



Diseño de la Muestra ECE

PROCESO DE MUESTREO

San José, Costa Rica
2009

INDICE

JUSTIFICACIÓN	4
1. POBLACIÓN DE ESTUDIO	5
2. MARCO MUESTRAL	6
3. DISEÑO DE LA MUESTRA	7
3.1 Dominios de Estudio	7
3.2 Estratificación	¡Error! Marcador no definido.
3.3 Tamaño de Muestra	8
3.4 Asignación Geográfica de la Muestra	9
3.5 Esquema de Rotación	11
3.6 Distribución Continua de la Muestra	12
3.7. Selección de la Muestra	15
4. FACTORES DE EXPANSIÓN	15
5. ERROR DE MUESTREO	16
6. ACTUALIZACIÓN DE LA MUESTRA	16

-

JUSTIFICACIÓN

El Instituto Nacional de Estadística y Censos (INEC) realiza anualmente la Encuesta de Hogares de Propósitos Múltiples (EHPM) que investiga las características de empleo, desempleo e ingresos de la población, así como otras variables socioeconómicas y sociodemográficas relacionadas con el mercado laboral. Actualmente está a las puertas de una reformulación de dicha encuesta que incorpora mejoras en el marco conceptual y metodológico con lo cual se espera contar con mayor precisión en los indicadores actuales de empleo, desempleo, ingresos y pobreza. Sin embargo, la periodicidad de esta encuesta no permite tener información con la frecuencia y oportunidad que se requiere, por esta razón, en el año 2006 el Banco Central de Costa Rica (BCCR) propuso al INEC la elaboración de una serie de requerimientos estadísticos para suministrar en el corto y mediano plazo, entre ellos la elaboración de una encuesta de empleo con periodicidad trimestral que permitiera medir la tendencia y las variaciones cíclicas y estacionales del mercado de trabajo, y generar información estadística actualizada de la fuerza de trabajo y los ingresos percibidos por los miembros del hogar con el fin de mantener un flujo continuo de indicadores del mercado laboral. Desde ese momento a la fecha se ha venido trabajando en forma conjunta con los funcionarios del banco en la formulación del proyecto de la Encuesta Trimestral de Población Económicamente Activa (ECE).

Objetivo General

Generar información estadística actualizada sobre las condiciones socioeconómicas y socio-laborales de la población costarricense con el fin de que el BCCR calcule los indicadores del mercado laboral.

Objetivos Específicos

- Proveer información estadística oportuna sobre la población ocupada, desocupada, volumen de empleo medido en horas trabajadas o puestos de trabajo equivalentes a tiempo completo.
- Estimar el monto y fuente de ingreso de los hogares y las personas residentes habituales de esos hogares.
- Obtener una serie de indicadores que permitan analizar la evolución, tendencia y las variaciones cíclicas y estacionales del mercado de trabajo.

Para la formulación del diseño muestral de la ECE se analizaron los niveles de precisión y tamaños de muestra para los principales indicadores del mercado de trabajo obtenidos de la EHPM; además, se contó con la asesoría técnica del señor Fernando Medina, consultor de la CEPAL; se realizaron pasantías al Instituto Nacional de Estadística (INE) de España, y al Instituto Nacional de Información y Desarrollo (INIDE) de Nicaragua; y se consultó los diseños muestrales de Encuestas Continuas elaborados por algunos países latinoamericanos.

1. POBLACIÓN DE ESTUDIO

La población que investiga la ECE 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.).

Por otro lado existe un interés del BCCR de obtener información de remuneraciones pagadas a trabajadores no residentes (RNH), por ejemplo, aquellos que participan en la zafra y recolección de otras cosechas o realizan otros trabajos, pero que residen en el país menos de un año, así como los pagos recibidos por personas residentes en su condición de trabajadores autónomos (ya sean empleadores o trabajadores por cuenta propia) que se dedican a la producción de servicios de mercado (consultores que prestan sus servicios en el exterior, por ejemplo), sin embargo, la incorporación de este grupo poblacional tiene dos implicaciones importantes: por un lado, dependiendo de la magnitud de esta población, puede afectar los indicadores tradicionales de empleo, desempleo e ingresos con lo cual se perdería la comparabilidad con los resultados de encuestas anteriores; por otro lado, incluirlos implicaría aumentar considerablemente los tamaños de muestra para captar esta subpoblación, y modificar los procedimientos de selección de viviendas. Esto sin lugar a dudas, afectaría los costos y los niveles de precisión inicialmente planteados.

Para evaluar el peso de este grupo poblacional se analizaron los resultados de la EHPM 2007, se estimó en 0.7% la población extranjera con menos de un año de residir en el país, y de esos un 0.3% tienen menos de seis meses. En el listado del Registro de Edificios y Viviendas (REV)¹ de la EHPM 2007, un 1.8% de estructuras son reportadas como viviendas con RNH, y un 3.3% son estructuras de uso transitorio: baches 0.5%, viviendas de veraneo 0.7%, y un 2.1% no se especifica el tipo de estructura transitoria. Esto significa que utilizando el REV como marco para llegar a este grupo ocupacional se estaría incluyendo a lo más un 5% de estructuras. Una alternativa para aumentar la cobertura es ampliar el área de estudio a los segmentos adyacentes a los seleccionados en la encuesta en donde se identifiquen viviendas con RNH o estructuras de uso transitorio exclusivamente para trabajo².

Como un paso preliminar durante el levantamiento del REV en los segmentos seleccionados se levantó una pequeña encuesta en estructuras identificadas como baches para caracterizar esta subpoblación ([Anexo 1](#)), y evaluar la importancia e inclusión dentro de la población de estudio. Para no afectar la comparabilidad con los indicadores tradicionales de empleo, desempleo e ingresos, en el caso de que se tome la decisión de incluir este grupo poblacional en la encuesta, se recomienda utilizar un filtro para identificarlos de tal manera que se puedan incluir o excluir en la estimación de estos indicadores. Además, para no afectar los tamaños de muestra y la precisión esperada en los principales indicadores de la ECE, se recomienda la selección de una muestra adicional de viviendas en los segmentos en donde se identifiquen viviendas con RNH o estructuras de uso transitorio exclusivamente para trabajo, o efectuar un censo o barrido en todas las estructuras que aparezcan

1 /El REV es el listado de estructura dentro de un segmento con su respectivo mapa para su ubicación geográfica y del cual se seleccionan las viviendas de la muestra.

