

# ESTUDIO SOBRE EL CONSUMO DE LEÑA INDUSTRIAL AÑOS 2010 a 2012



## ÍNDICE

- Universo de Estudio
- Definición de empresas Muy Grandes
- Marco Muestral
- Unidad de análisis
- Diseño muestral (definición de los ponderadores primarios)
  - Marco muestral
  - Diseño
  - Relevamiento y no respuesta
  - Muestra efectiva para el estudio de leña años 2010, 2011 y 2012
- Estadísticos descriptivos muestrales del consumo de leña
- Definición de los ponderadores finales para la expansión de los resultados
- No respuesta – Imputación de los datos faltantes del consumo de leña 2010 y 2012.
- Resultados
  - Participación de los subsectores en el consumo de leña total de la industria
- Precisión de las estimaciones

## Universo de Estudio

El universo de estudio lo componen el conjunto de todas las industrias del país (sector Formal), con personal ocupado promedio en el año (PO) mayor o igual a 10 (según BPS) que se encuentran en diferentes clases de actividad que se caracterizan por la necesidad de consumir leña (o un sustituto energético) en su proceso productivo. Es así que se realizó una clasificación del universo dependiendo de las características de producción de las industrias que lo forman, y se definieron subsectores, agrupando empresas según su clase de actividad económica principal siguiendo la clasificación Industrial CIU Rev. 4. En el siguiente cuadro se presentan los diferentes subsectores industriales definidos, así como las clases de actividad que los conforman:

Subsector	CIU Rev. 4					
	1011	1012	1013			
Frigoríficos	1011	1012	1013			
Lácteos	1050					
Otras Industrias alimenticias	1020	1030	1040	1072	1079	1080
Molinos	1061					
Bebidas y Tabaco	1103	1104	1200			
Textil y Cueros	1311	1312	1391	1392	1511	1512
Madera, papel y Celulosa	1610	1621	1701	1702	1709	
Química, Caucho, Plástico, Cemento y Cerámica	2012	2211	2391	2392	2394	

El universo de estudio entonces fue reducido al no incluirse las actividades económicas a cuatro dígitos según la CIU Rev. 4, que por diferentes motivos no se consideraron relevantes, es decir, clases que por su proceso de producción no consumen leña o si lo hacen, su consumo es muy menor y el peso de dichas clases en el consumo total de la Industria es muy reducido.

Cabe destacar, que cuando el consumo anual de leña de una industria es mayor o igual a 10.000 mil toneladas la misma es incorporada de forma forzosa al estudio, sin importar si la clase de actividad es alguna de las excluidas o si su PO es menor a 10.

En este estudio se dejaron de lado las industrias de la clase 1071, correspondiente principalmente a panaderías, dado que estas tienen un comportamiento que no se asemeja en su estructura a las demás industrias. Estas son consideradas por separado.

## Definición de empresas Muy Grandes (MG)

A partir de un estudio preliminar de las encuestas anuales realizadas por la DNE en períodos anteriores (existen registros desde 1985, pero se utilizaron desde el año 2006 al presente), se identifican empresas que pueden denominarse Muy Grandes (MG) debido a su consumo anual de leña.

Estas empresas se consideran por separado dado que su comportamiento es atípico y puede ser independiente al del resto de las empresas de la rama económica a considerar, es decir, se representan a sí

mismas (Forzosas). Es así que se definen a quienes consumen anualmente más de 10.000 toneladas de leña como empresas MG.

Por otra parte, las empresas incluidas en los tramos de Personal Ocupado (PO) “7” u “8”, correspondiente a un PO de 200 a 499 y mayor a 500 respectivamente, también son consideradas de inclusión forzosa, aunque las mismas pueden no consumir leña.

## Marco Muestral

El *marco muestral* está compuesto por las empresas provenientes del Registro Permanente de Actividad Económica (RPAE) para el año 2011 que realiza de forma anual el Instituto Nacional de Estadística (INE).

El RPAE está conformado por las empresas incluidas en los registros administrativos de la Dirección Nacional de Impositiva (DGI) y del Banco de Previsión Social (BPS). A su vez, la información contenida en el RPAE ha sido depurada por la DNE en base a encuestas específicas como ser, la Encuesta de Leña o la encuesta Industrial del año 2011 y en base a los registros provenientes de UTE.

## Unidad de Análisis

Hasta ahora la unidad tanto de análisis como de muestreo en todos los estudios realizados por la DNE ha sido el establecimiento. Esto se debe principalmente porque el marco muestral utilizado para los distintos estudios provenía de los registros de UTE los cuales se encuentran a nivel de establecimiento (más precisamente, a nivel de los contadores).

Ahora, *la unidad de análisis es la empresa*, la cual puede verse como un conglomerado (*cluster*) de establecimientos y de actividades económicas.

La decisión de utilizar la empresa como unidad de análisis es basada en que la información utilizada tanto para la definición del Universo de estudio como para la expansión de los resultados se encuentra definida a nivel de empresa y no a nivel de establecimientos. A su vez, las empresas con más de un local representan una proporción del orden del 5% y generalmente son MG, es decir, son censadas.

## Diseño muestral (definición de los ponderadores primarios)

Toda la información que posee la DNE proviene de relevamientos en donde la selección de la muestra fue aleatoria, es decir, bajo algún mecanismo de selección que le anexo probabilidades de inclusión a los establecimientos incluidos en el marco muestral. Dichas probabilidades, generalmente, tenían una alta correlación con el consumo eléctrico del establecimiento y no tenía en cuenta el consumo de leña, dando como resultado que las estimaciones obtenidas tuvieran pobres niveles de precisión, ya que las probabilidades de inclusión de los establecimientos (definición del diseño muestral) no tenían correlación alguna con el consumo de leña.

A su vez, en el proceso o etapa de estimación, tampoco se utilizaba ninguna información auxiliar de forma de poder conformar un sistema de ponderadores que tuvieran en mayor o menor medida información sobre el consumo de leña. Es decir, variables auxiliares correlacionadas con la variable de interés que produjeran estimadores con menor varianza.

**La muestra utilizada para este estudio proviene de la encuesta industrial del año 2011** que se llevo a cabo por parte de la Dirección Nacional de Energía en los años 2012 y 2013. Dicha muestra fue depurada por parte de la DNE en base a información de estudios anteriores y a relevamientos de información específicos.

Por otra parte, se incluyeron empresas que no pertenecían a la muestra de la encuesta de industria y los cuales provienen de relevamientos anteriores (la encuesta de Leña propiamente dicha).

A continuación se hace una breve reseña del diseño muestral de la encuesta Industrial del año 2011.

## Marco muestral

El marco muestral se encuentra conformado por la base de consumo de electricidad de UTE (medida en kWh) de los establecimientos con actividad económica principal Industrial según el código CNAE para el año 2011.

El marco muestral, no tiene un enlace perfecto con la población objeto de estudio, es más, existen establecimientos que su actividad económica no es industrial y a su vez, no incluye a otros establecimientos que si pueden tenerla pero se encuentran clasificados en otras actividad económicas. Dichos establecimientos no se encuentran incluidos en el marco muestral, y por ende, su probabilidad de ser seleccionados en la muestra es nula. Lo anterior, da como resultado que las estimaciones obtenidas sean sesgadas y hace vital el uso de información auxiliar en el proceso de estimación para poder minimizar dicho sesgo.

## Diseño

El diseño es aleatorio, estratificado con asignación óptima, utilizando como variable auxiliar el consumo eléctrico en el período de referencia según los registros de UTE.

Los estratos del diseño reconocen dos dimensiones:

- Una medida de tamaño (en términos de consumo eléctrico)
- La actividad económica principal del establecimiento (definiéndose 12 subsectores).

Se definieron 4 estratos de tamaño, en donde los límites de los mismos son los siguientes:

Estrato	Consumo (kWh)
Pequeñas	menos de 30.000
Medianas	entre 30.000 y 200.000
Grandes	entre 200.000 y 300.000
Muy grandes	más de 300.000

El estrato de tamaño 4 es de inclusión forzosa a priori, es decir, es censado (aunque la asignación óptima no lo requiera). Teniendo en cuenta la interacción de las dos variables se conformaron 48 estratos, en donde la asignación del tamaño de muestra por estrato se realizó de forma óptima utilizando como variable auxiliar el consumo eléctrico.

