



Minnesota Population Center
University of Minnesota

Notice to all users:

This document was scanned and processed at the University of Minnesota's [Minnesota Population Center](#). The document is part of the Center's [IPUMS International](#) project initiative (NSF SBR-9907416) to preserve and disseminate census microdata and metadata, including enumeration forms, instructions, codebooks, and similar documentation. Most of these materials were scanned from the archives of the [United Nations Statistics Division](#). We are grateful to the Statistical Division for permission to construct an electronic archive of this invaluable collection. For more information on the Center or the IPUMS International projects, please click on the highlighted names above.

This document was scanned using a [Fujitsu M4097D scanner](#) at 400 dpi in 1-bit mode. The documentation was processed through either Adobe's [Acrobat Capture 3.0](#) or ParaVision/[XMLCities PageGenie Pro 4.0](#) software, and then edited with Adobe's [Acrobat 4.05](#) software to create this PDF document. Roman script documents were processed using Acrobat Capture's or PageGenie Pro's optical character recognition filters to extract as much electronic text from the pages as possible, but no attempt has been made to check the resulting document's spelling or grammar. You may perform text searches on these documents, but due to limitations in character recognition technology and the quality of the individual documents, be aware that some words or phrases may not show up in the search even if they appear to be in the document.

All of the documents have thumbnail views and bookmarks embedded in them to aid the user in finding specific pages as quickly as possible (these are visible from the tabs in the left column of this document). Each document is stored in a PDF "Searchable Image (Exact)" format, which preserves the original scan quality while embedding electronic text (if available) behind the scanned image. The PDF file's security settings allow the user to print and select text; however, neither IPUMS International nor the Minnesota Population Center accept responsibility for any modifications made to these documents. The usual disclaimers apply.

If you have any questions about this document or the IPUMS International enumeration form project, please contact us by e-mail at mpc@pop.umn.edu.



1985.3

LA MUESTRA

XV CENSO NACIONAL DE POBLACION Y IV DE VIVIENDA

PROCEDIMIENTOS DE MUESTREO PARA EL CENSO 1985

(DOCUMENTO PRELIMINAR DE TRABAJO)

UNIDAD DE MUESTRAS

David Negill (US Census Bureau)
Marta Romero

BOGOTA, SEPTIEMBRE DE 1985

CONTENIDO

I. INTRODUCCION

II. MUESTRA DE AMPLIACION

1. Objetivos Generales de la Muestra
2. Objetivos Especificos
3. Definición de Variables y Formulario
4. Método de Muestreo
5. Tamaño de la muestra
6. Método de selección
7. Método de expansión
8. Cálculo de errores

III. MUESTRA DE COBERTURA

1. Objetivos de la Muestra
2. Variables y Formulario
3. Método de Muestreo
4. Tamaño de la Muestra
5. Método de Selección
6. Marco de Muestreo
7. Ajuste de Cobertura
8. Error de Muestreo en la Estimación de Cobertura

INTRODUCCION

Las actividades de muestreo que tradicionalmente se han aplicado en los eventos censales del país, se han restringido al diseño de Muestras de cobertura, que indican el cubrimiento logrado por el censo, y a las denominadas muestras de avance que permiten emitir resultados censales con mayor rapidez en razón que se procesa solo parte de los formularios censales.

Con base en los altos costos que implica recoger y procesar información detallada a todas y cada una de las personas del país, para el censo de 1985 se optó por una nueva estrategia, apoyada fundamentalmente en el uso de la técnica estadística del muestreo.

La estrategia para el XV Censo Nacional de Población y IV de Vivienda consiste en aplicar dos tipos de formularios. Un formulario que contiene las características mínimas de la población y de la vivienda, el cual se aplicará a toda la población y otro formulario con características más detalladas, que se aplicará solo a una parte de la población.

A partir de la estrategia anterior se puede definir el primer objetivo básico de la actividad de muestreo, consiste en obtener información detallada sobre características de población y vivienda, que no sea contemplada en el formulario básico, sin incurrir en un gran volumen de formularios, pero si garantizando una repre-

sentatividad de la información recogida. Para el cumplimiento de esta actividad se ha diseñado la denominada muestra MUESTRA DE AMPLIACION.

El segundo objetivo se enmarca dentro de la necesidad de estimar la omisión en que pueda incurrir el censo. Esta estimación permite en primera instancia evaluar el operativo censal, y también suministrar factores cuantitativos de ajuste para la información censal. Con el fin de obtener estas estimaciones se ha diseñado la MUESTRA DE COBERTURA.

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.

El documento que aquí se presenta se ocupa de ilustrar los procedimientos y metodologías involucradas en la planeación, diseño y ejecución de los dos tipos de muestra mencionados.

CAPITULO II

MUESTRA DE AMPLIACION

1. Objetivos Generales de la Muestra

- A) 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.
- B) Suministrar estimadores razonablemente precisos para áreas administrativas más amplias como son los municipios.
- C) Suministrar estimadores (a partir del formulario ampliado), para subgrupos de 5.000 habitantes o más, con un error de más o menos el 4% basado en un nivel de confianza del 95%.
- D) Aprovechar el operativo censal para aplicar un diseño de muestra que permita para algunas variables, tener la elasticidad de manejo que está implícita en la propiedad hereditaria de la información censal.

2. Objetivos Específicos 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. Definición 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 (Anexo No. 2), así:

CAPITULO I UBICACION GEOGRAFICA: Con el cual el empadronador se ubicará en una determinada manzana (urbana) o en un sector (rural), del municipio.

CAPITULO II IDENTIFICACION DE LA VIVIENDA: Para ubicar una vivienda dentro de la manzana o sector.

CAPITULO III DATOS DE LA VIVIENDA: Indaga por las

características principales de la vivienda, materiales con que está construida, conexión con los servicios públicos, etc.

CAPITULO IV TOTAL DE PERSONAS: El cual registra las personas que son residentes habituales de la vivienda.

CAPITULO V DATOS DE POBLACION: Para cada una de las personas registradas en la vivienda se indagan cerca de 25 Items.

4. Método de Muestreo

El método de muestreo a ser 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 representan.
- 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 esta determinado por diversos factores:

- Número de elementos en la población (N)
- Varianza de la variable en los elementos de la población:
(σ^2)
- Definición de la probabilidad con que se define ceptar 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 tienen un atributo cualquiera, se tiene que:

$$P = \frac{\sum_{j=1}^n x_j}{n} \quad \text{Donde } p \text{ es un estimador de } P \ (P \equiv p)$$

$$p = \frac{\sum_{i=1}^N x_i}{N}$$

Donde: N = número de elementos en la población de estudio.

N_i = elemento i -ésimo en la población.

X_i = es el valor de la variable x para el elemento i -ésimo de la población.

$\sum_{i=1}^N$ = 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 $x_j = \begin{cases} 1 & \text{si el individuo tiene el atributo} \\ 0 & \text{si el individuo no tiene el atributo.} \end{cases}$

p = proporción de elementos que poseen el atributo x , en la muestra.

