

M7/277.10  
Ej 2

**DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADISTICA**  
**DIRECCION GENERAL DE ANALISIS SOCIO-ECONOMICO**

doc. 3

**EL MUESTREO EN EL XV CENSO DE POBLACION Y**  
**IV DE VIVIENDA COLOMBIA 1985**

Preparado por:  
**NELLY SOFIA DUQUE**  
**DIVISION DE CENSOS Y**  
**PROYECTOS ESPECIFICOS**

Tema 2

**BOGOTA, ABRIL DE 1989**

# I N D I C E

	Pág.
INTRODUCCION	4
ASPECTO TECNICOS DE LA MUESTRA DE AMPLIACION	5
1    Objetivos Generales de la Muestra	5
2    Objetivos especificos de la muestra	5
3    Definición de variables y formulario	5
4    Método de muestreo	6
5    Tamaño de la muestra	7
5.1 Variables de la Vivienda	7
5.2 Variables de Población	8
5.3 Evaluación del tamaño de Muestra propuesta	10
6    METODO DE SELECCION	10
6.1 Selección urbana	10
6.2 Selección rural	11
7    CALCULO DE LOS PESOS PARA LA MUESTRA DE AMPLIACION	12
7.1 Pesos para la unidades de vivienda	12
7.2 Pesos para las personas	12
METODOLOGIA DE LA MUESTRA DE COBERTURA	14
1    Objetivos de la Muestra Cobertura	14
2    Metodología general	14
3    Diseño de la muestra	15
4    Tamaño de la muestra	16

5	Selección de municipios	17
6	Selección de segmentos de cabeceras	17
7	Selección de segmentos en el resto	18
8	Contenido de la encuesta	19
9	Pareamiento preliminar	19
10	Pareamiento definitivo	20
11	Método general para el cálculo de cobertura	20

Este documento se preparó para ser  
presentado en el seminario sobre "USO DEL  
MUESTREO EN LAS ETAPAS PREPARACION,  
UTILIZACION INFORMACION CENSAL".

## I N T R O D U C C I O N

Para la realización del XV Censo de Población y IV de Vivienda de 1985 se optó por la utilización de una nueva estrategia, apoyada fundamentalmente en el uso de la técnica estadística del Muestreo.

La estrategia consistió en aplicar dos tipos de formularios. Un formulario básico que contiene las características mínimas de la población y de la vivienda, el cual se aplicó a toda la población y un formulario ampliado con características más detalladas, que se aplicó solo a una parte de la población.

Para la aplicación del segundo formulario fue necesario diseñar la denominada Muestra de Ampliación, que tenía como objetivo obtener información detallada sobre características de población y vivienda, que no se contemplaba en el formulario básico, sin incurrir en un gran volumen de formularios, pero sí garantizando una representatividad de la información recogida.

Con el fin de estimar la omisión en que pueda incurrir el Censo, se diseñó la Muestra de Cobertura esta estimación permite en primera instancia evaluar el operativo censal, y también suministrar factores cuantitativos de ajuste para la información censal.

Para la realización de estimaciones a partir de información muestral, un requisito fundamental es que la muestra haya sido seleccionada probabilísticamente. Esto permite calcular los denominados errores de muestreo, mediante los cuales se mide la precisión de las estimaciones.

En este documento se presentan procedimiento y metodologías involucrados en la planeación, diseño y ejecución de los dos tipos de muestra mencionados.

## ASPECTOS TECNICOS DE LA MUESTRA DE AMPLIACION

### 1. OBJETIVOS GENERALES DE LA MUESTRA

- a) Generar información tradicionalmente censal, con mayor oportunidad, a un menor costo.
- b) Recoger información sobre características de la vivienda y de comportamiento demográfico y económico de la población, con mayor detalle que el contemplado en el formulario básico censal.
- c) Suministrar estimadores razonablemente precisos para áreas administrativas más amplias como son los municipios.
- d) Suministrar estimadores para subgrupos de 15.000 habitantes o más, con un error de más o menos el 4% basado en un nivel de confianza del 95%.

