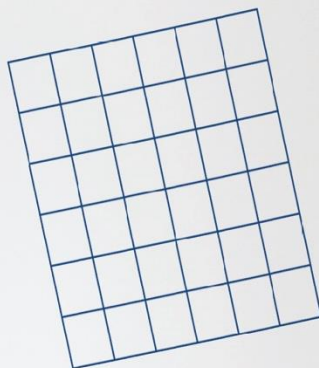


**Encuesta Nacional
Multipropósito de
Hogares
(Seguimiento al
Plan Nacional de
Desarrollo)**

**METODOLOGÍA DE
DISEÑO MUESTRAL**



Roberto Castillo

DIRECTOR EJECUTIVO (e)

Magaly Paredes

SUBDIRECTORA GENERAL (e)

Markus Nabernegg

COORDINADOR GENERAL TÉCNICO DE PRODUCCIÓN ESTADÍSTICA

Christian Garcés

DIRECTOR DE INFRAESTRUCTURA ESTADÍSTICA Y MUESTREO

Equipo técnico:

Francisco Céspedes Navas, Angel Gaibor Costa, Christian Garcés Mayorga.

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 Nacional Multipropósito de Hogares (Seguimiento al Plan Nacional de Desarrollo). Instituto Nacional de Estadística y Censos, Quito-Ecuador.

Contenido

1. Introducción.....	4
2. Antecedentes.....	5
3. Diseño Muestral ESPDN	6
3.1. Determinación del Marco de Muestreo.....	6
3.2. Universo de investigación	6
3.3. Unidad de observación	6
3.4. Unidad de análisis	6
3.5. Cobertura geográfica	7
3.6. Plan Muestral	7
3.7. Dominios de estudio	7
3.8. Variables de diseño	7
3.9. Tamaño muestral.....	7
3.10. Selección de la muestra	10
3.11. Control de cobertura de campo y muestral.....	10
3.12. Factores de expansión.....	11
3.12.1. Ponderación de primera etapa (conglomerado).....	12
3.12.2. Ajuste por no respuesta a nivel de conglomerado	12
3.12.3. Ponderación de segunda etapa (vivienda)	13
3.12.4. Ajuste por no respuesta a nivel de vivienda	13
3.12.5. Ponderación de tercera etapa (persona)	13
3.12.6. Ajuste por no respuesta a nivel de persona	14
3.12.7. Ponderadores muestrales totales	14
3.12.8. Factores de expansión ajustados (población)	14
3.12.9. Postestratificación.....	15
3.13. 2.4. Estimación de características	16
3.14. Error de muestreo	16
BIBLIOGRAFÍA	18

1. Introducción

El Instituto Nacional de Estadísticas y Censos (INEC), organismo coordinador de la Producción Estadística en el Ecuador, y en busca de monitorear los diferentes aspectos sociales, demográficos, salud y económicos de la población, realizará anualmente a partir de este período la encuesta de Seguimiento al Plan Nacional de Desarrollo dirigido a hogares delimitado a nivel nacional.

La Dirección de Estadísticas Sociodemográficas como responsable del levantamiento de la información en campo y los procesos operativos anexos, solicita el apoyo de la Dirección de Infraestructura Estadística y Muestreo en la generación del diseño muestral para la "Encuesta Nacional Multipropósito de Hogares (Seguimiento al Plan Nacional de Desarrollo)", cuya operación estadística se realiza a nivel nacional, tomando en cuenta las diferentes temáticas consideradas y aprobadas en las comisiones estadísticas.

El proceso que se realiza para la generación del diseño muestral consta de tres componentes:

En el primer componente se toma en cuenta:

- Determinación del marco de muestreo
- El universo de la investigación
- Las unidades de muestreo y análisis
- Cobertura geográfica
- Cobertura temática

En el segundo componente:

- Metodología para el diseño muestral
- Determinación de los dominios de estudio
- Determinación del tamaño de la muestra
- Selección de la muestra
- Validación de la cobertura obtenida en campo

Por último, se realiza:

- El cálculo de los factores de expansión
- Post-estratificación
- Estimación de características
- Error de muestreo

Esta última etapa incluye la post estratificación como técnica para realizar ajustes sobre la estructura poblacional por edades y sexo, esto debido a que existen casos donde los informantes declaran la edad con aproximaciones, conocida como atracción a los dígitos.