\sum_J^n = indica la suma sobre todos los elementos de la muestra ($J = 1, 2, \dots, n$).

Se demuestra que el error de muestreo de la proporción (σ_p) es:

$$\sigma_p^2 = \frac{N - n}{N} \cdot \frac{Pq}{n} \quad \text{donde } Q = 1 - p$$

$$\sigma_p^2 = \frac{N - n}{N} \cdot \frac{Pq}{n_1} = (1 - f) \cdot \frac{pq}{n}$$

donde $f = \frac{n}{N}$ la fracción de muestreo.

Si f es despreciable, entonces:

$$\sigma_p^2 \cong \frac{pq}{n}$$

El error relativo se define como:

$$RV_p = \frac{\sigma_p}{p} \times 100$$

El parámetro (RV) cuantifica el error de muestreo en términos relativos y al no estar afectado por la unidad de medida que se utilice, permite interpretar fácilmente la confiabilidad de una estimación así:

RV	Tipo de Estimación
Menor al 2%	Excelente
2% - 5%	Buena
5% - 10%	Aceptable
10% - 20%	Regular
Más de 20%	Deficiente

El cuadro No. 1 muestra para diversos tamaños del universo, en los cuales se toma una muestra del 10%, los intervalos de confianza del 95%, que corresponden a varias proporciones a estimar y su error relativo. Los cálculos del intervalo de estimación de una proporción se realizan como:

$$\left(P \pm k \sqrt{\frac{Pq}{n-1}} \right)$$

donde K = cuantila de la distribución normal, para un nivel dado de probabilidad de aceptación.

El error absoluto de muestreo (e) es la semilongitud del intervalo de confianza:

$$e = k \sqrt{\frac{pq}{n-1}} = N \sqrt{p}$$

luego $\sqrt{p} = e/k$

A partir del cual se puede estimar el error relativo de mues-

CUADRO Nro. 1
INTERVALOS DE CONFIANZA DE 95% Y ERRORES DE MUESTREO PARA ESTIMACION DE PROPORCIONES
BASADOS EN UNA MUESTRA DE 10% PARA DIFERENTES TAMAÑOS DE POBLACION

$$(P \pm 1.96 \frac{\sigma_p}{\sqrt{n}}) = (P \pm 1.96 \sqrt{\frac{pq}{n}})$$

$$R.V = \frac{\sigma_p}{P} \times 100$$

N	n	P=10%			P= 20%			P=30%		
		LI -	LS	R.V	LI -	LS	R.V	LI -	LS	R.V
500	50	1.60%	18.40%	42.85%	8.80%	31.20%	28.57%	17.17%	42.83%	21.82%
1000	100	4.09%	15.91%	30.15%	12.12%	27.88%	20.10%	20.97%	39.03%	15.35%
2000	200	5.83%	14.17%	21.27%	14.44%	25.56%	14.18%	23.63%	36.37%	10.83%
5000	500	7.37%	12.63%	13.43%	16.49%	23.51%	8.95%	25.98%	34.02%	6.84%
10000	1000	8.14%	11.86%	9.49%	17.52%	22.48%	6.33%	27.16%	32.84%	4.83%
20000	2000	8.69%	11.32%	6.71%	18.25%	21.75%	4.47%	27.99%	32.01%	3.42%
30000	3000	8.93%	11.07%	5.48%	18.57%	21.43%	3.65%	28.36%	31.64%	2.79%
40000	4000	9.07%	10.93%	4.74%	18.76%	21.24%	3.16%	28.58%	31.42%	2.41%
50000	5000	9.17%	10.83%	4.24%	18.89%	21.11%	2.83%	28.73%	31.27%	2.16%
60000	6000	9.24%	10.76%	3.87%	18.99%	21.01%	2.58%	28.84%	31.16%	1.97%
70000	7000	9.30%	10.70%	3.59%	19.06%	20.94%	2.39%	28.93%	31.07%	1.83%
80000	8000	9.34%	10.66%	3.35%	19.12%	20.88%	2.24%	29.00%	31.00%	1.71%
90000	9000	9.38%	10.62%	3.16%	19.17%	20.83%	2.11%	29.05%	30.95%	1.61%
100000	10000	9.41%	10.59%	3.00%	19.22%	20.78%	2.00%	29.10%	30.90%	1.53%
120000	12000	9.46%	10.54%	2.74%	19.28%	20.72%	1.83%	29.18%	30.82%	1.39%
140000	14000	9.50%	10.50%	2.54%	19.34%	20.66%	1.69%	29.24%	30.76%	1.29%
160000	16000	9.54%	10.47%	2.37%	19.38%	20.62%	1.58%	29.29%	30.71%	1.21%
180000	18000	9.56%	10.44%	2.24%	19.42%	20.58%	1.49%	29.33%	30.67%	1.14%
200000	20000	9.58%	10.42%	2.12%	19.45%	20.55%	1.41%	29.36%	30.64%	1.08%
250000	25000	9.63%	10.37%	1.90%	19.50%	20.50%	1.26%	29.43%	30.57%	0.97%
300000	30000	9.69%	10.34%	1.73%	19.55%	20.45%	1.15%	29.48%	30.52%	0.88%
400000	40000	9.71%	10.29%	1.50%	19.61%	20.39%	1.00%	29.55%	30.45%	0.76%
500000	50000	9.74%	10.26%	1.34%	19.65%	20.35%	0.89%	29.60%	30.40%	0.68%

CUADRO Nro. 1
INTERVALOS DE CONFIANZA DE 95% Y ERRORES DE MUESTREO PARA ESTIMACION DE PROPORCIONES
BASADOS EN UNA MUESTRA DE 10% PARA DIFERENTES TAMAÑOS DE POBLACION

