



---

# Encuesta Nacional de Participación Cultural 2017

---

**Informe metodológico Factor de Expansión**

**INSTITUTO NACIONAL DE ESTADÍSTICAS**

**Marzo / 2018**



**SUBDIRECCIÓN TÉCNICA**  
**DEPARTAMENTO DE INVESTIGACIÓN Y DESARROLLO**  
**SECCIÓN DE ESTADÍSTICAS SOCIALES**

---

Encuesta Nacional de Participación Cultural 2017.  
Informe Metodológico Factor de Expansión.  
Producto N°4.2/Versión 02.  
Instituto Nacional de Estadísticas, Chile.  
28 de marzo de 2018.

## ÍNDICE

I.	Desarrollo de Factores de Expansión .....	5
I.1.	Ponderador de selección de Unidades de Primera Etapa (Manzanas) .....	7
I.2.	Ajuste por omisión de conglomerados.....	8
I.3.	Ponderador de selección de Unidades de Segunda Etapa (Viviendas) .....	9
I.4.	Ajuste por elegibilidad.....	10
I.4.1.	Ajuste por elegibilidad desconocida.....	11
I.4.2.	Ajuste por no elegibilidad.....	11
I.5.	Suavizamiento del ponderador ajustado por elegibilidad .....	12
I.6.	Ajuste por no respuesta. ....	14
I.7.	Ponderador de Selección de Unidades de Tercera Etapa (Personas) .....	19
I.8.	Suavizamiento ponderador de selección de personas.....	20
I.9.	Ajuste para calibrar al total poblacional por región, sexo y tramo etario. ....	22
II.	Estimadores de precisión .....	24
II.1.	Estimación de la varianza, cálculo de estratos y conglomerados de varianza.....	24
II.1.1.	Creación de pseudos-estratos. ....	25
II.1.2.	Creación de pseudo-conglomerados.....	25
II.2.	Representatividad de las estimaciones.....	26
II.3.	Estimación de índice de participación cultural y errores efectivos.....	28
III.	ANEXOS.....	29
III.1.	Metodología de suavizamiento de factores de expansión. ....	29

## ÍNDICE DE TABLAS

Tabla 1. Distribución de viviendas seleccionadas por condición de elegibilidad.....	12
Tabla 2. Bondad de ajuste del modelo de respuesta aplicado.....	16
Tabla 3. Estimación de coeficientes del modelo de respuesta aplicado.....	17
Tabla 4. Razón de ajuste según decil.....	18
Tabla 5. Proyecciones de Población por región y sexo para personas de 15 años y más.....	22
Tabla 6. Total de Estratos, conglomerados, varstrat y varunit según región.....	26
Tabla 7. Criterio de representatividad .....	27
Tabla 8. Estimación índice de participación cultural y errores efectivos.....	28
Tabla 9. Dimensiones incluidas en el índice de participación cultural.....	30
Tabla 10. Distribución de frecuencia del Índice de participación cultural.....	31
Tabla 11. Estadísticas descriptivas para el índice de participación cultural.....	32
Tabla 12. Test fiabilidad construcción de indicador.....	32
Tabla 13. Niveles de participación en actividades culturales.....	32
Tabla 14. Frecuencia de varunit según número de viviendas que lo componen.....	35

## I. Desarrollo de Factores de Expansión

La Encuesta Nacional de Participación Cultural 2017 constituye un estudio descriptivo e inferencial, que describe la magnitud y tendencias en relación al consumo de servicios culturales que manifiestan las personas residentes de las áreas urbanas del país. Para esto, se utilizó un diseño muestral probabilístico, estratificado geográficamente por región y tamaño de las manzanas según el número de viviendas que poseen<sup>1</sup>.

El diseño es trietápico, es decir, la selección de unidades se realiza en tres etapas. Las manzanas, definidas como las unidades de primera etapa, fueron seleccionadas de forma sistemática, al interior de cada estrato de muestreo; por su parte las viviendas, unidades de segunda etapa, fueron seleccionadas con igual probabilidad de selección dentro de cada manzana; mientras que los informantes, unidades de tercera etapa, se seleccionaron mediante un método aleatorio simple automatizado por un dispositivo electrónico utilizado en el levantamiento de la información según el número de personas elegibles al interior de cada vivienda. En este sentido, como la población objetivo son las personas de 15 años y más, la selección del informante cumple con esta misma característica.

