

Documento metodológico Encuesta de Hogares de Propósitos Múltiples

Área de Censos y
Encuestas



Tabla de contenido

A. Introducción	3
B. Objetivos	3
C. Características del Diseño Muestral	3
1. Población de estudio	3
2. Marco muestral de viviendas	3
3. Dominios de Estudio	5
4. Diseño Muestral	5
5. Tamaño y distribución de la muestra	6
D. Sobre las estimaciones de los resultados	6
1. Cambio en la estimación de la población total	7
2. Procedimiento para la estimación de los valores poblacionales	7
E. Trabajo de campo	8
1. Organización y ejecución	8
2. Período de referencia	9
3. Duración del trabajo de campo	9
F. Procesamiento de los datos	9
1. Crítica y codificación	9
2. Entrada de datos	10
3. Limpieza de los datos	10
4. Creación de la Base de Datos	10
G. Sobre los resultados	10
1. Difusión y Publicaciones	10
2. Contenido temático	11
3. Sobre la comparabilidad y el uso de las series de datos 1987 - 1999 y 2000 en adelante	11
4. Confiabilidad de los datos	12

A. Introducción

La Encuesta de Hogares de Propósitos Múltiples (EHPM), que se realiza de forma permanente desde 1986, es un programa especializado en la recolección de información para la producción de estadísticas relacionadas con el empleo, el desempleo, los ingresos y otras características sociales y económicas de la población.

B. Objetivos

La Encuesta de Hogares de Propósitos Múltiples, realizada en julio de cada año, tiene como objetivos principales los siguientes: (a) mantener un flujo continuo de estadísticas relacionadas con la fuerza de trabajo, el empleo, el desempleo, el subempleo y los ingresos, y de otras variables socioeconómicas necesarias para el establecimiento de políticas y para la formulación de planes orientados al desarrollo económico y social del país, así como para la evaluación de sus efectos; (b) proveer información periódica, sistemática y oportuna en los períodos intercensales referente, entre otras, a las variables mencionadas; y (c) servir de fuente de información a instituciones gubernamentales, universitarias o de investigación, interesadas en temas relativos a la población y el empleo.

C. Características del Diseño Muestral

1. Población de estudio

La población de estudio de la Encuesta de Hogares de Propósitos Múltiples está definida como el conjunto de todas las viviendas particulares existentes en el país y de sus ocupantes, que son residentes habituales en esas viviendas. Se excluye del estudio a la población residente en las viviendas colectivas (cárceles, conventos, asilos, residencias colectivas para estudiantiles y trabajadores, hospitales y hoteles).

2. Marco muestral de viviendas

Los marcos muestrales de viviendas son los instrumentos utilizados para la selección de muestras de viviendas. Para seleccionar las muestras de las EHPM, entre 1987 y 1998 se empleó un marco muestral de viviendas construido en 1986 con la cartografía y la información de los Censos Nacionales de Población y Vivienda de 1984 (MMV-86).

Considerando que el crecimiento de la población desde esa fecha y los cambios ocurridos en su distribución geográfica, caracterizada especialmente por un proceso de creciente urbanización del país, había modificado la situación existente y por consiguiente provocado la desactualización del MMV-86, en 1997, la entonces Dirección General de Estadística y Censos inició acciones orientadas

a la elaboración de un nuevo marco muestral que contemplara apropiadamente el crecimiento en el número de viviendas y los cambios en la condición urbano-rural de las áreas o segmentos que lo componen.

Como parte de este proceso se construyó un nuevo marco (MMV-98), con base en la actualización cartográfica realizada en el período 1993-1997, actividad que formó parte de los preparativos del censo que finalmente se llevó a cabo en julio del año 2000¹. Este MMV presenta una nueva delimitación de segmentos, incluye un conteo del número de viviendas dentro de cada segmento y algunas variables que pueden ser utilizadas como criterios de estratificación: la región de planificación², el grado de urbanización (urbana, periferia urbana, rural concentrado y rural disperso), y un indicador del nivel socioeconómico (bajo, medio-bajo, medio, medio-alto, alto), construido a partir de una calificación visual de las viviendas y su entorno que se realizó para cada uno de los segmentos de la zona urbana y de la periferia urbana del país³.