$$(P \pm 1.96 \sigma_p) = (P \pm 1.96 \sqrt{\frac{PQ}{n}})$$

$$R.V. = \frac{\sigma_p}{P} \times 100$$

		P=40%				P=50%				P=60%			
N	n	LI	-	LS	R.V	LI	-	LS	R.V	LI	-	LS	R.V
500	50	26.28%		53.72%	17.50%	36.00%		64.00%	14.29%	46.28%		73.72%	11.66%
1000	100	30.35%		49.65%	12.31%	40.15%		59.85%	10.05%	50.35%		69.65%	8.21%
2000	200	33.19%		46.81%	8.68%	43.05%		56.95%	7.09%	53.19%		66.81%	5.79%
5000	500	35.70%		44.30%	5.48%	45.61%		54.39%	4.48%	55.70%		64.30%	3.66%
10000	1000	36.96%		43.04%	3.87%	46.90%		53.10%	3.16%	56.96%		63.04%	2.58%
20000	2000	37.85%		42.15%	2.74%	47.81%		52.19%	2.24%	57.85%		62.15%	1.83%
30000	3000	38.25%		41.75%	2.24%	48.21%		51.79%	1.83%	58.25%		61.35%	1.49%
40000	4000	38.48%		41.52%	1.94%	48.45%		51.55%	1.58%	58.48%		61.52%	1.29%
50000	5000	38.64%		41.36%	1.73%	48.61%		51.39%	1.41%	58.64%		61.36%	1.15%
60000	6000	38.76%		41.24%	1.58%	48.73%		51.27%	1.29%	58.76%		61.24%	1.05%
70000	7000	38.85%		41.15%	1.46%	48.83%		51.17%	1.20%	58.85%		61.15%	0.98%
80000	8000	38.93%		41.07%	1.37%	48.90%		51.10%	1.12%	58.93%		61.07%	0.91%
90000	9000	38.99%		41.01%	1.29%	48.97%		51.03%	1.05%	58.99%		61.01%	0.86%
100000	10000	39.04%		40.96%	1.22%	49.02%		50.98%	1.00%	59.04%		60.96%	0.82%
120000	12000	39.12%		40.88%	1.12%	49.11%		50.89%	0.91%	59.12%		60.88%	0.75%
140000	14000	39.19%		40.81%	1.04%	49.17%		50.83%	0.85%	59.19%		60.81%	0.69%
160000	16000	39.24%		40.76%	0.97%	49.23%		50.77%	0.79%	59.24%		60.76%	0.65%
180000	18000	39.28%		40.72%	0.91%	49.27%		50.73%	0.75%	59.28%		60.72%	0.61%
200000	20000	39.32%		40.68%	0.87%	49.31%		50.69%	0.71%	59.32%		60.68%	0.58%
250000	25000	39.39%		40.61%	0.77%	49.38%		50.62%	0.63%	59.39%		60.61%	0.52%
300000	30000	39.45%		40.55%	0.71%	49.43%		50.57%	0.58%	59.45%		60.55%	0.47%
400000	40000	39.52%		40.48%	0.61%	49.51%		50.49%	0.50%	59.52%		60.48%	0.41%
500000	50000	39.57%		40.43%	0.55%	49.56%		50.44%	0.45%	59.57%		60.43%	0.37%

CUADRO Nro. 1
INTERVALOS DE CONFIANZA DE 95% Y ERRORES DE MUESTREO PARA ESTIMACION DE PROPORCIONES
BASADOS EN UNA MUESTRA DE 10% PARA DIFERENTES TAMAÑOS DE POBLACION

$$(P \pm 1.96 \sigma_p) = (P \pm 1.96 \sqrt{\frac{PQ}{n}})$$

$$R.V. = \frac{\sigma_p}{P} \times 100$$

		P=70%			P=80%			P=90%		
N	n	LI	LS	R.V	LI	LS	R.V	LI	LS	R.V
500	50	51.17%	82.83%	9.35%	68.80%	91.20%	7.14%	81.60%	98.40%	4.76%
1000	100	60.97%	79.03%	6.58%	72.12%	87.88%	5.03%	84.09%	95.91%	3.35%
2000	200	63.63%	76.37%	4.64%	74.49%	85.51%	3.54%	85.83%	94.17%	2.36%
5000	500	65.98%	74.02%	2.93%	76.49%	83.51%	2.24%	87.37%	92.63%	1.49%
10000	1000	67.16%	72.84%	2.07%	77.52%	82.48%	1.58%	88.14%	91.86%	1.05%
20000	2000	67.99%	72.01%	1.46%	78.25%	81.75%	1.12%	88.87%	91.32%	0.75%
30000	3000	68.36%	71.64%	1.20%	78.57%	81.43%	0.91%	89.93%	91.07%	0.61%
40000	4000	68.58%	71.42%	1.04%	78.76%	81.24%	0.79%	89.67%	90.93%	0.53%
50000	5000	68.73%	71.27%	0.93%	78.89%	81.11%	0.71%	89.17%	90.83%	0.47%
60000	6000	68.84%	71.16%	0.85%	78.99%	81.01%	0.65%	89.24%	90.76%	0.43%
70000	7000	68.93%	71.07%	0.80%	79.06%	80.94%	0.60%	89.30%	90.70%	0.40%
80000	8000	69.00%	71.00%	0.73%	79.12%	80.88%	1.12%	89.34%	90.66%	0.37%
90000	9000	69.05%	70.95%	0.69%	79.17%	80.83%	0.53%	89.38%	90.62%	0.35%
100000	10000	69.10%	70.90%	0.65%	79.22%	80.78%	0.50%	89.41%	90.59%	0.33%
120000	12000	69.18%	70.82%	0.60%	78.28%	80.72%	0.46%	89.46%	90.54%	0.30%
140000	14000	69.24%	70.76%	0.55%	79.34%	80.66%	0.42%	89.50%	90.50%	0.28%
160000	16000	69.29%	70.71%	0.52%	79.38%	80.62%	0.40%	89.54%	90.46%	0.26%
180000	18000	69.33%	70.67%	0.49%	79.42%	80.58%	0.37%	89.56%	90.44%	0.25%
200000	20000	69.36%	70.64%	0.46%	79.45%	80.55%	0.35%	89.58%	90.42%	0.24%
250000	25000	69.43%	70.57%	0.41%	79.50%	80.50%	0.32%	89.63%	90.37%	0.21%
300000	30000	69.48%	70.52%	0.38%	78.55%	84.50%	0.29%	89.66%	90.34%	0.19%
400000	40000	69.55%	70.45%	0.33%	79.61%	80.39%	0.25%	89.71%	90.29%	0.17%
500000	50000	69.60%	70.40%	0.29%	79.65%	80.35%	0.22%	89.74%	90.26%	0.15%

treo:

$$RV = \sqrt{p/p} \times 100$$

En el cuadro No. 1 se puede observar que para poblaciones mayores de 5.000 unidades, se puede esperar un error de muestreo menor al 13% en la estimación de proporciones del 10%.

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.

En los estudios de población realizados por el DANE (1980-1982), se trabajaron algunas ciudades, con tamaño de muestra equivalente al 10% de población. El cuadro No. 2 muestra los resultados del cálculo de los errores de muestreo generados en la estimación de tasas como el desempleo y la fecundidad, para diferentes tamaños de población.

Cuadro 2: ERRORES RELATIVOS DE MUESTREO PARA TASAS O PROPORCIONES
ESTIMADOS EN UNA MUESTRA DEL 10% (29 CIUDADES)
FTE: ESTUDIOS DE POBLACION DANE (1980-1983)

