

Cálculo del tamaño y selección de la muestra

Unidades Educativas



DINEM

JULIO· 2025

Subtítulo 01

1. **Introducción**

El Índice de Precios al Consumidor (IPC) en Ecuador ha evolucionado constantemente desde su primera base en 1950-1951, adaptándose a las dinámicas del consumo y a los cambios socioeconómicos del país. Considerando la relevancia y utilidad del indicador y fundamentándose en las recomendaciones internacionales de actualizar el índice por lo menos una vez cada cinco años para mantener la representatividad del índice con respecto al patrón de consumo de los hogares, con el proyecto Cambio de Base del Sistema de Indicadores de Precios al Consumidor en el Ecuador (CAB-SIPCE), se busca realizar el séptimo cambio de base del IPC.

El IPC es un índice mensual y nacional, que permite medir los cambios en el tiempo del nivel general de los precios de consumo final de bienes y servicios de los hogares residentes en el área urbana del país.

Se calcula para un periodo de referencia dado y para nueve ciudades autorepresentadas.

El Índice de Precios al Productor de Disponibilidad Nacional (IPP-DN) es un indicador económico mensual investigado y calculado por el Instituto Nacional de Estadística y Censos (INEC). Su principal variable de investigación es el precio al productor de los productos investigados en los establecimientos económicos a nivel nacional exceptuando la región insular, en los sectores de: agricultura, ganadería, silvicultura, pesca y manufactura. Actualmente el Índice de Precios al Productor de Disponibilidad Nacional (IPP-DN) se desarrolla mediante un muestreo dirigido o intencional en su totalidad. En el marco del proceso de certificación del IPP-DN, ejecutado en el año 2017, se emitieron diversas observaciones (acciones de mejora) relacionadas con cada una de las fases de producción de esta operación estadística. Particularmente, en las fases de metodología sólida y diseño, se evidenció la necesidad de evaluar la implementación de un diseño de muestreo probabilístico a partir de un marco de muestreo en donde se incluyan a todas las empresas productoras, así como evaluar el uso de material cartográfico.

En consecuencia, el Proyecto Cambio de Base del Sistema de índices de Precios al Productor (CAB-SIPP) actualiza integralmente la metodología del IPP-DN año base 2015=100; y, prueba la aplicación del diseño muestral probabilístico; adicional, entre estas actualizaciones se redefine el nombre del indicador como Índice Nacional de Precios Productor (INPP), por lo que a lo largo del documento lo citaremos de esa manera.

En atención a la Acción de Mejora y en lo que compete a la Dirección de Infraestructura Estadística y Muestreo (DINEM), este informe presenta el ejercicio referente a la evaluación exhaustiva de la viabilidad de implementar o no un diseño muestral probabilístico. Es importante recalcar que el INPP investiga los productos definidos en la canasta de investigación de bienes y servicios, por lo que, lo ideal sería trabajar con un marco muestral detallado a nivel de producto y, a su vez, con este insumo elaborar un diseño muestral cuya unidad de análisis esté definida a nivel de producto, sin embargo, no se cuenta con dicha información. En tal sentido, el presente ejercicio pretende experimentar con el único insumo disponible que es el Registro Estadístico de Empresas 2022 (REEM-2022), usando como unidad de análisis la empresa y considerando los códigos CIIU (Clasificación Industrial Internacional Uniforme). Dicho código alfanumérico identifica y clasifica las actividades económicas de una entidad, los códigos CIIU a utilizar están disponibles en el REEM-2022 y se consideraran los catalogados desde uno hasta seis dígitos, lo que se pretende con ello es generar una aproximación al insumo ideal.

Es importante señalar que previamente se realizó una prueba piloto considerando empresas que pertenecen a varios sectores económicos; obteniéndose como resultado que el sector manufacturero presenta los mejores niveles de cobertura, por lo que el equipo responsable del CAB-SIPP solicita considerar únicamente dicho sector para el presente ejercicio. A tal efecto, se definen tres etapas: Inicialmente se evaluarán los insumos que nos permitan definir un adecuado marco de muestreo, universo de investigación, unidades de muestreo y cobertura temática. Posteriormente, se procede con el análisis de la metodología para la definición de un adecuado diseño muestral, definición de dominios de estudio, determinación del tamaño de la muestra y selección de la muestra. Finalmente, se evaluará la cobertura con un análisis que nos permita evidenciar los resultados que se han alcanzado.

1. **Planificación**

Objetivo General

Calculo del tamaño de muestra y selección aleatoria de instituciones educativas.

Tipo de operación estadística

La operación estadística es una encuesta dirigida a unidades educativas.

Cobertura temática

Las unidades educativas particulares con al menos 3 niveles educativos, ubicadas en el área urbana de los cantones: Manta, Machala, Santo Domingo, Esmeraldas, Ambato, Cuenca y Loja.