### 2. OBJETIVOS ESPECIFICOS DE LA MUESTRA

- a) Estimar el comportamiento de los tres componentes del crecimiento demográfico: Fecundidad, Mortalidad, Migración.
- b) Estimar las características básicas de la población económicamente activa, ocupados, desocupados.
- c) Proveer con mayor detalle que en la información censal, estimaciones sobre características de la vivienda.

### 3. DEFINICION DE VARIABLES Y FORMULARIO

Con base en los objetivos específicos de la muestra, la organización censal llegó a la conformación de un formulario, el cual comprende las variables a ser investigadas, en el universo de viviendas y en el de la población. Este comprende cinco capítulos así:

Capítulo I. **Ubicación geográfica:** Con el cual el empadronador se ubicará en un determinada

manzana (urbana) o en un sector (rural) del municipio.

Capítulo II. **Identificación de la vivienda:** para ubicar una vivienda dentro de la manzana o sector.

Capítulo III. **Datos de la vivienda:** Indaga por las características principales de la vivienda, materiales con que está constituida, conexión con los servicios públicos, etc.

Capítulo IV. **Total de personas:** El cual registra las personas que son residentes habituales de la vivienda.

Capítulo V. **Datos de Población:** Para cada una de las personas registradas en la vivienda se indagan cerca de 25 ítems.

#### 4. METODO DE MUESTREO

El método de muestreo utilizado es el sistemático de viviendas, con recolección CO-CENSAL, es decir, simultáneamente a la recolección censal.

Las principales razones para la escogencia de este método se pueden resumir así:

- a) Utilizar como marco de muestra los recuentos precensales, con las ventajas de actualidad que ello representa.
- b) Aprovechar la estructura organizacional del censo básico para la recolección.
- c) Obtener una muestra de tamaño y dispersión tal, que garantice una buena representatividad a niveles grandes de desagregación (municipios pequeños y principales sectores de las grandes ciudades).

## 5. TAMAÑO DE LA MUESTRA

El tamaño de una muestra está determinado por diversos factores:

- Número de elementos en la población (N).
- Varianza de la variable en los elementos de la población ( $\sigma$ ).
- Definición de la probabilidad con que se define aceptar las estimaciones (K).
- Error de muestreo (e).
- Los recursos financieros existentes (\*).

Dado que con el diseño de muestreo se busca optimizar los recursos existentes, estos factores se analizan según su influencia en el error de muestreo.

### 5.1 Variables de la Vivienda

Para el caso de viviendas, el muestreo sistemático se puede asimilar a un muestreo aleatorio simple donde la unidad de selección es la vivienda, y donde para estimar el porcentaje de elementos que de un universo tiene un atributo cualquiera, se tiene que:

$$p = \frac{\sum_j^n X_j}{n} \quad \text{Donde } X \text{ es un estimador de } P$$

$$P = \frac{\sum_i^N X_i}{N}$$

Donde:



- $N$  = Número de elementos en la población de estudio.
- $X_i$  = Es el valor de la variable  $x$  para el elemento  $i$ -ésimo.
- $N_i$  = Indica la suma sobre todos los elementos de la población ( $i = 1, 2, \dots, N$ ).
- $P$  = Proporción de elementos que poseen el atributo  $x$  en la población.
- $n_j$  = Elemento  $j$ -ésimo en la muestra.
- $n$  = Total de elementos en la muestra.
- $X_j$  = Valor de la variable  $x$  para el elemento  $j$ -ésimo.

Donde:

- $X_j$  = 1 si el individuo tiene el atributo  
0 si el individuo no tiene atributo
- $p$  = Proporción de elementos que poseen el atributo  $x$ , en la muestra.
- $n_j$  = Indica la suma sobre todos los elementos de la muestra ( $J = 1, 2, \dots, n$ ).

## 5.2 Variables de Población

En el caso de los datos sobre población, el diseño de la muestra corresponde a un muestreo de conglomerados de tamaño desigual, en donde las unidades de observación son todas las personas residentes habituales de la vivienda.