CIUDAD	TAMANO DE LA POBLACION 10 %	DESEMPLEO		FECUNDIDAD		% SEPARADOS		% DE SOLTEROS		% EDUCAC. PRIMARIA		% EDUC. SUPERIOR		ANALFABETISMO	
		TASA	C.V.	TASA	C.V.	%	C.V.	%	C.V.	%	C.V.	%	C.V.	TASA	C.V.
VILLA MARIA	1000	14.21	12.1	97.17	16.53	1.21	32.23	62.29	2.35	53.63	5.33	5	23	8.14	14.37
FLORIDA BLAN	1590	6.87	6.84	101.57	5.35	1.06	8.06	62.6	0.62	53.03	1.87	3.7	10.27	9.58	4.91
ROSADERO	1696	10.1	17.3	114.28	12.16	5.09	14.54	62.33	1.94	48.31	5.38	7.05	25.25	11.46	11.08
LEICIA	2275	6.17	14.75	168.13	9.69	3.21	12.77	61.45	1.29	54.72	2.54	3.41	15.25	16.91	6.33
IRON	2402	4.12	19.9	104.02	12.96	2.09	16.17	61.72	1.41	55.07	3.11	3.66	21.86	11.85	8.19
SEVILLA	3145	16.2	7.22	124.25	11.75	5.44	7.9	59.35	1.41	60.41	2.5	1.49	27.52	19.44	7.1
PIOMACHA	3401	9.65	11.19	146.42	8.4	3.43	11.37	63.45	1.36	48.7	2.87	3.73	14.75	20.56	6.42
SCALAPCA	3543	9.06	9.16	92.32	10.61	3.49	10.03	58.3	1.17	53.36	2.81	4.49	14.92	14.74	6.24
FLOPENCIA	4914	6.29	11.13	114.96	9.39	2.43	8.64	62.88	0.97	53.93	2.13	3.09	15.86	16.71	4.79
CIENAGA	5510	11.16	7.8	141.16	8.35	3.42	9.06	61.85	0.86	53.28	2.27	3.12	14.64	19.63	6.42
BUGA	7418	7.1	8.87	93.41	7.39	5.03	5.57	58.5	0.85	53.79	1.88	4.29	10.25	11.93	5.78
CARTAGO	8234	8.14	8.11	93.43	6.85	4.23	5.91	60.38	0.86	56.21	1.49	3.11	11.25	12.22	5.87
TULUA	8548	8.07	6.94	109.28	6.68	3.79	6.33	59.15	0.79	53.34	2.1	3.59	9.47	12.34	6.56
SINCELEJO	9793	9.11	7.24	93.86	6.18	3.48	6.03	62.41	0.62	48.98	1.51	3.09	10.68	21.82	3.8
SOLEDAD	11160	12.04	4.98	109.85	5.82	4.54	4.85	60.59	0.58	51.13	1.49	2.86	8.74	11.89	5.72
BARRANCA BER	11738	13.16	4.79	97.08	5.87	2.73	6.59	61.5	5.67	51.03	1.41	2.74	9.12	14.14	4.1
VALLEDUPAR	12157	10.56	5.02	113.32	5.03	5.77	4.16	62.83	0.57	52.6	1.12	3.59	0.96	19.13	4.13
MONTERIA	13273	8.47	5.67	95.49	5.8	4.38	4.79	62.56	0.62	48.42	1.67	5.4	9.07	16.78	5.01
NEIVA	13804	6.36	4.08	89.35	5.15	2.96	6.08	61.7	0.58	49.17	1.38	5.26	6.65	13.06	3.83
ARMENIA	14695	9.66	5.28	68	5.88	3.58	5.31	58.89	0.59	47.27	1.67	7.9	5.32	9.91	3.94
SANTA MARTA	14762	8.6	6.74	95.48	5.13	4.5	4.44	62	0.56	46.68	1.89	7.41	7.96	11.59	4.23
PASTO	15991	6.93	5.63	94.89	4.95	2.54	5.51	63.35	0.49	49.39	1.54	7.64	5.5	9.6	4.06
IBAGUE	20312	6.13	6.04	80.06	4.7	2.81	4.63	59.66	0.5	48.62	1.77	7	7.57	10.9	3.94
MANTALES	24506	10.15	4.01	72.17	4.49	2.3	4.35	61.97	0.45	47.11	1.32	8.35	5.03	8.33	3.72
PEREIRA	26445	9.27	3.56	80.17	3.99	2.71	4.06	60.31	0.43	51.99	0.85	5.17	4.64	9.26	2.59
BUCAPANGA	29622	6.47	4.48	7.59	4.09	2.6	4.23	62.3	0.4	46.85	1.13	8.6	3.84	8.68	2.88
CUOTA	34704	5.7	4.21	95.69	3.24	2.08	3.85	62.44	0.35	54.2	0.66	4.02	3.73	12.24	1.96
BEAQUILLA	76376	9.14	2.41	95.87	2.25	3.88	2.06	60.51	0.25	45.27	0.71	9.61	1.98	8.52	2
BOGOTA	311250	5.88	1.16	84.1	1.17	0.82	2.87	62.7	0.2	41.9	0.33	14.77	0.67	6.73	0.71

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 vivienda es una variable aleatoria. Para el cálculo de la varianza de una tasa, en muestreo de conglomerados se utiliza la siguiente formula:

$$V(r) = \frac{1}{x^2} [var(y) + r^2 var(x) - 2r cov(y, x)]$$

donde:

$$r = \frac{y}{x} = \frac{\sum_{\alpha}^a y_{\alpha}}{\sum_{\alpha}^a x_{\alpha}} \quad \alpha = 1, 2, \dots, a$$

a = total de conglomerados en la muestra

y_{α} = suma de los valores de la variable aleatoria Y en el α -ésimo conglomerados.

x_{α} = suma de los valores de la variable aleatoria X en el α -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. $Var(y) = \frac{1}{a} \sum_{\alpha}^a y_{\alpha}^2 - \bar{y}^2$

VAR (X) = Varianza entre conglomerados respecto a la variable X. $Var(x) = \frac{1}{a} \sum_{\alpha}^a x_{\alpha}^2 - \bar{x}^2$

COV (y,x) = Covarianza entre las variables X e Y
 $= [(\sum X_{\alpha} Y_{\alpha} / n) - \bar{x} \bar{y}]$

Para efectos del cálculo pertinente, la formula se puede simplificar a lo siguiente:

$$V(r) = \frac{1}{x^2} \left[\frac{a}{a-1} \sum_{\alpha} (y_{\alpha} - r X_{\alpha})^2 \right]$$

Para obtener el error estandar, se calcula la raíz cuadrada de $V(r)$, que es la medición en términos absolutos del error de estimación.

La medición en términos relativos se define como:

$$RV = \frac{\sqrt{V(r)}}{r} \times 100$$

5.3 Evaluación del Tamaño de Muestra Propuesto

Con base en los dos resultados anteriores, el marco teórico del sistemático y la aplicación real de 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 5.000 unidades o más, con un razonable error, de muestreo en la estimación.

6. Método de Selección

La selección se realizará 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, les será aplicado el formulario de muestra.

6.1 Selección Urbana.

Con base en el listado de recuento preliminar de viviendas particulares del censo, el Jefe de cada Centro de Operaciones seleccionará sistemáticamente 1 de cada 10 viviendas, a partir de un arranque aleatorio, que será un número entre 1 y 10, previamente sorteado.

El operativo censal prevee tener dos tipos de recolectores, uno para el formulario Básico y otro para el Ampliado. A las viviendas seleccionadas se les transcribirá la información de identificación a otro formato, a partir del cual el recolector asignado, aplica a dichas viviendas el formulario denominado Ampliado.

En caso de que el recolector censal encuentre más de una vivienda, dentro de la vivienda identificada en el recuento censal, aplicará a todas y cada una de ellas el formulario básico. Igualmente hará el recolector asignado a las viviendas de la muestra de ampliación.