2/ Se haría uso de las **secciones** construidas en el Marco Muestral de Viviendas 2000 (**MMV-00**) con un tamaño de cinco segmentos adyacentes por sección.

RNH. Esto debe considerarse a la luz de los costos adicionales que implicaría la inclusión de esta población así como su peso en las estimaciones según los objetivos del BCCR.

2. MARCO MUESTRAL

El marco muestral de viviendas (MMV-2000) para la selección de la muestra se construyó a partir de la información de los Censos Nacionales de Población y Vivienda efectuados en el año 2000. El MMV-00 corresponde al tipo llamado Marco de Áreas, que se caracteriza porque sus unidades son superficies geográficas con límites claramente definidos. En un marco de áreas los hogares se encuentran asociados a las respectivas viviendas individuales ubicadas dentro de estas unidades, esto hace posible tener una medida muy aproximada de las probabilidades de selección de los hogares.

En el caso del MMV-00 de Costa Rica, 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. En total lo conforman 17 496 segmentos clasificados como urbanos un 55.2% y como rurales 44.8%, y 1 029 867 viviendas, 61.2% en zonas urbanas y 38.8% en zonas rurales ([Cuadro 1](#)).

Además del conjunto de segmentos censales, el MMV-00 se complementa con información demográfica y socioeconómica proveniente de los censos nacionales, y con mapas cartográficos que permiten la ubicación, ordenamiento y estratificación de cada uno de estos segmentos.

CUADRO 1
DISTRIBUCIÓN DE SEGMENTOS Y VIVENDAS DEL MARCO MUESTRAL
SEGÚN REGIÓN DE PLANIFICACIÓN Y ZONA

Región y Zona	Segmentos		Viviendas	
	Total	%	Total	%
Costa Rica	17.496	100,0	1.029.867	100,0
Urbano	9.656	55,2	630.119	61,2
Rural	7.840	44,8	399.748	38,8
Central	10.503	60,0	647.507	62,9
Urbano	7.670	73,0	500.590	77,3
Rural	2.833	27,0	146.917	22,7
Chorotega	1.481	8,5	83.076	8,1
Urbano	452	30,5	30.339	36,5
Rural	1.029	69,5	52.737	63,5
Pacífico Central	1.086	6,2	61.302	6,0
Urbano	499	45,9	30.550	49,8
Rural	587	54,1	30.752	50,2
Brunca	1.589	9,1	82.851	8,0
Urbano	364	22,9	25.103	30,3
Rural	1.225	77,1	57.748	69,7
Huetar Atlántica	1.839	10,5	102.006	9,9
Urbano	545	29,6	34.884	34,2
Rural	1.294	70,4	67.122	65,8
Huetar Norte	998	5,7	53.125	5,2
Urbano	126	12,6	8.653	16,3
Rural	872	87,4	44.472	83,7

3. DISEÑO DE LA MUESTRA

En la elaboración del diseño de la muestra se consideró los objetivos de la encuesta, los requerimientos de información del BCCR, y se utilizó la información proveniente de la EHPM 2006 y 2007 sobre los principales indicadores de empleo y desempleo y sus niveles de precisión.

3.1 Dominios de Estudio

Los dominios de estudio son subpoblaciones generalmente relacionadas a áreas geográficas para las cuales se requiere estimaciones con una precisión conocida. Es interés del BCCR obtener estimaciones precisas a nivel nacional, y aceptables por zona urbana y rural para las principales variables de empleo, desempleo e ingresos³. Para estos dominios de estudios se esperan coeficientes de variación (CV) menores al 10% para la tasa de desempleo abierto (TDA), sin embargo, para algunas desagregaciones importantes de la población ocupada, como son la clasificación por rama de actividad o sector institucional, se pueden presentar limitaciones de estimación para algunas

3/ El coeficiente de variación (CV) o error relativo es una medida de la precisión de una estimación. Si el CV es menor al 10% se dice que la estimación es bastante precisa, entre 10% y 15% aceptable, y mayor al 15% es no aceptable.

categorías con niveles de precisión no aceptables ($CV > 15\%$), y que deberán tratarse con cautela (Cuadro 2).

CUADRO 2
NIVELES DE PRECISIÓN ESPERADOS PARA LOS PRINCIPALES
INDICADORES POR DOMINIO DE ESTUDIO.

Indicador	Nacional	Zona
Pea	Si	Si
Desocupados	Si	Si
Puestos de Trabajo	Si	Si
Ocupados por Horas Trabajadas	Si	Si
Total de Horas Trabajadas	Si	Si
TNP	Si	Si
TDA	Si	Si
Pobreza	Si	Si
Ingreso Per Cápita Hogar	Si	Si
Rama de Actividad	R	No
Grupo Ocupacional	Si	R
Categoría Ocupacional	Si	R
Sector Institucional	R	No

Si : $CV \leq 10$

R : $10 < CV \leq 15$

No : $CV > 15$

3.2 Estratificación

Los estratos son subpoblaciones con alguna característica en común que se desea estén representadas en la muestra con el objetivo de incrementar la precisión general de los estimadores de la encuesta. Los estratos pueden coincidir o no con los dominios de estudio, y no necesariamente se requiere niveles de precisión específicos para estas subpoblaciones.

El MMV-00 se estratificó por región de planificación, zona, y en el caso de la zona urbana de la región Central se estratificó por nivel socioeconómico según una estimación de los niveles de ingreso per cápita de los segmentos [1].

3.3 Tamaño de Muestra

Para determinar el tamaño de muestra se tomó en consideración los requerimientos estadísticos del BCCR sobre la ECE, y se definió un grupo de indicadores prioritarios por rama de actividad, sector, categoría y grupo institucional:

- Total de ocupados
- Total de Horas Trabajadas
- Promedio de Horas Trabajadas
- Total de Puestos de Trabajo⁴
- Ingreso Promedio Primario

La precisión estadística para cada uno de estos indicadores se midió con el coeficiente de variación (CV) de acuerdo a los siguientes criterios:

- Aceptable (Si): estimaciones con CV menores o iguales al 10% ($CV \leq 10\%$).
- Regular (R): estimaciones con CV entre un 10 y un 15% ($10\% < CV \leq 15\%$).
- No aceptables (No): estimaciones con CV mayores al 15% ($CV > 15\%$).