2. Antecedentes

Los Planes Nacionales de Desarrollo pasados se monitoreaban con la Encuesta Nacional de Empleo, Desempleo y Subempleo-ENEMDU, esta operación estadística, en sus inicios fue diseñada para obtener información del mercado laboral, condición de actividad de la población, pobreza y extrema por ingresos, rama de actividad entre otros. Sin embargo en el transcurso del tiempo se ha incrementado a la encuesta temas como Educación, TICS, Datos de la Vivienda y el Hogar, Cobertura de Programas Sociales, Percepción de la calidad de los Servicios Públicos, Participación Ciudadana y Seguridad Ciudadana que respondía al interés de medir continuamente los indicadores para el monitoreo de la metas planteadas en los planes de desarrollo. La información que se obtiene de la ENEMDU para los temas diferentes a indicadores laborales, no asegura una significancia estadística¹ para todas las desagregaciones que se tienen en las fichas metodológicas de los indicadores del Plan Nacional de Desarrollo.

Fue necesario disponer de una operación estadística multi-temática diseñada para el monitoreo de los indicadores planteados en el Plan Nacional de Desarrollo, que garantice la significancia estadística al menos en su dominios de estudio.

En este contexto, en el 2018, se realizó la primera investigación de la Encuesta Nacional Multipropósito de Hogares (Seguimiento al Plan Nacional de Desarrollo), con una representatividad nacional, urbano, rural, la misma que permitirá conservar una línea de continuidad en los diagnósticos de la población, a través de los indicadores que fueron monitoreados con información de la Encuesta Nacional de Empleo, Desempleo y Subempleo – ENEMDU.

Con estos antecedentes, el propósito del presente documento es realizar la descripción técnica del diseño muestral propuesto la ejecución de la Encuesta Multipropósito, posterior a ello se definirán los aspectos importantes en la comparabilidad de las estimaciones a partir de la línea base ENEMDU diciembre 2017.

¹ Cumplimiento de requisitos estadísticos mínimos (n muestra, coeficiente de variación, efecto de diseño, etc.).

3. Diseño Muestral ESPDN

3.1. Determinación del Marco de Muestreo.

El marco de la muestra se basa en los resultados definitivos y la cartografía del VII Censo de Población y VI de Vivienda del 2010 (CPV-2010); este Marco tiene la característica de ser constituido por áreas geográficas, que tienen límites perfectamente definidos e identificables sobre el terreno.

El marco ha tenido actualizaciones parciales en función a los cambios presentados en las de unidades de observación (viviendas), en determinadas áreas geográficas, 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.

Se han ejecutado en el periodo 2014 – 2017 diferentes procesos de actualización cartográfica con diferente alcance, las cuales se detallan a continuación:

- | | |
|---|-------------------------|
| • Encuesta Condiciones de Vida 2013-2014: | 2.425 sectores censales |
| • Actualización ENEMDU 2014: | 5.564 sectores censales |
| • Proyecto 2015: | 548 sectores censales |
| • Actualización 2017: | 1.779 sectores censales |

3.2. Universo de investigación

El universo de estudio de la Encuesta Multipropósito 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).

3.3. Unidad de observación

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

3.4. 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.