En caso de viviendas totalmente omitidas en el recuento previo, se construirá 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 disponible para llevar a cabo la selección. Se optó por llevar aquí un operativo simultáneo al censo con las siguientes actividades:

- A) Asignación de empadronadores respetando los límites de sección definidos en la cartografía existente.
- B) El empadronador censal recogerá simultáneamente el formulario Básico y Ampliado.
- C) La selección de viviendas para el ampliado se ejecutará sobre el terreno, seleccionando 1 de cada 10 viviendas censales, a partir del arranque aleatorio que le haya sido asignado.
- D) El supervisor del trabajo de campo, hará una revisión detallada del sistema de recorrido y de selección de la muestra para evitar que el empadronador introduzca sesgos en la selección de la muestra.
- E) El empadronador colocará en la cartografía y en el formulario de recuento, el número de la vivienda que se

lista en el recorrido. Esta numeración ayudará a la identificación de viviendas en la muestra de cobertura.

7. Métodos de Expansión

El factor de expansión F se define como:

$$F = \frac{1}{\text{Fracción de muestreo}} = \frac{1}{f} = \frac{N}{n}$$

donde:

$$f = \frac{n}{N} = \text{fracción de muestreo}$$

n = Número de elementos en la muestra

N = Número de elementos en el universo.

Con el diseño de muestra propuesto el factor de expansión debe ser 10 para todos los centros de operaciones y áreas urbano-rural de los municipios del país. Sin embargo, para controlar los posibles errores de empadronamiento y de selección de la muestra se deberán calcular los factores de expansión finalmente obtenidos, a nivel de cada municipio y separando urbano y rural. (Centro Poblado)

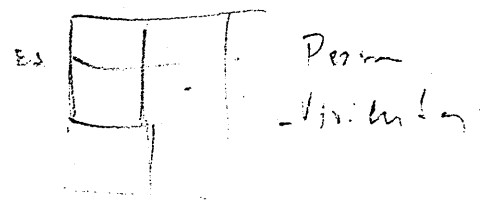
Estos factores se calcularán después de concluido el proceso de inconsistencias, y los totales del censo se obtendrán por

agregación de los formularios Básicos y Ampliados en razón a que a la vivienda que se le aplica Ampliado no se le aplica el Básico dado que el Ampliado lo contiene.

En general el factor de expansión se define para formulario Ampliado y para formulario Básico así:

$$FA = \frac{\frac{N_B}{B} + \frac{N_A}{A}}{\frac{N_A}{A}} \times \text{Ajuste de Cobertura} \quad (p/Depth) \quad \text{Ejemplo y Efecto}$$

$$FB = \text{Ajuste de Cobertura}$$



Donde: F_A = Factor de expansión para las variables de los formularios Ampliados.

F_B = Factor expansión para las variables de los formularios Básicos.

N_B = Número de formularios Básicos.

N_A = Número de formularios ampliados según muestra.

8. Cálculo de Errores

El cálculo del error de muestreo en la estimación de un total de variable (de vivienda o de población) se puede definir en

términos generales así:

$$\text{Dado que: } \hat{X}_A = N \cdot C \cdot \bar{X}$$

$$\text{Entonces: } \sqrt{\hat{X}_A}^2 = N^2 \cdot \sqrt{(C \cdot \bar{X})}^2$$

donde: \hat{X}_A = Estimación del total de la variable, obtenida con formulario **A**mpliado.

N = Número de viviendas o de personas (estimado con la unión de los Básicos y los Ampliados):

C = Ajuste por cobertura.

\bar{X} = Promedio de la variable

$\sqrt{\hat{X}_A}^2$ = Error de muestreo (con la muestra de ampliación) en la estimación del total de una variable X .

Se entiende $\sqrt{C \cdot \bar{X}}^2$ = Error de muestreo en la estimación de un producto de dos variables aleatorias: la Cobertura y el Promedio.

Dependiendo de los niveles a los cuales se obtengan estimaciones confiables de cobertura, los métodos de expansión y cálculo de errores de la muestra de ampliación podrán sufrir algunas modificaciones.

CAPITULO III

MUESTRA DE COBERTURA

1. Objetivos de la Muestra de Cobertura

- 1.1 Evaluar el operativo censal en función de las estimaciones de la cobertura.
- 1.2 Proporcionar estimaciones razonablemente precisas de la cobertura censal con el siguiente orden de prioridad: Nacional, urbano-rural, regional, departamental, ciudades principales.
- 1.3 Evaluar la cobertura en viviendas y población.
- 1.4 Evaluar la magnitud de los errores de contenido en la información por edad y sexo de la población.
- 1.5 Aplicar la muestra de cobertura inmediatamente concluido el censo, para disminuir los efectos de recordación del informante.

2. Variables y Formularios

Con base en la necesidad de evaluar la cobertura en viviendas y población y adicionalmente los errores de contenido en la información básica de la población, las variables a incluir en el formulario se clasificaron así:

- A) De identificación de la vivienda (ubicación, y tipo de vivienda).
- B) Las pertinentes a edad y sexo de los residentes habituales de la vivienda, en el momento del censo y en el momento de la cobertura.
- C) Los controles necesarios para determinar la población migrante.
- D) Controles adicionales sobre si la vivienda fue censada o no, en función de la existencia de una calcomanía censal y/o del concepto de los residentes en el momento de aplicar la cobertura.

3. Método de Muestreo en la Cobertura

Esta muestra se asimila a una muestra de áreas, las cuales se seleccionana probabilísticamente hasta llegar a identificar segmentos de viviendas.

En la parte urbana estos segmentos estan constituidos por manzanas o parte de manzanas, y en la parte rural por áreas; donde existan aproximadamente 25 viviendas.

También se define esta muestra como una muestra estratificada, de afijación igual por estrato y con distribución proporcional dentro del estrato.

Los estratos son cada uno de los 22 departamentos y las ciudades de Bogotá D. E., Cali y Medellín. Dentro de cada uno de ellos se asignará la muestra proporcional a la distribución urbano-rural de su población.

Las razones para definir este método de muestreo son las siguientes:

- A) Si el objetivo de la muestra fuera obtener estimaciones de cobertura solo a nivel nacional, la asignación proporcional por departamento sería la más adecuada.
- B) Sin embargo para un tamaño nacional dado, con la asignación proporcional, el número de unidades por departamento difiere radicalmente y por tanto la precisión de las estimaciones de cobertura por departamento serían también muy diferentes.
- C) En razón a que se requiere en lo posible dar estimación de cobertura por departamento y para las principales ciudades, se prefiere buscar precisiones similares por departamento.
- D) Con la distribución de la muestra en el departamento según la proporción de población urbana y rural, se busca en lo posible, la muestra sea autoponderada por departamento.

4. Tamaño de la Muestra

En la determinación del tamaño de muestra de cobertura es necesario tomar en cuenta tanto los errores de muestreo como los de no muestreo, así como los recursos físicos y financieros que la organización del censo tiene presupuestados.

