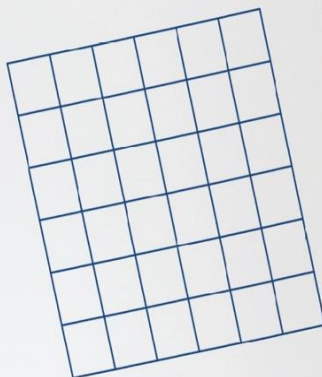


**Documento
Metodológico de
Diseño Muestral
de la Encuesta
Multipropósito**

JUNIO 2020



Diego Andrade
DIRECTOR EJECUTIVO

Víctor Bucheli
SUBDIRECTOR GENERAL

Markus Nabernegg
COORDINADOR GENERAL TÉCNICO DE PRODUCCIÓN ESTADÍSTICA

Christian Garcés
DIRECTOR DE INFRAESTRUCTURA ESTADÍSTICA Y MUESTREO

Equipo técnico:
Francisco Céspedes Navas

Propiedad Institucional
©INEC
Instituto Nacional de Estadística y Censos
Juan Larrea N15-36 y José Riofrío. Casilla postal 135 C
Telf: (02) 2555-701 / 2529-858

Citar como:

INEC (2018). Metodología de Diseño Muestral de la Encuesta Multipropósito. Instituto Nacional de Estadística y Censos, Quito-Ecuador.

Contenidos

1. Introducción	5
2. Diseño Muestral	6
2.1. Determinación del Marco de Muestreo.....	6
2.2. Universo de investigación.....	7
2.3. Unidad de observación	7
2.4. Unidad de análisis	7
2.5. Cobertura geográfica.....	7
2.6. Diseño Muestral.....	7
2.7. Equilibrio de las Unidades Primarias de Muestreo (UPM)	7
2.8. Estratificación del Marco de Muestreo.....	9
2.9. Dominios de estudio	11
2.10. Tamaño y selección de la muestra	12
2.10.1. Tamaño de muestra de personas	12
2.10.2. Tamaño de muestra de viviendas.....	13
2.10.3. Tamaño de muestra de UPM.....	13
2.10.4. Selección de la muestra.....	15
2.11. Factores de expansión	15
2.11.1. Ponderación de selección de primera etapa.....	15
2.11.2. Tasa de no respuesta de primera etapa de selección.....	16
2.11.3. Ponderación de selección de segunda etapa.....	16
2.11.4. Ajuste de no respuesta de viviendas	17
2.11.5. Ajuste de no elegibilidad de viviendas	17
2.11.6. Ponderación de tercera etapa	18
2.11.7. Ajuste por no respuesta para personas 16 años y más.....	18
2.11.8. Factor de expansión final.....	19
2.11.9. Postestratificación.....	19
2.12. Estimación de características.....	20
2.13. Error de muestreo	20
2.14. Métodos de estimación de errores para diseños muestrales complejos	21
Bibliografía.....	23

Tablas

Tabla 1: Ciudades auto representadas ENEMDU - viviendas totales y viviendas ocupadas por sector censal – CPV 2010

Tabla 2: Dimensiones y variables de estratificación

Tabla 3: variables de investigación - 17 indicadores priorizados

Tabla 4: Tamaño de muestra a nivel de viviendas y conglomerados por área geográfica

Tabla 5. Variables requeridas para la declaración del diseño muestral

1. Introducción

El Instituto Nacional de Estadística y Censos – INEC, como ente rector de la estadística nacional es el encargado de llevar a cabo una serie de encuestas que ayudan a la planificación nacional, además de la producción estadística del país. En el marco del monitoreo constante de los indicadores que forman parte de Plan Nacional de Desarrollo, Objetivos de Desarrollo Sostenible, se presenta la “Encuesta Multipropósito 2019”, en su segunda etapa para continuar con el monitoreo de los indicadores de avance en cada uno de los ejes de seguimiento.

Para cumplir con los propósitos de la planificación nacional e ir adecuando las diversas necesidades que se van presentando en el camino de la investigación, el Instituto Nacional de Estadística y Censos – INEC, ha tomado en cuenta un plan de mejoras adaptadas al Modelo de Producción Estadística, es por ello que es necesario actualizar la información de manera continua y dar cumplimiento a los diversos compromisos con la sociedad demandante de información de calidad y oportuna.

El Diseño Muestral es uno de los procesos medulares en el desarrollo de la investigación estadística, es por ello, la necesidad de elaborar un diseño muestral que cumpla con cada uno de los detalles demandado por los diversos usuarios, en este contexto, para el año 2019 dando cumplimiento a este compromiso de considerar un tamaño de muestra suficiente, mediante la información de la Encuesta Multipropósito 2018 y el indicador de desnutrición de la encuesta de condiciones de vida “ECV-2014”, se actualizó el tamaño de muestra para obtener los indicadores a nivel nacional urbano y rural.

