practicarse a diferentes niveles de agregación (estrato, región, país), una vez decidido ese nivel los factores de ajuste se aplican a todas las unidades que ese nivel agrupa. Las estimaciones para totales por estrato, se pueden extender para los niveles regionales y nacionales a través de sucesivas agregaciones.

La estimación de la expansión muestral y la proyección independiente deben estar próximas, en caso contrario, si se presentan grandes diferencias, pueden dar indicios de una cobertura insuficiente o de error en las propias proyecciones.

De acuerdo a los puntos mencionados anteriormente, las encuestas por muestreo hacen uso de estimadores de razón para mantener la coherencia con los datos provenientes de las proyecciones de población, utilizando estas como una cota superior a nivel total y para mantener la estructura interna de población.

No obstante lo anterior, y dado que el programa de EHPM no dispone de un marco de nuevas construcciones, el diseño muestral de la encuesta contempla un ajuste según las proyecciones oficiales de población⁴, lo cual permite compensar las variaciones del marco muestral producto del crecimiento poblacional.

4/ Centro Latinoamericano de Demografía, CELADE. Estimaciones y Proyecciones Regionales de Población. San José. 1988.

Para este ajuste, en un inicio se consideró utilizar las nuevas proyecciones de población elaboradas por el Programa Centroamericano de Población y el Área de Estadística y Censos⁵, pero a las puertas de los próximos Censos Nacionales de Población y Vivienda en el 2000, cuya información servirá de insumo para la actualización de los supuestos en que se basan las proyecciones y para efectuar los ajustes necesarios en el diseño de la encuesta, se decidió hacer la menor cantidad de cambios en el presente rediseño y de esta forma mantener en lo posible la comparabilidad de serie de datos de la encuesta.

El ajuste por proyecciones toma en consideración las diferencias existentes en las variables demográficas y laborales que investiga la encuesta entre regiones de planificación. El procedimiento es el siguiente:

a) Tasa de crecimiento de las proyecciones de población

La tasa de crecimiento de las proyecciones de población es la razón de la población estimada para el año t entre la población estimada del año anterior.

$$T_{Ph} = \frac{y_{Ph}(t)}{y_{Ph}(t-1)}$$

donde:

$h_{ipa}(t)$: población estimada por proyecciones de población (P) en el año t (o t-1), para la h-ésima región.

T_{Ph} : tasa de crecimiento de las proyecciones de población la h-ésima región.

Esta tasa de crecimiento se aplica a la población estimada el año anterior de la encuesta de hogares (t-1), para obtener una estimación de la población que se espera en el presente año.

5/ Programa Centroamericano de Población, PCP; Área de Estadística y Censos. Estimaciones y

$$\mathbf{y}'_{PHh} = \mathbf{y}_{Hh}(\mathbf{t} - 1) * \mathbf{T}_{Ph}$$

donde:

$\mathbf{y}_{Hh}(\mathbf{t}-1)$: población estimada de la EHPM (H) en el año anterior para la h-ésima región.

$\mathbf{y}'_{PHh}(\mathbf{t})$: población estimada que se espera obtener en el año t de la encuesta según tasa de crecimiento de proyecciones de población para la h-ésima región.

b) Estimación natural de la encuesta

La población estimada de la encuesta para el año t sin considerar el ajuste por proyecciones de población se obtiene multiplicando el dato muestral de este año por el factor final de expansión.

$$\mathbf{y}'_{Hh}(\mathbf{t}) = \sum_{i=1}^2 \mathbf{n}_{hi}(\mathbf{t}) * \mathbf{F}_{hi}^*$$

donde:

$\mathbf{n}_{hi}(\mathbf{t})$: tamaño de muestra en el año t para la h-ésima región y en la i-ésima zona.

\mathbf{F}_{hi}^* : factor final de expansión para la h-ésima región y la i-ésima zona.

$\mathbf{y}'_{Hh}(\mathbf{t})$: población estimada de la encuesta sin ajuste de proyección para la h-ésima región.

c) Factor de Ajuste por proyecciones

El factor de ajuste se obtiene al dividir para cada región la población que se espera obtener en el presente año, según la tasa de crecimiento de las proyecciones, entre la estimación natural de la encuesta en el presente año.

$$T_h = \frac{y'_{PHh}(t)}{y'_{Hh}(t)}$$

donde:

T_h : tasa de ajuste del factor final de expansión de la encuesta por proyecciones de población para la h-ésima región.