3.5. Cobertura geográfica

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

3.6. Plan Muestral

El diseño muestral para esta investigación corresponde a un muestreo probabilístico con tres etapas de selección:

- **Primera etapa:** selección de Unidades Primarias de Muestreo (conglomerados) por estrato.
- **Segunda etapa:** selección de viviendas ocupadas dentro de cada uno de los conglomerados seleccionados en la primera etapa.
- **Tercera etapa:** Selección de una persona de 16 años y más dentro de cada hogar dentro de las viviendas seleccionadas en la segunda etapa

3.7. Dominios de estudio

Los dominios de estudio para la Encuesta Multipropósito 2018 son los siguientes:

- Nacional
- Nacional Urbano
- Nacional Rural
- Provincial (Proporción de la población que vive en hogares con acceso a servicios básicos)

3.8. Variables de diseño

El tamaño de la muestra se calcula en función a su variable de diseño, esta consideración se realizó posterior a la conciliación con el resto de indicadores que se presentaron dentro del plan de seguimiento, de esta manera se optimizó el tamaño para satisfacer los objetivos planteados por parte de Encuesta Multipropósito.

3.9. Tamaño muestral

Para la determinación del tamaño de la muestra se requiere establecer la característica o características a estimar, el nivel de confianza y la precisión requeridas, de tal manera que los resultados obtenidos no sean demasiado costosos y/o imprecisos.

En este sentido, las variables de investigación que se utilizaron para determinar el tamaño de la muestra son 19 indicadores que fueron identificados de manera previa y socializada con los diversos actores interinstitucionales, dos de ellos tienen una desagregación rural.

Tabla 1: Variables de Investigación

Nº	Variables de investigación
1	Confianza en el Consejo de la Judicatura
2	Confianza en la Policía Nacional
3	Confianza en la Defensoría Pública
4	Confianza en la Fiscalía General del Estado
5	Hogares con acceso a agua por red pública y saneamiento adecuado
6	Índice de confianza en las fuerzas armadas
7	Índice de percepción de la calidad de los servicios públicos en general
8	Porcentaje de hogares que dispone de una instalación para lavarse las manos con agua y jabón(básico)
9	Porcentaje de hogares que usa servicios de saneamiento básico
10	Porcentaje de hogares que utiliza suministros seguros de agua para beber
11	Porcentaje de hogares que utiliza suministros seguros de agua para beber y saneamiento básico
12	Porcentaje de la población que realiza actividad física más de 3,5 horas a la semana
13	Porcentaje de personas entre 18 y 29 años de edad con bachillerato completo
14	Proporción de la población que vive en hogares con acceso a servicios básicos
15	Tasa neta de asistencia a educación general básica
16	Tasa neta de asistencia ajustada a bachillerato en el área rural
17	Tasa neta de asistencia bachillerato ajustada
18	Tasa de pobreza multidimensional (TPM)
19	Tasa de pobreza multidimensional en el área rural

Fuente: INEC- Metodología Encuesta Multipropósito

La fórmula para la determinación del tamaño de la muestra corresponde a un muestreo probabilístico para el estimador de proporciones.

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

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 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.

A partir del marco de muestreo constituido por 107.014 Unidades Primarias de Muestre (UPM), se ha diseñado una muestra de tamaño igual a 26.928 viviendas y 2.244 UPM, con el 12% de error relativo y un nivel de confianza del 95%, estos parámetros se establecieron para garantizar que las estimaciones resultado de la encuesta sean representativos en las áreas geográficas y grupos poblacionales requerido por los indicadores de diseño, que son parte del Plan Nacional de Desarrollo.

Como paso siguiente 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) y el porcentaje de personas con la característica de interés ($r =$ grupo poblacional), de la siguiente forma:

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

Los hogares y las personas se observan a partir de las UPM. En este paso final es necesario calcular el número de UPM's 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 será de 12 viviendas por UPM, es decir la misma de la ENEMDU diciembre 2017.