Para determinar los tamaños de muestra por trimestre se realizó una simulación con los turnos de rotación de la EHPM⁵, se analizaron los niveles de precisión para los indicadores propuestos para cuatro tamaños de muestra: 3000, 6000, 9000 y 12000 viviendas, que equivalen aproximadamente a un 25%, 50%, 75% y 100% de la muestra de la EHPM ([Anexo 1](#)). Además se evaluaron los costos estimados y los requerimientos operativos para los diferentes tamaños de muestra.

De acuerdo a estos resultados, tomando en consideración los objetivos del BCCR, y la capacidad técnica del INEC de satisfacer a futuro los requerimientos del banco, se consideró que una muestra de 9000 viviendas era recomendable desde el punto de vista de precisión, costos y capacidad operativa. De cada segmento censal se seleccionará una muestra de 12 viviendas para un total de 750 segmentos a nivel nacional.

3.4 Asignación Geográfica de la Muestra

Generalmente se utiliza la TDA como la variable de diseño para calcular el tamaño y la distribución de la muestra. Esta variable tiene un comportamiento diferencial por región de planificación y zona, y requiere más muestra en las regiones o zonas donde esta tasa sea más baja. Es evidente que utilizar la TDA como criterio para la asignación de la muestra favorecerá en mayor medida a las zonas rurales por la poca frecuencia de esta subpoblación en estas zonas, sin embargo, como el objetivo de la ECE es obtener indicadores sobre la fuerza de trabajo más que estimar la tasa o nivel de desocupación, se utilizó la variable número de ocupados en lugar de la TDA para la afijación de la muestra de manera de tal forma que se privilegie más las zonas urbanas donde se tiene una mayor concentración de la población ocupada.

Como la zona urbana y rural se plantean como dominios de estudio la muestra se distribuyó en primer término por zona y posteriormente por región de planificación dentro de cada zona. Para la afijación de la muestra se utilizó la información de la EHPM simulando dos muestras de aproximadamente 9000 viviendas cada una, y se obtuvo la precisión relativa (CV) para el número

4/ Son las jornadas de tiempo completo (40 horas) trabajadas por las personas ocupadas.

5 / La EHPM investiga cada año aproximadamente 12000 viviendas y la muestra está dividida en cuatro turnos de rotación de aproximadamente 3000 viviendas cada uno que representan un 25% de la muestra total.

de ocupados por zona y región de planificación. En la distribución de la muestra se consideró tres criterios de asignación diferentes: proporcional, Neyman y de Potencia con nivel de $\alpha=0.5$ [2].

$$\text{Proporcional: } n_h = n * \frac{N_h}{\sum N_h}$$

$$\text{Neyman: } n_h = n * \frac{N_h S_h}{\sum N_h S_h}$$

$$\text{Potencia: } n_h = n * \frac{N_h^\alpha S_h}{\sum N_h^\alpha S_h} \quad \alpha=0.5$$

La asignación de Potencia mostró una mejor distribución por región de planificación en términos de CV más homogéneos. La asignación de Potencia asigna cerca de un 55% de la muestra a la zona urbana con CV menores al 15% por región dentro de cada zona, mientras que la asignación de Neyman privilegia considerablemente a la zona urbana con un 71% de las viviendas pero con CV mayores al 15% para el resto de las regiones fuera de la región Central (Cuadro 3).

CUADRO 3
DISTRIBUCIÓN DE LA MUESTRA POR CRITERIO DE AFIJACIÓN Y NIVELES DE PRECISIÓN ESPERADOS PARA LOS PRINCIPALES INDICADORES POR ZONA Y REGIÓN.

ZONA Y REGIÓN	EHPM			PROPORCIONAL			NEYMAN			POTENCIA ($\alpha=0,5$)		
	Muestra	Relativo	CV _h	Muestra	Relativo	CVe	Muestra	Relativo	CVe	Muestra	Relativo	CVe
OCUPADOS	8.895	100,0	2,9	9.000	100,0	2,9	9.000	100,0	2,9	9.000	100,0	2,9
Urbana	3.642	40,9	3,9	5.507	61,2	3,2	6.428	71,4	3,0	4.928	54,8	3,4
Central	2.216	60,9	4,5	4.375	48,6	3,2	6.145	68,3	2,7	1.021	11,3	6,7
Chorotega	366	10,1	15,9	265	2,9	18,7	74	0,8	35,5	885	9,8	10,2
Pac. Central	303	8,3	6,8	267	3,0	7,2	29	0,3	21,8	379	4,2	6,1
Brunca	288	7,9	14,8	219	2,4	17,0	42	0,5	39,1	750	8,3	9,2
H. Atlántica	249	6,8	18,5	305	3,4	16,8	124	1,4	26,2	1.105	12,3	8,8
H. Norte	220	6,0	26,6	76	0,8	45,3	14	0,2	107,0	788	8,8	14,0
Rural	5.253	59,1	4,1	3.493	38,8	5,0	2.572	28,6	5,8	4.072	45,2	4,6
Central	1.962	37,3	5,7	1.284	14,3	7,1	1.338	14,9	6,9	558	6,2	10,7
Chorotega	689	13,1	10,7	461	5,1	13,1	200	2,2	19,8	625	6,9	11,2
Pac. Central	577	11,0	12,8	269	3,0	18,7	78	0,9	34,8	571	6,3	12,9
Brunca	768	14,6	14,9	505	5,6	18,3	349	3,9	22,1	910	10,1	13,7
H. Atlántica	834	15,9	10,7	587	6,5	12,7	394	4,4	15,5	705	7,8	11,6
H. Norte	425	8,1	13,1	389	4,3	13,7	213	2,4	18,4	702	7,8	10,2

Con el fin de garantizar niveles de precisión aceptables para otros indicadores importantes, se analizó la distribución de la muestra para otras variables como son la TDA, el número de desocupados, el número de puestos de trabajo y el ingreso per cápita. De todas las distribuciones resultantes por criterio de afijación, el criterio de potencia presenta una mejor distribución en términos de CV más homogéneos (Anexo 2).