El tamaño de muestra teórico se fijó en 608 establecimientos para estimar el total del consumo eléctrico de la industria con un error del 1.4% para un nivel de confianza del 95 %. Finalmente, para cada uno de los 12 sectores definidos los errores se encuentran en el orden del 5% para el mismo nivel de confianza.

## Relevamiento y NO respuesta

La tasa de no respuesta se situó en el orden del 12%, configurando así un tamaño de muestra efectivo de 540 establecimientos. La no respuesta obtenida se debió en parte a: Problemas de cobertura en el marco muestral. Establecimientos que no pertenecen al universo de estudio; Información geográfica insuficiente para la localización del establecimiento (omisos); Rechazos de establecimientos que no quisieron proporcionar la información. En el siguiente cuadro se presenta la distribución de los establecimientos en la muestra por estrato de tamaño según subsector de actividad.

Sub sector	Estrato tamaño (consumo eléctrico)				
	Total	Muy grandes	Grandes	Medianas	Pequeñas
<b>Total</b>	<b>540</b>	<b>124</b>	<b>301</b>	<b>71</b>	<b>44</b>
Frigoríficos	55	23	31	1	0
Lácteos	29	12	16	1	0
Molinos	37	16	20	1	0
Otras Alimenticias	116	14	53	31	18
Bebidas y Tabaco	27	10	12	3	2
Textil	25	5	17	2	1
Cuero	13	4	6	3	0
Madera	21	8	7	3	3
Papel	21	3	14	2	2
Química, Caucho y Plástico	115	16	87	9	3
Cemento	7	6	1	0	0
Otras Manufactureras	74	7	37	15	15

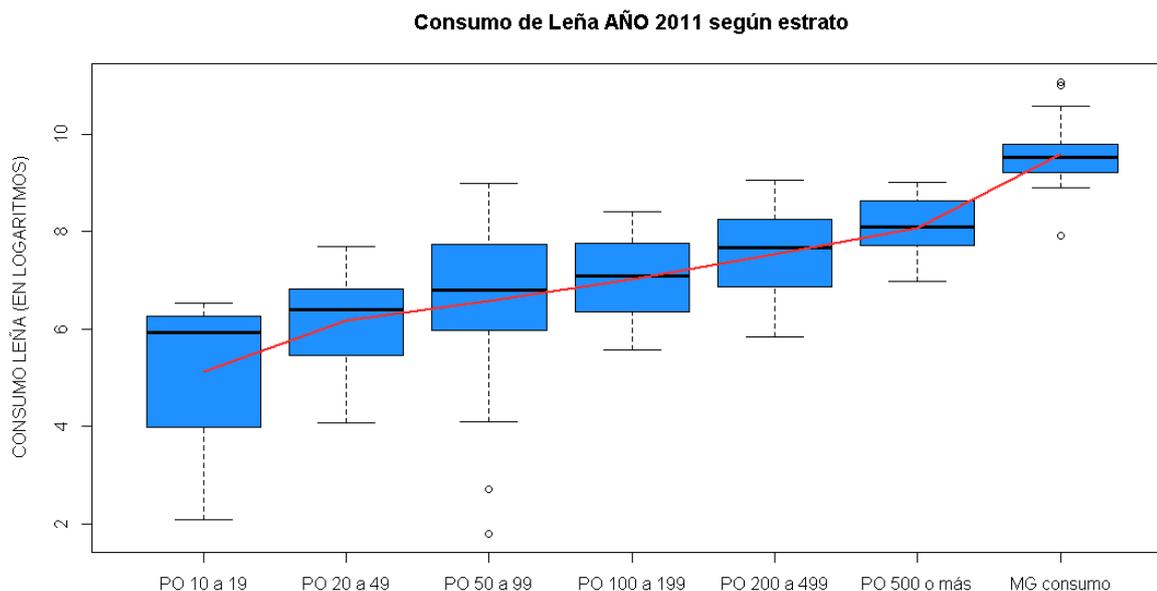
## Muestra efectiva para el estudio de leña años 2010, 2011 y 2012

Para el estudio de leña se definieron seis estratos a posteriori.

- Estrato 1.** Empresas con PO de 10 a 19.
- Estrato 2.** Empresas con PO de 20 a 49.
- Estrato 3.** Empresas con PO de 50 a 99.
- Estrato 4.** Empresas con PO de 100 a 199.
- Estrato 5.** Empresas con PO más de 200.
- Estrato 6.** Empresas con consumo de leña anual mayor o igual a 10.000 toneladas en algún año independientemente del personal ocupado que posea.

La definición de los estratos en términos del personal ocupado se debe a que existe una relación positiva con el consumo de leña, es decir, a medida que la cantidad de empleados aumenta en promedio (o generalmente) el consumo de leña también aumenta.

En la siguiente figura se presentan los diagramas de caja del consumo de leña por empresa (en logaritmos) según estrato.



Los estratos 5 y 6, conforman las empresas denominadas anteriormente como MG y los mismos son **estratos forzosos**, es decir, son censados y los ponderadores de las empresas incluidas en dichos estratos es uno. Cabe destacar que esto implica que las empresas que no consumen leña son igualmente censadas y su monitoreo a lo largo del tiempo es fundamental por su condición de “únicas” para poder mantener estimaciones estables y poder captar posibles cambios grandes en el futuro si las mismas comienzan a consumir leña (ver diagrama de caja anterior).

Por otra parte, los estratos 1, 2, 3 y 4 conforman las empresas denominadas “no MG” y se **asume** un diseño aleatorio simple sin reposición (SI) de forma independiente en cada uno de los estratos del diseño.

Entonces, los ponderadores primarios (o provenientes del diseño muestral) se definen como:

$$a_k = \begin{cases} \frac{N_h}{n_h}, & \text{si la empresa } k \text{ es NO MG} \\ 1, & \text{si la empresa } k \text{ es MG} \end{cases}$$

donde

$a_k$  es el ponderador primario de la  $k$ -ésima empresa

$N_h$  es la cantidad de empresas del estrato  $h$  según el RPAE

$n_h$  es el tamaño de muestra efectivo en el estrato  $h$ .

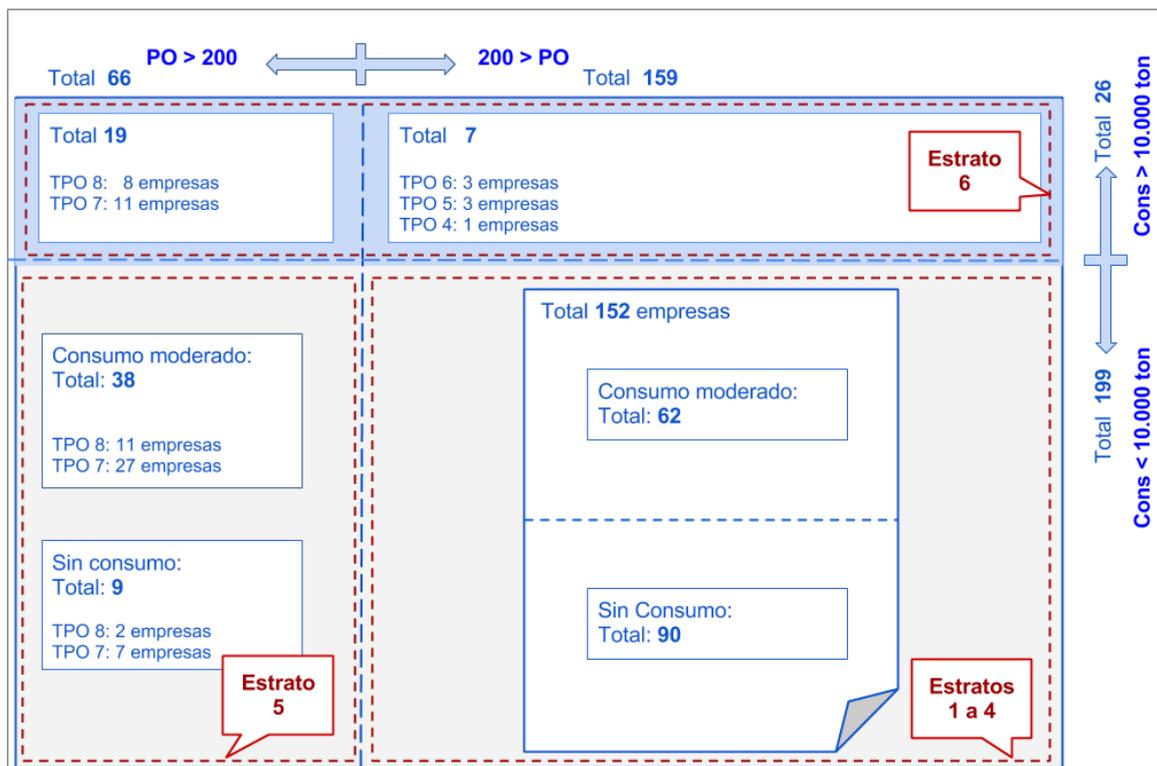
Finalmente, la muestra propiamente dicha para este estudio se conforma por:

- 1- Teniendo en cuenta la agrupación de los establecimientos en empresas (unidad de análisis).
- 2- Los subsectores definidos para el estudio, el personal ocupado de la empresa (proveniente del RPAE).
- 3- Otras empresas para las que se cuenta con información sobre su consumo de leña proveniente de la encuesta de Leña de años anteriores.