El MMV de 1998 quedó constituido por 16 106 segmentos y 826 521 viviendas con la distribución por Regiones de Planificación que se aprecia en el cuadro 1. Con este nuevo marco se seleccionaron las muestras utilizadas en la encuesta de 1999 y las subsiguientes.

CUADRO 1

Número de segmentos y viviendas del Marco Muestral de Viviendas de 1998 por zona, según región de planificación

Región de planificación	Total		Zona			
	Segmentos	Viviendas	Urbana		Rural	
			Segmentos	Viviendas	Segmentos	Viviendas
Costa Rica	16 106	826 521	7 781	463 594	8 325	362 927
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	50 061	461	26 359	531	23 702
Brunca	1 452	64 715	273	16 653	1 179	48 062
Huetar Atlántica	1 669	78 349	439	25 936	1 230	52 413
Huetar Norte	905	40 598	116	7 199	789	33 399

Actualmente, se trabaja en la elaboración de un nuevo marco muestral que toma como referencia la cartografía y la información del Censo 2000, y que se espera aplicar en próximas encuestas, una vez concluida la preparación y el diseño de una nueva muestra.

La regionalización del país utilizada para esta Encuesta se ilustra en el mapa de la siguiente página.

¹ Lo deseable habría sido elaborar el nuevo Marco Muestral de Viviendas con los datos de un nuevo censo, no obstante, ante la incertidumbre respecto a la fecha en que finalmente se iba a realizar, se decidió elaborar el MMV-98 con base en las actualizaciones cartográficas realizadas entre 1993 y 1997.

² Según Decreto N° 160068-PLAN del 15 de febrero de 1985, La Gaceta N° 59 del 26 de marzo de 1985.

³ Asignación del Nivel Socioeconómico. Proceso de Muestreo, 1998.



3. Dominios de Estudio

Los dominios de estudio para los cuales se pueden obtener estimaciones con un nivel de confianza conocido son los siguientes:

- Región Central urbana
- Región Central rural
- Resto del país urbano
- Resto del país rural

Además la muestra permite obtener estimaciones de algunas características de la población, para las otras regiones de planificación: Pacífico Central, Chorotega, Brunca, Huetar Atlántica y Huetar Norte.

4. Diseño Muestral

El diseño muestral corresponde a un diseño probabilístico de áreas, estratificado y bietápico. Es de áreas debido a que las probabilidades de selección están asociadas a los segmentos censales los cuales son áreas geográficas debidamente delimitadas; es estratificado porque para la distribución y selección de la muestra se definieron doce estratos de interés - cada región de planificación dividida por zona urbana y rural - con la finalidad de tener una mejor representatividad de estas áreas y aumentar así la eficiencia

relativa del diseño; es bietápico ya que en una primera etapa se seleccionan segmentos censales o Unidades Primarias de Muestreo (UPM), y en una segunda etapa se seleccionan viviendas o Unidades Secundarias de Muestreo (USM) dentro de las UPM seleccionadas en la primera etapa.

El diseño muestral establece que la muestra sea autoponderada dentro de cada estrato. Por ello, como los segmentos se seleccionaron con igual probabilidad, para mantener la autoponderación se toma una fracción fija de viviendas en la segunda etapa de selección: un cuarto (1/4) en los segmentos urbanos y un tercio (1/3) en los segmentos rurales.

5. Tamaño y distribución de la muestra

El cálculo del tamaño de muestra se hace bajo el supuesto de muestreo aleatorio simple con un ajuste por efecto de diseño por tratarse de una muestra compleja⁴. El cálculo del tamaño de muestra y su distribución se efectuó de la siguiente manera:

a. Se procedió a calcular el tamaño de muestra necesario para obtener estimaciones de una tasa de desempleo del 5%, con un margen de error del 1%, un nivel de confianza del 95% y un efecto del diseño de 2,45.

b. Dado que se establecieron cuatro dominios de estudio (Región Central Urbana, Región Central Rural, Resto del País Urbano y Resto del País Rural), el tamaño de muestra para el total del país corresponde al cálculo anterior multiplicado por cuatro.

c. Considerando el interés del país en la obtención de información sobre el trabajo de los menores por medio de la aplicación de módulos especiales, se asignó o distribuyó la muestra nacional por zona urbana y rural de acuerdo a la variabilidad de la tasa de participación económica de la población de 5 a 17 años, estimada a partir del módulo de trabajo infantil aplicado en el año 1995.

d. La muestra resultante en cada zona, se distribuyó por región, de acuerdo a la variabilidad relativa del número de desempleados.