3.5 Esquema de Rotación

Un esquema de rotación consiste en la división de la muestra con el objetivo de ir renovando una parte de la muestra de viviendas de una encuesta a otra. Tradicionalmente el interés en este proceso ha sido el disminuir el cansancio de los entrevistados, y por consiguiente disminuir los posibles rechazos que se presenten por este motivo.

Los esquemas de rotación se utilizan también en encuestas tipo panel cuando hay interés de estimar los cambios entre periodos de encuestas. En años recientes, el INEC le ha dado un mayor uso a estos grupos o paneles de rotación para análisis longitudinales sobre la evolución del empleo y el desempleo, y el impacto de las políticas sociales y económicas en las condiciones de vida de los hogares y las personas.

En cada período de encuesta habrá una fracción de viviendas que se repiten y otra fracción de viviendas que ingresan por primera vez sustituyendo a una fracción similar que sale de la encuesta. El esquema de rotación que se utilice afectará la precisión en los estimadores de cambio en los estudios de panel, en este tipo de mediciones la ganancia en precisión de las muestras traslapadas es directamente proporcional al porcentaje de traslape debido que el efecto de correlación es mayor [3]. Por ejemplo, en el caso de que se desee estimar el cambio de una proporción en dos encuestas consecutivas se tiene:

$$d = p_t - p_{t-1}$$

$$\text{Var}(d) = S_{p_t}^2 + S_{p_{t-1}}^2 - 2 \text{cov}(p_t, p_{t-1})$$

$$\text{Var}(d) = S_{p_t}^2 + S_{p_{t-1}}^2 - 2 \text{cov}(p_t, p_{t-1}) = 2S_p^2(1 - f\rho)$$

$$\text{si } S_{p_t}^2 = S_{p_{t-1}}^2 = S_p^2$$

$$\text{Var}(d) = 2S_p^2(1 - f\rho)$$

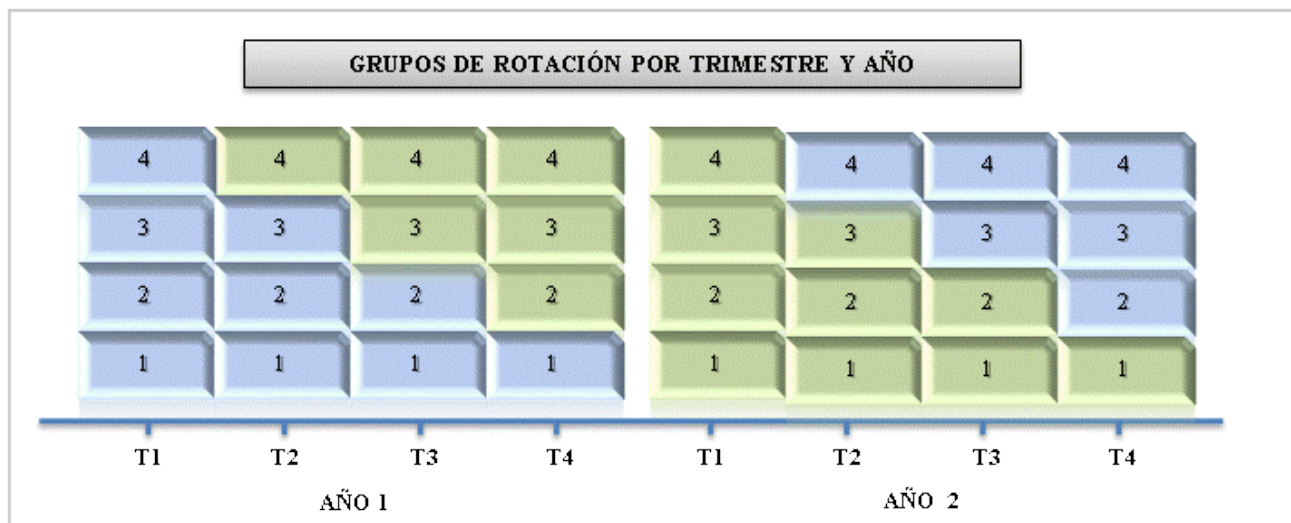
Donde **d** es el estimador de cambio entre la proporción estimada (**p**) en el período *t* y *t* - 1, **Var (d)** es la variancia de ese estimador, **S²** es la variancia de la proporción en cada período de encuesta, **cov (p_t, p_{t-1})** es la covariancia en las estimaciones entre encuestas, **f** es la fracción de traslapamiento y **ρ** es el coeficiente de correlación intraclase. Se puede apreciar que la variancia para la diferencia de dos encuestas consecutivas es siempre igual a la variancia de una encuesta multiplicada por **2(1-fρ)**, entre mayor sea **f** o **ρ** la precisión del estimador será mayor, en otras palabras, **Var (d)** menor. En el caso extremo si **f=0**, no hay traslape entre las encuestas, como es el caso de muestras independientes, la **Var (d)** será igual a dos veces la variancia de una encuesta.

Con este razonamiento se analizaron dos porcentajes de traslape, del 50% y 75% por trimestre. Con un porcentaje del 50% la muestra se agota más rápidamente, una vivienda permanecerá en la muestra únicamente dos períodos de encuestas. Con un porcentaje de rotación del 25% una vivienda

permanecerá en la muestra por cuatro encuestas consecutivas, y la correlación intraclase (ρ) será mucho mayor favoreciendo la precisión de los estimadores de cambio (Esquema 1).

ESQUEMA 1

ESQUEMA TRIMESTRAL DE ROTACIÓN DE UN 25% LA MUESTRA DE VIVIENDAS PARA LOS PRIMEROS DOS AÑOS DE ENCUESTA.