A medida que aumenta el tamaño de muestra disminuye el error de muestreo pero existe mayor posibilidad de aumentar el error de no muestreo en razón a que se dificulta el control de calidad del operativo.

Debido a la magnitud del operativo censal, los resultados del mismo están sujetos a significativos errores de cobertura y a errores ajenos al muestreo, por tanto se hace necesario buscar una muestra de cobertura que minimice los errores ajenos al muestreo, lo cual es posible en la medida en que se utilice una muestra relativamente pequeña y manejable en el terreno.

Para evaluar entonces el tamaño de muestra se debe tener en cuenta los errores de muestreo implicados en el diseño.

4.1 Errores de muestreo en las estimaciones de viviendas.

Como se trata de una muestra de segmentos de viviendas, se debe tener en cuenta el efecto de conglomeración de las viviendas en los segmentos, lo cual introduce alguna

ineficiencia en la muestra y conduce a un aumento en la varianza.

Este aumento se mide a través de la denominada correlación intraclase ($RHO = \delta$), que es la medida de homogeneidad dentro de los conglomerados. La varianza, aumentada debido al muestreo de segmentos es:

$$\sigma_{sg}^2 = \sigma_{MAS}^2 [1 + \delta(n-1)]$$

Donde: n = Número promedio de viviendas en el conglomerado

δ = Efecto de conglomeramiento (correlación intraclásica).

σ_{MAS}^2 = Varianza en el diseño muestreo aleatorio simple.

σ_{sg}^2 = Varianza en el diseño con segmentos de viviendas.

Una vez concluida la encuesta se puede y se debe calcular el tamaño de las correlaciones intraclase y las varianzas para las características más importantes, pero como marco teórico se utilizara la experiencia de muchos países, donde el conglomeramiento de viviendas en un segmento de 10 unidades produce un, estimado de 0.025 así, para un departamento

$$\sigma_{MAS}^2 (viv) = \frac{PQ}{1000}$$

Si se pretende estimar un parámetro de omisión supuesto de $P=0.05$.

$$\sigma_{MAS(VIV)}^2 = \frac{0.05(0.95)}{1000} = 4.75 \times 10^{-5}$$

La conglomeración de viviendas en segmentos de 25 unidades, bajo el supuesto de un $\delta = 0.025$, produciría una varianza de

$$\sigma_{seg}^2 = \sigma_{MAS}^2 [1 + (n-1)\delta]$$

$$\sigma_{seg}^2 = \sigma_{MAS}^2 [1 + (25-1)0.025]$$

$$\sigma_{seg}^2 = 1.6 \sigma_{MAS}^2$$

$$\sigma_{seg}^2 = 7.6 \times 10^{-6}$$

$$\sigma_{seg}^2 = 8.7178 \times 10^{-3}$$

lo cual produce un error relativo (al estimar el $P=0.05$) equivalente de:

$$RV = \frac{\sigma_{seg}}{\text{cobertura}} = \frac{8.717 \times 10^{-3}}{0.05} = 17.4\%$$

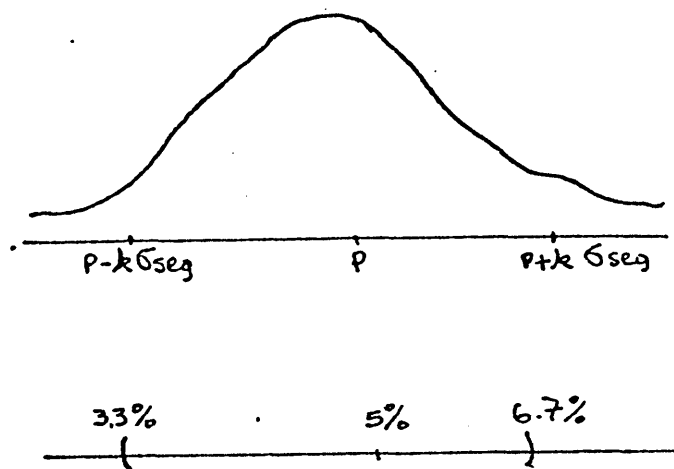
A partir de este valor el intervalo de estimación de p en función del error absoluto sería:

$$e = k \sigma_{seg}$$

$$= 1.96 (8.717 \times 10^{-3})$$

$$= 0.017$$

Luego:



Con base en este resultado, se prevee que una vez analizadas las correspondientes correlaciones intraclase, si el error relativo aumenta, las estimaciones de cobertura solo se obtendrían a nivel de regiones.

Para las estimaciones a nivel regional disminuiría el error de muestreo, dado que se aumentarían los tamaños de muestra.

Al aplicar las formulas pertinentes se tiene que:

CUADRO No. 3

ERROR DE MUESTREO ESPERADO POR REGION

P = 0.05 = 95%

Región	Departamentos	No. de Dptos.	No.de Viv. en muestra	% error de muestreo. P=0.05	Intervalo para P=95 % LI- LS
1. Atlántica	Atlántico Bolívar, Cesar, Córdoba, Guajira, Magdalena, Sucre.	7	7000	5.21	4.5 - 5.5
2. Central	Antioquia*, Caldas, Huila, Quindío, Risaralda, Tolima	6	7000	5.21	4.5 - 5.5
3. Oriental	Boyacá, Cundinamarca, Meta, Nte. Santander, Santander, Caquetá	6	6000	5.63	4.4 - 5.6
4. Pacífica	Cauca, Chocó, Nariño, Valle, **	4	5000	6.17	4.4 - 5.6
5. Bogotá	Bogotá D.E	1	2000	9.75	4.0 - 6.0
T O T A L		24	26000	2.70	4.7 - 5.3

* Incluye la ciudad de Medellín con 1.000 viviendas.

** Incluye la ciudad de Cali con 1.000 viviendas.

LI = Límite inferior

LS = Igual límite superior

Se concluye entonces que para la estimación de la cobertura a nivel de las regiones se puede aceptar una muestra mayor o igual de 2.000 unidades de vivienda, pues el error relativo de muestreo genera niveles aceptables en la estimación.

4.2 Errores de Muestreo en la Estimación de Personas.

La estimación del error de muestreo implicado en un tamaño de muestra, para el caso de las personas que a su vez esta conglomerada en viviendas, implica la estimación de las correspondientes correlaciones intraclase.

El Marco Teórico indica que :

- A) En una muestra aleatoria simple de n personas, la varianza de una estimación simple insesgada de una proporción (P) es :

$$\sigma_{mas}^2(p) = \frac{pq}{n}$$

- B) Cuando se trata sin embargo de una muestra de viviendas (que contiene personas), la conglomeración de personas conduce a un aumento de la varianza cuantificado por la correlación intraclase que mide la homogeneidad dentro del conglomerado así:

$$\sigma_{viv}^2 = \sigma_{mas}^2 [1 + (\bar{n} - 1) \rho]$$

Donde n = número promedio de personas en la vivienda.

δ = medición de la homogeneidad dentro de la vivienda

Con base en experiencias en otros países esta correlación intraclase, para el caso de 5 personas en promedio por vivienda, se estima en 0.25. La aplicación al caso concreto de $P=10\%$ se ilustra en el capítulo correspondiente a la muestra de ampliación.