El factor de expansión se interpreta como la cantidad de unidades en la población que representa una unidad de la muestra, y es calculado como el inverso de la probabilidad de selección de las unidades de muestreo. Atendiendo al diseño muestral, el cálculo del factor de expansión se puede separar en tres partes que dan cuenta de cada etapa, es decir, un ponderador que da cuenta de la selección de las manzanas; otro que da cuenta de la selección de las viviendas y; finalmente, otro que da cuenta de la selección del informante dentro de la vivienda.

Adicionalmente, por tratarse de un diseño complejo es necesario incorporar una serie de ajustes que permiten mejorar la precisión de las estimaciones y evitar posibles sesgos.

---

<sup>1</sup> Según el número de viviendas las manzanas se clasifican en 5 grupos de tamaño, excluyendo aquellas manzanas que tengan siete viviendas o menos.

En consecuencia, la metodología de cálculo de los factores de expansión consiste en la aplicación secuencial de siete ponderadores o ajustes:

1. Ponderador de selección de unidades de primera etapa.
2. Ajuste por omisión de conglomerados.
3. Ponderador de selección de unidades de segunda etapa.
4. Ajuste por elegibilidad de las viviendas (elegibilidad desconocida y no elegibilidad).
5. Ajuste por no respuesta.
6. Ponderador de selección de unidades de tercera etapa.
7. Ajuste para calibrar al total de población según proyecciones por región, sexo y tramo de edad.

### **I.1. Ponderador de selección de Unidades de Primera Etapa (Manzanas)**

---

La selección de las unidades de primera etapa (manzanas), en cada región y dentro de cada grupo de tamaño, se efectuó en forma aleatoria, sistemática y con igual probabilidad. El ponderador de selección de estas unidades, se calcula como el inverso de sus probabilidades de selección dentro de cada estrato de muestreo, conformado por región y grupo de tamaño, y viene dado por la siguiente fórmula:

$$w_{rgi} = \frac{N_{rg}}{n_{rg}^{teo}}$$

Donde:

$w_{rgi}$  : Ponderador de selección del conglomerado (manzana)  $i$  del grupo tamaño  $g$  de la región  $r$ .

$r$  : Índice correspondiente a la región.

$g$  : Índice correspondiente al grupo de tamaño de la manzana.

$i$  : Índice de manzana.

$N_{rg}$  : Número de manzanas en el Marco Muestral 2015 (MM2015) del grupo tamaño  $g$  en la región  $r$ .

$n_{rg}^{teo}$  : Número de manzanas seleccionadas del grupo tamaño  $g$  de la región  $r$ .

El ponderador de selección de conglomerados puede ser interpretado como el total de unidades (manzanas) de la población, es decir, cada manzana presente en la muestra representa a  $w_{rgi}$  manzanas en la población. Del mismo modo, la suma de los ponderadores de selección de todas las manzanas en la muestra, corresponde al total de las manzanas existentes en el marco muestral.

## 1.2. Ajuste por omisión de conglomerados

---

Previo al proceso de selección de la muestra de conglomerados (manzanas) son excluidos algunos por razones operativas y de costo; en consecuencia, para compensar la falta de cobertura provocada, se incluye el ajuste por omisión de conglomerados. Para este caso, las manzanas excluidas corresponden a todas aquellas que pertenecen al grupo tamaño “0”, es decir, que contienen menos de 8 viviendas.

La forma empleada para el cálculo de este ajuste se obtiene como la razón entre:

- El total de viviendas contenidas en las manzanas del MM2015 a nivel de región, considerando sólo las comunas con presencia en la muestra (considerando las manzanas de todos los grupos de tamaño).
- El total de viviendas según MM2015 en las manzanas seleccionadas, estimadas a partir del ponderador de selección de unidades de primera etapa ( $w_{rgi}$ ) a nivel de región.

Dicho procedimiento se calcula a partir de la siguiente fórmula:

$$\tilde{R}_r = \frac{M_r}{\sum_{i \in \theta_r} M_{ri} \cdot w_{rgi}}$$

Y luego, se obtiene el factor ajustado por omisión de conglomerados según:

$$w'_{rgi} = w_{rgi} \cdot \tilde{R}_r$$

Donde:

- $w'_{rgi}$  : Ponderador de selección del conglomerado  $i$  del grupo tamaño  $g$  de la región  $r$  ajustado por omisión de conglomerados.
- $M_r$  : Es el total de viviendas en las manzanas registradas en el MM2015 en la región  $r$ , para las comunas con presencia en la muestra.
- $\tilde{R}_r$  : Razón de ajuste por omisión de conglomerados en la región  $r$ .
- $M_{ri}$  : Es el total de viviendas en la manzana  $i$  en la región  $r$  perteneciente a la muestra registradas en el MM2015.
- $\theta_r$  : Es el listado de manzanas seleccionadas en la región  $r$ .