CUADRO 8
CÁLCULO DEL AJUSTE POR PROYECCIONES DE POBLACIÓN

Región	EHPM 1998 (1)	PROYECCIÓN 1998 (2)	PROYECCIÓN 1999 (3)	TASA (4) (3)-(2)	EHPM 1/ 1999 (5) (1)*(4)	EHPM 2/ 1999 (6)	AJUSTE (7) (5)/(6)
Costa Rica	3,340,909	3,594,827	3,671,959		3,412,614	3,293,282	1.03623
Central	2,105,573	2,309,232	2,355,668	1.02011	2,147,913	2,085,524	1.02992
Urbano	1,199,329	n.a.	n.a.	n.a.	1,223,446	1,199,635	1.01985
Rural	906,244	n.a.	n.a.	n.a.	924,467	885,889	1.04355
Chorotega	253,188	263,945	265,991	1.00775	255,151	270,830	0.94211
Pac. Central	178,033	171,517	173,443	1.01123	180,032	193,214	0.93178
Brunca	328,991	344,337	353,061	1.02534	337,326	284,012	1.18772
H. Atlántica	306,123	307,686	319,758	1.03924	318,134	301,600	1.05482
H. Norte	169,001	198,110	204,037	1.02992	174,057	158,102	1.10092

1/ Población esperada para 1999 según crecimiento de proyecciones.

2/ Estimación natural de población para 1999 que da la encuesta.

d) Ajuste del Factor de Expansión Final

El ajuste al factor de expansión resulta de multiplicar el factor de expansión inicial por la tasa de ajuste obtenida por proyecciones de población. Aunque el factor de expansión se calculó por zona dentro de cada región de planificación, se aplica la misma tasa de ajuste regional por zona, con excepción de la región Central que se efectúa por zona:

$$F_{Phi} = F_{hi} * T_h$$

donde:

F_{hi}^* : factor final de expansión la h-ésima región y la i-ésima zona.

F_{Phi}^* : factor final de expansión ajustado por proyecciones de población para la h-ésima región y la i-ésima zona

e) Aplicación del Factor Final de Expansión ajustado por Proyecciones

El factor de expansión final ajustado por proyecciones se aplica a todas las unidades (viviendas o personas) dentro de cada zona y región de planificación, luego se van agregando los datos para obtener las estimaciones para cada subpoblación de interés, como se verá más adelante.

3.6.6 Estimación de la media

La estimación de una media está dada por la razón de dos variables X y Y. Si el ajuste por proyecciones se hace para subpoblaciones específicas, por ejemplo: población masculina y femenina, población menor de 12 años y población de 12 años y más, el estimador de la media está dado por:

$$\begin{aligned}\bar{Y}_{hi(s)} &= \frac{\hat{Y}_{hi(s)}}{\hat{N}_{hi(s)}} \bar{N}_{hi(s)} = \frac{\hat{Y}_{hi(s)}}{\hat{X}_{hi(s)}} \bar{X}_{hi(s)} \\ &= \hat{R}_{hi(s)} \bar{X}_{hi(s)}\end{aligned}$$

donde :

- $\hat{R}_{hi(s)}$: estimador de razón para una subpoblación (s) en la i-ésima zona de la h-ésima región.
- $\hat{Y}_{hi(s)}$: total estimado para la correspondiente subpoblación (s) en la i-ésima zona de la h-ésima región.
- $\hat{N}_{hi(s)}$: el total de personas de la subpoblación (s) en la i-ésima zona de la h-ésima región, estimado a partir de la muestra = $\hat{X}_{hi(s)}$.

