

Informe de aplicabilidad de muestreo probabilístico para el INPP



DINEM

12· Marzo· 2025

Subtítulo 01

1. **Introducción**

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

Evaluar la factibilidad de la aplicación de un diseño muestral probabilístico para el sector “manufactura” de la operación estadística Índice Nacional de Precios Productor (INPP).

Tipo de operación estadística

La operación estadística es una encuesta dirigida a empresas.

Cobertura temática

Las empresas consideradas para el presente estudio son aquellas cuya sección económica es:

* C: Industrias manufactureras

Las demás secciones no serán objeto de estudio.

Se consideran los siguientes tamaños de empresas:

* Empresas de tamaño Pequeña empresa (2)
* Mediana empresa “A” (3)
* Mediana empresa “B” (4)
* Grande empresa (5).

Universo de estudio

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

Unidad de observación y de análisis

Empresas.

Cobertura geográfica

La cobertura geográfica está definida por todas las empresas que se encuentran en el Ecuador.

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

Tipos y etapas del diseño muestral

Muestreo probabilístico estratificado de elementos con selección aleatoria.

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.

Los dominios de estudio para el análisis muestral están definidos por el sector económico (Código CIIU4.1 - Sección - 1 dígito) agrupados por el tamaño de la empresa a nivel nacional, para este caso se considera únicamente la sección C-Industrias manufactureras. Se considera un grupo de inclusión forzosa, es decir, un subconjunto de empresas que formarán parte de la muestra con probabilidad uno (1), este grupo corresponde a aquellas empresas catalogadas como “Grande Empresa” (tamaño 5). De esta manera, los dominios de estudio definidos son: 2C, 3C, 4C y 5C (Tabla 1).

## 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 definido por las empresas que constan en el Registro de Empresas Estadístico de Empresas 2022 (REEM-2022). Para la construcción del marco muestral se consideraron las actividades económicas mediante el Código CIIU Rev 4.1 - 6 dígitos, dichos códigos consideran una correspondencia a nivel de los productos definidos en la canasta del INPP que fue elaborada por el equipo a cargo del Cambio de Base del Sistema de Índices de Precios al Productor (CAB-SIPP). Además, con los Código CIIU Rev 4 - 6 dígitos, se construye una estratificación dentro de cada dominio.

Dichas actividades económicas están agrupadas por una actividad principal:

* C: Industrias manufactureras

Una vez revisada la estructura del REEM-2022 que está conformado por 1’242.483 empresas. Para la construcción del marco muestral se seleccionaron aquellas que cumplan las siguientes condiciones:

* Empresas de tamaño Pequeña empresa (2)
* Mediana empresa “A” (3)
* Mediana empresa “B” (4)
* Grande empresa (5).
* Se excluyen las empresas que constan como “empresas no ubicadas” y en estado “inactivo” en el REEM-2022.

El marco de muestreo queda constituido de la siguiente forma:

Tabla 1. Marco de muestreo

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Código Sección** | **TAMAÑO DE EMPRESA** | | | | **Total** |
| **Pequeña** | **Mediana**  **“A”**  **(3C)** | **Mediana**  **“B”**  **(4C)** | **Grande** |
| **Empresa**  **(2C)** | **Empresa**  **(5C)** |
| **C** | 620 | 176 | 145 | 241 | 1182 |

*Fuente: REEM-2022*

## Variables de diseño

Para el análisis del tamaño muestral se ha considerado como variable de diseño “Ventas Totales 2022” que es un dato anual y ha sido obtenido del REEM-2022.

## 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 “Ventas Totales”, dicha variable está disponible en el REEM-2022. La fórmula para dicho cálculo se presenta a continuación:

* = Tamaño de la muestra por dominio.
* = 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.

Para definir la tasa de no respuesta (TNR) en cada uno de los dominios, se usó como referencia las TNR´s resultantes de la Encuesta Estructural Empresarial (ENESEM-2021). Para los dominios en los que no se especifica una TNR se realizó un promedio acorde a los tamaños de las empresas y actividad principal.

La muestra resultante obtenida para cada dominio de estudio se presenta en la Tabla 2, obteniéndose un tamaño muestral de 672 empresas.

Tabla 2. Tamaño muestral

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Código Sección** | **Tamaño muestral** | | | | **Total** |
| **Pequeña** | **Mediana**  **“A”**  **(3C)** | **Mediana**  **“B”**  **(4C)** | **Grande** |
| **Empresa**  **(2C)** | **Empresa**  **(5C)** |
| **C** | 186 | 133 | 112 | 241 | 672 |

*Fuente: Gestión de Diseño Muestral - DINEM*

## Distribución y selección de la muestra

Con la estratificación realizada dentro de cada uno de los dominios a nivel de Código CIIU Rev 4.1 - 6 dígitos, se realiza una distribución proporcional al tamaño (PPT) considerando un mínimo muestral en cada estrato. Con esto se busca obtener la participación de cada uno de los Códigos CIIU Rev 4 – 6 dígitos en la muestra. 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. El objetivo de la estratificación nos permite mejorar la eficiencia del diseño maestral mediante la disminución de la varianza de los estimadores.

En el presente ejercicio se tiene un total de 50 estratos para el dominio “2C”, 41 estratos para el dominio “3C” y 42 estratos para el dominio “4C” (Tabla 3), cabe mencionar que para el dominio “5C” no se consideran estratos ya que es un dominio de inclusión forzosa.

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. **Control de cobertura muestral y de campo**

La cobertura se refiere a la información que se obtuvo en campo y que ha sido proporcionada por los informantes de las empresas.