### 1.3. Ponderador de selección de Unidades de Segunda Etapa (Viviendas)

---

El ponderador de selección de unidades de segunda etapa (viviendas) dentro de cada manzana seleccionada, refleja cuántas viviendas en la población representa la vivienda seleccionada en la muestra. En términos generales se calcula como el inverso de la probabilidad de selección de la vivienda, condicionado a la selección de la manzana dentro del estrato de muestreo al que pertenece.

$$w_{rgij} = \frac{M'_{rgi}}{m_{rgi}^{teo}}$$

Donde:

- $w_{rgij}$  : Ponderador de selección de la vivienda  $j$  del conglomerado  $i$  del grupo tamaño  $g$  dentro de la región  $r$ .
- $M'_{rgi}$  : Número actualizado (según enumeración) de viviendas en la manzana  $i$  en el grupo tamaño  $g$  dentro de la región  $r$ .
- $m_{rgi}^{teo}$  : Número de viviendas seleccionadas en la manzana  $i$  en el grupo tamaño  $g$  dentro de la región  $r$ .

Así, el factor que da cuenta de la selección de manzanas y viviendas, ajustado por omisión de conglomerados viene dado por:

$$w''_{rgij} = w'_{rgi} \cdot w_{rgij}$$

Siendo:

- $w'_{rgi}$  : Ponderador de selección del conglomerado  $i$  del grupo tamaño  $g$  dentro de la región  $r$ , ajustado por omisión de conglomerados.
- $w_{rgij}$  : Ponderador de selección de la vivienda  $j$  del conglomerado  $i$  del grupo tamaño  $g$  dentro de la región  $r$ .

#### I.4. Ajuste por elegibilidad

---

La población objetivo del estudio está compuesta por personas de 15 años y más que habitan en viviendas particulares en las zonas urbanas de todas las regiones del país y, la selección de la muestra se basa en el listado de viviendas que provee el MM2015. Sin embargo, a pesar de todos los esfuerzos que se realizan por mantener este marco muestral válido para la correcta selección de las unidades, siempre existe la posibilidad de errores de inclusión (o de exclusión) debido a su desactualización.

Es por ello que, al término del trabajo de campo, todas las viviendas seleccionadas inicialmente se clasifican en tres grandes grupos, estos son:

- **Viviendas elegibles:** Corresponde a edificaciones identificadas como viviendas particulares.
- **Viviendas no elegibles:** Corresponde a edificaciones identificadas como negocios, viviendas colectivas, viviendas deshabitadas, viviendas de veraneo, viviendas destruidas, entre otros.
- **Viviendas de elegibilidad desconocida:** Corresponde a edificaciones en que no se pudo determinar su estado. Este es el caso, por ejemplo, de unidades que nunca fueron enviadas a terreno, o viviendas que no se pudieron localizar, y otros casos similares.

El ponderador de selección de viviendas tiene valores válidos para las viviendas elegibles, no elegibles y de elegibilidad desconocida. Por esta razón, es importante ajustar las probabilidades de selección de las viviendas seleccionadas para incorporar el hecho de que una proporción de éstas resulta ser no elegibles, y para otras viviendas simplemente, se desconoce su elegibilidad.

A continuación, se describen los ajustes aplicados en caso de viviendas de elegibilidad desconocida; y de no elegibles.

#### 1.4.1. Ajuste por elegibilidad desconocida.

Este ajuste consiste en redistribuir los pesos acumulados por las viviendas de elegibilidad desconocida entre las viviendas con elegibilidad conocida (elegibles y no elegibles) dentro de cada región.

La fórmula utilizada es la siguiente:

$$\tilde{R}_{r, know} = \frac{\sum_{i \in \theta_r} \sum_{j \in \varphi_i} w''_{rgij}}{\sum_{i \in \theta_r} \sum_{j \in \varphi_{i, know}} w''_{rgij}}$$

Donde

$\varphi_i$  : Es el conjunto de viviendas seleccionadas en la manzana  $i$ .