De esta forma se obtuvo un tamaño de muestra inicial de 10 890 viviendas, el cual ha aumentado por el crecimiento en el número de viviendas. La muestra se distribuye entre los diferentes estratos de la siguiente forma:

D. Sobre las estimaciones de los resultados

⁴ Una muestra se considera compleja cuando utiliza más de una etapa de selección, como en el caso de la EHPM.

1. Cambio en la estimación de la población total

Los datos que actualmente suministra la Encuesta de Hogares de Propósitos Múltiples (EHPM), son estimaciones que se obtienen a partir de una muestra de viviendas, seleccionada de un marco muestral elaborado en el año 1998.

Es común que las estimaciones obtenidas en este tipo de investigaciones por muestreo deban someterse a revisiones periódicas con el fin de identificar y corregir problemas de subestimación de la población, que se generan en el transcurrir del tiempo, como consecuencia de los cambios demográficos y de la distribución espacial de la población (i. e. en la estructura urbano-rural y regional).

En el año 2001, al disponerse de los resultados censales, se determinó que al año 2000 (año del censo), la EHPM presentaba una subestimación de la población total de un 9,3%; pues estimó una población de 3 486 048, mientras que el Censo de Población enumeró un total de 3 810 179 personas en todo el país.

Además se observó una modificación en la estructura urbano rural de la encuesta, la cual estaba sobreestimando la población rural y subestimando la urbana.

Esta situación motivó a recalcular las estimaciones derivadas de la EHPM para el año 2000 haciendo coincidir los totales de población del Censo y de la Encuesta, tanto para cada región de planificación, como para las zonas urbana y rural del país.

Para las estimaciones del año 2001 y subsiguientes se aplica el mismo procedimiento, utilizando como población de referencia el cálculo obtenido de sumar los nacimientos y restar las defunciones del período julio - junio de cada año, a las cifras del Censo 2000 y aplicar un ajuste por migración.

Este cambio en el procedimiento de estimación afecta la comparación de los resultados con los de los años anteriores al 2000, por eso, se recomienda al lector, leer la sección F.2 “Sobre la comparabilidad y uso de las series de datos 1987-1999 y 2000-2003”.

2. Procedimiento para la estimación de los valores poblacionales

El procedimiento de estimación de la Encuesta consiste en multiplicar el resultado de la muestra por un factor de expansión. Este procedimiento se aplica para cada uno de los doce estratos de selección (las seis regiones por zona urbana y rural) de la siguiente forma:

a) Factor de expansión inicial: se calcula el factor de expansión como el inverso de la fracción final de muestreo para cada estrato. La fracción final de muestreo es el producto de la fracción de muestreo en la primera etapa y la fracción de muestreo en la segunda etapa de selección.

b) **Ajuste por crecimiento:** se calcula un ajuste por crecimiento de viviendas que consiste en la razón del número de viviendas en el marco actualizado por el crecimiento observado en los segmentos o UPMs de la muestra (número inicial más las nuevas viviendas) entre el número original de viviendas en el marco.

c) **Ajuste por no respuesta:** se calcula un ajuste por no respuesta que consiste en la razón del número total de viviendas seleccionadas sujetas a entrevista entre el número de estas viviendas que efectivamente respondieron la entrevista.

d) **Ajuste de población:** se hace con el fin de corregir la subestimación mencionada en el apartado I, punto D.1. Este ajuste consiste en la razón de la población proyectada con base en la información censal y la estimación de la encuesta.

e) **Factor de expansión final:** el factor de expansión final se obtiene como el producto del factor de expansión inicial, el ajuste de crecimiento, el ajuste por no respuesta y el ajuste de población.

Para obtener la estimación de población para las diferentes variables que investiga la encuesta, se multiplica la variable o característica de cada persona por su correspondiente factor de expansión, y el resultado se agrega para cada ámbito geográfico: región, zona urbana o rural y nacional.

En el caso de las estimaciones sobre los hogares, el factor de expansión que se aplica a un hogar es el factor de expansión correspondiente al jefe de ese hogar. Las características del hogar se multiplican por este factor de expansión y el resultado se agrega de igual forma para cada ámbito geográfico.