De esta forma, el esquema de rotación del nuevo DMEC-09 corresponde a una rotación de la muestra de viviendas en un 25% de las UPM o segmentos seleccionados. Para esto es necesario dividir la muestra de segmentos dentro de cada estrato cuatro réplicas o grupos de rotación, cada uno de igual tamaño y con el mismo diseño muestral. Antes de cada trimestre de encuesta, se cambia o rota la selección de viviendas dentro de uno de estos grupos de rotación, de esta forma un 75% de la muestra de viviendas coincidirá entre una encuesta y otra, y a los cuatro trimestres de encuesta la muestra de viviendas se renovará en un 100%.

3.6 Distribución Continua de la Muestra

Una encuesta continua consiste en una serie vinculada de encuestas periódicas que proporcionan para un período específico de tiempo estimaciones independientes de los indicadores del mercado de trabajo, y permiten medir las variaciones cíclicas y estacionales de esos indicadores [4].

La característica fundamental de una encuesta continua es que el trabajo de campo se realiza en forma ininterrumpida a lo largo del período de encuesta, razón por la cual la información se obtiene mediante un período de referencia móvil (lapso de tiempo inmediatamente anterior a la entrevista) y por lo tanto diferente de un informante a otro dependiendo de la fecha de la entrevista.

El período de la encuesta continua puede dividirse en “subperíodos” (meses, semanas), cada uno de los cuales comprenderá una muestra representativa del período completo de la encuesta. La división de la muestra en subperíodos permite una cobertura mejor, en el sentido de tener una muestra más representativa para estudiar las variaciones cíclicas o estacionales ocurridas durante el período completo. Con el trabajo de campo distribuido de esta forma se mejoran los resultados agregados para el período completo, dado que las variaciones estacionales y demás variaciones temporales durante el período de la encuesta quedan cubiertas de una forma más equilibrada. La principal

desventaja está en que eleva los costos de la encuesta y dificulta el trabajo de campo porque implica un recorrido continuo del país durante todo el período y subperíodos de encuesta.

Para garantizar la distribución continua de la muestra durante todo el trimestre, se realizó una distribución balanceada del total de segmentos por réplica, región y zona entre las 13 semanas del trimestre de la siguiente forma:

- El tamaño de muestra trimestral asignado a cada región y zona se dividió en cuatro paneles o réplicas independientes de tamaño similar que corresponden a los cuatro turnos de rotación de la muestra.
- Teniendo en cuenta que la periodicidad de la encuesta es trimestral, la muestra de cada réplica se distribuyó de manera uniforme en las 13 semanas del trimestre para garantizar una mejor representatividad de cada región y zona, y cubrir de forma equilibrada las variaciones estacionales y coyunturales a lo interno de las regiones.
- Finalmente, después de algunos ajustes por réplica y por semana dentro del trimestre, el tamaño de muestra quedó en 752 segmentos (Cuadro 4), o 9024 viviendas (Cuadro 5), con 182 segmentos por réplica que corresponden al 25% de la muestra total (Cuadro 6).

CUADRO 4
DISTRIBUCIÓN DE LA MUESTRA DE SEGMENTOS POR SEMANA DENTRO DEL TRIMESTRE
SEGÚN REGIÓN Y ZONA.

Estrato	Muestra de Segmentos		Subréplica dentro del Trimestre												
	Absoluto	Relativo	1	2	3	4	5	6	7	8	9	10	11	12	13
Costa Rica	752	100,0	58	58	59	58	56	59	59	57	58	58	58	57	57
Urbano	409	54,4	32	32	31	31	30	32	33	31	31	32	32	31	31
Rural	343	45,6	26	26	28	27	26	27	26	26	27	26	26	26	26
Central	128	17,0	10	10	10	10	9	10	10	10	10	10	10	10	9
Urbano	80	62,5	6	6	6	6	6	7	7	6	6	7	6	6	5
Bajo	36	28,1	3	2	3	3	3	3	3	3	3	3	3	2	2
Medio	28	21,9	2	2	2	2	2	2	3	2	2	3	2	2	2
Alto	16	12,5	1	2	1	1	1	2	1	1	1	1	1	2	1
Rural	48	37,5	4	4	4	4	3	3	3	4	4	3	4	4	4
Chorotega	128	17,0	9	9	10	10	10	10	10	10	10	10	10	10	10
Urbano	76	59,4	5	5	6	6	6	6	6	6	6	6	6	6	6
Rural	52	40,6	4	4	4	4	4	4	4	4	4	4	4	4	4
Pacífico Central	76	10,1	6	6	6	6	5	6	6	5	6	6	6	6	6
Urbano	29	38,2	3	2	2	2	2	2	3	2	2	2	2	2	3
Rural	47	61,8	3	4	4	4	3	4	3	3	4	4	4	4	3
Brunca	140	18,6	11	11	11	11	11	11	11	11	10	11	10	10	11
Urbano	64	45,7	5	5	5	5	5	5	5	5	4	5	5	5	5
Rural	76	54,3	6	6	6	6	6	6	6	6	6	6	5	5	6
Huetar Atlántica	152	20,2	12	12	12	11	12	12	12	11	12	11	12	11	12
Urbano	92	60,5	7	8	7	7	7	7	7	6	8	6	7	7	8
Rural	60	39,5	5	4	5	4	5	5	5	5	4	5	5	4	4
Huetar Norte	128	17,0	10	10	10	10	9	10	10	10	10	10	10	10	9
Urbano	68	53,1	6	6	5	5	4	5	5	6	5	6	6	5	4
Rural	60	46,9	4	4	5	5	5	5	5	4	5	4	4	5	5

3.7. Selección de la Muestra

La muestra de segmentos se seleccionó con probabilidad proporcional al tamaño (PPT) dentro de cada región de planificación, zona y réplica, como medida de tamaño se tomó el número de viviendas ocupadas por segmento. Dentro de cada segmento se seleccionan 12 viviendas con probabilidades iguales (PI), en conglomerados de tres viviendas ocupadas.