Universo de estudio

Se consideran aquellas unidades educativas que cumplen con los aspectos definidos en torno a la cobertura temática.

Unidad de observación y de análisis

Unidades educativas.

Cobertura geográfica

La cobertura geográfica está definida por todas las unidades educativas particulares que se encuentran en el área urbana de los siguientes cantones: Manta, Machala, Santo Domingo, Esmeraldas, Ambato, Cuenca y Loja.

1. **Metodología del marco muestral y el diseño de la muestra**

Tipos y etapas del diseño muestral

Muestreo probabilístico con selección aleatoria de elementos.

Dominios de estudio

Un dominio de diseño consiste en una subpoblación que puede identificarse en el marco de muestreo y, por lo tanto, puede ser manejada independientemente en el tamaño de muestra, procedimientos de muestreo y demás.

Para el presente diseño de muestra se definió un solo dominio de estudio para el cálculo del tamaño muestral, posterior a ello se procede con la distribución en cada uno de los cantones que se pretende llevar la encuesta.

Los dominios de estudio para el análisis muestral están definidos por el cantón al que pertenece cada unidad educativa, de esta manera, por cada cantón se tendrá un tamaño de muestra. Se considera un grupo de inclusión forzosa, es decir, un subconjunto de unidades educativas que formarán parte de la muestra con probabilidad uno (1), este grupo corresponde a instituciones catalogadas como “indispensables” para el estudio por su cantidad de estudiantes e importancia en el entorno educativo de cada cantón, este listado ha sido elaborado por el equipo DECON - CAB-SIPCE.

## Marco muestral

Un marco muestral es una lista de los elementos que forman el universo que será objeto de estudio, dichos elementos son llamados unidades muestrales que tienen definidas características que nos permitan identificarlos. Con el marco muestral definido correctamente podremos extraer la población que se va a investigar, es decir, la muestra que se analizará. Para realizar un muestreo probabilístico es sumamente necesario construir un marco muestral ya que permite encontrar la unidad que va a ser el objeto de estudio. El marco muestral debe ser preciso, libre de omisiones y duplicaciones. Los datos y las unidades deben cubrir a toda la población e identificarlas correctamente.

El marco de muestreo que se considera está construido a partir del listado de unidades educativas que fue elaborado por el equipo CAB-SIPCE que usó como insumo la información de precios de pensiones y matrículas de Unidades Educativas (registros administrativos) del Ministerio de Educación (MINEDUC).

Para la construcción del marco muestral se seleccionaron aquellas unidades educativas que cumplan las siguientes condiciones:

* Las instituciones particulares con al menos 3 niveles educativos que se encuentren ubicadas en el área urbana de Manta, Machala, Santo Domingo, Esmeraldas, Ambato, Cuenca y Loja.
* Se excluyen unidades educativas que no disponen de información en las variables costo de matrícula y/o costo de pensión.

El marco de muestreo queda constituido por 186 unidades educativas.

## Variables de diseño

Para el análisis del tamaño muestral se ha considerado como variable de diseño el “costo medio de matrícula” que es un dato actualizado en cada año lectivo de las instituciones educativas.

## 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 requerida de tal manera que los resultados obtenidos no sean demasiado costosos y/o imprecisos.

Bajo este criterio, para determinar el tamaño de la muestra se usa como variable de diseño “costo medio de matrícula”, dicha variable está disponible en el marco muestral. La fórmula para dicho cálculo se presenta a continuación:

* = Tamaño de la muestra.
* = Tamaño del dominio i.
* = Cuasivarianza de la variable de diseño para el dominio i.
* e = Error relativo máximo admisible.
* z = Coeficiente que representa el nivel de seguridad o confianza.
* = Variable de diseño (ventas totales) en el dominio i.
* = Tasa de no respuesta del dominio i.

Se cuenta con listado de unidades educativas que fueron definidas como de inclusión forzosa. De esta manera el tamaño muestral resultante está constituido por el cálculo a nivel nacional y las unidades educativas de inclusión forzosa, con lo que el tamaño resultante es de 86 unidades educativas.

Tabla 2. Tamaño muestral

|  |  |
| --- | --- |
| **Cantón** | **Tamaño** |
| AMBATO | 15 |
| CUENCA | 16 |
| ESMERALDAS | 6 |
| LOJA | 7 |
| MACHALA | 13 |
| MANTA | 17 |
| SANTO DOMINGO | 12 |
| **Total general** | **86** |

*Fuente: Gestión de Diseño Muestral – DINEM*

*Inclsuión forsoza*

|  |  |
| --- | --- |
| **Etiquetas de fila** | **Cuenta de cantón** |
| AMBATO | 3 |
| CUENCA | 4 |
| ESMERALDAS | 1 |
| LOJA | 2 |
| MACHALA | 3 |
| MANTA | 6 |
| SANTO DOMINGO | 3 |
| **Total general** | **22** |

## Distribución y selección de la muestra

Se definen como estratos los cantones en los que se lleva a cabo la encuesta: Manta, Machala, Santo Domingo, Esmeraldas, Ambato, Cuenca y Loja. Con la estratificación definida se realiza una distribución proporcional al tamaño (PPT) considerando un mínimo muestral en cada estrato. Con esto se garantiza la participación de cada uno de los cantones que son de interés para la encuesta. La estratificación se refiere a la subdivisión de una población determinada en subconjuntos con características propias. Esta acción se lleva a cabo como una etapa previa a la selección de la muestra, utilizando variables que aportan información para todas las unidades de la población.