El Modelo de Producción Estadística, sigue procesos estándares por cada una de las operaciones estadísticas, por esta razón el diseño muestral de la Encuesta Multipropósito 2019, sigue varias etapas desarrolladas en el presente documento.

2. Diseño Muestral

2.1. Determinación del Marco de Muestreo.

El Marco de Muestreo se basa en los resultados definitivos y la cartografía del VII Censo de Población y VI de Vivienda del 2010 (CPV-2010); siendo constituido por áreas geográficas, que tienen límites perfectamente definidos e identificables sobre el terreno, de este modo se puede ubicar cada una de las unidades de selección, que están definidas mediante un listado completo definido por dos fuentes de información: la primera es el precenso que es la unidad geográfica que nos permite identificar a las unidades primarias de muestreo (UPM), y a su vez en la UPM encontramos a las viviendas; la segunda es el censo de población, el cual por sus características y variables, se utiliza para estratificar el Marco de Muestreo por completo.

El Marco de Muestreo ha tenido actualizaciones parciales, en determinadas áreas geográficas, en función a los cambios presentados en las de unidades de observación (viviendas), dependiendo de las operaciones estadísticas a realizarse, debido principalmente a los movimientos y dinámica demográfica propios de la población, así como los cambios en estructuras habitacionales; que en su conjunto hacen necesario un mantenimiento continuo del Marco de Muestreo.

En resumen, el Marco de Muestreo se basa en los resultados definitivos y la cartografía del VII Censo de Población y VI de Vivienda del 2010 (CPV-2010), complementado con la información parcialmente actualizada por distintos procesos cartográficos que ha llevado el INEC en el período 2014-2017. Para más detalle dirigirse al documento de Marco de Muestreo para Encuestas a Hogares¹.

¹https://www.ecuadrencifras.gob.ec/documentos/web-inec/EMPLEO/2018/Disenio_Muestral_2018/SIEH%20-MMM.pdf

2.2. Universo de investigación

El universo de estudio de la encuesta multipropósito 2019 son personas de 5 y más años de edad, residentes en los hogares del Ecuador, exceptuando la población que reside en viviendas colectivas, viviendas flotantes y población indigente (sin techo).

2.3. Unidad de observación

La unidad de observación son todas las viviendas particulares ocupadas por personas que se encuentran en territorio nacional, las mismas que tienen ligada su identificación geográfica mediante fuentes cartográficas.

2.4. Unidad de análisis

Debido a que se trata de una encuesta de propósitos múltiples existen dos unidades de análisis: el hogar particular y las personas miembros de hogar divididas en los diferentes grupos de interés de la operación estadística.

2.5. Cobertura geográfica

La cobertura geográfica está definida por las viviendas ocupadas que se encuentren ubicadas dentro del territorio ecuatoriano incluyendo la región insular.

2.6. Diseño Muestral

El diseño muestral para esta investigación corresponde a un muestreo probabilístico con dos etapas de selección para el formulario general y una tercera etapa para una sección específica, detallada de la siguiente manera:

- **Primera etapa:** selección de Unidades Primarias de Muestreo por estrato.
- **Segunda etapa:** selección de viviendas ocupadas dentro de cada uno de las UPM seleccionadas en la primera etapa.
- **Tercera etapa:** selección de personas de 16 años y más dentro del hogar.

2.7. Equilibrio de las Unidades Primarias de Muestreo (UPM)

El muestreo es probabilístico y bi-etápico; la Unidad Primaria de Muestreo (UPM) es el conglomerado y la Unidad Secundaria de Muestreo (USM) son las viviendas ocupadas.

El diseño vigente contempla una actualización dentro de la construcción de la Unidad Primaria de Muestreo – (UPM). Para años anteriores, incluida la ENEMDU diciembre 2017,

se realizaba la selección de "sectores censales" en función a un criterio operativo para la ejecución y levantamiento de información del Censo de Población y Vivienda 2010. Sin embargo, en función al crecimiento y disminución de la población en ciertas áreas geográficas por el transcurso del tiempo, estas UPM pasaron a ser heterogéneas en cuanto al número de viviendas ocupadas que tienen dentro de sus límites (originalmente, en promedio, 150 viviendas ocupadas en el área amanzanada y 80 en el área dispersa), situación que traería como consecuencia probabilidades de selección desbalanceadas en la segunda etapa (Tabla 1).