Adicionalmente se deberá tener en cuenta el conglomeramiento de las viviendas en los segmentos.

La aplicación de las fórmulas anteriores, en un departamento con una muestra de 1.000 unidades de vivienda, seleccionadas en segmentos de 25 viviendas, las cuales en promedio contengan 5 personas por vivienda, los errores de muestreo debido al diseño expuesto, al estimar una proporción del 5% se estiman así:

La varianza de la estimación si la muestra de personas fuera en un diseño de muestreo aleatorio simple.

$$\sigma_{mas}^2 (personas) = \frac{PQ}{n} = \frac{0.05(0.95)}{5000} = 9.5 \times 10^{-6}$$

La varianza, consecuencia de la conglomeración de las personas en la vivienda, supuesto un efecto de conglomeración = 0.025, el cual ha sido obtenido por experiencia de otros países, es:

$$\sigma_{viv}^2 = \sigma_{mas}^2 [1 + (\bar{n}-1)\delta]$$

$$\sigma_{viv}^2 = \sigma_{mas}^2 [1 + (5-1) 0.25]$$

$$= 2 \sigma_{mas}^2$$

$$= 2 (9.5 \times 10^{-6})$$

$$= 1.9 \times 10^{-5}$$

La estimación de la varianza final deberá tener en cuenta el efecto de conglomeramiento de la vivienda en los segmentos. Si se supone un efecto $\delta = 0.025$ (valor obtenido por experiencias de otros países con selección de conglomerados de 10 viviendas) se tiene que:

$$\sigma_{segmento}^2 = \sigma_{viv}^2 [1 + (n-1)\delta]$$

Donde n = Tamaño del conglomerado el cual se plantea en 25 unidades de vivienda.

$$\sigma_{seg}^2 = \sigma_{viv}^2 [1 + (25-1) 0.025]$$

$$\sigma_{seg}^2 = 1.6 \sigma_{viv}^2$$

$$\sigma_{seg}^2 = 3.04 \times 10^{-5}$$

Con base en lo cual para un $P = 10\%$ se puede esperar un error de muestreo en la estimación equivalente a:

$$RV_{(P)} = \frac{\sigma_p(seg)}{p} = \frac{5.51 \times 10^{-3}}{0.10} = 5.5\%$$

Resultado que se considera aceptable. Sin embargo se debe prever que este valor aumente una vez calculada las reales correlaciones intraclase para segmentos de 25 viviendas.

4.3 Tamaño del Segmento.

En función de ilustrar la decisión del tamaño de los segmentos de muestra, se analizó como marco de referencia, el número de viviendas por manzana, en las diferentes ciudades donde el DANE realizó los denominados Estudios de Población. Allí se observa un promedio total de 20.7 viviendas por manzana, con una variación que va desde 11.3 (Guajira) hasta 26.1 (Antioquia).

Teniendo en cuenta que para un tamaño dado de muestra, a menor tamaño del segmento la muestra queda más dispersa, lo cual implica mayores recursos, se decidió entonces tomar un tamaño promedio de 25 viviendas por segmento en el área urbana y en el área rural.

5. Método de Selección

En función de la información existente, las etapas de selección difieren para el área rural y para el área urbana. La explicación se realiza utilizando la terminología cartográfica del DANE para identificación de áreas geográficas.

5.1 Areas Urbanas.

En el método de selección se seguirán las siguientes etapas:

1. Delimitación cartográfica de los sectores. Toda área dentro del perímetro urbano debe pertenecer a uno y solo a un sector.
2. Asignación de probabilidades de selección a los sectores. Esta probabilidad se asigna proporcional al número de viviendas esperado en el sector. La base podrá ser los Recuentos del Precenso, los Estudios de Población o el Resultado del Censo 1973, según la disponibilidad existente para cada caso.
3. Selección sistemática de sectores.
4. Delimitación cartográfica de secciones en cada uno de los sectores seleccionados. Se busca proveer de una identificación única de manzanas, áreas no amanzanadas y áreas sin desarrollar.
5. Identificación de segmentos de 25 viviendas.
6. Asignación de probabilidades a cada segmento (manzana o similar) proporcionales al número de viviendas.
7. Identificación cartográfica y/o en terreno de los segmentos seleccionados.
8. Selección sistemática de segmentos..

5.2 Areas Rurales.

Este método debe tener en cuenta que la cartografía existente esta actualizada solo al año 73, al igual que los recuentos de viviendas.

1. Delimitación cartográfica de los sectores rurales.
2. Asignación de probabilidad de selección del sector, con base en la información de viviendas por sector de la cinta magnética del censo 1973.
3. Selección sistemática de sectores.
4. División del área total del sector seleccionado, con el siguiente indicador:

$$\begin{array}{rcl} \text{No. de segmentos} & & \text{No. de viviendas totales} \\ \text{en el sector} & = & \text{del sector} \\ & & \hline & & \text{No. de viviendas por} \\ & & \text{segmento.} \end{array}$$

Es decir se obtiene el número de segmentos de área que contiene cada sector seleccionado según su número de viviendas. Con base en este indicador, las áreas que se visiten para recoger la muestra, tendrán diferente extensión de un sector a otro, dependiendo del volumen de población del sector y su superficie.

5. Selección aleatoria del segmento de viviendas para la muestra, dentro del sector seleccionado.

5.3 Procedimientos de Submuestreo Rural.

Como la selección de segmentos se hace con base en la cartografía y recuentos de 1973, los segmentos seleccionados en oficina pueden tener en el momento del Censo un número de viviendas muy superior a 250 será necesario subdividir las áreas seleccionadas así:

Viviendas en el terreno	Número de segmentos	Fracción de selección
Menos de 12	0	Unir a otra área
12 - 38	1	1
38 - 62	2	1/2
62 - 88	3	1/3

En la actualidad se están desarrollando actualizaciones cartográficas, con base en fotointerpretación, para delimitar los segmentos en oficina.

6. Marcos de Muestreo

Dado que la muestra probabilística de cobertura no requiere ser expandida al marco de selección, se han definido los siguientes procedimientos en el uso de la información.

6.1 En el área urbana se podrán utilizar los recuentos urbanos de viviendas elaborados en la etapa precensal, para una gran parte de municipios.

En aquellos en los que no se disponga de esta información, se recurrirá al último recuento de viviendas existente en el DANE, o en los recuentos elaborados para los estudios de población del DANE (1980-1982).

- 6.2 En el área rural, dado que solo existe información del número de viviendas en el censo del año 1973, la selección se elaborará con base en los datos correspondientes que existen en las magnéticas del censo.

7. Ajuste de Cobertura

Con base en recomendaciones de funcionarios del Bureau of The Census, se ha llegado a la conclusión de que el mejor tipo de evaluación de cobertura implica el uso de un sistema de estimación dual, este impone el apareamiento de registros con los de la muestra de cobertura, caso por caso.

Con base en esta metodología se define que:

A) La estimación de población total se puede calcular así:

$$\hat{N} = \frac{N_1 \hat{N}_2}{\hat{M}}$$

Donde: \hat{N} = Población total estimada.