Como la vivienda es la unidad de selección para la muestra sistemática del 10%, el conglomerado estará constituido por el conjunto de las personas residentes en cada una de las viviendas seleccionadas.

La estimación de una característica de la población se enmarca, por lo general, en estimadores del tipo tasas, debido a que el volumen de personas por viviendas es una variable aleatoria. Para el cálculo de la varianza de una tasa, en muestreo de conglomerados se utiliza la siguiente fórmula:

$$V(r) = \frac{1}{2} \frac{1}{X} \left[ \text{Var}(y) + r^2 \text{Var}(x) - 2r \text{COV}(y, X) \right]$$

Donde:

$$r = \frac{y}{x} = \frac{\sum_{a=1}^a y_a}{\sum_{a=1}^a x_a} \quad = 1, 2, \dots, a$$

$a$  = Total de conglomerados en la muestra

$y_a$  = Suma de los valores de la variable aleatoria  $Y$  en el  $a$ -ésimo conglomerado

$x_a$  = Suma de los valores de la variable aleatoria  $X$  en el  $a$ -ésimo conglomerado

$Y$  = Suma de los valores de la variable aleatoria  $Y$

$X$  = Suma de los valores de la variable aleatoria  $X$

VAR(Y) = Varianza entre conglomerados respecto a la variable Y

$$\text{Var}(y) = \frac{1}{a} \sum_a y^2 - \bar{y}^2$$

VAR(X) = Varianza entre conglomerados respecto a la variable X

$$\text{Var}(x) = \frac{1}{a} \sum_a x^2 - \bar{x}^2$$

### 5.3 Evaluación del tamaño de muestra propuesto

Con base en el marco teórico anterior, los resultados obtenidos en los estudios de población en Colombia (1980-1982), se concluye que un tamaño de muestra del 10% de población basados en niveles de confianza del 95%, permite estimadores para subgrupos de 15.000 unidades o más, con un razonable error de muestreo en la estimación.

## 6. METODO DE SELECCION

La selección se realizó, sistemáticamente escogiendo una de cada 10 viviendas censales, a partir de un arranque aleatorio, tanto en lo urbano, como en lo rural. A la vivienda censal y a todos los individuos residentes habituales de la misma, se les aplicó el formulario ampliado.

### 6.1 Selección urbana

Con base en el listado de recuento preliminar de viviendas particulares del censo, el Jefe de cada Centro de Operaciones seleccionó sistemáticamente 1 de cada 10 viviendas, a partir de un arranque aleatorio, entre 1 y

10, previamente seleccionado al azar.

El operativo censal contó con dos tipos de recolectores, uno para el formulario básico y otro para el ampliado. A las viviendas seleccionadas se les transcribió la información de identificación a otro formato, a partir del cual el recolector asignado, aplicó a dichas viviendas el formulario denominado ampliado.

En caso de viviendas totalmente omitidas en el recuento previo, se construyó con todas ellas un listado aparte, para seleccionar en él la muestra del 10%, con un arranque aleatorio.

## 6.2 Selección Rural

La selección de muestra rural con los mismos pasos metodológicos que los realizados en la urbana, implicaría altísimos costos en razón de la dispersión de las viviendas en el área rural, la desactualización cartográfica existente, y el tiempo disponibles para llevar a cabo la selección; se optó por llevar aquí un operativo cocensal con las siguientes actividades:

- a) Asignación de empadronadores respetando los límites de sección definidos en la cartografía existente.
- b) El empadronador censal recogió simultáneamente el formulario básico y ampliado.
- c) La selección de viviendas para el ampliado se efectuó sobre el terreno, seleccionando 1 de cada 10 viviendas censales, a partir del arranque aleatorio que le fue asignado al empadronador.
- d) El supervisor del trabajo de campo, realizó una revisión detallada del sistema de recorrido y de selección de la muestra para controlar que el empadronador introdujera sesgos en la selección de la muestra.
- e) El empadronador colocó en la cartografía y en el formulario de recuento, el número de la vivienda que se lista en el recorrido. Esta numeración ayuda a la identificación de viviendas en la muestra de cobertura.