La muestra se distribuye de la siguiente manera:

Tabla 2. Distribución de la muestra por provincias

Dominios	Provincias	Número de viviendas	Número de UPM
1	Azuay	804	67
2	Bolívar	1.068	89
3	Cañar	804	67
4	Carchi	876	73
5	Cotopaxi	1.356	113
6	Chimborazo	816	68
7	El Oro	1.116	93
8	Esmeraldas	1.440	120
9	Guayas	1.632	136
10	Imbabura	1.572	131
11	Loja	804	67
12	Los Ríos	1.812	151
13	Manabí	1.164	97
14	Morona Santiago	1.344	112

15	Napo	804	67
16	Pastaza	1.404	117
17	Pichincha (Excepto Cantón Quito)	1.452	121
18	Tungurahua	804	67
19	Zamora Chinchipe	804	67
20	Galápagos	552	46
21	Sucumbíos	1.632	136
22	Orellana	1.164	97
23	Santo Domingo de los Tsáchilas	804	67
24	Santa Elena	804	67
90	Zonas no Delimitadas	96	8
TOTAL		26.928	2.244

Fuente: INEC- Diseño Muestral

Este tamaño de muestra garantiza representatividad a nivel nacional urbano y nacional rural de la totalidad de indicadores utilizados como diseño.

3.10. Selección de la muestra

En primera instancia, la selección de las UPM's 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.

Para satisfacer los módulos que fueron planificados para personas de 16 años y más, se seleccionó mediante el método del cumpleaños más próximo; método que garantiza la aleatoriedad en la selección de la persona y que además facilita su aplicación en campo.

3.11. Control de cobertura de campo y muestral

La cobertura hace referencia a la información obtenida en campo y proporcionada por los informantes de las viviendas seleccionadas puede estar constituida de dos formas:

Tabla 3. Estructura de condición de levantamiento de información

Condición	Novedad
Efectiva	Completa
	Rechazo
	Nadie en casa
	Vivienda Temporal
No efectiva	Vivienda desocupada
	Vivienda en construcción
	Vivienda inhabitable o destruida
	Vivienda convertida en negocio
	Otra razón

Fuente: Encuesta Multipropósito

Se debe considerar que la cobertura está en función de cada uno de los dominios y se toma en cuenta variables como:

- universo,
- muestra mínima requerida; y,
- muestra final.

Posteriormente, se realizan seguimientos de cobertura con el fin de verificar el total de viviendas efectivas, ya que de éste dependerán las posteriores estimaciones en los indicadores.

Para corresponder con el número de UPM efectivas con mínimos 8 viviendas con información completa, se excluyó a las UPM que no cumplieran con esta característica, además por cuestiones de calidad en la información y por efectos de validación se excluyó a las viviendas que no cumplieran con la totalidad de información en los formularios respectivos, cabe mencionar que el máximo de viviendas por UPM fue 12 viviendas más 4 reemplazos.

Se obtuvo un total de 2.811 UPM como total, pero por las características mínimas anteriormente mencionadas se concluyó con 2.445 UPM con viviendas efectivas y con los parámetros de calidad respectiva.

3.12. Factores de expansión

El procedimiento de ponderación general para la Encuesta Multipropósito requiere de tres etapas.

- **La primera etapa** consiste en calcular una ponderación de conglomerado. De ser necesario, se aplica un ajuste por no respuesta a nivel de conglomerado.
- **La segunda etapa** consiste en calcular una ponderación de viviendas dentro del conglomerado. De ser necesario se aplica un ajuste por no respuesta a nivel de vivienda.
- **La tercera etapa (formularios F1 y F2)** consiste en calcular una ponderación a nivel de la persona seleccionada, realizando el cálculo a nivel de vivienda para posteriormente inferir por población de 16 años y más, con respecto al número de personas en el hogar.

Una de las novedades a considerar dentro del cálculo de los factores de expansión es la falta de respuesta, la misma que se produce cuando no llegan a obtenerse respuestas de algunas de las unidades de la muestra. Resulta útil pensar en la población de la muestra dividida en dos conjuntos, uno formado por todas las unidades de la muestra de las que se ha obtenido respuesta y otro por todas las unidades de la muestra de las que no ha podido obtenerse respuesta.