Tabla 5: Ciudades auto representadas ENEMDU - viviendas totales y viviendas ocupadas por sector censal – CPV 2010

DOMINIO		Viv_tot	Viv_ocu
Ambato	Media	104,7	100,7
	Mínimo	27,0	23,0
	Máximo	219,0	218,0
Cuenca	Media	114,7	111,1
	Mínimo	35,0	35,0
	Máximo	226,0	225,0
Guayaquil	Media	128,5	125,6
	Mínimo	1,0	1,0
	Máximo	247,0	240,0
Machala	Media	119,2	116,0
	Mínimo	6,0	6,0
	Máximo	211,0	207,0
Quito	Media	115,3	114,1
	Mínimo	4,0	4,0
	Máximo	678,0	677,0
Total	Media	120,8	118,4
	Mínimo	1,0	1,0
	Máximo	678,0	677,0

Para resolver este problema se utiliza el "Algoritmo de generación de conglomerados para necesidades operativas -AGCNO". Este algoritmo garantiza que el número de viviendas por conglomerado se encuentre entre un número t de viviendas ocupadas hasta un número $2t$; de esta manera, se asegura que los factores de expansión (ponderación) de diseño pertenezcan a un intervalo que va desde s hasta $2s$. El número de viviendas ocupadas que tiene una UPM actualmente es de 30 a 59 viviendas, tanto para el área amanzanada como para el área dispersa.

Una de las responsabilidades del INEC es la mejora continua de sus procesos con la finalidad de seguir generando información de calidad. En este contexto, la construcción de nuevas UPM responde a la identificación y consiguiente solución a un problema específico de las UPM antiguas: la dispersión en el tamaño en número de viviendas. Esta diferencia fue solventada de diferentes formas a lo largo del tiempo, hasta la

introducción definitiva de las nuevas UPM en el año 2018.

Con respecto a un posible impacto los indicadores de seguimiento del Plan Nacional de Desarrollo, cabe recalcar que los estimadores muestrales dependen del diseño muestral y los procesos de selección. A pesar de que estos cambien, si se preservan las características deseadas en un estimador (insesgado o asintóticamente insesgado y de varianza mínima, en otras palabras, un Error Cuadrático Medio bajo), estos son comparables.

2.8. Estratificación del Marco de Muestreo

Al diseñar una encuesta de hogares, una técnica muy extendida es la estratificación de la población que se desea encuestar previa a la selección de la muestra. Esta sirve para clasificar a la población en subpoblaciones (estratos) basándose en información auxiliar conocida sobre toda la población. Después, dentro de cada estrato se seleccionan, probabilística e independientemente, las unidades que formarán parte de la muestra.

Con la estratificación se persiguen diversos fines:

- Dar estimaciones separadas para ciertas subpoblaciones.
- Agrupar unidades homogéneas entre sí para mejorar la precisión de las estimaciones globales.
- Posibilidad de utilizar diferentes métodos de muestreo en los distintos estratos.

Existen varias metodologías que permiten, mediante técnicas matemáticas, la creación de subconjuntos aplicando el análisis de variables cuantitativas como cualitativas. Dado que la fuente principal es el Censo de Población y Vivienda 2010 (CPV 2010), donde sus variables son de característica cualitativa, se ha tomado como alternativa técnica aplicar el **ESCALAMIENTO ÓPTIMO CON BASE EN EL ANÁLISIS DE COMPONENTES PRINCIPALES NO LINEALES**; la cual se basa en asignar cuantificaciones y orden a las variables que son parte del proceso de estratificación.

Como resultado de la aplicación y cálculo del análisis de componentes principales se obtienen tres dimensiones y un total de veinte y uno variables de estratificación

Tabla 6: Dimensiones y variables de estratificación

VIVIENDA	Tipo de viviendas
	Vías de acceso
	Materiales del techo
	Materiales del piso
	Materiales de las paredes
	Procedencia del agua
	Agua que recibe la vivienda
	Servicio higiénico (eliminación de excretas)
	Eliminación de Basura
HOGAR	Disponibilidad de servicio higiénico del hogar
	Disponibilidad de ducha
	Teléfono convencional
	Teléfono celular
	Internet
	Computador
	TV cable
	Equipamiento del Hogar
	Tenencia de la vivienda
POBLACIÓN	Nivel de instrucción
	Grado
	Características económicas

El Marco de Muestreo utilizado para la Multipropósito se lo ha dividido por dominios de estudio, y dentro de ellos sus correspondientes UPM, a las cuales se asignó un estrato tomando principalmente en entre otras características geográficas, socio-económicas y socio-demográficas, con la finalidad de mejorar la precisión y exactitud (minimizar varianza) y mantener la comparabilidad de los de los estimadores con respecto a la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU).