$\varphi_{i, know}$  : Es el conjunto de viviendas que resultan de elegibilidad conocida (elegible y no elegible) dentro de la manzana  $i$ .

$\theta_r$  : Es el conjunto de manzanas seleccionadas a nivel de región.

Luego el factor ajustado por elegibilidad desconocida viene dado por la siguiente expresión:

$$w'''_{rgij} = w''_{rgij} \cdot \tilde{R}_{r, know}$$

Este ajuste se realiza con el objeto de que las viviendas con elegibilidad desconocida puedan ser representadas por las viviendas cuya elegibilidad es conocida.

#### 1.4.2. Ajuste por no elegibilidad

Las viviendas cuyo estatus sea “no elegible” (oficinas de empresas, viviendas abandonadas, viviendas de veraneo, viviendas demolidas, entre otras) no forman parte de la población objetivo, por lo que a partir de esta etapa y para fines del cálculo del factor de expansión se excluyen las unidades clasificadas con este tipo de elegibilidad, lo cual se implementa colocando un valor blanco (missing) en el ponderador de selección de viviendas corregido por elegibilidad desconocida, a las viviendas que poseen clasificación “no elegibles”.

La tabla 1 muestra la distribución de las viviendas según su condición de elegibilidad, en donde 2,12% resultan ser no elegibles; y 1,12% de elegibilidad desconocida.

**Tabla 1. Distribución de viviendas seleccionadas por condición de elegibilidad.**

Condición de Elegibilidad	Frecuencia	Porcentaje
Total de viviendas	17.745	100
Elegible	17.169	96,75
No elegible	377	2,12
Elegibilidad desconocida	199	1,12

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

### 1.5. Suavizamiento del ponderador ajustado por elegibilidad

Debido a la incorporación de diferentes ajustes a los ponderadores iniciales de las unidades, se identifican unidades con factores de expansión atípicos, es por ello que se hace necesario la incorporación de un procedimiento de suavizado de los factores, a objeto de compensar la variabilidad introducida. Dicho procedimiento consiste en truncar los factores de expansión que se ubican por encima de un umbral definido, para luego redistribuir la diferencia resultante entre los factores asociadas a las unidades que se encuentran por debajo de éste umbral, de manera proporcional<sup>2</sup>.

El factor suavizado, se describe mediante la siguiente expresión:

$$w_{rgij}^{t*} = \begin{cases} w_{rgij}''' \cdot \frac{(\sum_{g \in \Psi_r} \sum_{i \in \Theta_g} \sum_{j \in \Omega_i} w_{rgij}''' - \sum_{g \in \Psi_r} \sum_{i \in \Theta_g} \sum_{w_{rgij}''' > M_{ra}} w_{rgij}''' M_{ra})}{\sum_{g \in \Psi_r} \sum_{i \in \Theta_g} \sum_{w_{rgij}''' \leq M_{ra}} w_{rgij}'''} & ; \text{ si } w_{rgij}''' \leq M_{ra} \\ M_{ra} & ; \text{ si } w_{rgij}''' > M_{ra} \end{cases}$$

Donde:

$w_{rgij}^{t*}$  : Factor de expansión suavizado.

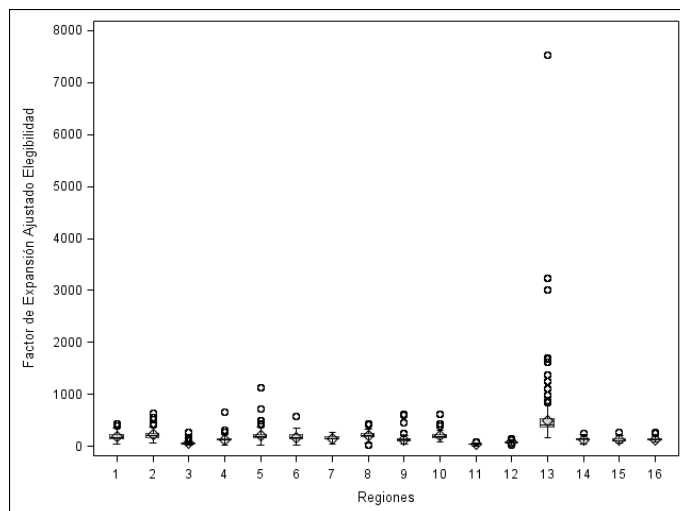
$a$  : Punto de corte en el cual se minimiza el error cuadrático medio, del índice de participación cultural. En esta caso se prueba con los valores entre 4 y 10, resultando como corte más favorable  $a = 4$ .