No se recomienda la sustitución de viviendas por los sesgos que pueden presentarse. Por un lado se presenta el sesgo de selección debido a que las viviendas seleccionadas para sustitución tendrán doble probabilidad de ser seleccionadas: la probabilidad de selección entre las primeras doce viviendas, y posteriormente la probabilidad de selección dado que no fueron seleccionadas entre las doce primeras. Por otro lado, la sustitución de viviendas puede agregar más problemas de los que realmente resuelve, está ampliamente comprobado que las viviendas que no responden en primera instancia pueden diferir en forma importante de las que si responden, lo cual induce a sesgos de estimación, para reducir estos sesgos, la mejor forma de enfrentar la no respuesta es evitándola [5]. En este sentido es mejor dedicar esfuerzos a la calidad del dato obtenido y de mantener las tasas de respuesta en niveles aceptables (menores al 10%), y en caso de situaciones especiales diseñar procedimientos ad-hoc para la no respuesta, por ejemplo utilizar el cuestionario del trimestre anterior en caso de que la vivienda ya haya sido entrevistada anteriormente, o la selección previa de una muestra mayor de viviendas en segmentos con bajas tasas de respuesta.

Además, la selección de viviendas adicionales para reemplazo dificulta el control de la selección y del esquema de rotación de viviendas entre encuestas.

4. FACTORES DE EXPANSIÓN

Si la población permaneciera estática desde la construcción del marco muestral de viviendas hasta la fecha de la encuesta, el factor de expansión de las viviendas sería simplemente el inverso de las probabilidades iniciales de selección, sin embargo, es usual que ocurran cambios en la estructura poblacional desde el momento en que se construyó el marco hasta el momento en que se realizó la encuesta. Para corregir esas diferencias normalmente se consideran tres tipos de ajuste a los factores de expansión: por no respuesta, por desactualización de la muestra, y por proyecciones de población.

El ajuste por no respuesta corrige las diferencias en las probabilidades de selección por la no respuesta de las viviendas seleccionadas. El ajuste consiste en darle un mayor peso a las viviendas que respondieron para compensar las viviendas que no respondieron.

El ajuste por desactualización corrige de manera parcial los problemas de desactualización del MMV-00 debido a que se efectúa únicamente en los segmentos seleccionados. El ajuste consiste en actualizar las probabilidades de selección de las viviendas entrevistadas según los cambios poblacionales observados dentro de cada segmento de la muestra. Sin embargo, este ajuste no compensa la desactualización general del marco, en este sentido, una práctica común en las encuestas de hogares es utilizar estimaciones de población exógenas, como son las proyecciones de población, para corregir la desactualización general del marco.

Existen muchas posiciones en cuanto el nivel de ajuste por proyecciones de población, ya sea que se haga a nivel total, para ciertos niveles de desagregación geográfica, o para grupos específicos de

población por sexo y edad. Algunos investigadores opinan que las proyecciones demográficas no son tan rigurosas ni tan exactas para poder utilizarlas como referencia válida para las encuestas, a menos que se efectúe un conteo de población cada cinco años con el fin de actualizar con más precisión las proyecciones de población[6]. Otros argumentan que las encuestas de hogares no deberían deferir de las proyecciones oficiales de población, y si es así, en muchas ocasiones proporcionan mejores cifras de la estructura poblacional por sexo y edad que las proyecciones demográficas, sobre todo si se garantiza la representatividad y calidad de los datos muestrales [7].

Por lo anterior y como el principal objetivo de la ECE es dar estimaciones para el total país, por zona urbana y rural, y para algunos grupos de población hasta donde los niveles de precisión lo permitan, el ajuste por proyecciones de población se hace a nivel nacional para mantener la coherencia con las demás estimaciones de población que brinda el INEC, y se mantiene la estructura por zona, sexo y grupos de edad que proporcione cada encuesta.

5. ERROR DE MUESTREO

Anteriormente la disponibilidad de software para el cálculo de errores de muestreo en encuestas complejas⁶ estaba limitado a unos cuantos programas estadísticos, sin embargo, actualmente la disponibilidad de programas estadísticos que han incluido dentro de sus rutinas el cálculo de errores de muestreo para encuestas complejas es mucho mayor y son más amigables en cuanto al procesamiento. Estos programas utilizan métodos robustos para el cálculo de los errores de muestreo como por ejemplo el método de Conglomerados Últimos (IMPS), y el método de linearización de Series de Taylor (SPSS, STATA, SAS), cuyos resultados no difieren significativamente entre sí [8].

El cálculo de los errores de muestreo para la ETE se efectuará en el programa estadística IMPS para las principales estimaciones de la encuesta, su uso y publicación estará limitada según los niveles de precisión alcanzados.

6. ACTUALIZACIÓN DE LA MUESTRA

Los segmentos de la muestra permanecerán de forma indefinida en la muestra, salvo situaciones particulares en que se reporte una baja considerable de viviendas ocupada en algún segmento, agotamiento de la muestra de viviendas, o se presenten problemas de alta peligrosidad o dificultades de acceso.

Cada trimestre se renovará o rotarán las viviendas seleccionadas dentro del 25% de los segmentos de la muestra, con esto se evitará el cansancio de los entrevistados y se renovará parcialmente la muestra entre trimestres, y totalmente cada año.

En la etapa de trabajo de campo se efectuará un recorrido completo del segmento para verificar la condición o cambio de estatus de las viviendas (ocupadas, desocupadas, abandonadas o en construcción), y la aparición de viviendas nuevas con el fin de actualizar las probabilidades de selección para la encuesta en ese trimestre, y asignarles a las nuevas viviendas una probabilidad conocida de ser seleccionadas para la próxima encuesta.

6 / Una encuesta se considera compleja cuando incluye dos o más etapas de selección.

REFERENCIAS

1. ENIG 2004: Asignación de la Muestra. Documento de Trabajo. INEC.2002.
2. INEC. Resultado sobre el Análisis del Tamaño de Muestra. Documento de Trabajo. 2008
3. Medina, Fernando. Las Encuestas de Panel. CEPAL.
4. Husmanns, Mehran y Verma. Encuestas de Población Económicamente Activa, Empleo, Desempleo y Subempleo. Manual de la OIT sobre conceptos y métodos.1990.
5. Medina & Galván. Imputación de datos: teoría y Práctica. 2007.
6. IBGE. Proyecciones de Población y los Ajustes de las Estimaciones de Encuestas Domiciliares: solución o un problema. 1990.
7. DGEC. Diseños Metodológicos de la Encuesta de Hogares de Propósitos Múltiples. El Salvador.1998.
8. Siller & Tompkins. The Big Four: Analyzing Complex Sample Survey Data Using SAS, SPSS, STATA, and SUDAAN. National Center for Health Statistics. USA.