Diferenciar entre área urbana y rural es necesario, lo cual se considera como estrato implícito dentro del Marco de Muestreo. Para ello, se toma en cuenta la división o

identificación de centros poblados¹ (criterio poblacional) sugerido por la Comunidad Andina de Naciones (CAN); en el – “Seminario: Censos 2000 de Población y Vivienda de los Países Andinos”, clasificándolos de la siguiente manera:

Rural:

- Población dispersa y centros poblados con menos de 2000 habitantes.

Urbano:

- 2.000 a 4.999 habitantes
- 5.000 a 9.999 habitantes
- 10.000 a 19.999 habitantes
- 20.000 a 49.999 habitantes
- 50.000 a 99.999 habitantes
- 100.000 a 199.999 habitantes
- 200.000 a 499.999 habitantes
- 500.000 a 999.999 habitantes
- 1.000.000 y más habitantes

Para las encuestas a hogares se considera esta clasificación de área urbana y rural, por ende, la estratificación también usa este criterio para la asignación de los estratos.

La identificación de estrato para cada UPM a nivel nacional respeta los límites geográficos de las jurisdicciones a las cuales pertenecen, para que no existan rupturas con la asignación cartográfica preestablecida.

Sugiero se haga referencia al documento de la estratificación Panchito.

2.9. Dominios de estudio

Los dominios de estudio para la encuesta multipropósitos 2019 son los siguientes:

- Nacional
- Nacional Urbano
- Nacional Rural

¹ Es todo lugar del territorio nacional rural o urbano, identificado mediante un nombre y habitado con ánimo de permanencia. Sus habitantes se encuentran vinculados por intereses comunes de carácter económico, social, cultural e histórico. Dichos centros poblados pueden acceder, según sus atributos, a categorías como: caserío, pueblo, villa, ciudad y metrópoli.

2.10. Tamaño y selección de la muestra

A partir de las expresiones de tamaño de muestra para muestras complejas, se calcula el tamaño de muestra necesario para lograr una precisión adecuada en la inferencia.

2.10.1. Tamaño de muestra de personas

Para la determinación del tamaño de la muestra se requiere establecer la característica o características a estimar, el nivel de confianza y la precisión requeridos.

En este sentido, la variable de investigación, es la desnutrición que se obtuvo de la encuesta de condiciones de vida "ECV-2014", la cual presenta este indicador con los detalles en sus estadísticos para la construcción del tamaño actual, que se detalla a continuación:

Tabla 7: Porcentaje de niños menores de 5 años con desnutrición crónica

Desagregación	Estimación	Error Estándar	I. Inferior	I. Superior	CV	Efecto de Diseño
Nacional	23,9%	0,01	0,23	0,25	2,56	2,2925
Urbana	19,7%	0,01	0,18	0,21	4,09	2,9980
Rural	31,9%	0,01	0,30	0,34	2,54	1,1646

Los valores poblacionales de esta variable se tomaron de la base de datos de la Encuesta de Condiciones de Vida "ECV – 2014".

La expresión matemática utilizada para el cálculo de tamaño de muestra es:

$$n_k \geq \frac{z_{1-\frac{\alpha}{2}}^2 N_k p(1-p)}{(e_r p)^2 (N_k - 1) + z_{1-\frac{\alpha}{2}}^2 p(1-p)} \times \frac{deff}{1 - tnr} \quad (1)$$

donde:

- n_k = Tamaño de la muestra de personas para el dominio de estudio k
- $z_{1-\frac{\alpha}{2}}$ = Nivel de confianza correspondiente al percentil α .
- N_k = Total de personas de 0 a 5 años en el dominio de estudio k .
- p = Probabilidad de incidencia de la variable a investigar
- e_r = error relativo.
- $deff$ = efecto de diseño.
- tnr = tasa de no respuesta.

Los parámetros que se utilizan para el cálculo aplicado son recomendados para los estudios de este tipo, cuyo nivel de confianza es de 95%, error relativo de 15%, tasa de

no respuesta 20%, la estrategia asegura el rigor estadístico de la muestra, además nos permite la posibilidad de comparar entre sí los estudios ejecutados.