$M_{ra}$  : Producto entre el punto de corte y la media del factor de expansión a nivel de región.

<sup>2</sup> La Metodología explícita, se describe con mayores detalles en el Anexo 1: "Metodología de suavizamiento de los factores de expansión".

Los gráficos 1 y 2 muestran la distribución de los valores del factor de expansión antes y después del suavizamiento.

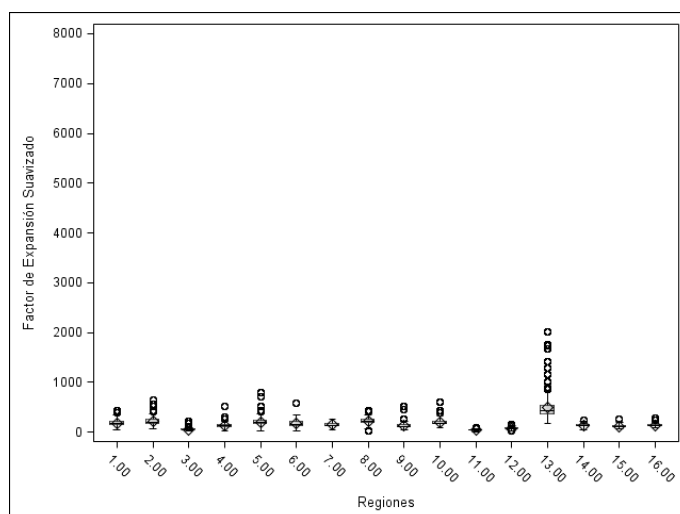
**Gráfico 1. Factor de Expansión ajustado por elegibilidad según región**



Fuente: Elaborado por el Instituto Nacional de Estadísticas.

En el gráfico 1, se aprecia la presencia de factores de expansión atípicos que sobrepasan el valor de 3.000, específicamente en la región metropolitana, observándose también que la mayoría de los factores no superan el valor 2.000.

**Gráfico 2. Factor de Expansión ajustado por no elegibilidad suavizado según región.**



Fuente: Elaborado por el Instituto Nacional de Estadísticas.

Luego del suavizamiento, se truncan el 0,4% de los registros, lo que provoca que los valores de los factores de expansión no superen las 3.000 unidades.

## I.6. Ajuste por no respuesta.

---

En esta etapa sólo son consideradas las viviendas elegibles y que cumplen con las características de pertenecer a la población objetivo, sin embargo, se presenta el caso de que algunos residentes de las viviendas elegibles no deseen participar, por lo que se origina la no respuesta.

Ante esto, surge la necesidad de aplicar un ajuste por no respuesta con el objetivo de lograr que las unidades que no responden sean representadas por las que sí, previendo no introducir sesgo debido a la posibilidad de que exista relación entre la no respuesta y la variable de interés.

Es por ello que, para compensar la pérdida de unidades se realiza un ajuste del factor a través del método denominado “Propensity Score”, que permite estimar mediante un modelo de regresión lineal la probabilidad de responder para las viviendas elegibles, y con esto crear grupos homogéneos en cuanto a la probabilidad de respuesta para, posteriormente realizar la redistribución de los pesos.

Aunque sería ideal realizar dicho ajuste a nivel de vivienda, resulta imposible, ya que el nivel mínimo para el cual se dispone de información que permite configurar el modelo, es el conglomerado, lo que origina que éste ajuste sea realizado a nivel de la unidad primaria de muestreo (manzana).

Con esto, se aplica un modelo de regresión lineal en donde la variable dependiente corresponde a la tasa de respuesta a nivel de manzana, mientras que las variables explicativas utilizadas presentan diferentes niveles de desagregación, tal como se presenta a continuación:

- A nivel de manzana:
  1. Distancia a una comisaría.
  2. Distancia a áreas verdes.
  3. Logaritmo del avalúo fiscal.
  4. Grupo de tamaño.
  5. Clasificación socio-económica.
  6. Decil de ingreso.

- A nivel de distrito censal:
  1. Número de establecimientos educacionales.
  2. Número de establecimientos de salud.
  