3.6.7 Estimación del Total

El estimador de razón para los totales subpoblacionales se obtiene a partir de la proyección exógena del número de total de personas en la subpoblación, por tanto, el estimador del total de ocupados o desocupados, pertenecientes a la fuerza laboral u otra variable está dado por:

$$\begin{aligned}\hat{Y}_{hi(s)} &= \frac{\hat{Y}_{hi(s)}}{\hat{X}_{hi(s)}} \mathbf{X}_{hi(s)} \\ &= \frac{\mathbf{F} \mathbf{y}_{hi(s)}}{\mathbf{F} \mathbf{x}_{hi(s)}} \mathbf{X}_{hi(s)} \\ &= \frac{\mathbf{M}_{hi}^* \mathbf{y}_{hi(s)}}{\mathbf{m}_{hi}} \mathbf{X}_{hi(s)} \\ &= \frac{\mathbf{M}_{hi}^*}{\mathbf{m}_{hi}} \mathbf{x}_{hi(s)}\end{aligned}$$

$$= \frac{\mathbf{y}_{hi(s)}}{\mathbf{x}_{hi(s)}} \mathbf{X}_{hi(s)}$$

Los estimadores naturales $\hat{Y}_{hi(s)}$ y $\hat{X}_{hi(s)}$ se definen como se hizo en el punto 3.3.4:

$$\hat{Y}_{hi} = \frac{N_{hi}}{n_{hi}} \sum_{i=1}^{n_{hi}} \hat{Y}_{hij} = \frac{N_{hi}}{n_{hi}} \sum_{i=1}^{n_{hi}} \frac{1}{f_{hi}} y_{hij} = \frac{M^*_{hi}}{m_{hi}} y_{hi} = F_{hi} y_{hi}$$

$$\hat{X}_{hi} = \frac{N_{hi}}{n_{hi}} \sum_{i=1}^{n_{hi}} \hat{X}_{hij} = \frac{N_{hi}}{n_{hi}} \sum_{i=1}^{n_{hi}} \frac{1}{f_{hi}} x_{hij} = \frac{M^*_{hi}}{m_{hi}} x_{hi} = F_{hi} x_{hi}$$

Este es un estimador sesgado del total pero con una variabilidad menor ⁶ que el estimador simple $\hat{Y}_s = N \bar{y}_s$. La variancia estimada está dada por:

$$\hat{V}(\hat{y}_{hi(s)}) = N^2_{hi(s)} \frac{(1 - f_{hi(s)}) \sum_{j=1}^n M^2_{hi(s)} (\bar{y}_{hij(s)} - \bar{y}_{hi(s)})^2}{n_{hi(s)} (n_{hi(s)} - 1)}$$

Los totales regionales y nacionales se obtienen agregando los valores subpoblacionales, como se especifica a continuación:

- a) **Totales para Subpoblaciones:** se obtiene agregando las estimaciones para una subpoblación específica primero por zona urbana y rural, y después por región:

$$\hat{Y}_{h(s)} = \hat{Y}_{h1(s)} + \hat{Y}_{h2(s)} \quad \text{se agrega por zona}$$

$$\hat{Y}_{(s)} = \sum_{h=1}^6 \hat{Y}_{h(s)} \quad \text{se agrega por región}$$

- b) **Totales Regionales y Nacionales:** en el caso de sólo dos grupos subpoblacionales (hombre y

⁶/ Siempre que se cumplan las condiciones planteadas.

mujer, menor de 12 años y mayor o igual a 12 años), los totales se obtienen primero agregando las estimaciones para las dos subpoblaciones, luego agregando por zona y de último por región.

$$\hat{Y}_{hi} = \hat{R}_{hi(1)} X_{hi(1)} + \hat{R}_{hi(2)} X_{hi(2)} \quad \text{se agrega por subpoblación}$$

$$\hat{Y}_h = \hat{Y}_{h1} + \hat{Y}_{h2} \quad \text{se agrega por zona}$$

$$\hat{Y} = \sum_{h=1}^6 \hat{Y}_h \quad \text{se agrega por región}$$

3.7 Cálculo de Errores Muestrales

El cálculo de las variancias en un diseño bietápico con muestreo de UPM seleccionadas sin reposición resulta complejo y presenta una mayor sensibilidad frente a los distintos procedimientos de selección. Lo complicado de las fórmulas del muestreo sin reemplazo hacen más atractivo el uso del muestreo con reemplazo, y, aunque la variancia del muestreo con reemplazo es mayor, en muchas ocasiones, aunque el muestreo utilizado sea sin reemplazo por razones prácticas se estima la variancia como si fuera con reemplazo y se interpreta como una cota superior de la variancia del muestreo sin reemplazo.