2.10.2. Tamaño de muestra de viviendas

Es necesario calcular el número total de viviendas que deben ser seleccionados para lograr entrevistar a todas las personas que serán observadas en el punto anterior. El número de viviendas que deben ser seleccionados estará determinado por la muestra de personas (n_k), número promedio de personas por hogar (b_k) y el porcentaje de personas con la característica de interés ($r = 0$ a 5 años), de la siguiente forma:

$$n_{vk} = \frac{n_k}{r_k * b_k}$$

2.10.3. Tamaño de muestra de UPM

Los hogares y las personas se observan a partir de las UPM. En este paso final es necesario calcular el número de UPM que deben ser seleccionadas en el muestreo a partir de la relación:

$$n_{UPMk} = \frac{n_{vk}}{\text{Carga técnica operativa}}$$

La carga técnica operativa se refiere al número de viviendas asignadas a cada encuestador como carga de trabajo; que fue el resultado de un previo análisis de correlación intraclase donde se pudo verificar, mediante simulaciones matemáticas, el número de observaciones necesarias para minimizar la varianza dentro de cada conglomerado. Como resultado de este procedimiento se definió que la carga operativa es de 12 viviendas y 4 viviendas de reemplazo por UPM.

El total de UPM seleccionadas de acuerdo a la temporalidad de la Encuesta Multipropósito 2019 es:

Tabla 8: Tamaño de muestra a nivel de viviendas y UPM por área geográfica

Código	Provincia	Viviendas	UPM
1	Azuay	636	53
2	Bolívar	216	18
3	Cañar	228	19
4	Carchi	168	14
5	Cotopaxi	384	32
6	Chimborazo	420	35
7	El Oro	492	41
8	Esmeraldas	384	32
9	Guayas	2628	219
10	Imbabura	348	29
11	Loja	396	33
12	Los Ríos	612	51
13	Manabí	996	83
14	Morona Santiago	132	11
15	Napo	120	10
16	Pastaza	144	12
17	Pichincha	2064	172
18	Tungurahua	516	43
19	Zamora Chinchipe	156	13
20	Galápagos	72	6
21	Sucumbíos	180	15
22	Orellana	168	14
23	Sto. Domingo de los Tsáchilas	336	28
24	Santa Elena	204	17
90	Zonas no Delimitadas	72	6
	TOTAL	12072	1006

La muestra será distribuida proporcionalmente por cada estrato dentro de cada dominio de estudio.

$$n_{ki} = n_k \cdot \frac{N_{ki}}{N_k}$$

donde:

- n_{ki} = Tamaño de la muestra para el estrato i del dominio de estudio k
- n_k = Tamaño de la muestra para el dominio de estudio k
- N_{ki} = Total de viviendas ocupadas en el estrato i del dominio de estudio k
- N_k = Total de viviendas ocupadas en el dominio de estudio k .

2.10.4. Selección de la muestra

En primera instancia, la selección de las UPM que forman parte de la muestra, de acuerdo al tamaño establecido, se realiza de manera independiente en cada uno de los dominios de forma aleatoria, asignando a cada UPM igual probabilidad de ser seleccionada. De la misma forma, la selección de viviendas es aleatoria dentro de cada UPM seleccionada.

2.11. Factores de expansión

El procedimiento de ponderación general para la encuesta multipropósito requiere de dos etapas para el formulario general y una etapa adicional para la población específica.

- **La primera etapa** consiste en calcular una ponderación de UPM. De ser necesario, se aplica un ajuste por no respuesta a nivel de UPM.
- **La segunda etapa** consiste en calcular una ponderación de viviendas dentro de la UPM. De ser necesario se aplica un ajuste por no respuesta a nivel de vivienda.
- **La tercera etapa** consiste en calcular una ponderación de las personas de 16 años y más. De ser necesario se aplica un ajuste por no respuesta a nivel de personas de 16 años y más a la ponderación de la población mencionada. La ponderación de personas de 16 años y más se calcula de manera independiente para cada hogar.

2.11.1. Ponderación de selección de primera etapa

La ponderación de primera etapa representa el inverso de la probabilidad de selección de la primera etapa asignada a cada UPM.

Una vez realizado el equilibrio en el número de viviendas que constituyen las UPM, es posible obtener la probabilidad de selección de cada UPM de una manera aleatoria simple dentro de cada estrato. La ponderación de primera etapa (PPE) está definida por:

$$ppe_{hi} = \frac{m_h}{M_h}$$

donde:

m_h = número de UPM del estrato h en la muestra.

M_h = número de UPM del estrato h en el marco.

2.11.2. Tasa de no respuesta de primera etapa de selección

Para realizar este ajuste, es necesario definir la tasa de no respuesta a nivel de UPM, la cual mide la proporción de UPM seleccionadas que efectivamente participaron en la encuesta.

El ajuste por no respuesta a nivel de UPM se calcula como sigue:

$$TNR_h = \frac{m_{h_efec}}{m_h}$$

donde:

TNR_h = Tasa de no respuesta de la primera etapa de selección en el estrato h .

m_{h_ef} = Número de UPM efectivas de la muestra en el estrato h .

m_h = Número de UPM de la muestra en el estrato h .