La tasa de no respuesta puede estimarse con exactitud si se llevan a cabo recuentos de todos los elementos que cumplen los requisitos que se incluyen en la muestra. La tasa de respuesta en una encuesta se define como el coeficiente entre el número de cuestionarios completados por las unidades y el número total de unidades de la muestra que cumplen los requisitos². La falta de respuesta puede deberse a la ausencia del domicilio de las personas

² Algunas unidades incluidas en la muestra puede que se hallen fuera del ámbito de la encuesta y no cumplan los requisitos. Tal podría ser el caso, por ejemplo, de las viviendas vacías, en ruina o abandonadas.

seleccionadas, a los cambios en la condición de ocupación de la vivienda, a la negación de dichas personas a participar o a su incapacidad para responder a las preguntas. Otra posible causa de la falta de respuesta puede ser la imposibilidad de llevar a cabo la encuesta en determinadas zonas por razones climatológicas, dificultades del terreno o cuestiones de seguridad.

El procedimiento a seguir para el cálculo de los factores de expansión se describe a continuación:

3.12.1. Ponderación de primera etapa (conglomerado)

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

Como se mencionó en apartados anteriores, una vez realizado el equilibrio en el número de viviendas que constituyen las UPM, es posible obtener la probabilidad de selección de cada conglomerado de una manera aleatoria simple dentro de cada estrato.

La ponderación básica de primera etapa para el i -ésimo conglomerado muestreado en el estrato k queda definida por:

$$PB_c^{ki} = \frac{C_k}{c_k},$$

donde C_k es el número de conglomerados en el estrato k , c_k es el número de conglomerados muestreados en el estrato k .

3.12.2. Ajuste por no respuesta a nivel de conglomerado

Para realizar este ajuste, es necesario definir la tasa de no respuesta a nivel de conglomerado, la cual mide la proporción de conglomerados seleccionados que efectivamente participaron en la encuesta. El cálculo de la tasa de no respuesta usa el siguiente término, derivado de la recolección de datos:

$$c_{kex} = \text{número de conglomerados seleccionados que fueron excluidos en el estrato } k$$

La tasa de no respuesta viene dada por:

$$TNR_k = \frac{c_k - c_{kex}}{c_k}$$

Las ponderaciones de la primera etapa se calculan para los conglomerados seleccionados. Si existieran conglomerados excluidos de la investigación se debe recalcular la ponderación de primera etapa para el i -ésimo conglomerado investigado en el estrato k , queda definida por:

$$PF_c^{ki} = \frac{PB_c^{ki}}{TNR_k}$$

3.12.3. Ponderación de segunda etapa (vivienda)

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 para la vivienda j en el conglomerado i viene dada por:

$$PB_v^{ij} = \frac{V_i}{v_i}$$

donde V_i es el número de viviendas en el conglomerado i , v_i es el número de viviendas muestreadas en el conglomerado i .

3.12.4. Ajuste por no respuesta a nivel de vivienda

La tasa de no respuesta a nivel de vivienda, mide la proporción de viviendas seleccionadas que efectivamente participaron en la encuesta. El cálculo de la tasa de respuesta usa el siguiente término, derivado de la recolección de datos:

$$v_{iex} = \text{número de viviendas seleccionados que fueron excluidos en el conglomerado } c$$

La tasa de no respuesta viene dada por:

$$TNR_c = \frac{v_i - v_{iex}}{v_i}$$

Las ponderaciones de segunda etapa se calculan para los conglomerados seleccionados. En el caso de que existan viviendas excluidas de la investigación, se debe recalcular la ponderación de primera etapa para la j -ésima vivienda en el conglomerado i , queda definida por:

$$PF_v^{ij} = \frac{PB_v^{ij}}{TNR_i}$$

3.12.5. Ponderación de tercera etapa (persona)

La ponderación de tercera etapa representa el inverso de la tercera probabilidad de selección asignada a cada persona. Para el caso de la Encuesta Multipropósito se seleccionó una persona de 16 años y más³ con igual probabilidad de selección, por lo que la ponderación básica de segunda etapa para la persona j en la vivienda i viene dada por:

³ Método de selección: cumpleaños próximo en persona de 16 años y más miembro del hogar.

$$PD_p^{ij} = \frac{P_i}{p_i}$$

donde P_i es el número de personas en la vivienda/hogar i , p_i es el número de personas muestreadas en el conglomerado i .