En diseños polietápicos complejos, cuando se estiman muchas características poblacionales, la aplicación de las fórmulas anteriores para los estimadores de variancias puede ocasionar tediosos problemas de cálculo, por ello se han desarrollado programas especiales para la estimación de variancias. Para el cálculo de errores muestrales de la EHPM se utilizará el módulo CENVAR⁷ del programa estadístico IMPS (Integrated Microcomputer Processing System). El IMPS es un paquete estadístico para entrada, edición, tabulación, estimación, análisis y divulgación de datos de censos y encuestas por muestreo. El CENVAR ha sido desarrollado específicamente para el cálculo de errores

⁷ / CENVAR: Variance Calculation System. 1995.

muestrales en diseños complejos, para cada parámetro estimado CENVAR produce una tabla de salida o reporte con los siguientes indicadores:

1. El valor estimado del parámetro.
2. El error estándar.
3. El coeficiente de variación (C.V.).
4. El intervalo de confianza del 95 por ciento.
5. El efecto del diseño (Deff).
6. El número de observaciones (n).

El CENVAR utiliza el método de conglomerados últimos para el cálculo de los errores muestrales. El concepto de conglomerado último permite considerar el muestreo polietápico como un caso especial del muestreo de conglomerados sin submuestreo, es decir una etapa. Se denomina conglomerado último al conjunto de unidades muestrales de última etapa que pertenecen a una unidad primaria cualquiera que sea el número de etapas efectuadas dentro de ella. La variancia estimada esta basada en desviaciones entre totales de UPM. El submuestreo dentro de una UPM ha de ser independiente del efectuado en cualquier otra, la muestra ha de tener por lo menos dos unidades primarias.

La aplicación del método de conglomerados últimos es muy simple y conveniente cuando no se necesitan estimaciones separadas de las contribuciones a la variancia debidas a las distintas etapas de muestreo.

Las fórmulas de conglomerados últimos pueden ser usadas tanto para selección PPT o igual probabilidad de selección, en una sola etapa o múltiples etapas. La clave es definir apropiadamente las UPM y los ponderadores o factores de expansión para los elementos:

$$\hat{V}(\hat{Y}_A) = \sum_{h=1}^L \left[\frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{Y}_{Ahi} - \frac{\hat{Y}_{Ah}}{n_h} \right)^2 \right]$$

donde

$$\hat{Y}_{Ahi} = \sum_{j \in A} w_{hij} y_{hij}$$

$$\hat{Y}_{Ah} = \sum_{i=1}^{n_h} \sum_{j \in A} w_{hij} y_{hij}$$

$$\hat{Y}_A = \sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j \in A} w_{hij} y_{hij}$$

\hat{Y}_A : es el total estimado en la subpoblación A.

A: subconjunto de elementos que poseen el atributo de interés.

y_{hij} : total de valor observado para el j-ésimo elemento en la i-ésima UPM en h-ésimo estrato.

w_{hij} : factor de expansión final incluyendo todas las etapas de selección.

Esta técnica se puede aplicar al estimador de razón:

$$\hat{R}_A = \frac{\hat{Y}_A}{\hat{X}_A}$$

donde

\hat{Y}_A : es el total estimado para la variable Y en la subpoblación A

\hat{X}_A : es el total estimado para la variable X en la subpoblación A.

la estimación de la variancia es:

$$\hat{V}(\hat{R}_A) = \frac{1}{\hat{X}_A^2} [\hat{V}(\hat{Y}_A) + \hat{R}_A^2 \hat{V}(\hat{X}_A) - 2\hat{R}_A \text{cov}(\hat{X}_A, \hat{Y}_A)]$$

donde

$$\text{cov}(\hat{X}_A, \hat{Y}_A) = \sum_h \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} [\hat{X}_{Ahi} - \frac{\hat{X}_{Ah}}{n_h}] [\hat{Y}_{Ahi} - \frac{\hat{Y}_{Ah}}{n_h}]$$

El cálculo de errores muestrales se hace para las principales estimaciones que brinda la encuesta junto con una explicación de su significado y recomendaciones en cuanto a su uso e interpretación.

3.8 Rotación de la Muestra

No se ha planteado dentro de los objetivos de la encuesta hacer estimaciones de cambio entre dos períodos o estimaciones de promedios para dos períodos, para las cuales se requiere un traslape de la muestra o muestras independientes respectivamente. El interés en el proceso de rotación de la EHPM se enfoca a disminuir el cansancio de los entrevistados. En este sentido el esquema de rotación que se plantea es efectuar una rotación de la muestra de viviendas en un 25% de las UPM, se hace únicamente con el fin de evitar el cansancio de los entrevistados y por consiguiente disminuir los posibles rechazos que se puedan presentar por este motivo.