En el siguiente cuadro se presenta la distribución de las empresas en la muestra por tramo de personal ocupado y sector de actividad

Sector	Sector	Estratos (estudio Leña)							
		NO MG (ALEATORIAS)				MG (FORZOSAS)			
		Tramo de Personal ocupado promedio							Consumo más de 10.000ton
		10 a 19	20 a 49	50 a 99	100 a 199	200 a 499	500 o más		
<b>Total</b>	225	16	57	49	30	45	21	7	
<b>Bebidas y tabaco</b>	18	2	4	4	2	1	2	3	
<b>Frigoríficos</b>	57	0	11	10	7	15	12	2	
<b>Lácteos</b>	24	0	7	6	4	5	1	1	
<b>Madera, papel y celulosa</b>	22	1	8	4	3	6	0	0	
<b>Molinos</b>	22	2	5	8	2	4	1	0	
<b>Otras Industrias Alim.</b>	41	9	12	7	6	5	2	0	
<b>Química, caucho, plástico, cemento y cerámica</b>	12	0	5	3	0	3	0	1	
<b>Textil y cueros</b>	29	2	5	7	6	6	3	0	

En el siguiente diagrama se observa esquemáticamente como se distribuyen las empresas según los estratos definidos por PO y Consumo de leña anual.

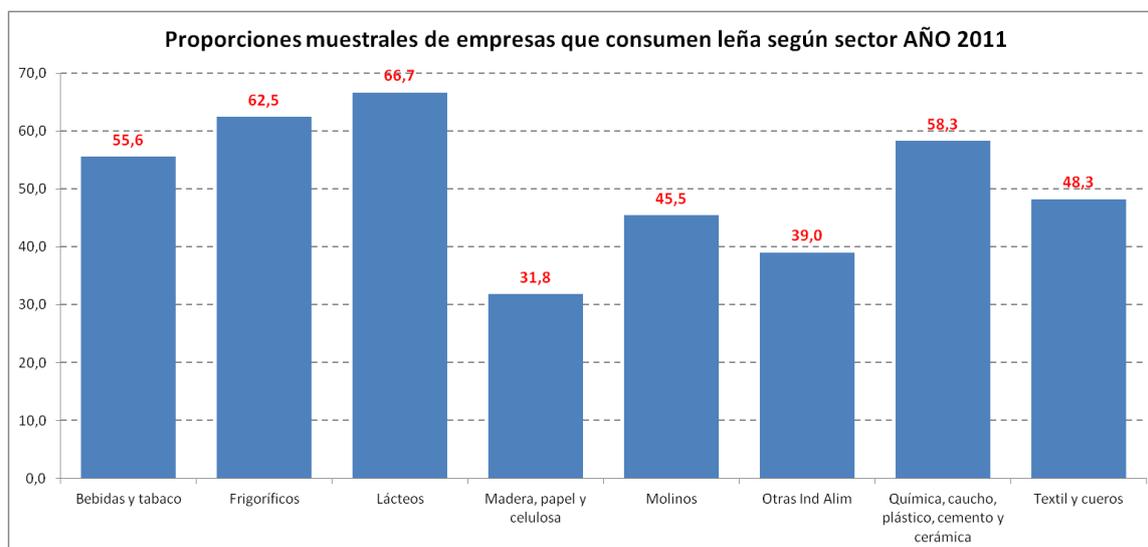


## Estadísticos descriptivos MUESTRALES del consumo de leña

El objetivo es describir brevemente la información auxiliar utilizada para producir estimaciones más eficientes y como se encuentra relacionada con la variable de estudio, es decir, con el consumo de leña.

La proporción de empresas que consumen leña depende del subsector de actividad económica en el cual se encuentren, es decir, depende del proceso productivo, en donde la *moda* se encuentra en el sector Lácteos.

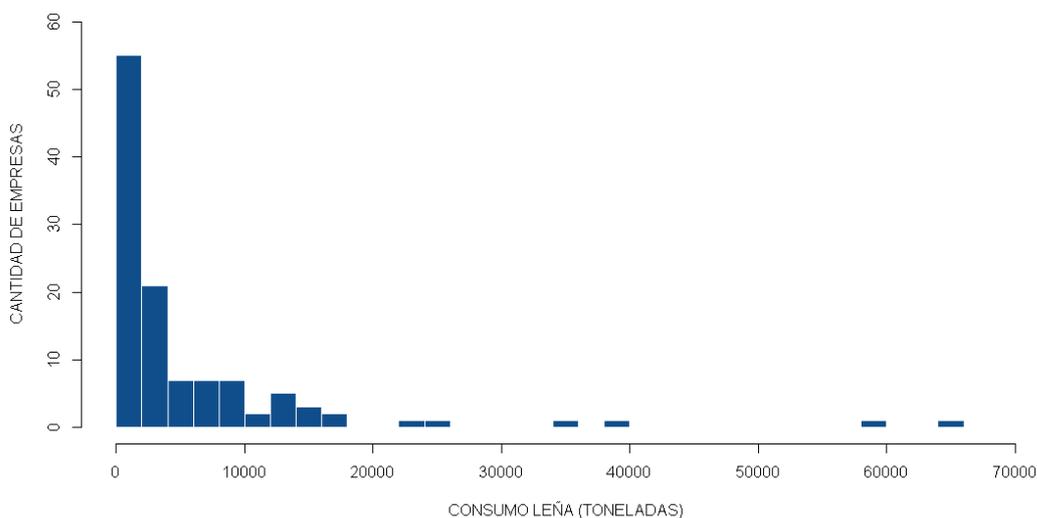
En la siguiente figura se presentan las proporciones muestrales de empresas que consumen leña según sector para el año 2011.



El consumo de leña presenta claramente una asimetría positiva, en donde la mayoría de las empresas que consumen leña, tiene un consumo pequeño o intermedio y unas pocas empresas presentan un consumo grande o muy grande.

En la siguiente figura se presenta la distribución de las empresas que consumen leña en el año 2011 y los deciles de consumo para el mismo año

### DISTRIBUCIÓN DEL CONSUMO LEÑA AÑO 2011

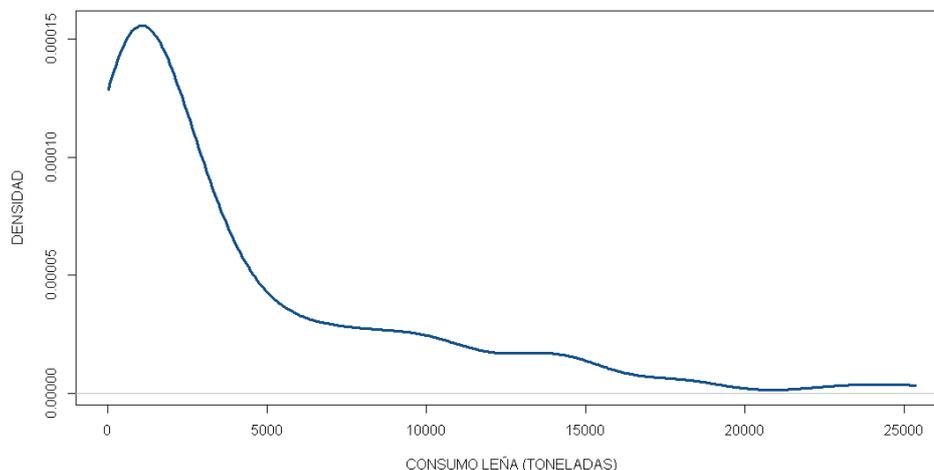


Deciles de consumo AÑO 2011								
10%	20%	30%	40%	50%	60%	70%	80%	90%
276	559	810	1215	2169	2867	4896	8776	13776

La mediana de consumo de leña se encuentra en 2169 toneladas, es decir, el 50 % de las empresas que consumieron leña en el año 2011 tuvieron un consumo menor igual a 2169 toneladas. La asimetría es aún más clara si tenemos en cuenta que el consumo promedio se encuentra en 5750 toneladas y cuanto se aleja de la mediana.