3.12.6. Ajuste por no respuesta a nivel de persona

La tasa de no respuesta a nivel de persona, mide la proporción de personas de 16 años y más seleccionadas que efectivamente participaron en la encuesta. El cálculo de la tasa de respuesta usa el siguiente término, derivado de la recolección de datos:

$$p_{iex} = \text{número de personas seleccionados que fueron excluidos en la vivienda/hogar } c$$

La tasa de no respuesta viene dada por:

$$TNR_c = \frac{p_i - p_{iex}}{p_i}$$

Las ponderaciones de tercera etapa se calculan para los viviendas/hogar seleccionados. En el caso de que existan personas excluidas de la investigación, se debe recalcular la ponderación de segunda etapa para la j -ésima persona en la vivienda i , queda definida por:

$$PD_p^{ij} = \frac{PB_p^{ij}}{TNR_i}$$

3.12.7. Ponderadores muestrales totales

El ponderador muestral total es el producto del ponderador final de primera etapa por el ponderador final de segunda etapa. Para la vivienda j investigada en el conglomerado i del estrato k , el ponderador muestral es:

$$P_T^j = PF_c^{ki} * PB_v^{ij} * PD_p^{ij}$$

El uso de ponderadores muestrales es crítico para obtener estimadores a partir de encuestas que usan técnicas distintas del muestreo aleatorio simple.

3.12.8. Factores de expansión ajustados (población)

Para concluir con los ajustes, a la ponderación calculada en el apartado anterior, se realiza un ajuste que toma en cuenta la proyección oficial de la población.

Este cálculo inicia luego de determinar la población expandida con este primer ajuste. El factor de ajuste por la proyección de la población se calcula de la siguiente manera:

$$\text{Ajuste proyección} = \frac{\text{Población proyección esperada para el periodo en estudio}}{\text{Población expandida de la encuesta}}$$

Este ajuste afecta a cada conglomerado i , por dominios de estimación.

3.12.9. Post-estratificación

El factor de expansión, para un estudio basado en una muestra, es la cifra que actuando como un multiplicador, permite llevar los datos muestrales a la población (expandir la muestra).

La post-estratificación es un procedimiento similar a la estratificación del Marco Muestral, con la diferencia que la post-estratificación es realizada después de haber obtenido los resultados del estudio realizado. Como se señaló anteriormente, se estableció que la estratificación constituye una herramienta para la selección de la muestra para nuestro estudio y que ésta se realiza en las primeras etapas del diseño.

La post-estratificación se realiza una vez que se cuenta con los datos obtenidos en campo. Dado que antes de realizar el estudio no se conocen con detalle muchas características de la población, que luego de terminado el levantamiento se cuenta con variables tales como la edad, sexo, entre otras; las mismas que permiten afinar y re-estratificar nuevos grupos con propósitos diversos, es decir, que para la post-estratificación se utilizan variables obtenidas en campo.

De igual manera que en la estratificación tradicional, la post-estratificación consiste en la creación de nuevos grupos de elementos denominados post-estratos y éstos poseen características homogéneas al interior y heterogéneas al exterior de los mismos, basados en variables investigadas en nuestro estudio.

Luego de la post-estratificación, cada uno de los elementos dentro de los datos muestrales será identificado como perteneciente a un post-estrato, según los criterios establecidos, que para el caso de la Encuesta Multipropósito son la población por área, el dominio de estudio, el sexo y los grupos de edad.

El ajuste de los factores de expansión por post-estratificación consiste en la creación de un componente extra (δ)⁴ en la fórmula de cálculo, para ajustar la población estimada obtenida en el paso anterior, en subgrupos o post-estratos que para el caso de la ENEMDU son el sexo (hombre y mujer) y grupos de edad de interés (menores de 15 años, 15 a 64 años, 65 años y más). Este ajuste se lo realiza utilizando como dato o parámetro conocido a

⁴ Este componente "delta" permitirá el ajuste de los factores de expansión a fin de que la Población expandida en el estudio coincida con el dato real, que en este caso es una proyección de la población a la fecha de ejecución de la encuesta.

la proyección de población para la fecha de realización de la encuesta, desagregada en los post-estratos construidos.