El esquema propuesto es dividir en un estrato las UPM en grupos de rotación y llevar un control sobre cual grupo de UPM les corresponde rotar en un año determinado de la encuesta y de cuantos años ha permanecido una misma muestra de viviendas dentro de una UPM en la encuesta.

Para conformar los grupos de rotación se determina para un estrato el número (n_i) de UPM a las cuales les corresponde rotar según el porcentaje propuesto de rotación (25%), y para cada período de la encuesta hacer una selección aleatoria de las siguientes n_i UPM que les corresponde rotar, de esta forma la selección se hace sobre aquellas UPM cuya muestra de viviendas ha permanecido mayor tiempo en la encuesta.

3.8.1 Esquema de Rotación

Según el porcentaje de rotación propuesto de un 25%, el número de UPM (n_i) a las cuales se les debe aplicar el programa de rotación en un año de la encuesta según la región y zona, se determina por:

$$n_i = n * 0.25$$

donde n corresponde al número de UPM en una zona de selección de la muestra, por ejemplo la zona urbana de la región Central. El esquema de rotación es el siguiente:

1. En el primer año de la encuesta (1999) la variable de control de rotación de segmentos tendrá el valor de 1 para todos los segmentos.
2. En el segundo año de la encuesta (2000), se selecciona en forma aleatoria n_i de las n UPM, y una vez que se les aplique el programa de rotación de viviendas se les asigna, o en este caso se les mantiene, un 1 en la variable de rotación. A las restantes UPM se les asigna un 2 en la variable de rotación indicando que es el segundo año que permanecen en la encuesta.
3. En el tercer año (2001), se seleccionan n_i UPM de las restantes UPM que tienen un dos en la variable de rotación, después de efectuar la rotación en estas n_i UPM se les asigna un 1 en la variable de rotación y al resto de las UPM se les incrementa en una unidad el valor de la variable de control de rotación. De esta forma para el tercer año se tiene:

ESQUEMA DE ROTACIÓN DE UPM

Número de UPM (25%)	Variable de Control de Rotación		
	1999	2000	2001
n_i	1	1*	2
n_i	1	2	1*
n_i	1	2	3
n_i	1	2	3

* UPM seleccionadas aleatoriamente a las que se les aplica rotación.

4. Para el cuarto año, se selecciona n_i UPM de las restantes UPM que tienen un 3 en la variable de control, después de efectuar la rotación se asigna un 1 en la variable de rotación a estas n_i UPM y al resto de UPM se les incrementa en una unidad.

5. Para el siguiente año, les corresponde rotar a las restantes n_i UPM que tienen un 4 en la variable de rotación y después se les asigna un 1 en esta variable. Al resto de UPM se incrementa en una unidad el valor de la variable de rotación como sigue:

ESQUEMA DE ROTACIÓN DE UPM

Número de UPM (25%)	Valor de la variable de Control de Rotación por año				
	1999	2000	2001	2002	2003
n_i	1	1*	2	3	4
n_i	1	2	1*	2	3
n_i	1	2	3	1*	2
n_i	1	2	3	4	1*

* UPM seleccionadas aleatoriamente a las que se les aplica rotación.

6. En los siguientes años se sigue el mismo procedimiento del punto anterior.

En los primeros cuatro años se van a rotar segmentos que tienen menos de 4 años de estar en la muestra, después del cuarto año se rotan aquellos segmentos que tienen cuatro años de estar en la muestra (con código 4 en la variable de control de rotación).

3.8.2 Grupos de Rotación

Según el número de segmentos a rotar por región de planificación y zona se formaran cuatro grupos de rotación, así, se seleccionan desde el inicio de la nueva muestra los segmentos que les corresponden rotar en los primeros cuatro años de encuesta.

