### Metodología del marco muestral y el diseño de la muestra

***Nota:*** *Esta sección se deberá incluir solo en el caso de operaciones estadísticas que requirieron la aplicación de un proceso de muestreo probabilístico y/o no probabilístico. De lo contrario, si la operación estadística se trata de cálculo de índices, reemplace por la sección “Metodología de construcción de canastas y diseño de la muestra”; o si se trata de estadísticas basadas en registros administrativos, pase a la sección 3 “Metodología de recolección/recopilación”.*



## Tipos y etapas del diseño muestral

Muestreo probabilístico 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 los sectores económicos (Código CIIU4 - Sección - 1 dígito) agrupados por el tamaño de la empresa a nivel nacional. Dentro de cada uno de los dominios de estudio 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).

## 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 - 6 dígitos, dichos códigos consideran una correspondencia a nivel de los productos definidos en la canasta del índice del Precio al Productor (IPP) que fue elaborada por el equipo a cargo (Equipo-Cambio de año base del Índice 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” 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”** | **Mediana “B”** | **Grande** |
| **empresa** | **empresa** |
| **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 las “Ventas Totales 2022” que es un dato anual y ha sido obtenido del REEM-202. Cabe señalar que se usa esta variable a falta de información con respecto a las ventas totales o precios de cada uno de los productos que recoge el IPP.

## 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:

* ni = Tamaño de la muestra por dominio.
* Ni = Tamaño del dominio i.
* Si = Cuasivarianza del dominio i.
* e = Error relativo máximo admisible.
* z = Coeficiente que representa el nivel de seguridad o confianza.
* vi = ventas totales en el dominio i.
* tnri = 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.

Tabla 2. Tamaño muestral

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Código Sección** | **Tamaño muestral** | | | | **Total** |
| **Pequeña** | **Mediana “A”** | **Mediana “B”** | **Grande** |
| **empresa** | **empresa** |
| **C** | 186 | 133 | 112 | 241 | 672 |

*Fuente: Elaboración propia*

## Selección de la muestra

Con la estratificación realizada dentro de cada uno de los dominios a nivel de Código CIIU Rev 4 - 6 dígitos, se realiza una distribución proporcional al tamaño (PPT) considerando un mínimo muestral en cada estrato. Con esto se busca obtener la representatividad 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. La distribución por estratos está construida de la siguiente manera:

Donde:

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

## Control de cobertura de campo y muestral

<<Descripción de cómo se realizó el control de cobertura de campo y muestral, siga las especificaciones de las págs. 29 - 30 del *“Manual para documentar la metodología de una operación estadística”*>>

## Factores de expansión

<<Descripción de cómo se construyeron y/o ajustaron los factores de expansión, siga las especificaciones de la pág. 30 del *“Manual para documentar la metodología de una operación estadística”*>>

## Evaluación de diseño muestral

<<Descripción de los niveles de confianza y precisión de las estimaciones, siga las especificaciones de la pág. 30 del *“Manual para documentar la metodología de una operación estadística”*>>

## Rotación de la muestra

<<Descripción del procedimiento y los métodos utilizados para la rotación de la muestra, siga las especificaciones de la pág. 30 del *“Manual para documentar la metodología de una operación estadística”*>>