N_1 = Población enumerada en el censo

\hat{N}_2 = Estimación de la población enumerada en la muestra de cobertura.

\hat{M} = Estimación de casos apareados en la muestra de cobertura (población enumerada en el censo y en la muestra).

B) El Factor de Cobertura se define como:

$$W_1 = \frac{\hat{M}}{\hat{N}_2}$$

c) La estimación de M y N se definen así:

$$\hat{M} = \sum_{i=1}^m F_i M_i$$

$$\hat{N}_2 = \sum_{i=1}^m F_i n_i$$

Donde: $F_i = \frac{N_i}{n_i}$ El factor de expansión del segmento i (i = 1, 2, ..., m).

$i = 1, 2, \dots, m$

$m =$ Total de segmentos en la muestra

$M_i =$ Número de casos apareados (enumerados en el censo y en la muestra), en el segmento i .

$N_i =$ Número de casos enumerados en la muestra en el segmento.

Las estimaciones del sistema dual involucra: (1)

- A) Obtener datos independientemente para una muestra de individuos en una población, por dos sistemas separados el Censo y la muestra de evaluación.
- B) Determinar cuales de los casos en el estudio de evaluación también se informaron en el censo. La determinación se hace pareando los resultados de la evaluación con los cuestionarios censales.
- C) Usar la proporción de personas en el estudio de evaluación, que se informaron en el censo, como estimación de la proporción de todas las personas informadas en el censo.

(1) Fuente: Estudio de caso para los censos de población y habitación POPSTAN Capítulo A-13 pag. 205

Los resultados de la estimación del sistema dual se ilustran en el cuadro No. . Las anotaciones representan el número de unidades (personas o viviendas) cubiertas en el censo en la muestra de cobertura.

CUADRO No.
MODELO DE ESTIMACION DE SISTEMA DUAL

Muestra de cobertura	TOTAL	C E N S O	
		Informados	No Informados
Total	(N)	N_1	V_i
Informados	N_2	M	$U_2 = N_2 - M$
No Informados	(V_2)	$U_1 = N_1 - M$	

Las anotaciones entre paréntesis son desconocidas. Las otras anotaciones pueden determinarse del censo, de la muestra de cobertura, o del apareamiento.

M = Casos apareados (casos para los que la información en el estudio de evaluación calza con la información censal).

N_2 = Total de casos en la muestra de cobertura.

V_i = Casos que se informan en el censo pero no en el estudio de cobertura (lo determina el apareamiento).

U_1 = Casos que se informan en el estudio de cobertura pero no en el censo (determinado por el apareamiento).

N = Total de casos en el Censo

Esencialmente, el procedimiento de sistema dual para estimar lo completo del empadronamiento es el mismo que el procedimiento para estimar cualquier otro porcentaje cuando faltan algunas de las respuestas.

Analogamente el porcentaje de personas enumeradas en el censo se basa en los casos de la muestra para los cuales se da información, es decir, los N casos en la muestra de cobertura. Este porcentaje se trata como si se aplicará a toda la población incluyendo los V casos omitidos en la muestra de cobertura.

El procedimiento para evaluar la cobertura de población, lista todas las personas que viven (o están) en las viviendas en el momento de la encuesta de cobertura, más las personas que vivían en el segmento de muestreo el día del censo. Los que no se mudaron y los emigrantes (no los inmigrantes) seorean con los retornos censales. La proporción de los emigrantes oreados y no oreados se aplica al número de inmigrantes para derivar el número oreado y no oreado de los que se trasladaron.

8. Errores de Muestreo en la Estimación de Cobertura

El factor de cobertura (W_i) se define como:

$$w_i = \frac{\hat{M}}{\hat{N}_2}$$

Tanto \hat{M} como \hat{N}_2 son estimaciones, por lo tanto w_i corresponde al cociente de 2 variables aleatorias, con lo cual el error de muestreo relativo RV_w es:

$$RV_w^2 = RV_{\hat{M}}^2 + RV_{\hat{N}_2}^2 - 2RV_{\hat{M}\hat{N}_2}$$

donde: RV_w^2 = Error relativo cuadrático de w

$RV_{\hat{M}}^2$ = Error relativo cuadrático de \hat{M}

$RV_{\hat{N}_2}^2$ = Error relativo cuadrático de \hat{N}_2

$RV_{\hat{M}\hat{N}_2}^2$ = Covarianza relativa entre \hat{M} y \hat{N}_2

Los errores relativos cuadráticos son:

$$RV_{\hat{M}}^2 = \frac{S_{\hat{M}}^2}{\hat{M}^2}$$

$$RV_{\hat{N}_2}^2 = \frac{S_{\hat{N}_2}^2}{\hat{N}_2^2}$$

Donde: $S_{\hat{M}}^2 =$ Error de muestreo en la estimación de \hat{M}

$S_{\hat{N}_2}^2 =$ Error de muestreo en la estimación de \hat{N}_2

Para el cálculo de los errores de muestreo se utilizará el método de las medias muestras replicadas que consiste en dividir la muestra de m segmentos en dos medias muestras de $m/2$ segmentos cada una. En este caso de error de muestreo en la estimación de \hat{M} es:

$$S_{\hat{M}}^2 = \sum_g^{m/2} F_g^2 [M_{g1} - M_{g2} - W(n_{g1} - n_{g2})]^2$$

Donde: $S_{\hat{M}}^2$: Error de muestreo en la estimación de \hat{M} .

$F_g = \frac{N_g}{n_g}$ Factor de expansión del segmento g

$$g = 1, 2, \dots, m/2$$

$\frac{m}{2}$: Número de segmentos en cada media muestra.

M_{g1} = Número de casos apareados en el segmento g de la primera media muestra.

M_{g2} = Número de casos apareados en el segmento g de la segunda media muestra.

n_{g1} = Número de casos empadronados en la muestra de cobertura en el segmento g de la primera media muestra.

n_{g2} = Número de casos empadronados en la muestra de cobertura en el segmento g de la segunda media muestra.

De la misma forma se tiene que:

$$S_{\hat{N}_2}^2 = \sum_{g=1}^{m/2} F_g^2 [n_{g1} - n_{g2} - \bar{n} (n_{g1} - n_{g2})]^2$$

$$\text{con: } \bar{n} = \frac{\hat{N}_2}{N_1} = \frac{\sum_{g=1}^{m/2} \sum_{j=1}^2 n_{gj} F_g}{N_1}$$

Finalmente:

$$RV_{\hat{N}_1 \hat{N}_2} = \frac{S_{\hat{N}_1 \hat{N}_2}}{\hat{N}_1 \hat{N}_2}$$

Donde:

$$S_{\hat{N}_1 \hat{N}_2} = \sum_{g=1}^{m/2} U_g Z_g$$

$$\text{con: } U_g = F_g [M_{g1} - M_{g2} - \bar{w} (n_{g1} - n_{g2})]$$

$$Z_g = F_g [n_{g1} - n_{g2} - \bar{n} (n_{g1} - n_{g2})]$$