Cada selección n_i formará un grupo de rotación. Las primeras n_1 UPM seleccionadas aleatoriamente formarán el grupo de rotación 1 y les corresponderá rotar el segundo año de encuesta (2000), las segundas n_2 UPM seleccionadas aleatoriamente formarán el grupo de rotación 2 y les corresponderá rotar el tercer año de encuesta (2001), las tercera n_3 UPM seleccionadas aleatoriamente formarán el grupo de rotación 3 y les corresponderá rotar el cuarto año de encuesta (2002), y por

último, las restantes n_4 UPM formarán el grupo de rotación 4 y les corresponderá rotar el quinto año de encuesta (2003). Para el sexto año de la encuesta (2004) volverá a rotar el primer grupo de rotación y así sucesivamente como se aprecia en el siguiente cuadro:

GRUPOS DE ROTACIÓN

Grupo de Rotación	Año de Rotación				
	2000	2001	2002	2003	2004 ...
1	*				*
2		*			
3			*		
4				*	

3.8.3 Procedimiento y Control de Rotación

En el momento de efectuar la rotación de la muestra de viviendas de un segmento se debe seguir un procedimiento que permita decidir y controlar cuales viviendas, o conglomerados de viviendas, salen de la muestra y cuáles son las viviendas que entran en la muestra, en este sentido se va a utilizar un sistema de rotación circular hacia atrás.

Como el arranque aleatorio es un número entero entre uno y el intervalo de selección, cuando se efectuó la rotación de la muestra de viviendas el procedimiento a seguir es restar una unidad al arranque aleatorio, lo que implica seleccionar el conglomerado anterior al que estaba seleccionado. En caso de que el conglomerado que estaba seleccionado sea el primero, el siguiente conglomerado a seleccionar será igual al valor del intervalo de selección. Por ejemplo, si para un segmento el intervalo de selección es 3.89, el arranque inicial aleatorio debe ser un número entre 1 y 4, digamos que es 3, en la tercera rotación de ese segmento el primer conglomerado seleccionado sería el 4:

Arranque Inicial	Primera Rotación	Segunda Rotación	Tercera Rotación
3	2	1	4

El resto de los conglomerados a seleccionar quedan determinados por este primer conglomerado seleccionado.

La ventaja de este esquema de rotación es que con el procedimiento de actualización de UPM que se propone, el cual consiste en hacer una nueva selección de UPM cuando se reporte un alto crecimiento dentro de una UPM, se puede llevar un mejor control sobre a cual UPM se le ha aplicado anteriormente el procedimiento de rotación y además se puede saber cuantos años ha permanecido una misma muestra de viviendas de una UPM en la encuesta.

En el caso de que una UPM experimente grandes cambios, y por consiguiente se hace necesario efectuar una nueva selección de viviendas, se le cambiará el valor de la variable de rotación por un 1, no importa el número de veces que ha permanecido en la encuesta. Esta UPM tendrá que permanecer cuatro años en la encuesta antes de volver a aplicarle el programa de rotación, a menos que experimente nuevamente grandes crecimientos que amerite una nueva selección.

De acuerdo al número de UPM determinado en la muestra por región y zona, a continuación se presenta el número de UPM (n_i) que es necesario seleccionar en cada período de la encuesta para aplicarles el programa de rotación:

**NÚMERO DE SEGMENTOS A SELECCIONAR POR TURNO DE ROTACIÓN
SEGÚN REGIÓN Y ZONA**

REGIÓN	TOTAL DE UPM	URBANA		RURAL	
		N	Selecciones de n_i	n	Selecciones de n_i
Costa Rica	726	322		404	
Central	334	200	4 selec. de 50 UPM	134	2 selec. de 33 UPM
Chorotega	82	8	4 selec. de 7 UPM	54	2 selec. de 13 UPM
Pac. Central	74	32	4 selec. de 8 UPM	42	2 selec. de 10 UPM
Brunca	92	24	4 selec. de 6 UPM	68	4 selec. de 17 UPM
H. Atlántica	92	26	2 selec. de 6 UPM	66	2 selec. de 16 UPM
H. Norte	52	12	4 selec. de 3 UPM	40	4 selec. de 10 UPM

Para un segmento específico, por ejemplo un segmento urbano de 60 viviendas y que no presente grandes cambios durante los años de ejecución de la encuesta, se tiene:

- N=60 total de viviendas en el segmento
- C= 60/3 = 20 total de conglomerados en el segmento
- n = 60/4 = 15 viviendas en la muestra
- c = 15/3 = 5 conglomerados en la muestra
- I = C/c = 20/5 = 4 intervalo de selección de conglomerados
- a = 3 arranque aleatorio entre 1 y 4