E. Trabajo de campo

1. Organización y ejecución

Para la ejecución del trabajo de campo se contratan cerca de 66 entrevistadores, quienes recibieron capacitación durante las dos últimas semanas del mes de junio de cada año.

El trabajo de campo se organiza en dos etapas. En la primera se recolecta la información de la Región Central, formada por el Área Metropolitana de San José y zonas circunvecinas de las provincias de Alajuela, Cartago y Heredia, dividida para efectos del trabajo de campo en 22 zonas.

La segunda etapa abarca el resto del país, dividido en 20 zonas.

En cada zona se cuenta con un equipo de trabajo compuesto por un supervisor, tres entrevistadores, un chofer y un vehículo asignado durante el período de recolección de los datos. Además, durante la segunda etapa dos grupos de trabajo permanecen en la Región Central para realizar las entrevistas pendientes de la primera etapa.

2. Período de referencia

El período de referencia para captar la información sobre empleo es la semana anterior a la fecha de la entrevista, en tanto que la captación de la información sobre desempleo se refiere a las cinco semanas anteriores a tal fecha.

3. Duración del trabajo de campo

La recolección de datos se efectúa en dos etapas:

Primera etapa: segunda y tercera semana del mes de julio.

Segunda etapa: cuarta semana del mes de julio y primera semana del mes de agosto.

F. Procesamiento de los datos

El procesamiento de la encuesta se inicia una vez recolectada la información de la Región Central, es decir, dos semanas después del comienzo del trabajo de campo y se extiende por un período de tres meses hasta la difusión de los primeros resultados definitivos.

El proceso está compuesto por tres grandes etapas: la crítica y codificación de las boletas, la entrada de datos y la limpieza y depuración de los mismos. Una vez concluidas estas fases se procede a la creación de variables y a la tabulación de los resultados.

1. Crítica y codificación

La crítica consiste en revisar manualmente, boleta por boleta, la información que viene del campo recolectada por los entrevistadores, para detectar errores generales o faltantes de datos, que de ser posible son corregidos o completados (por medio de llamadas telefónicas), de tal forma que asegure una buena calidad de la información y se faciliten los siguientes procesos.

La codificación implica asignar códigos numéricos a las respuestas que no están precodificadas, de tal forma que se facilite la digitación. Por ejemplo, la ocupación y el país de nacimiento, son variables que requieren pasar por este proceso.

Este proceso inicia a las dos semanas de haber empezado el trabajo de campo y es finalizado en cinco semanas.

2. Entrada de datos

Conforme la información de las boletas de la EHPM va siendo criticada y codificada, se procede con la etapa de digitación, que consiste en la entrada de los datos por medio de una red de microcomputadoras a una base de datos. Esta fase dio inicio a la semana y media de empezada la Crítica y finaliza en cuatro semanas.

En la captura de la información se utiliza el programa CPro, específicamente el módulo CENTRY (Census Data Entry), el cual además permite la verificación, corrección y modificación de los datos.

3. Limpieza de los datos

La limpieza o depuración de los datos consiste en la detección o corrección de inconsistencias o errores entre los datos de un mismo cuestionario. Para esto se debe contar con un Plan de Inconsistencias que especifica cuales pueden ser los errores que se encuentren al combinar preguntas. De encontrarse un error la boleta debe ser revisada manualmente y el dato corregido según corresponda.

La detección de los errores se realiza por medio de un programa informático basado en el Plan de Inconsistencias y elaborado en un módulo que para estos afectos tiene el CPro. El programa genera en pantalla un listado con los errores, los cuales son revisados por los mismos supervisores de la encuesta en las boletas y corregidos directamente en pantalla.

4. Creación de la Base de Datos

Para la obtención de la base de datos final es necesario realizar procedimientos específicos para la creación de variables y validación de la información. En este proceso se utiliza el Plan de Creación de variables que contempla las variables básicas para poder generar los tabulados propuestos.

Finalmente se genera un archivo de datos que puede ser trabajado en programas para generación de estadísticas como IMPS, SPSS y otros similares.

G. Sobre los resultados

1. Difusión y Publicaciones

La información obtenida de la Encuesta de Hogares se presenta al público, básicamente a través de tres fuentes, estas son los medios de comunicación social, medios impresos (boletines, tabulados y publicaciones) y también por medios electrónicos.