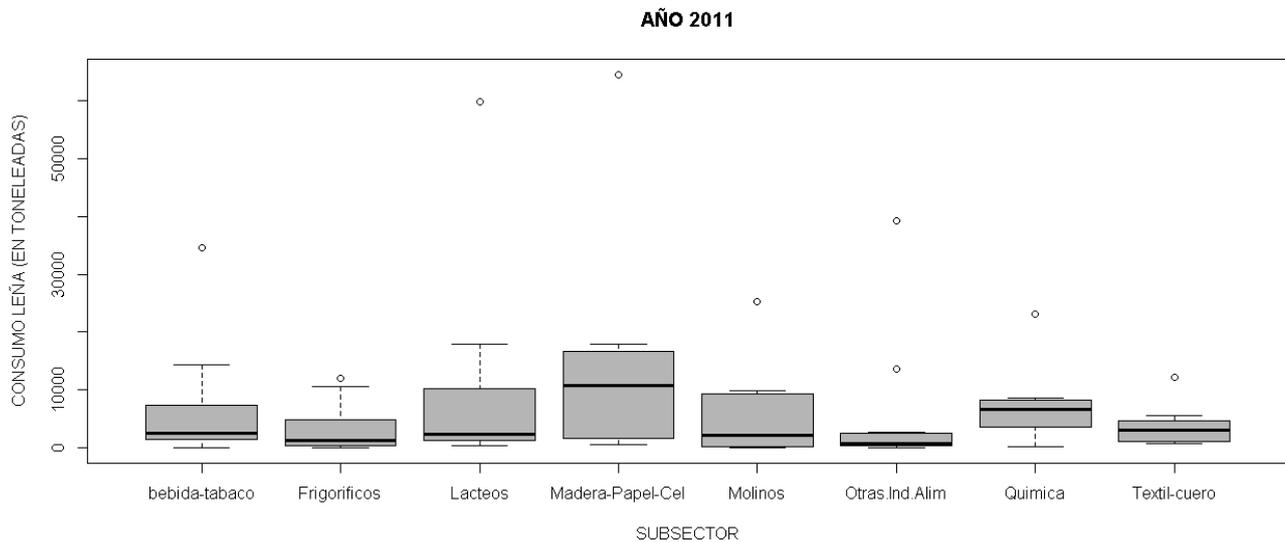
En la siguiente figura se presenta la densidad estimada del consumo de leña para el año 2011 (sin tener en cuenta a FANAPEL y a CONAPROLE) utilizando un *kernell gaussiano*

### DENSIDAD ESTIMADA DEL CONSUMO DE LEÑA AÑO 2011



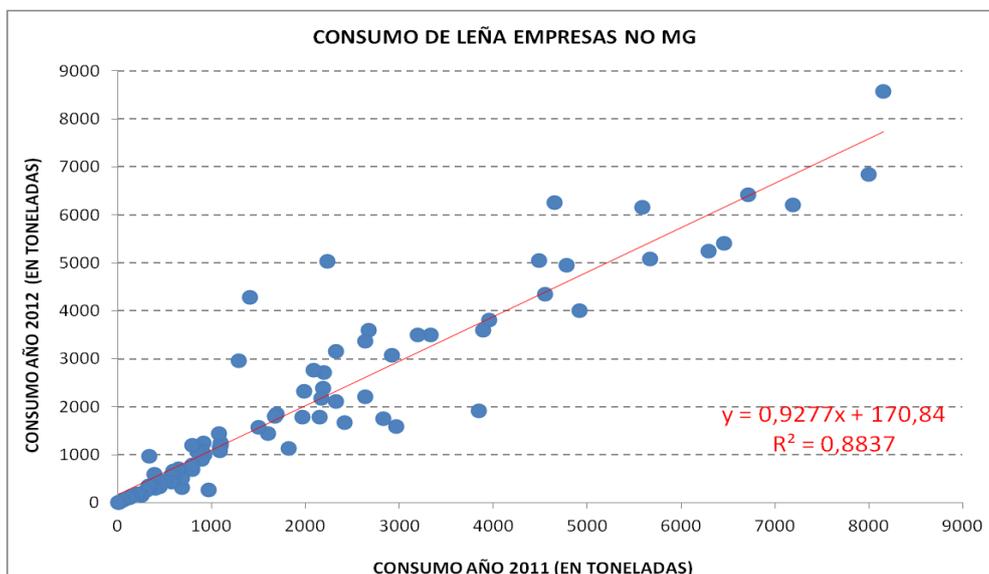
Por otra parte, los consumos de leña no solo están correlacionados con el tamaño de las empresas en términos de su personal ocupado. También se encuentran relacionados con el subsector de actividad.

En la siguiente figura se presenta los diagramas de caja del consumo de leña según subsector para el año 2011.



Como se observa en la figura anterior, la mayoría de los subsectores presentan asimetrías positivas en los consumos de leña, en donde unas pocas empresas dentro del subsector tienen consumos grandes o muy grandes y la mayoría consumos pequeños o medianos. El único sector que presenta una asimetría negativa es **Química, caucho, plástico, cemento y cerámica**, donde la mayoría de las empresas tienen consumos grandes o muy grandes (siempre respecto al subsector) y unas pocas empresas consumos pequeños o medianos.

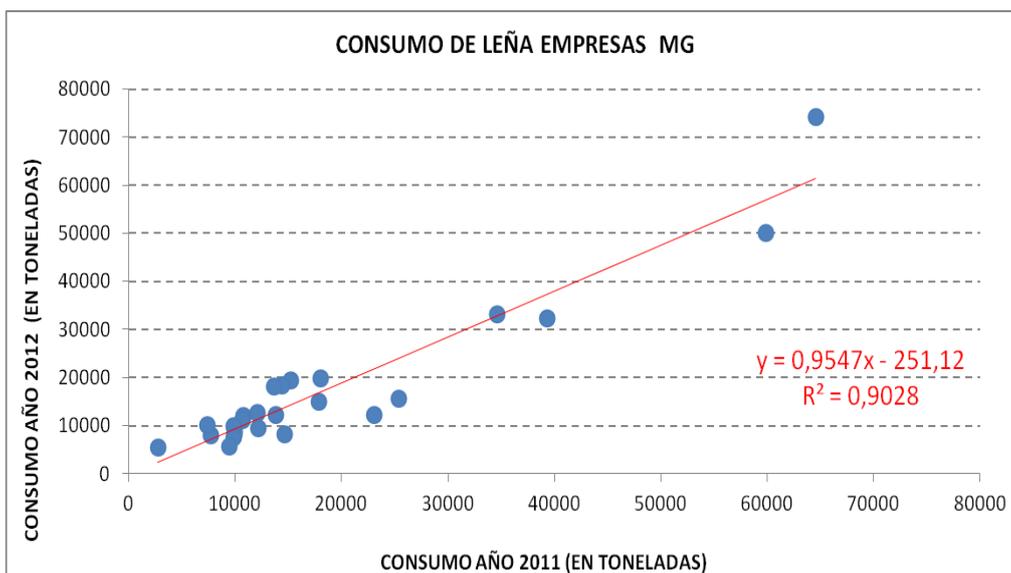
Como es de esperarse, el consumo de un año a otro se encuentra altamente correlacionado. Si bien las empresas analizadas presentan grandes fluctuaciones (debido a cambios de tecnologías, sustitución de fuentes energéticas, u otros) a lo largo del tiempo el consumo en un año depende mayoritariamente del consumo del año anterior.



Teniendo en cuenta la recta de regresión estimada para las empresas “NO MG”, podemos apreciar que por cada tonelada extra consumida de leña por las empresas en el año 2011 en promedio (o generalmente) se consumió 0.93 toneladas aproximadamente en el año 2012, es decir, una reducción del 7% entre 2011 y 2012.

Por otra parte, teniendo en cuenta el coeficiente de determinación de la regresión ( $R^2$ ) se concluye que el 88% de la **variación** del consumo de leña en el 2012 es explicado por la **variación** en el consumo de leña de las empresas en el 2011, dejando un 12% de la variación que no es explicada por el consumo del año anterior y se debe a múltiples factores que no fueron tenidos en cuenta en el modelo de regresión planteado, como ser factores económicos.