ANEXO 1**POSIBLES NIVELES DE DESAGREGACIÓN PARA PRINCIPALES INDICADORES
SEGÚN NIVELES DE PRECISIÓN ESPERADOS PARA DIFERENTES TAMAÑOS DE MUESTRA**

Indicador	Tamaño de Muestra Viviendas			
	3000	6000	9000	12000
Pea	Si	Si	Si	Si
Urbano	R	Si	Si	Si
Rural	R	Si	Si	Si
Desocupados	R	Si	Si	Si
Urbano	No	R	Si	Si
Rural	No	R	Si	Si
Puestos de Trabajo	Si	Si	Si	Si
Urbano	R	Si	Si	Si
Rural	R	Si	Si	Si
Ocupados por Horas Trabajadas	R	Si	Si	Si
Urbano	R	Si	Si	Si
Rural	R	Si	Si	Si
Total de Horas Trabajadas	Si	Si	Si	Si
Urbano	R	Si	Si	Si
Rural	R	Si	Si	Si
TNP	Si	Si	Si	Si
Urbano	Si	Si	Si	Si
Rural	Si	Si	Si	Si
TDA	Si	Si	Si	Si
Urbano	R	Si	Si	Si
Rural	R	Si	Si	Si
Pobreza	Si	Si	Si	Si
Urbano	R	Si	Si	Si
Rural	R	Si	Si	Si
Ingreso Per Cápita Hogar	Si	Si	Si	Si
Urbano	Si	Si	Si	Si
Rural	Si	Si	Si	Si
Rama de Actividad				
Agricultura, ganadería, caza y silvicultura	R	Si	Si	Si
Pesca	No	No	No	No
Explotación de minas y canteras	No	No	No	No
Industrias manufactureras	R	Si	Si	Si

Indicador	Tamaño de Muestra Viviendas			
	3000	6000	9000	12000
Suministro de electricidad, gas y agua	No	No	R	R
Construcción	R	Si	Si	Si
Comerc por may y men, repar vehíc autom, motoc, etc	R	Si	Si	Si
Hoteles y restaurantes	R	Si	Si	Si
Transporte, almacenamiento y comunicaciones	No	R	Si	Si
Intermediación financiera	No	R	Si	Si
Actividades inmobiliarias, empresariales y de alquiler	R	Si	Si	Si
Admin pública y defensa; planes de seguridad social	R	R	Si	Si
Enseñanza	R	Si	Si	Si
Servicios sociales y de salud	No	R	Si	Si
Otras activid de serv comun, social y person	No	R	Si	Si
Hogares privados con servicios doméstico	R	Si	Si	Si
Organizaciones y organos extraterritoriales	No	No	No	No
Grupo Ocupacional				
Nivel direct adm públ y empr priv	No	R	R	Si
Nivel prof, científ e intelect	R	Si	Si	Si
Nivel técnico y profesional medio	R	Si	Si	Si
Apoyo administrativo	R	Si	Si	Si
Venta locales prestac serv direct a personas	R	Si	Si	Si
Agropecuarias, agrícolas y pesqueras "calificadas"	R	Si	Si	Si
Prod artes, const, mecán, art gráf y manuf "calificadas"	R	Si	Si	Si
Montaje y operación de instalaciones y de máquinas	R	R	Si	Si
Ocupaciones no calificadas	Si	Si	Si	Si
Categoría Ocupacional				
Patrono o socio activo	R	Si	Si	Si
Trabajador cuenta propia	Si	Si	Si	Si
Empleado del estado	R	Si	Si	Si
Empleado empresa privada	Si	Si	Si	Si
Servidor doméstico	R	Si	Si	Si
Familiar no remunerado	No	R	R	Si
Sector Institucional				
Gobierno Central	R	Si	Si	Si
Insti.Autonom y Semiauton de servic	R	Si	Si	Si
Empresas Públicas No Financieras	No	R	Si	Si
Empresas Públicas Financieras	No	No	R	R
Empresas Públicas Constituidas en SA	No	No	No	No
Municipalidades y Consejos de Distrito	No	No	R	R

Indicador	Tamaño de Muestra Viviendas			
	3000	6000	9000	12000
Sector Privado	Si	Si	Si	Si
Organismos Internacionales	No	No	No	No

No : $CV > 15$

R : $10 < CV \leq 15$

Si : $CV \leq 10$

[ANEXO 2](#)