Este componente de ajuste resulta del cociente de la población proyectada entre la población expandida.

Para el cálculo del componente de ajuste se utilizó la siguiente fórmula:

$$\delta = \frac{U}{U_{hat}}$$

Donde:

U = Proyección al momento de la encuesta por post-estrato.

Uhat = Expansión de la población por medio de la encuesta por post-estrato.

Contando con los datos del número de personas proyectadas, el número de personas expandidas y el factor de expansión ajustado (población), se aplica la fórmula para cada uno de los post-estratos.

Con los valores obtenidos para los correspondientes componentes de ajuste (δ) para cada uno de los post-estratos, se proceda a multiplicar el factor originalmente calculado por el componente de ajuste, y el resultado obtenido es el nuevo factor de expansión ajustado o calibrado.

3.13. 2.4. Estimación de características

Una vez obtenidos los factores de expansión ajustados, se calcula los estimadores provenientes de la Encuesta Multipropósito, se utiliza el estimador de Horvitz-Thompson, el cual sirve para estimar el valor total de una característica determinada (empleo adecuado, desempleo, etc). Está dado por (Carl-Erik Särndal, 1992).

$$\hat{Y}_{mHT} = \sum_{i=1}^{C_m} \sum_{j=1}^{v_i} P_T^j y_{ij},$$

Donde:

- \hat{Y}_{mHT} = *estimador insesgado para el total Y_k de una variable y en el dominio de estudio m*
- C_m = *número de conglomerados investigadas en el dominio de estudio m.*
- v_i = *número de viviendas investigadas en el conglomerado i del dominio de estudio m*
- y_{ij} = *valor de la variable y para la vivienda j del conglomerado i del dominio de estudio m*

3.14. 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

tres 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}_{kHT} viene dado por la siguiente expresión:

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

Un estimador insesgado para la varianza está dado por:

$$\hat{V}_{2st}(\hat{t}_{\pi}) = \sum \sum_{s_I} \check{\Delta}_{Iij} \frac{\hat{t}_{i\pi}}{\pi_{Ii}} \frac{\hat{t}_{j\pi}}{\pi_{Ij}} + \sum_{s_I} \frac{\hat{V}_i}{\pi_{Ii}}$$

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

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

Dónde:

- i, j = Índice que recorre los conglomerados i, j en el dominio de estudio m .
- q, r = Índices que recorren las viviendas del conglomerado i en el dominio de estudio m .
- π_{Ii} = Probabilidad de selección de la Etapa I para el i – ésimo conglomerado en el dominio de estudio m .
- $\pi_{IIq|i}$ = Probabilidad de selección de la Etapa II para la q – ésima vivienda, dada el i – ésimo conglomerado.
- $\check{\Delta}_{Iij}$ = Cantidad Δ expandida asociada a los conglomerados i, j .
- $\check{\Delta}_{IIqr|i}$ = Cantidad Δ expandida asociada a las viviendas q, r dada la selección del i – ésimo conglomerado.

BIBLIOGRAFÍA

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

Educación, M. d. (2016). *Archivo Maestro de Instituciones Educativas*. Obtenido de <https://educacion.gob.ec/amie/>


Educación, M. d. (s.f.). *Archivo Maestro de Instituciones Educativas*.

Hidiroglu, M., Särndal, C., & Binder, D. (1995). Weighting and Estimation in Business Surveys. En Cox, Binder, Chinnappa, Christianson, Colledge, & Kott, *Business Survey Methods* (págs. 477-502). Jhon Wiley & Sons.


Peña, D. (2002). *ANÁLISIS DE DATOS MULTIVARIANTES*.

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


**CADA
HECHO
DE TU
VIDA**
Cuenta

 @ecuadorencifras

 INEC/Ecuador

 @InecEcuador

 INECEcuador

 t.me/equadorencifras

 INEC Ecuador