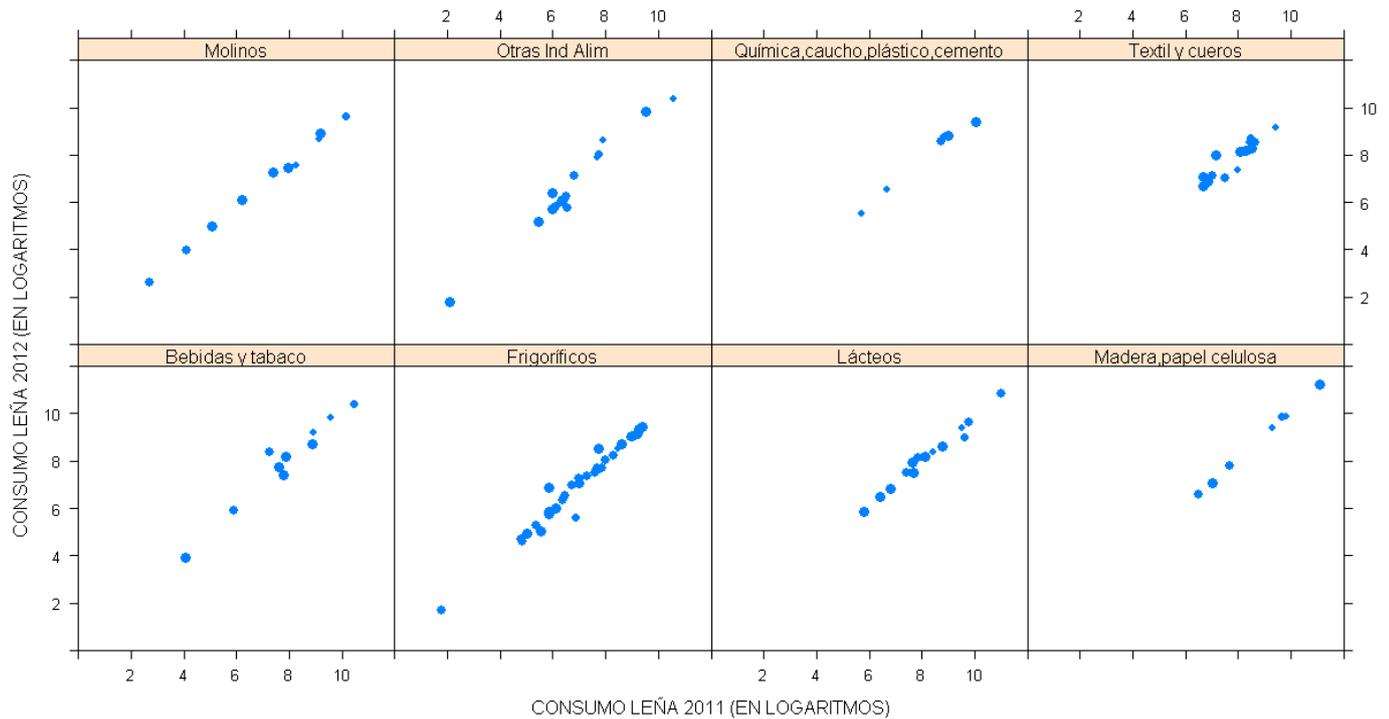
La misma situación se presenta para las empresas MG, en donde por cada tonelada extra consumida en el año 2011 en promedio se consumió 0.95 toneladas aproximadamente en el 2012.



El 90% de la variación en el consumo de leña en el año 2012 para las empresas MG es explicado por la variación en el consumo del año anterior.

**Las rectas de regresión estimadas anteriormente, los subsectores definidos y el tamaño de la empresa (en términos de su personal ocupado) dan una clara idea del uso de dicha información para poder imputar los datos faltantes y para la definición de los ponderadores finales para las expansiones de los resultados.**

En la siguiente figura se presentan los consumos para los años 2011 y 2012 (en logaritmos) según subsector. El tamaño de los puntos en los diagramas de dispersión es proporcional al tramo de personal ocupado de la empresa.



## Definición de los ponderadores FINALES para la expansión de los resultados

El uso de información auxiliar en toda estimación por muestreo es vital para obtener estimadores más eficientes. Para el caso de la estimación del consumo anual Industrial de leña, se utilizaron técnicas de calibración (estimadores calibrados), las cuales generan un nuevo sistema de ponderadores en base a la información auxiliar disponible.

La calibración, se trata de un procedimiento para calcular sistemas de ponderadores que estimen de forma más eficiente los parámetros poblacionales de interés. El objetivo es modificar los ponderadores primarios en base a la información auxiliar disponible.

Los ponderadores calibrados estiman sin error los totales de las variables auxiliares utilizadas para su cálculo, haciendo coincidir las estimaciones con los totales poblacionales de las variables auxiliares, lo que brinda coherencia a las estimaciones.

Si las variables auxiliares utilizadas son cualitativas (tramo de personal ocupado, si tiene o no caldera, actividad económica, etc.) a la calibración se le denomina post-estratificación.

La post-estratificación puede aplicarse de dos formas dependiendo de la información auxiliar disponible y del tamaño de muestra efectivo.

El primer caso es cuando se dispone de información a nivel de las celdas de una tabla de contingencia, por ejemplo, si se conoce la distribución de las empresas por tramo de personal ocupado y sector de actividad económica.

Sea  $X$  y  $Z$  dos variables cualitativas con  $I$  y  $J$  categorías respectivamente. La tabla de contingencia (o de doble entrada) que surge de cruzar estas dos variables tiene  $I \times J$  celdas.

$X \setminus Z$	$z_1$	.....	$z_j$	.....	$z_J$
$x_1$	$N_{11}$	.....	$N_{1j}$	.....	$N_{1J}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
$x_i$	$N_{i1}$	.....	$N_{ij}$	.....	$N_{iJ}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
$x_I$	$N_{I1}$	.....	$N_{Ij}$	.....	$N_{IJ}$

Para cada una de las celdas de la tabla, se obtiene una estimación del total poblacional  $N_{ij}$ , el cual representa la cantidad de empresas que presentan la categoría  $X = x_i$  e  $Z = z_j$  simultáneamente y la estimación de dicho total viene dada por

$$\hat{N}_{ij} = \sum_{k \in s_{ij}} a_k,$$

donde:

$a_k$  es el ponderador primario de la empresa  $k$  proveniente de la muestra

$s_{ij}$  es el subconjunto de las empresas incluidas en la muestra que pertenecen a la celda  $ij$ .

El ajuste de los ponderadores muestrales  $a_k$  se realiza haciendo el cociente entre el total  $N_{ij}$  y el estimado  $\hat{N}_{ij}$ .

El ponderador obtenido luego de la post estratificación se calcula como

$$w_k = \frac{N_{ij}}{\hat{N}_{ij}} a_k$$

El segundo caso, se da cuando **no se conocen** los totales poblacionales a nivel de las celdas, pero sí se conocen al nivel de las marginales de una tabla.

Esto se puede dar por ejemplo, cuando la información auxiliar proviene de distintas fuentes, lo cual imposibilita confeccionar una tabla cruzada, por ejemplo, si se utiliza información proveniente de los registros de UTE y del RPAE.

Otro motivo puede ser que haya pocas observaciones a nivel de las celdas, como es en el caso de la Encuesta Industrial del año 2011, generando que los estimadores obtenidos a nivel de celdas no sean estables o imposibles de calcular (celdas vacías en la muestra).

Lo anterior, obliga a realizar un **ajuste iterativo proporcional llamado RAKING**.

Al igual que el caso anterior, se parte de las estimaciones de los totales de las celdas, para así obtener estimaciones de las marginales de la tabla

$$\hat{N}_{i.} = \sum_j \hat{N}_{ij} \quad \text{y} \quad \hat{N}_{.j} = \sum_i \hat{N}_{ij}$$

$X \setminus Z$	$z_1$		$z_j$		$z_j$	Total
$x_1$	$N_{11}$		$N_{1j}$		$N_{1j}$	$N_{1.} = \sum N_{1j}$
$x_i$	$N_{i1}$		$N_{ij}$		$N_{ij}$	$N_{i.} = \sum N_{ij}$
$x_j$	$N_{j1}$		$N_{jj}$		$N_{jj}$	$N_{.j} = \sum N_{ij}$
Total	$N_{.1} = \sum N_{i1}$		$N_{.j} = \sum N_{ij}$		$N_{.j} = \sum N_{ij}$	

El algoritmo puede comenzar ajustando las filas o las columnas.