Los resultados básicos son presentados a los medios de comunicación a través de una conferencia de prensa. Luego de lo cual se observa cobertura en prensa, radio y televisión.

Respecto a medios impresos, se producen dos boletines y una publicación con datos sobre aspectos metodológicos, conceptos y definiciones, además de gráficos y cuadros que describen los principales resultados

1. Boletín Cifras Básicas sobre Pobreza e Ingresos
2. Boletín Cifras Básicas sobre Fuerza de Trabajo
3. Publicación titulada “Encuesta de Hogares de Propósitos Múltiples, Principales Resultados”

Estos boletines y la publicación están disponibles en el Centro de Información, en forma impresa, en disquete y en la página web del Instituto (www.inec.go.cr).

Adicionalmente se generan alrededor de 120 cuadros contemplados en el Plan Básico de Tabulados, los cuales están a disposición de los usuarios en el Centro de Información del INEC.

2. Contenido temático

Para el contenido de las publicaciones se seleccionan temas asociados a la población total según los diferentes grupos de edad, fuerza de trabajo, empleo, desempleo, subempleo, ingresos, pobreza, inactividad y niveles de escolaridad. Se incluyen así mismo otras variables a nivel de hogar, tales como el número de miembros, ocupados, jefatura femenina, ingreso per cápita y distribución porcentual del ingreso, entre otros. La información permite además caracterizar a la población económicamente activa según el sector institucional, la categoría ocupacional, el número de horas trabajadas, el grupo ocupacional y el sector o rama de actividad económica. En general las variables se encuentran desagregadas por sexo, por región de planificación y por zona urbana y rural del país.

3. Sobre la comparabilidad y el uso de las series de datos 1987- 1999 y 2000 en adelante

Los cambios efectuados en la estimación de la población se han aplicado a las Encuestas a partir del 2000, por ello, las series de datos correspondientes a los períodos 1987-1999 y 2000 en adelante no son comparables en sus valores absolutos. Sí se considera factible comparar las tasas de participación económica (ocupación, neta de participación, desempleo y subempleo) así como los porcentajes de hogares en situación de pobreza, debido a que el análisis realizado con los datos del 2000, antes y después de aplicar la nueva ponderación, mostró que estos indicadores no presentaban cambios de magnitud importante. Sin embargo, no se recomienda la comparación de las variables relacionadas con los ingresos, los sectores de actividad y las ocupaciones, debido a que sí mostraron ser más susceptibles al cambio en las ponderaciones urbano- rural.

Igualmente importante es tener presente que las cifras recalculadas para el año 2000, difieren de las publicadas por el INEC en boletines, tabulados y otras publicaciones editadas antes de efectuar el mencionado ajuste, específicamente en las ediciones

anteriores al 2001, razón por la cual se sugiere (en caso de trabajar con cifras del 2000) asegurarse de utilizar las cifras recalculadas para ese año.

Otro aspecto que se desea destacar es que a partir de la Encuesta de Hogares del 2001, se adoptaron las más recientes nomenclaturas internacionales para la clasificación de las variables económicas.

4. Confiabilidad de los datos

La muestra utilizada en la encuesta es una de todas las muestras posibles del mismo tamaño que podrían haberse seleccionado utilizando el mismo diseño muestral; cada muestra proporciona una estimación del valor poblacional que se desea conocer. El error de muestreo representa una medida de la dispersión o variabilidad de las estimaciones de todas las muestras posibles con respecto a ese valor poblacional que se desea estimar y es de hecho una medida de la precisión de las estimaciones.

Debido a que se selecciona sólo una muestra de todas las muestras posibles, el error de muestreo no es posible calcularlo directamente sino que se estima por medio del error estándar que es una medida del nivel de precisión de las estimaciones de la encuesta. En este sentido, se recomienda consultar la publicación Principales Resultados de la Encuesta de Hogares, apartado III sobre los resultados de los “errores de muestreo” tales como coeficiente de variación, error estándar y límites de confianza los cuales han sido calculados para las principales variables.

Para analizar las estimaciones, es importante tomar en consideración que las celdas con valores pequeños representan características poco frecuentes en la población y, por tanto, los resultados deben ser tratados con cuidado.