Las ponderaciones de primera etapa se calculan para las UPM originalmente muestreadas y las UPM de reemplazo que participaron. Cualquier UPM que ya no era elegible fue removida del cálculo del ajuste por no respuesta y queda definido por:

$$ppe_{f_{hi}} = ppe_{hi} * TNR_h$$

donde:

$ppe_{f_{hi}}$ = Probabilidad de selección final de la primera etapa para la UPM i del estrato h .

2.11.3. Ponderación de selección de segunda etapa

La ponderación de segunda etapa representa el inverso de la segunda probabilidad de selección asignada a cada vivienda. Las viviendas fueron seleccionadas con igual probabilidad de selección, por lo que la ponderación básica de segunda etapa está dada por:

$$pse_{hij} = \frac{n_i}{N_i}$$

donde:

pse_{hij} = Probabilidad de segunda etapa para la vivienda j en la UPM i del estrato h .

n_i = Número de viviendas visitadas en la UPM i .

N_i = Total de viviendas de la UPM i .

Para cada UPM, n_i toma un valor de 12 a 16 dependiendo del número de viviendas de reemplazo investigadas.

Adicionalmente es necesario ajustar las probabilidades de inclusión mediante la proporción de viviendas seleccionadas que efectivamente participaron en la encuesta (efectivas), las viviendas rechazo, nadie en casa (no respuesta) y por cambio de ocupación (no elegibilidad), se realiza este ajuste a nivel de estrato.

2.11.4. Ajuste de no respuesta de viviendas

El cálculo del ajuste por no respuesta se realizó por estrato y está dada por la siguiente expresión:

$$Aj_{nr_h} = \frac{n_h}{n_h - n_{hnr} - n_{hnc} - n_{h_ex}}$$

donde:

Aj_{nr_h} = Ajuste por no respuesta para las viviendas del estrato h .

n_h = Número de viviendas visitadas en el estrato h .

n_{hnr} = Número de viviendas que no respondieron en el estrato h .

n_{hnc} = Número de viviendas con nadie en casa en el estrato h .

n_{h_ex} = Número de viviendas no elegibles en el estrato h .

2.11.5. Ajuste de no elegibilidad de viviendas

Al igual que el ajuste de no respuesta, el ajuste de no elegibilidad se realizó por estrato y está dado por la siguiente expresión:

$$Aj_{ne_h} = \frac{n_h - n_{h_ex}}{n_h}$$

donde:

Aj_{ne_h} = Ajuste por no elegibilidad de viviendas del estrato h .

n_h = Número de viviendas visitadas en el estrato h .

n_{h_ex} = Número de viviendas no elegibles en el estrato h .

Dado que la tasa de respuesta a nivel de vivienda se calcula a nivel de estrato, todas las UPM del mismo estrato comparten la misma corrección de cobertura.

La ponderación de segunda etapa se calcula para las viviendas seleccionadas dentro de las UPM incluidos en la muestra. En el caso de que existan viviendas excluidas de la investigación, se debe recalcular su probabilidad de selección. Por lo tanto, la ponderación de segunda etapa para la j -ésima vivienda en la UPM i del estrato h queda definida por:

$$pse_{f_{hij}} = \frac{pse_{hij}}{A_{nr_h} * A_{nr_h}}$$

donde:

$pse_{f_{hij}}$ = Probabilidad de selección final de segunda etapa para la UPM i del estrato h .

2.11.6. Ponderación de tercera etapa

La ponderación de tercera etapa representa el inverso de la probabilidad de selección de la tercera etapa asignada a personas de 16 años y más. Puesto que dentro de cada hogar de la vivienda j de la UPM i en el estrato h se selecciona únicamente a una persona con estas características, se simplifica la notación con subíndice k . Por lo tanto, la ponderación básica de tercera etapa para la persona seleccionada de 16 años o más k es:

$$pte_k = \frac{1}{M_k}$$

donde:

pte_k = Probabilidad de tercera etapa para la persona k de la vivienda i , UPM j y estrato h .

M_k = Número de personas elegibles en el hogar k de la vivienda i , UPM j y estrato h .

2.11.7. Ajuste por no respuesta para personas 16 años y más

Las ponderaciones de tercera etapa se calculan para las personas de 16 años y más seleccionadas dentro del hogar, por lo que se debe ajustar su probabilidad de selección. El ajuste se realizó por estrato y queda definido por:

$$A_{j_{hk}} = \frac{p_{16h}}{\sum_h \frac{1}{pte_k}}$$

donde:

Aj_{hk} = Ajuste por no respuesta a personas de 16 años y más del estrato h .

p_{16h} = Total de personas de 16 años y más elegibles en el estrato h .