Si se comienza ajustando por las filas se obtendrá un primer juego de ponderadores que estime sin error la marginal de las filas. Entonces, se obtiene un ponderador

$$w_k^1 = \frac{N_{.i}}{N_{i.}} a_k$$

Ahora, con los ponderadores  $w_k^1$  se obtienen nuevas estimaciones para los totales de las filas y columnas de la tabla. Los totales de las filas ajustan sin error, pero no tiene porque suceder lo mismo con los totales de las columnas.

Entonces, se obtiene una nueva estimación para los totales de las celdas  $\hat{N}_{ij}^1 = \sum_{k \in S_{ij}} w_k^1$  y por ende para las marginales de las columnas  $\hat{N}_{.j}^1 = \sum_i \hat{N}_{ij}^1$ .

Ahora, se realiza un segundo ajuste para las columnas  $w_k^2 = \frac{N_{.j}}{\hat{N}_{.j}^1} w_k^1$ , con lo cual, se estima de forma exacta las columnas pero no las marginales de las filas.

Este algoritmo se repite hasta las que las diferencias entre las marginales estimadas y las poblacionales sean menores que un error fijado.

Con este procedimiento, se obtiene un nuevo juego de ponderadores  $w_k$  que tienen la característica de estimar sin error el total de las marginales de las variables utilizadas.

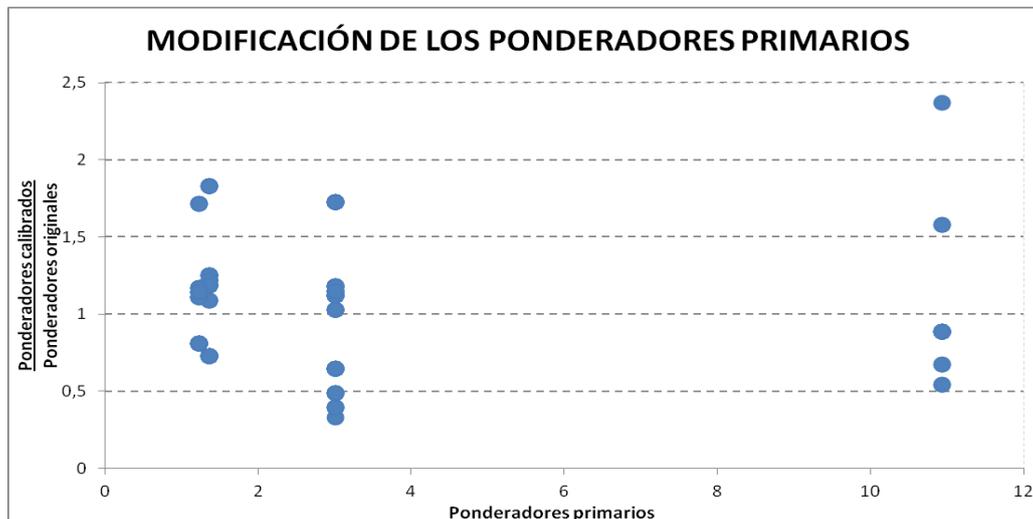
Para el caso de la estimación del consumo industrial de leña se utilizaron la cantidad de empresas por tramo de personal ocupado según subsector.

Subsector (X)	Tramo de personal ocupado (Z)				TOTAL
	10 a 19	20 a 49	50 a 99	100 a 199	
Bebidas y tabaco	8	7	3	2	20
Frigoríficos	13	28	12	10	63
Lácteos	14	17	6	4	41
Madera, papel y celulosa	43	31	4	6	84
Molinos	8	11	7	3	29
Otras Industrias Alimenticias	48	50	19	5	122
Química, caucho, plástico, cemento y cerámica	9	7	4	0	20
Textil y cueros	32	21	12	7	72
<b>Total</b>	<b>175</b>	<b>172</b>	<b>67</b>	<b>37</b>	<b>451</b>

Fuente: Elaboración propia a partir del Registro permanente de actividad económica (RPAE)

La calibración se realiza utilizando las empresas en la muestra clasificadas como NO MG y utilizando los datos del marco del RPAE eliminando las empresas MG, es decir, aquellas en donde su personal ocupado es mayor o igual a 100 y/o su consumo es mayor a 10.000 toneladas en algún período.

En el siguiente gráfico se presenta la modificación de los ponderadores primarios producto de la calibración



Los ponderadores calibrados obtenidos para el año 2011 se utilizan también para brindar estimaciones para los años 2010 y 2012.

## NO respuesta Imputación de los datos faltantes del consumo de leña 2010 -2012

La no respuesta, es la regla más que la excepción en las encuestas por muestreo. Generalmente, en la encuestas de panel<sup>1</sup> el desgastamiento de los informantes es algo común, es decir, no respuesta ocasional.

Por otra parte, el fallecimiento de unidades también es algo que comúnmente ocurre, ya sea, por empresas que cesaron sus actividades o porque sus atributos no cumplen los requisitos necesarios para pertenecer al universo de estudio. Por ejemplo, cambios en su actividad económica principal o porque los consumos de leña anuales han disminuido mucho en el correr del tiempo.

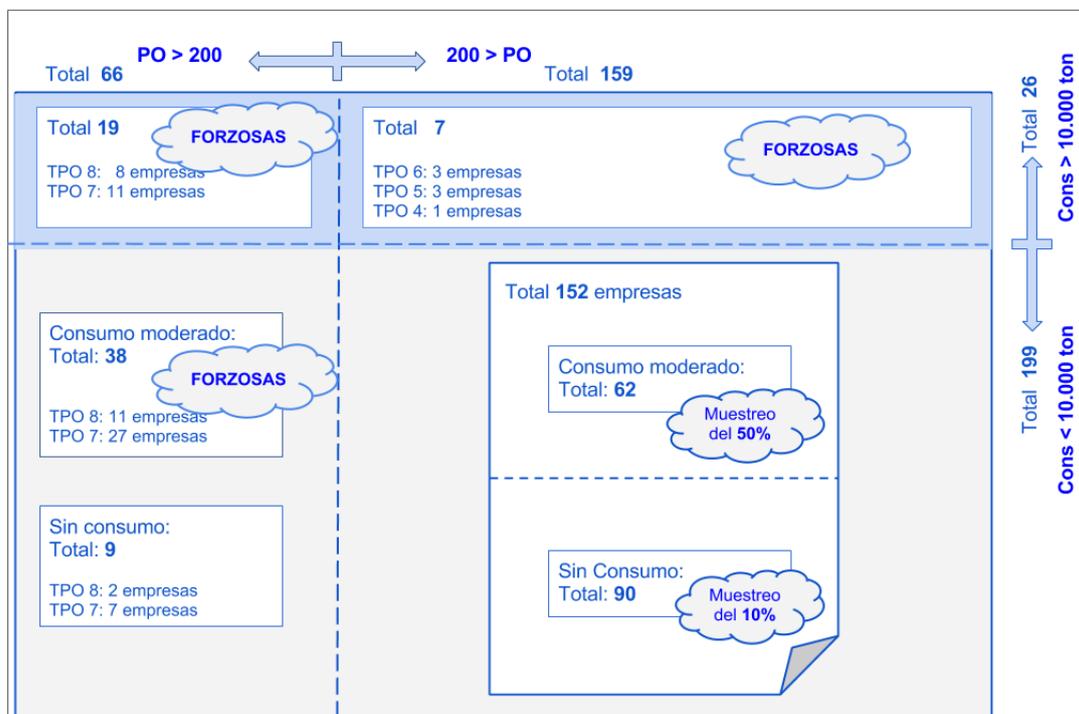
La no respuesta introduce sesgo e incrementa la variabilidad de los estimadores. Lo que implica formular estrategias para la imputación de los datos faltantes o generar sistemas de ponderadores que tengan en cuenta a las empresas omisas en el proceso.

Para el año base de este estudio, 2011, no existe no respuesta. En cambio, para los años 2010 y 2012 no se tienen datos para todas las empresas. Lo anterior, implicó realizar una nueva etapa de relevamiento de información sobre los consumos de aquellas empresas omisas de forma de intentar reducir la no respuesta.