## Seguimiento de cobertura a nivel de empresa

Considerando como objeto de análisis a las empresas, la información levantada puede estar clasificada de dos formas:

Tabla 4. Estructura de condición de levantamiento de información por empresa

|  |  |
| --- | --- |
| **Efectividad de la empresa** | **Resultado de entrevista** |
| **Efectiva** | Investigada |
| Ubicada |
| **No Efectiva** | No ubicada |
| Rechazo |
| Liquidada |
| Embargada |
| Absorbida |
| Fusionada |
| Inactiva |

*Fuente: Gestión de Diseño Muestral - DINEM*

Un empresa “Investigada” la catalogamos como aquella empresa que dispone de uno de los productos que se investigan y por tanto, brindó la información que requiere la encuesta; mientras que una empresa “Ubicada” la catalogamos como aquella empresa que se encontró pero no dispone de ninguno de los productos que se investiga, por lo que, no se recoge información alguna. El resumen de estas dos categorías se muestra en la Tabla 5.

Tabla 5. Efectividad a nivel de empresas

|  |  |
| --- | --- |
| **Efectivas** | **Empresas** |
| Investigada | 255 |
| Ubicada | 316 |
| **Total** | **571** |

*Fuente: Gestión de Diseño Muestral – DINEM*

Cabe mencionar que la cobertura evalúa la efectividad por cada uno de los dominios de estudio comparando los resultados obtenidos en campo con el total de empresas obtenidas en la muestra. En la Tabla 6 se puede observar que existen 571 empresas efectivas que representan un 84,97% de la cobertura.

Tabla 6. Seguimiento de cobertura por dominio de estudio.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dominio** | **Muestra** | **Efectivas** | **Cobertura** |
| 2C | 186 | 153 | 82,26% |
| 3C | 133 | 112 | 84,21% |
| 4C | 112 | 102 | 91,07% |
| 5C | 241 | 204 | 84,65% |
| **Total** | 672 | 571 | 84,97% |

*Fuente: Gestión de Diseño Muestral – DINEM*

## Seguimiento de cobertura a nivel de producto

Como se ha mencionado en los apartados anteriores, el INPP tiene como objetivo levantar información a nivel de productos, en ese sentido, a continuación evaluamos la cobertura considerando el producto producido por cada empresa que ha sido investigada.

En la Tabla 5 del apartado anterior podemos apreciar que existen 255 empresas investigadas, es decir, 255 empresas efectivas a nivel de producto. La efectividad a nivel de producto desglosada por dominio de estudio lo podemos observar en la Tabla 7, notemos que la cobertura alcanza un 37,95% de la muestra total, siendo el dominio 2C el más deficiente con una cobertura del 23,66%.

Tabla 7. Seguimiento de cobertura por dominio a nivel de producto

|  |  |  |  |
| --- | --- | --- | --- |
| **Dominio** | **Muestra** | **Efectivas** | **Cobertura** |
| 2C | 186 | 44 | 23,66% |
| 3C | 133 | 34 | 25,56% |
| 4C | 112 | 41 | 36,61% |
| 5C | 241 | 136 | 56,43% |
| **Total** | 672 | 255 | 37,95% |

*Gestión de Diseño Muestral - DINEM*

Por otro lado, de las 255 empresas efectivas se tiene que el 32,94% (Tabla 8) disponen estrictamente de uno de los productos investigados. Es importante revisar este dato ya que está directamente relacionado con la participación a nivel de producto, esto lo evidenciamos posteriormente en la Tabla 9.

Tabla 8. Seguimiento de cobertura cantidad de productos por empresa

|  |  |  |
| --- | --- | --- |
| **Cantidad de Productos** | **Empresas** | **Porcentaje** |
| 1 | 84 | 32,94% |
| 2 | 70 | 27,45% |
| 3 | 44 | 17,25% |
| 4 | 25 | 9,80% |
| 5 | 32 | 12,55% |
| **Total** | 255 | 100,00% |

*Gestión de Diseño Muestral - DINEM*

Con el objetivo de evaluar la participación de cada producto en la cobertura, se presenta un resumen que muestra la frecuencia de tomas obtenidas, por ejemplo: existen 88 productos que cuentan con una sola toma, 53 productos que cuentan con dos tomas, etc.

Tabla 9. Participación por producto en la cobertura.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tomas** | **Total de Productos** | **% Productos** | **% Acumulado** |
| 13 | 1 | 0,44% | 0,44% |
| 12 | 1 | 0,44% | 0,88% |
| 11 | 1 | 0,44% | 1,32% |
| 10 | 3 | 1,32% | 2,64% |
| 9 | 5 | 2,20% | 4,84% |
| 8 | 1 | 0,44% | 5,28% |
| 7 | 4 | 1,76% | 7,04% |
| 6 | 9 | 3,96% | 11,00% |
| 5 | 11 | 4,85% | 15,85% |
| 4 | 16 | 7,05% | 22,90% |
| 3 | 34 | 14,98% | 37,88% |
| 2 | 53 | 23,35% | 61,23% |
| 1 | 88 | 38,77% | 100,00% |
| **Total** | **227** |  |  |

*Fuente: Gestión de Diseño Muestral – DINEM*

Se puede evidenciar en la Tabla 8 que de los 277 productos levantados, 88 de ellos aparecen una sola vez, es decir, el 38,77% de productos tienen una única participación en la información levantada y también podemos evidenciar que el 23,35% (53 productos) tienen una doble participación. Si solo consideramos los productos con una y dos participaciones se alcanzan un 62,12%, es decir, más de la mitad de productos están limitadas a dos tomas, lo que claramente indica que la recolección a nivel de producto es deficiente. Acorde a lo que actualmente levanta el IPP, lo deseable sería que al menos cada producto aparezca 5 veces en la cobertura, sin embargo, solo el 15,85% (porcentaje acumulado) de productos cumplen este particular.

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 |  |