$\sum_h \frac{1}{pte_k}$ = Total de personas de 16 años y más elegibles en los hogares en el estrato h .

La ponderación de tercera etapa se calcula para las personas de 16 años y más seleccionadas en los hogares de la muestra, la ponderación de tercera etapa para las personas de 16 años y más k queda definida por:

$$pte_{fk} = \frac{pte_k}{Aj_{nk}}$$

2.11.8. Factor de expansión final

Para el factor de expansión final se deben considerar las tres unidades de análisis: vivienda, hogar y personas de 16 años y más. Debido a lo antes mencionado, es necesario generar factores de expansión diferenciados.

Para viviendas y hogares el factor de expansión es:

$$fexp = \frac{1}{ppe_{f_{hi}} * pse_{f_{hi}}}$$

Mientras que la persona seleccionada el factor de expansión es:

$$fexp = \frac{1}{ppe_{f_{hi}} * pse_{f_{hi}} * pte_{fk}}$$

2.11.9. Postestratificación

Considerando que la muestra fue seleccionada a partir del Marco de Muestro para encuestas dirigidas a hogares del INEC, el cual está constituido en su mayoría por información que data del 2010 y que el periodo de levantamiento de información es 2019, existe una diferencia sustancial en la constitución de los dominios de estudio. Con el fin de solventar este inconveniente se utiliza un método denominado postestratificación.

Esta metodología tiene flexibilidad para realizar ajustes en el factor de expansión final, cuyas características deben ser completamente homologadas, de esta manera no se pierde los objetivos planteados en la encuesta de manera inicial, por lo tanto es

necesario utilizar las proyecciones de población generadas a partir del censo 2010, de esta manera se ajustan los totales poblacionales por los dominios planificados, de este modo la postestratificación está definida de manera general de la siguiente manera:

- Nacional urbano, nacional rural
- Sexo
- Grupo de edad (menor a 15 años y mayor o igual 15 años y más)
- Para la tercera etapa de selección el grupo de edad (mayor o igual 16 años y más)

Considerando las características demográficas de la población, por los dominios de estudio, se presenta el cálculo del factor de expansión por personas es:

$$fexp_{kp} = a_k * \frac{N_p}{\hat{N}_p}$$

donde:

$fexp_{kp}$ = Factor de expansión de persona k que pertenece al postestrato p .

a_k = Factor de expansión de persona k .

N_p = Tamaño de la población del postestrato p .

\hat{N}_p = Tamaño estimado de la población del postestrato p .

2.12. Estimación de características

Una vez obtenido los factores de expansión ajustados por cobertura, se calcula los estimadores de la encuesta multipropósito mediante el estimador de Horvitz-Thompson, el cual sirve para estimar el valor total de una característica determinada. Está dado por (Carl-Erik Särndal, 1992).

$$\hat{Y}_{mHT} = \sum_i \sum_j fexp_{kp} * y_k$$

donde:

\hat{Y}_{mHT} = Estimador insesgado para el total Y_k de una variable y en el dominio de estudio m .

i = Número de UPM investigadas en el dominio de estudio m .

j = Número de viviendas investigadas en la UPM i del dominio de estudio m .

y_k = Valor de la variable y para la unidad de análisis k .

2.13. Error de muestreo

Luego de realizar la estimación respectiva para la variable de interés a nivel de dominio de estudio, el error de muestreo podrá ser calculado a partir de la estimación de la varianza del estimador del total \hat{Y}_{mHT} . Para calcular adecuadamente los errores de muestreo de cada estimador, se debe tomar en cuenta los diferentes aspectos del diseño muestral, es decir, las dos etapas de muestreo, la estratificación presente en los dominios de estudio y los procesos de selección en cada una de las etapas.

Con todos estos elementos, el coeficiente de variación para el estimador \hat{Y}_{mHT} viene dado por la siguiente expresión:

$$CV(\hat{Y}_{mHT}) = \frac{\sqrt{\hat{V}_{2st} * (\hat{Y}_{mHT})}}{\hat{Y}_{mHT}}$$

Un estimador insesgado para la varianza está dado por:

$$\hat{V}_{2st}(\hat{t}_{\pi}) = \sum_{SI} \sum_{SI} \check{\Delta}_{ij} \frac{\hat{t}_{i\pi}}{\pi_{ii}} \frac{\hat{t}_{j\pi}}{\pi_{jj}} + \sum_{SI} \frac{\hat{V}_i}{\pi_{ii}}$$

En el cual el \hat{V}_i apropiado es:

$$\hat{V}_i = \sum_{SI} \sum_{SI} \check{\Delta}_{IIqr|i} \frac{\hat{t}_{iq\pi}}{\pi_{IIq|i}} \frac{\hat{t}_{ir\pi}}{\pi_{IIr|i}}$$

donde:

i, j = índice que recorre la UPM i, j en el dominio de estudio m .

q, r = índice que recorren las viviendas la UPM i, j en el dominio de estudio m .