## 7. CALCULO DE LOS PESOS PARA LA MUESTRA DE AMPLIACION

Puesto que las estimaciones de la muestra de ampliación incluyen tanto características de vivienda, como de personas, se calculan pesos (factores de expansión) diferentes para cada una. Se determinan pesos diferentes para unidades de vivienda y personas, en municipios o grupos de municipios de por lo menos 15.000 habitantes. En el caso del factor expansión para las personas, se calcularán proporciones (tasas) diferentes para cada grupo de edad y sexo dentro de estas áreas geográficas, de manera que las estimaciones de población de la muestra de ampliación para cada una de estas celdas post-estratificadas coincidan con las correspondientes cifras del censo con enumeración 100%.

### 7.1 Pesos para las unidades de vivienda

Se determina un sólo peso para todas las unidades de vivienda en cada área geográfica. Este peso se calcularía así:

$$W_{(H)} = \frac{M_i}{m_i}$$

Siendo:

$W_{(H)}$  = El peso asignado a cada unidad de vivienda en la  $i$ -ésima área geográfica.

$M_i$  = Número total de unidades de vivienda en la  $i$ -ésima área geográfica, con base en el 100% de la enumeración del censo.

$m_i$  = Número de unidades de vivienda, en la muestra de ampliación para la  $i$ -ésima área geográfica.

### 7.2 Pesos para las personas

En el caso de estimaciones de características de población de la muestra de ampliación, el procedimiento

de estimación implica la asignación de un peso individual a cada persona en la muestra, con base en los grupos de edad y sexo dentro del área geográfica. Los pesos se determinan de modo que las estimaciones por muestra de ampliación de la población, en cada grupo de edad y sexo dentro del área geográfica sean iguales a las correspondientes cifras del censo. Los 16 grupos de sexo y edad son:

Hombres y Mujeres según los siguientes grupos de edad:

0 - 4 años	25 - 34
5 - 9	35 - 44
10 - 14	45 - 60
15 - 24	60 y más

El peso para las personas en cada grupo dentro de un área geográfica se calcula así:

$$W_{(p)ij} = \frac{N_{ij}}{n_{ij}}$$

Siendo:

$W_{(p)ij}$  = Peso asignado a una persona de la muestra en el  $i$ -ésimo grupo de edad-sexo en la  $i$ -ésima área geográfica.

$N_{ij}$  = Número total de personas en el  $i$ -ésimo grupo de edad y sexo en la  $i$ -ésima área geográfica, con base en la enumeración 100% del censo.

$n_{ij}$  = Número de personas de la muestra en el  $i$ -ésimo grupo de edad y sexo en la  $i$ -ésima área geográfica.

Con base en los datos del Censo 1973, puede verse que a nivel nacional, el grupo más pequeño de edad y sexo, son los hombres de 60 o más años, que representan el 2.4% de la población. Puesto que por definición, las áreas geográficas para las cuales se calcularán pesos diferentes, tienen una población de 15,000 o más, el grupo más pequeño estimado tendría, por lo menos, alrededor de 360 personas, lo cual parece razonable para este tipo de estimación de razón.

## METODOLOGIA DE LA MUESTRA DE COBERTURA

### 1. OBJETIVOS DE LA MUESTRA DE COBERTURA

Calcular las tasas de cobertura del Censo de Población y Vivienda de 1985.

Para la población residente en viviendas evaluando la:

- Cobertura de viviendas
- Cobertura de población por grupos de edad y sexo

Se requieren estimaciones confiables a los siguientes niveles de desagregación geográfica:

- Total Nacional
- Nacional, cabecera y resto
- Regional total (5 regiones)
- Regional, cabecera y resto
- Cuatro principales ciudades (Bogotá, Medellín, Cali y Barranquilla)
- Departamental total

### 2. METODOLOGIA GENERAL

- Selección de una muestra probabilística de áreas representativa de cada departamento y las cuatro principales ciudades.
- Desarrollo de una encuesta post-censal en cada área seleccionada, utilizando las mismas definiciones censales y garantizando al máximo la independencia entre el Censo y la Encuesta.
- Pareamiento uno a uno de viviendas y personas enumeradas en el Censo y la Muestra para detectar las omisiones censales.