Tabla 3. Estratos por dominio de estudio

|  |  |
| --- | --- |
| **Dominio** | **Cantidad de Estratos** |
| 2C | 50 |
| 3C | 41 |
| 4C | 42 |

*Fuente: Gestión de Diseño Muestral – DINEM*

Para la distribución de la muestra en cada uno de los estratos se procede así:

Donde:

* = 𝑀𝑢𝑒𝑠𝑡𝑟𝑎 𝑝𝑎𝑟𝑎 𝑒𝑙 𝑒𝑠𝑡𝑟𝑎𝑡𝑜 ℎ 𝑑𝑒𝑙 𝑑𝑜𝑚𝑖𝑛𝑖𝑜 𝑑𝑒 𝑑𝑖𝑠𝑒ñ𝑜 𝑖.
* = 𝑇𝑎𝑚𝑎ñ𝑜 𝑑𝑒 𝑚𝑢𝑒𝑠𝑡𝑟𝑎 𝑑𝑒𝑙 𝑑𝑜𝑚𝑖𝑛𝑖𝑜 𝑑𝑒 𝑑𝑖𝑠eño 𝑖.
* = 𝑇𝑎𝑚𝑎ñ𝑜 𝑑𝑒𝑙 𝑒𝑠𝑡𝑟𝑎𝑡𝑜 ℎ 𝑑𝑒𝑙 𝑑𝑜𝑚𝑖𝑛𝑖𝑜 𝑑𝑒 𝑑𝑖𝑠𝑒ñ𝑜 𝑖.
* = 𝑁ú𝑚𝑒𝑟𝑜 𝑑𝑒 𝑒𝑠𝑡𝑟𝑎𝑡𝑜𝑠 𝑒𝑛 𝑒𝑙 𝑑𝑜𝑚𝑖𝑛𝑖𝑜 𝑑𝑒 𝑑𝑖𝑠𝑒ñ𝑜 𝑖.

1. **Conclusiones**

* La cobertura evaluada a nivel de empresa tiene un 84,97%. Este porcentaje de empresas evidencia que la selección de la muestra tomada a partir del marco muestral utilizado responde de manera favorable con todas su variables identificadoras y geográficas que permiten la correcta ubicación e identificación de la empresa.
* Centrándonos en un análisis a nivel de producto, se tiene que el 37,95% de empresas dispone de al menos uno de los productos que son objeto de estudio, lo cual representa un bajo nivel de efectividad a nivel de empresa-producto. Es importante señalar que de este porcentaje el 32,94% de empresas dispone de estrictamente uno de los productos que son objeto de investigación. Este comportamiento del levantamiento limita la cantidad de productos que se obtendrán al final, pues, la mayoría de empresas no aportan más que una toma, lo que reduce la efectividad a nivel de producto.
* La falta de un marco muestral a nivel de producto queda evidenciada, si bien la cobertura a nivel de empresa es alta, la investigación requiere profundizar en la recolección información a nivel de producto, en tal sentido, se tiene que el 62,12% de los productos participan a lo sumo dos veces, lo que indica que más de la mitad de productos investigados aparecen en la cobertura máximo dos veces. Siguiendo el mismo razonamiento se puede evidenciar que el 84,15% de los productos tienen una participación menor a 4 tomas, solo el 15,85% de productos investigados por INPP disponen de al menos cinco tomas. Estos resultados ponen en evidencia que la cobertura a nivel de producto es deficiente, dichos porcentajes nos permiten evidenciar que la participación a nivel de producto es mínima. En este sentido, se evidencia que el ejercicio experimental realizado no responde a los objetivos esperados por la encuesta, por lo que, con la cobertura actual no es viable definir o implementar un diseño muestral probabilístico definitivo que nos permita garantizar las necesidades establecidas a nivel de producto.

|  |  |  |
| --- | --- | --- |
| **FIRMAS DE RESPONSABILIDAD** | | |
| **Nombre y Apellido** | **Dirección** | **Firmas** |
| Elaborado por:  OMAR LLAMBO | DINEM |  |
| Revisado y aprobado por:  CHRISTIAN GARCÉS | DINEM |  |