ZONA Y REGIÓN	EHPM			PROPORCIONAL			NEYMAN			POTENCIA ($\alpha=0,5$)		
	Muestra	Relativo	CV _h	Muestra	Relativo	CVe	Muestra	Relativo	CVe	Muestra	Relativo	CVe
TDA	8.895	100,0	5,1	9.000	100,0	5,1	9.000	100,0	5,1	9.000	100,0	5,1
Urbana	3.642	40,9	6,6	5.507	61,2	5,4	5.185	57,6	5,6	4.623	51,4	5,9
Central	2.216	24,9	7,9	4.375	48,6	5,6	3.163	35,1	6,6	1.150	12,8	11,0
Chorotega	366	4,1	23,3	265	2,9	27,4	526	5,8	19,4	831	9,2	15,4
Pac. Central	303	3,4	17,5	267	3,0	18,6	468	5,2	14,1	626	7,0	12,2
Brunca	288	3,2	20,5	219	2,4	23,5	349	3,9	18,6	667	7,4	13,5
H. Atlántica	249	2,8	17,1	305	3,4	15,4	526	5,8	11,7	653	7,3	10,5
H. Norte	220	2,5	36,5	76	0,8	62,2	152	1,7	44,0	696	7,7	20,5
Rural	5.253	59,1	7,9	3.493	38,8	9,7	3.815	42,4	9,3	4.377	48,6	8,7
Central	1.962	22,1	11,6	1.284	14,3	14,4	1.401	15,6	13,8	722	8,0	19,2
Chorotega	689	7,7	15,9	461	5,1	19,4	518	5,8	18,3	590	6,6	17,1
Pac. Central	577	6,5	20,5	269	3,0	30,1	335	3,7	26,9	583	6,5	20,4
Brunca	768	8,6	21,6	505	5,6	26,6	752	8,4	21,8	841	9,3	20,6
H. Atlántica	834	9,4	14,7	587	6,5	17,5	518	5,8	18,6	615	6,8	17,1
H. Norte	425	4,8	30,0	389	4,3	31,4	292	3,2	36,2	1.026	11,4	19,3
DESOCUPADOS	8.895	100,0	6,0	9.000	100,0	5,9	9.000	100,0	5,9	9.000	100,0	5,9
Urbana	3.642	40,9	7,7	5.507	61,2	6,2	6.134	68,2	5,9	4.527	50,3	6,9
Central	2.216	24,9	9,1	4.375	48,6	6,4	5.913	65,7	5,5	1.107	12,3	12,8
Chorotega	366	4,1	35,6	265	2,9	41,8	74	0,8	79,1	1.072	11,9	20,8
Pac. Central	303	3,4	20,4	267	3,0	21,8	47	0,5	51,7	617	6,9	14,3
Brunca	288	3,2	27,5	219	2,4	31,5	32	0,4	82,0	753	8,4	17,0
H. Atlántica	249	2,8	17,6	305	3,4	15,9	63	0,7	35,1	569	6,3	11,7
H. Norte	220	2,5	25,4	76	0,8	43,3	4	0,0	197,9	408	4,5	18,6
Rural	5.253	59,1	9,5	3.493	38,8	11,6	2.866	31,8	12,9	4.473	49,7	10,3
Central	1.962	22,1	13,8	1.284	14,3	17,1	1.987	22,1	13,7	694	7,7	23,2
Chorotega	689	7,7	18,6	461	5,1	22,8	159	1,8	38,8	560	6,2	20,6
Pac. Central	577	6,5	27,2	269	3,0	39,8	64	0,7	81,5	624	6,9	26,1
Brunca	768	8,6	31,8	505	5,6	39,2	335	3,7	48,1	1.000	11,1	27,8
H. Atlántica	834	9,4	16,1	587	6,5	19,1	226	2,5	30,8	545	6,1	19,9
H. Norte	425	4,8	38,0	389	4,3	39,7	96	1,1	80,0	1.050	11,7	24,2
PUESTOS DE TRABAJO	8.895	100,0	2,6	9.000	100,0	2,6	9.000	100,0	2,6	9.000	100,0	2,6
Urbana	3.642	40,9	3,5	5.507	61,2	2,9	6.332	70,4	2,7	4.787	53,2	3,1
Central	2.216	24,9	4,6	4.375	48,6	3,2	6.049	67,2	2,8	986	11,0	6,8
Chorotega	366	4,1	17,5	265	2,9	20,6	82	0,9	37,0	932	10,4	11,0
Pac. Central	303	3,4	7,5	267	3,0	8,0	33	0,4	23,0	403	4,5	6,5
Brunca	288	3,2	15,3	219	2,4	17,5	43	0,5	39,5	739	8,2	9,5
H. Atlántica	249	2,8	16,9	305	3,4	15,3	112	1,2	25,3	966	10,7	8,6
H. Norte	220	2,5	26,8	76	0,8	45,7	14	0,2	106,1	762	8,5	14,4
Rural	5.253	59,1	3,9	3.493	38,8	4,8	2.668	29,6	5,5	4.213	46,8	4,3
Central	1.962	22,1	6,1	1.284	14,3	7,5	1.393	15,5	7,2	579	6,4	11,1
Chorotega	689	7,7	12,2	461	5,1	14,9	219	2,4	21,6	699	7,8	12,1
Pac. Central	577	6,5	12,9	269	3,0	18,8	81	0,9	34,3	562	6,2	13,0
Brunca	768	8,6	15,3	505	5,6	18,9	336	3,7	23,1	917	10,2	14,0
H. Atlántica	834	9,4	11,3	587	6,5	13,5	414	4,6	16,1	732	8,1	12,1
H. Norte	425	4,8	13,8	389	4,3	14,4	225	2,5	18,9	724	8,0	10,5
INGRESO PER CAPITA	8.895	100,0	3,1	9.000	100,0	3,1	9.000	100,0	3,1	9.000	100,0	3,1
Urbana	3.642	40,9	3,8	5.507	61,2	3,1	6.905	76,7	2,8	5.273	58,6	3,2
Central	2.216	24,9	4,7	4.375	48,6	3,3	5.097	56,6	3,1	1.687	18,7	5,4
Chorotega	366	4,1	9,5	265	2,9	11,1	437	4,9	8,7	837	9,3	6,3
Pac. Central	303	3,4	7,8	267	3,0	8,3	320	3,6	7,6	693	7,7	5,2
Brunca	288	3,2	9,4	219	2,4	10,8	406	4,5	7,9	758	8,4	5,8
H. Atlántica	249	2,8	9,6	305	3,4	8,7	518	5,8	6,7	915	10,2	5,0
H. Norte	220	2,5	8,1	76	0,8	13,8	126	1,4	10,7	383	4,3	6,1
Rural	5.253	59,1	3,4	3.493	38,8	4,2	2.095	23,3	5,4	3.727	41,4	4,0
Central	1.962	22,1	6,0	1.284	14,3	7,4	804	8,9	9,4	756	8,4	9,7
Chorotega	689	7,7	11,5	461	5,1	14,1	361	4,0	15,9	868	9,6	10,3
Pac. Central	577	6,5	9,2	269	3,0	13,5	214	2,4	15,2	532	5,9	9,6
Brunca	768	8,6	7,1	505	5,6	8,8	237	2,6	12,8	562	6,2	8,3
H. Atlántica	834	9,4	4,4	587	6,5	5,2	182	2,0	9,4	372	4,1	6,5
H. Norte	425	4,8	9,2	389	4,3	9,6	298	3,3	11,0	638	7,1	7,5

1/ Con base a una muestra de 9000 viviendas (tres réplicas)