Como primer paso, se verificaron que aquellas empresas omisas estuvieran aún en actividad.

<sup>1</sup> Encuesta por muestreo en donde el mismo individuo es seguido a lo largo del tiempo de forma de poder captar cambios en la población y micropoblaciones y proporcionar estimaciones tanto transversales como longitudinales.

Luego, debido a que los no respondientes pertenecen a grupos (o estratos) bien diferenciados se seleccionaron muestras independientes bajo un muestreo sistemático con arranque aleatorio ordenando a las empresas por su consumo y/o clase de actividad. Las tasas de muestreo en cada uno de estos grupos se observan en el siguiente diagrama.



Para las empresas de la muestra seleccionada de los no respondientes se destinaron suficientes recursos para captar la información. El relevamiento se realizó de forma telefónica y vía e-mail.

Luego, en base a los datos obtenidos con este nuevo relevamiento y a los que ya se tenían (muestra de respondientes) se imputaron los consumos de aquellas empresas omisas.

El método de imputación utilizado consiste en la media geométrica de las variaciones de los establecimientos NO omisos según subsector, y si es o no MG.

Sea  $\hat{y}_{t,k}$  el consumo de leña imputado para la k-ésima empresa en el año  $t$ ,

$$\hat{y}_{t,k} = \left[ \prod_{k \in r_{g,t}} \Delta_{k,t-1}^t \right]^{1/m_{g,t}}$$

donde:

$\Delta_{k,t-1}^t = \frac{y_{t,k}}{y_{t-1,k}}$ , es la variación o incremento relativo del consumo de leña de la empresa  $k$  en el año  $t$ , respecto al año  $t-1$ ,  $y_{t,k}$  y  $y_{t-1,k}$  es el consumo de leña informado por la empresa año  $t$  y  $t-1$  respectivamente.

$r_{g,t}$  es el conjunto de las empresas pertenecientes a la muestra que contestaron en ambos años y son pertenecientes al grupo  $g$  (interacción entre el subsector y si es MG).

$m_{g,t}$  es la cantidad de empresas respondientes pertenecientes al grupo  $g$ .

La elección de la media geométrica en vez de la media aritmética se debe a que la primera es más robusta que la segunda, es decir, no se ve influenciada por la presencia de variaciones atípicas en el consumo de leña, las cuales ocurren, por ejemplo, cuando un establecimiento incorpora una nueva caldera a leña y/o su consumo se reduce bruscamente por un tema productivo o económico, por ejemplo.

Finalmente, la estimación para el consumo total de leña en el sector industrial para un año  $t$  viene dada por

$$\hat{Y}_t = \sum_{h=1}^6 \sum_{k \in S_h} y_{t,k} w_k$$

donde

$y_{t,k}$  es el consumo de leña de la empresa  $k$  en el año  $t$  el cual puede ser conocido o imputado ( $\hat{y}_{t,k}$ ).

$w_k$  es el ponderador calibrado de la empresa  $k$ . Si la empresa  $k$  es MG (estrato 5 y 6)  $w_k = 1$

## Resultados

Luego de aplicada la metodología expresada en las diferentes secciones de este documento, se llegó a obtener una cantidad de leña consumida tanto para el año 2011, base del estudio, como para los años 2010 y 2012. Estos consumos de leña se obtienen desagregados en los ocho subsectores definidos. En el siguiente cuadro se observan para los diferentes años y subsectores los resultados obtenidos.

Subsector	Año 2010	Año 2011	Año 2012
<b>Total</b>	<b>608.597</b>	<b>730.194</b>	<b>685.881</b>
Bebidas y tabaco	67.593	74.656	82.445
Frigoríficos	115.590	112.506	114.627
Lácteos	91.508	142.717	123.508
Madera, papel y celulosa	124.005	127.275	145.956
Molinos	42.373	53.599	34.817
Otras Industrias Alimenticias	57.292	89.771	74.637
Química, caucho, plástico, cemento y cerámica	37.913	63.124	39.968
Textil y cueros	72.322	66.546	69.922

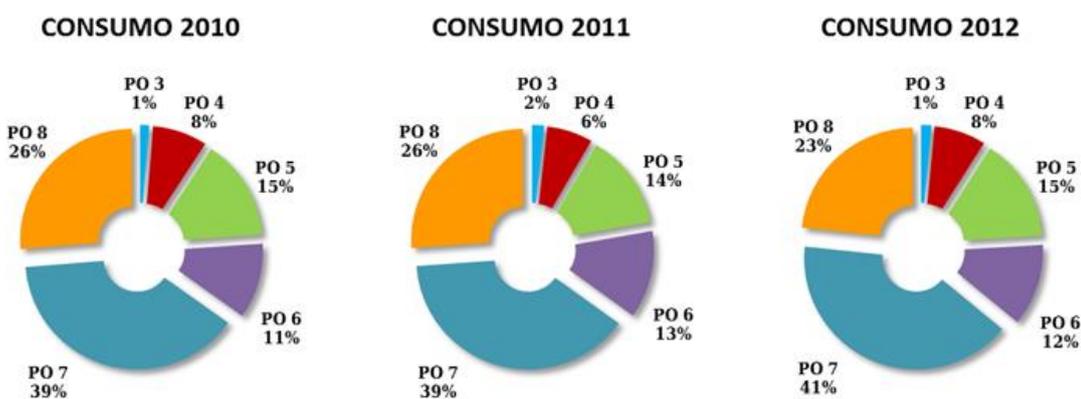
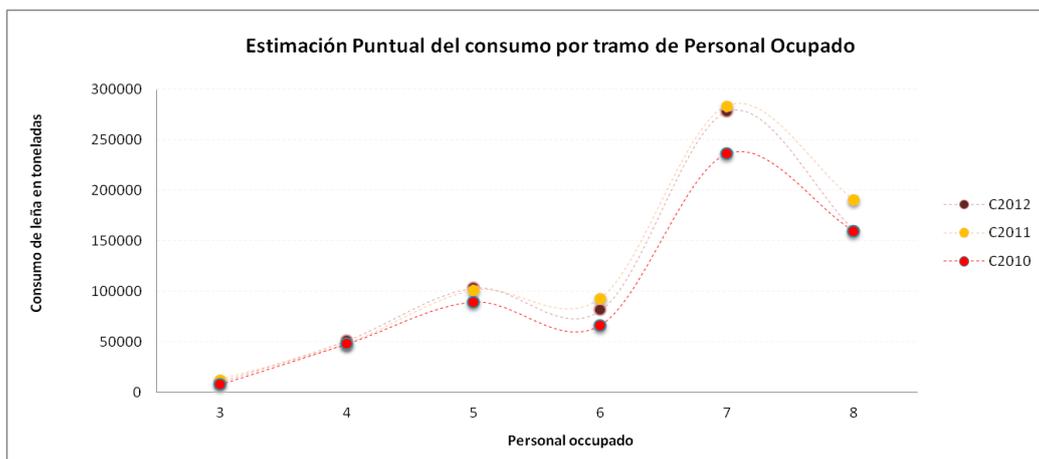
Esta información de consumos, a nivel de totales industriales, permite evaluar la calidad de la metodología de actualización que se llevaba adelante a partir de los estudios realizados en 2006 por la DNE. Así en el siguiente cuadro se presentan los totales y las variaciones interanuales con los dos métodos

Subsector	Año 2010	Año 2011		Año 2012	
	Total	Total	$\Delta(11/10)$	Total	$\Delta(12/11)$
Resultados del Estudio de Estimación de Consumos de leña 2011	608.597	730.194	▲ +20%	685.881	▼ -6%
Actualización con encuesta de leña industrial a partir de estudio de Consumos y Usos de la energía del año 2006	618.425	670.096	▲ +8%	635.281	▼ -5%

Es notoria la diferencia entre las variaciones detectadas con este estudio respecto a los resultados de la actualización anual de leña a partir de resultados del año 2006. Esto en parte se debe a que las hipótesis del estudio comenzaron a perder validez al igual que su metodología de actualización debido a cambios en el sector.

Por otra parte, para los totales hasta el año 2010 no eran considerables las diferencias entre las dos estimaciones, pero como se mencionó al inicio, el subsector correspondiente a panaderías fue estudiado por separado, lo cual implica, que con un consumo resultante en el estudio se debe aumentar en valores cercanos a las 50.000 toneladas, implicando que las diferencias sobre todo del año 2011 se vean acrecentadas.