π_{ii} = probabilidad de selección de etapa I para la i – ésima UPM en el dominio de estudio m .

$\pi_{IIq|i}$ = probabilidad de selección de etapa II para la q – ésima vivienda, dada la i – ésima UPM.

$\check{\Delta}_{ij}$ = cantidad Δ expandida asociada a la UPM i, j .

$\check{\Delta}_{IIqr|i}$ = cantidad Δ expandida asociada a las viviendas q, r dada la selección de la i – ésima UPM.

2.14. Métodos de estimación de errores para diseños muestrales complejos

Aunque la selección del diseño de muestreo y el estimador sean de libre elección para los investigadores, no lo es el cálculo de las medidas de confiabilidad y precisión. Dado que la base científica sobre la cual descansa el muestreo es la inferencia estadística, se deben respetar las normas básicas para la asignación y posterior cálculo del margen de

error, que constituye una medida unificada del error total de muestreo el cual cuantifica la incertidumbre acerca de las estimaciones en una encuesta.

Los métodos de estimación de los errores muestrales pueden clasificarse en cuatro categorías:

- a) Métodos exactos
- b) Métodos de la última UPM
- c) Aproximaciones por linealización
- d) Técnicas de replicación

Para la descripción de los métodos se ha tomado como referencia los textos de Kish y Frankel (1974), Wolter (1985) y Lehtonen y Pahkinen (1995). A continuación, se describe las principales características de cada uno de los métodos de estimación de errores para el muestreo complejo:

- Los métodos exactos pueden ser utilizadas para estimar totales, medias, tamaños y proporciones.
- La linealización de Taylor debe ser utilizada para estimar parámetros no lineales como razones, medias dentro de dominios, cuartiles o funciones de distribución.
- La técnica de la última UPM junto con la linealización de Taylor puede ser utilizada para estimar la varianza de los indicadores de interés de las encuestas dirigidas a hogares que tengan diseños muestrales complejos. Esta es la técnica que por defecto utiliza el software SPSS.
- Las técnicas de replicación pueden ser usadas para estimar eficientemente todos los parámetros de interés, sin importar su forma funcional.
- La comparación general entre los métodos de linealización y replicación, es que no generan resultados idénticos del error de muestreo, pero hay que señalar que existen estudios (Kish y Frankel, 1974) que concluyen que las diferencias presentadas no son significativas cuando se trata de grandes muestras.

El INEC utiliza para la estimación de los parámetros de interés y sus correspondientes errores de muestreo diversos programas estadísticos tales como SPSS, Stata y R. En virtud de las características de cada uno de los métodos, es la técnica de la última UPM en combinación con la linealización de Taylor, la cual induce a una muy buena aproximación del error muestral sobre los indicadores más importantes de las encuestas dirigidas a hogares, además de su facilidad de cálculo y replica. En este sentido, será

esta la técnica la utilizada para la estimación de los errores muestrales en la encuesta multipropósito 2019.

Las variables requeridas para declarar el diseño muestral en los programas estadísticos (SPSS, Stata y R) y ejecutar el cálculo de los errores de muestreo son presentadas en la Tabla 6, donde se describe las etiquetas de las variables identificadoras de las UPM, estratos y ponderación.

Tabla 5. Variables requeridas para la declaración del diseño muestral

Característica	Variable	Descripción
Unidad primaria de muestreo	upm	Agrupación de viviendas ocupadas en un número entre 30 a 60, próximas entre sí y con límites definidos.
Estratos	Estrato	Identificación de estrato muestral (aproximación clasificación socio-económica).
Ponderación	Fexp	Factor de expansión calibrado.

Fuente: Elaboración propia, INEC 2019


Bibliografía


Carl-Erik Särndal, B. S. (1992). *Model Assisted Survey Sampling* (1ra Edición ed.). Nueva York, Nueva York, USA: Springer-Verlag.

Kish, L. (1972). *Muestreo de Encuestas*. México: Trillas.

Särndal, C., Swensson, B., & Wretman, J. (2003). *Model Assisted Survey Sampling*, Nueva York, USA. New York: Springer.


**CADA
HECHO
DE TU
VIDA**
Cuenta

 @ecuadorencifras

 INEC/Ecuador

 @InecEcuador

 INECEcuador

 t.me/equadorencifras

 INEC Ecuador