- Revisita a las áreas con pareamientos dudosos
- Cálculo de las tasas de cobertura y sus errores de muestreo para obtener estimaciones de las tasas de cobertura y población ajustadas a los niveles de desagregación geográfica que resulten confiables.

### 3. DISEÑO DE LA MUESTRA

- Muestra probabilística de áreas constituidas por manzanas en las cabeceras municipales y secciones censales en el resto.
- Cada área está constituida por un segmento de 25 viviendas en promedio.
- La selección de cada área o segmento de muestreo se hace en varias etapas en forma independiente para cada departamento y separando entre cabeceras municipales y el resto.
- Cada área está constituida por un segmento de 25 viviendas en promedio.
- La selección de cada área o segmento de muestreo se hace en varias etapas en forma independiente para cada departamento y separando entre cabeceras municipales y el resto.
- Las etapas de selección para cada área son:
  - \* Selección de municipios (U P M)
  - \* Selección de sectores
  - \* Selección de secciones
  - \* Selección de manzanas (en cabeceras)
  - \* Submuestreo en la sección (en resto)
  - \* Submuestreo en la manzana (en cabeceras)



#### 4. TAMAÑO DE LA MUESTRA

- Si se supone que la tasa de no cobertura es de 5%, se trata de estimar una proporción ( $P = 5\%$ ).
- La muestra estará constituida por segmentos de 25 viviendas.
- Teniendo en cuenta estos diseños, si se toma una muestra de 1.000 viviendas (40 segmentos) por departamento, los errores de muestreo se estiman en:

Tipo de Estimación	Supuesto de No. Cobertura	Error Relativo de Muestreo	Intervalo de Confianza (95%)	
Viviendas	5%	17.4%	3,3%	6,7%
Personas	5%	11.03%	3,93%	6,08%

- Si las estimaciones de no cobertura de viviendas se hacen en las 5 regiones tradicionales, los errores de muestreo esperados son los siguientes:

Regiones	Departamentos	Viviendas en la Muestra	Error Relativo	Intervalo de Confianza 95%
Atlántica	7	7.000	5.21%	4.5 - 5.5
Central*	6	7.000	5.21%	4.5 - 5.5
Oriental	6	6.000	6.63%	4.4 - 5.6
Pacífica**	4	5.000	6.17%	4.4 - 5.6
Bogotá D.E.	1	2.000	9.75%	4.0 - 5.6
Total Nacional	24	27.000	2.70%	4.7 - 5.3

\* Incluye Medellín con 1.000 viviendas

\*\* Incluye Cali con 1.000 viviendas

## 5. SELECCION DE MUNICIPIOS

- Los 40 segmentos a seleccionar en cada departamento se distribuyen entre cabecera y resto en forma proporcional a la población de cada área en el departamento.
- Dentro de cada área se seleccionan al azar los municipios según el número de segmentos que le corresponda y en forma proporcional a la población proyectada de cada municipio.

## 6. SELECCION DE SEGMENTOS DE CABECERAS

- a) Para la selección en CABECERAS, se tomó como marco de muestreo el más actualizado disponible de los siguientes recuentos:
  - Recuento Procensal de viviendas
  - Estudios de población 1980 - 1983
  - Censo de 1973
- b) La selección de la manzana a visitar, se hizo al azar y en forma proporcional al número de viviendas en tres etapas, así:
  - Selección de sectores
  - Selección de secciones
  - Selección de manzanas
- c) En cada manzana seleccionada se elabora un recuento previo de viviendas, para realizar un submuestreo al azar del segmento final de alrededor de 25 viviendas contiguas a encuestar, así:

No. de Viviendas en la Manzana	Fracción de Submuestreo
Menos de 40	1
41 - 61	1/2
62 - 88	1/3
89 - 111	1/4
112 - 138	1/5
139 - 161	1/6

## 7. SELECCION DE SEGMENTOS EN EL RESTO

- a) Se procesaron los datos del Censo de 1973 para construir un marco de muestra a nivel de sector y sección rural y su número de viviendas.
- b) Selección al azar de secciones "rurales" en forma proporcional al número de viviendas y en dos etapas, así:
  - Selección de sectores
  - Selección de secciones
- c) División de la sección seleccionada en tantas áreas de muestreo de igual superficie como segmentos de 25 viviendas contiene la sección.
- d) Selección al azar de una de las áreas de muestreo y ubicación aleatoria del punto inicial para el trabajo de campo.
- e) A partir del punto inicial, el encuestador realiza un barrido del área de muestreo hasta completar 40 viviendas, o recorrer toda el área asignada sin sobrepasar 5 kilómetros cuadrados.

## 8. CONTENIDO DE LA ENCUESTA

- a) El trabajo de campo se realiza con dos formularios básicos a saber:
- Informe de Recorrido: con la identificación detallada y ordenada de las viviendas encuestadas.
  - Formulario de Cobertura: con la información requerida para la medición de cobertura de viviendas y población.
- b) El formulario de cobertura está dividido en cinco capítulos así:
- Capítulo 1: Con la ubicación geográfica del segmento
  - Capítulo 2: Con la identificación detallada de la vivienda.
  - Capítulo 3: Datos de la vivienda
  - Capítulo 4: Residentes habituales en el momento de la encuesta de cobertura y su clasificación entre no migrantes e inmigrantes entre el Censo y la Encuesta.
  - Capítulo 5: Emigrantes de la vivienda entre el censo y la encuesta de cobertura.

## 9. PAREAMIENTO PRELIMINAR

Para cada segmento se localizan las viviendas enumeradas en el censo y se hace un cotejo, uno a uno, entre los datos del censo y la encuesta de cobertura, así:

- Pareo de viviendas: Por direcciones y composición de la población
- Pareo de personas: Según nombres, sexo y edad

Para el pareo de edad se permiten las siguientes tolerancias:

Edad	Tolerancia
Hasta 9 años	- 1 + 1
10 - 19	- 2 + 2
20 - 39	- 6 + 6
40 - 59	- 10 + 10

Para las viviendas y personas que no pueden parearse, se diligencia un formulario de revisita que explica con detalle las discrepancias entre Censo y Cobertura que deben aclararse con una nueva visita a las viviendas.

#### 10. PAREAMIENTO DEFINITIVO

Con base en los resultados de la revisita, se hace el pareamiento definitivo de las viviendas y personas para clasificarlas en alguna de las siguientes categorías:

- Pareados
- Localización geográfica incorrecta en el Censo
- Duplicadas en el Censo
- Enumeradas en el Censo, pero con información insuficiente para el pareamiento.

#### 11. METODO GENERAL PARA EL CALCULO DE COBERTURA

El método general de estimación de la cobertura y la población corresponde a la teoría del sistema dual de estimación así:

$$\hat{P} = \frac{(\hat{N}_c - \hat{G} - \hat{E} - \hat{D} - \hat{I}) \times \hat{N}}{\hat{M}}$$

Donde:

- $\hat{P}$  = Estimación de la población total
- $\hat{N}_c$  = Población enumerada en el Censo
- $\hat{G}$  = Estimación de la población con localización geográfica incorrecta
- $\hat{E}$  = Estimación de la población erróneamente incluida en Censo
- $\hat{D}$  = Estimación de la población duplicada en Censo
- $\hat{I}$  = Estimación de la población con información insuficiente para el Pareo
- $\hat{N}$  = Estimación de la población que parea (incluida en el censo y en la Muestra de Cobertura)
- $\hat{M}$  = Estimación de la población total a partir de la muestra de cobertura

La tasa de cobertura c) es:  $C = \frac{\hat{M}}{\hat{N}}$