Este segmento estaría en la muestra 16 años antes de agotarse como se observa en el procedimiento de rotación siguiente:

PROCEDIMIENTO DE ROTACIÓN PARA UN SEGMENTO URBANO

Conglomerado	AÑOS DE LA ENCUESTA															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1								*	*	*	*					
2					*	*	*	*								
3	*	*	*	*												
4													*	*	*	*
5								*	*	*	*					
6					*	*	*	*								
7	*	*	*	*												
8													*	*	*	*
9								*	*	*	*					
10					*	*	*	*								
11	*	*	*	*												
12													*	*	*	*
13								*	*	*	*					
14					*	*	*	*								
15	*	*	*	*												
16													*	*	*	*
17								*	*	*	*					
18					*	*	*	*								
19	*	*	*	*												
20													*	*	*	*

Para un segmento rural de 40 viviendas y que igualmente no presente grandes cambios entre períodos de la encuesta se tiene:

$N=40$	total de viviendas en el segmento
$C= 40/3 = 13.33\sim 14$	total de conglomerados en el segmento
$n = 40/3 = 13.33\sim 13$	viviendas en la muestra
$c = 13/3 = 4.33\sim 4$	conglomerados en la muestra
$I = C/c = 14/4 = 3.5$	intervalo de selección de conglomerados
$a = 2$	arranque aleatorio entre 1 y 4

PROCEDIMIENTO DE ROTACIÓN PARA UN SEGMENTO RURAL

Conglomerado	AÑOS DE LA ENCUESTA												
	1	2	3	4	5	6	7	8	9	10	11	12	
1			*		*	*	*						
2	*	*	*	*									
3								*	*	*	*		
4													
5			*		*	*	*						
6	*	*	*	*									
7								*	*	*	*		
8			*		*	*	*						
9	*	*	*	*									
10								*	*	*	*		
11													
12			*		*	*	*						
13	*	*	*	*									
14								*	*	*	*		

Para efecto de control y seguimiento de la muestra y del proceso de rotación se recomienda crear un archivo con los segmentos de la muestra y en el que se incluya para cada segmento el número de viviendas del segmento, el intervalo de selección que se utilizó, el arranque aleatorio inicial, el año de rotación y el nuevo arranque que se utilizó en ese año de rotación.

4. Publicación de Errores Muestrales

En la mayoría de las encuestas se obtienen centenares o millares de estadísticas, las medias o totales de estas estadísticas son más fáciles de calcular que los errores estándar. Además, la presentación de todos los errores estándar duplicará el tamaño de las tablas, y complica mucho la presentación de cualquier informe de la encuesta. Por consiguiente no se suelen presentar los errores estándar de todas las medias o totales de la encuesta.

Si los resultados de la encuesta son proporciones y el efecto del diseño es cercano a uno, se puede aproximar el error estándar mediante $\sqrt{pq/n}$, elaborando una tabla o curva gráfica para su cálculo.

Cuando los resultados de la encuesta no son proporciones, el problema es menos fácil de resolver, y se tiene que buscar regularidades para construir un modelo para los errores:

- a) Determinar si el comportamiento de la desviación estándar (S_y), o de los coeficientes de variación (S_y / \bar{y}), es más uniforme.
- b) Buscar un grupo de variables con errores semejantes.
- c) Describir dos o más grupos de subclases con errores semejantes.

En muestras de encuestas complejas, el efecto del diseño es mayor que 1, y hay que tomar en cuenta por separado sus consecuencias en los errores estándar de diversas estadísticas. Al agrupar diversas estadísticas para un solo efecto del diseño se deben considerar dos aspectos:

- a) Para algunas características de la encuesta que son muy semejantes, el efecto del diseño puede diferir ampliamente.
- b) Para la misma característica, el efecto puede diferir para diversos tipos de subclases: edad, sexo, ocupación, etc.

En este sentido, la publicación de la encuesta se limita a brindar los errores muestrales para unos cuantos grupos, muchos grupos pueden confundir al lector, y la variación de los errores muestrales que se calculan es demasiado amplia como para que se hagan buenas separaciones empíricas de los grupos estadísticos. Las estadísticas para las cuales se calculan errores muestrales corresponden a las categorías de mayor importancia en la encuesta, errores muestrales para otros grupos y variables específicas se calculan a solicitud de los usuarios.