Anteriormente se observaba que en la muestra, al aumentar el personal ocupado había un aumento (generalmente o en promedio) en consumo de leña, esto es así en todas las ramas, pero en el universo de la industria son muchas más las empresas que se encuentran en el tramo de PO 7 que en el de PO 8, y es por este motivo que los consumos totales de leña aumentan paulatinamente con el aumento de PO hasta su máximo en PO igual a 7, dado que luego se reduce considerablemente el número de empresas.



En el siguiente cuadro se presenta el porcentaje de empresas que consumen leña según establecimiento para los años 2010, 2011 y 2012.

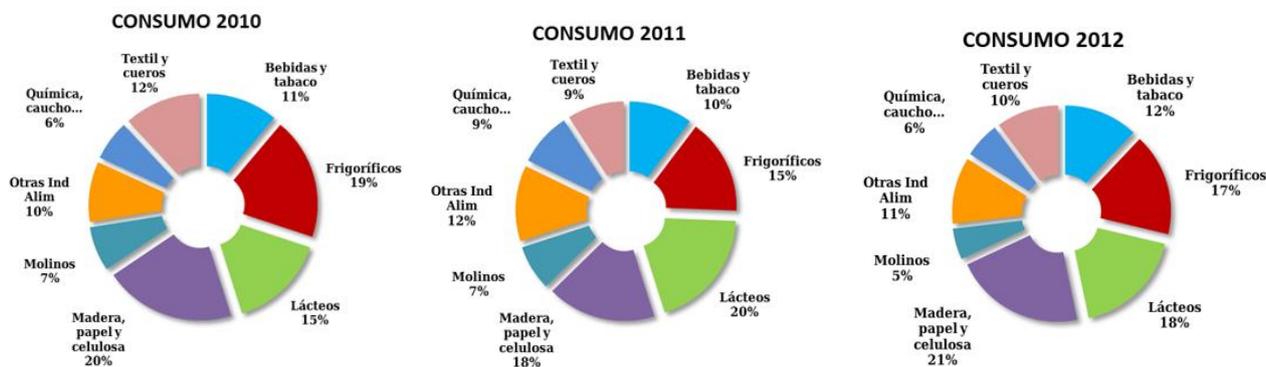
Sector	Año		
	2010	2011	2012
Bebidas y tabaco	53,5%	53,5%	53,5%
Frigoríficos	51,4%	51,4%	51,4%
Lácteos	52,7%	52,7%	52,7%
Madera, papel y celulosa	18,8%	18,8%	18,8%
Molinos	26,9%	26,9%	26,9%
Otras Industrias Alimenticias	34,4%	34,4%	33,6%
Química, caucho, plástico, cemento y cerámica	50,8%	50,8%	46,6%
Textil y cueros	31,5%	30,2%	31,5%

En general se observa que no hay grandes variaciones en períodos cortos de tiempo. Esto se debe a que la incorporación de nuevos emprendimientos, o los cambios tecnológicos en los existentes no son tan dinámicos como las variaciones de consumos. Esto se debe a que en general las industrias tienen la opción de utilizar un combustible sustituto en lugar de cambiar completamente la tecnología, ocasionando entonces una marcada estabilidad en cuanto al número de empresas que utiliza leña.

Los sectores con mayor penetración de la leña son Bebidas y Tabaco, Lácteos, Frigoríficos y el subsector que agrupa Química, Caucho, Plástico, Cemento y Cerámica, en donde las empresas que consumen leña son del mismo orden que las que no utilizan esta fuente energética.

### Participación de los subsectores en el consumo de leña total de la industria

El sector de Madera, Papel y Celulosa, a pesar de no estar constituido mayormente por empresas que consumen leña (menos del 19% lo hacen), es el de mayor consumo a nivel industrial en los tres años evaluados, con porcentajes que rondan el 20%. Los sectores que siguen en importancia son Frigoríficos, Lácteos y Bebidas y Tabacos. Estos cuatro sectores que ocupan los primeros puestos en consumo representan entre el 63 y el 68% del total industrial



## Precisión de las estimaciones

Los errores de estimación ocurren debido a que las **inferencias** de la población se realizan por medio de la información obtenida de una **muestra** de la misma. El diseño muestral, la variabilidad de los datos y el tamaño de muestra determinan el error de estimación. A su vez, diferentes métodos de estimación conllevan a diferentes errores.

El estimador utilizando los ponderadores calibrados  $w_k$  de la muestra  $s$  para el total de la variable de interés  $y$ , en la población  $U$  en el año  $t$

$$Y_t = \sum_{k \in U} y_{t,k},$$

viene dado por

$$\hat{Y}_t = \sum_{k \in s} w_k y_{t,k},$$

y un estimador de la varianza del mismo es

$$\hat{V}(\hat{Y}_t) = \sum_{h=1}^5 \frac{1}{n_h(n_h - 1)} (1 - f_h) \sum_{k \in s_h} (y_{t,k} w_k - \hat{Y}_{t,h})^2,$$

donde

$f_h = n_h/N_h$  es la tasa de muestreo en el estrato  $h$

$\hat{Y}_{t,h} = \sum_{k \in s_h} y_k w_k$  es la estimación puntual del consumo total de leña en el estrato  $h$  para el año  $t$ .

Para la estimación de cualquier subpoblación perteneciente al universo de estudio, por ejemplo, para uno de los subsectores definidos en este estudio, se debe cambiar en la ecuación anterior la variable  $y$ , por la variable  $y_d$  la cual toma los siguientes valores

$$y_{k,d} = \begin{cases} y_k, & \text{si } k \text{ pertenece al subsector } d \\ 0, & \text{en otro caso} \end{cases}$$

Lo anterior, implica que las estimaciones obtenidas a nivel de subsector tengan una menor precisión que a nivel de la población total, producto de la cantidad de ceros en la variable de interés  $y_d$  y por el hecho de no haber controlado el tamaño de muestra en el subsector, es decir, definir el subsector como un estrato a la hora de seleccionar la muestra.

Los intervalos de confianza para el total de la variable de interés  $y$  para un nivel del 95% se calculan de la siguiente forma

$$\hat{Y}_t \pm 1.96 \sqrt{\hat{V}(\hat{Y}_t)},$$

donde 1.96 es el valor de una normal estándar que acumula el 0.975 de probabilidad.

En el siguiente cuadro se presentan los intervalos de confianza al 95% para el consumo total de leña en cada uno de los años por subsector de actividad y para el total de la población utilizando el paquete *Survey* del software libre R.

Sector	Año 2010			Año 2011			Año 2012		
	Límite inferior	Estimación puntual	Límite superior	Límite inferior	Estimación puntual	Límite superior	Límite inferior	Estimación puntual	Límite superior
<b>Total</b>	<b>577.292</b>	<b>608.597</b>	<b>639.902</b>	<b>695.357</b>	<b>730.194</b>	<b>765.031</b>	<b>651.554</b>	<b>685.881</b>	<b>720.209</b>
Bebidas y tabaco	63.705	<b>67.593</b>	71.481	68.494	<b>74.656</b>	80.817	76.197	<b>82.445</b>	88.694
Frigoríficos	114.040	<b>115.590</b>	117.140	107.241	<b>112.506</b>	117.771	109.953	<b>114.627</b>	119.300
Lácteos	83.079	<b>91.508</b>	99.937	133.847	<b>142.717</b>	151.587	113.450	<b>123.508</b>	133.567
Madera, papel y celulosa	103.782	<b>124.005</b>	144.228	106.836	<b>127.275</b>	147.714	123.895	<b>145.956</b>	168.017
Molinos	40.524	<b>42.373</b>	44.222	51.750	<b>53.599</b>	55.448	33.153	<b>34.817</b>	36.481
Otras Ind Alim	47.540	<b>57.292</b>	67.045	75.368	<b>89.771</b>	104.175	63.531	<b>74.637</b>	85.744
Química, caucho, plástico, cemento y cerámica	25.036	<b>37.913</b>	50.791	45.968	<b>63.124</b>	80.280	25.205	<b>39.968</b>	54.730
Textil y cueros	54.184	<b>72.322</b>	90.460	48.888	<b>66.546</b>	84.204	51.447	<b>69.922</b>	88.398

Cabe aclarar, que la variabilidad proporcionada por el método de imputación no es tenida en cuenta a la hora de calcular los intervalos de confianza.

En los siguientes gráficos se presentan las estimaciones de consumo de cada subsector así como su respectivo intervalo de confianza (al 95%) para los años 2010, 2011 y 2012.