- A nivel de comuna:
  1. Tasas de carencia provenientes de la medición de pobreza multidimensional de Casen 2015, en total 19 tasas.
  2. Tasa de alfabetización.
  3. Tasa de ocupación.
  4. Número de hogares que reciben asignación familiar.
  5. Número de hogares con productos bancarios.
  6. Número de hogares que reciben subsidio.
  7. Número de viviendas habitadas por su propietario.
  
- A nivel de macrozona:
  1. Macrozona de pertenencia de cada manzana.

La especificación del modelo de estimación es la siguiente:

$$y_{ilhk} = \alpha + \beta X_i + \gamma Z_l + \delta W_h + \varphi M_k + \varepsilon$$

Donde:

$y_{ilhk}$  : Tasa de respuesta de la manzana “i”, perteneciente al distrito “l”, de la comuna “h”, de la

$X_i$  : Matriz de variables a nivel de manzana, para la manzana “i”.

$Z_l$  : Matriz de variables a nivel de distrito, para el distrito “l”.

$W_h$  : Matriz de variables a nivel de comuna, para la comuna “h”.

$M_k$  : Matriz de variables a nivel de macrozona, para la macrozona “k”.

Las variables que ingresan al modelo de regresión, son seleccionadas mediante el método *stepwise*, y luego, se modifican o se van excluyendo del modelo a objeto de tener una interpretación consistente y evitar problemas de multicolinealidad.

A continuación, se presentan las variables que conforman el modelo según nivel de desagregación:

A nivel de manzana fueron seleccionadas las variables:

- Distancia a comisaría.
- Logaritmo del avalúo fiscal.
- Grupo de tamaño 1.
- Condición socio-económica.

En particular, a la variable “distancia a comisaria” se le realizó una transformación logarítmica debido a la alta variabilidad presentada.

Por otro lado, a nivel de comuna fueron seleccionadas las variables:

- Tasa de rezago escolar.
- Tasa de carencia de adscripción al sistema de salud.
- Tasa de carencia en jubilaciones.
- Tasa de carencia en habitabilidad.
- Tasa de carencia en accesibilidad.
- Tasa de carencia en apoyo y participación.
- Tasa de carencia en seguridad.

Por último, a nivel de macrozona, la variable que ingresó al modelo es:

- Pertenencia a macrozona norte y macrozona centro.

La tabla 2 muestra la bondad de ajuste del modelo, con su correspondiente valor P y el  $R^2$  obtenido.

**Tabla 2. Bondad de ajuste del modelo de respuesta aplicado**

Fuente	Grados Libertad	Suma de cuadrados	Cuadrado de la media	F-Valor	Pr > F
Modelo	14	27,48	1,96	31,97	<.0001
Error	1.926	118,28	0,06		
Total corregido	1.940	145,76			
Estadísticos de Ajuste					
Raíz MSE	0,25	R-cuadrado	0,1886		
Media dependiente	0,7	R-Sq Ajust	0,1827		
Coef Var	35,63				

Fuente: Elaborado por el Instituto Nacional de Estadísticas.



Se aprecia en la tabla 2, que el modelo utilizado para explicar la tasa de respuesta en términos globales resulta ser significativo y que alcanza un  $R^2$  ajustado de 18,27%, lo que indica que el modelo logra explicar ese porcentaje de la variabilidad total de la tasa de respuesta. Este valor se considera aceptable dado que la finalidad es agrupar viviendas con características similares en cuanto a su probabilidad de responder la encuesta.

La tabla 3 muestra la estimación de los coeficientes de las variables explicativas.

**Tabla 3. Estimación de coeficientes del modelo de respuesta aplicado.**

Variable	Estimador	Constante de heterocedasticidad			Inflación de la varianza
		Error Estándar	Valor t	Pr >  t	
Intercepto	2,01	0,17	11,62	<0,0001	0,00
Logaritmo de la distancia a una comisaría	-0,01	0,01	-1,04	0,2979	1,06
Logaritmo del avalúo fiscal promedio	-0,08	0,01	-8,55	<0,0001	1,40
Manzana pequeña (Grupo 1)	-0,03	0,01	-2,44	0,0147	1,02
CSE Baja	0,05	0,01	3,82	0,0001	1,34
CSE Alta	-0,09	0,03	-3,30	0,0010	1,35
M, Norte	0,10	0,02	6,18	<0,0001	1,66
M, Centro	0,03	0,01	2,27	0,0234	1,38
Rezago Escolar	2,25	0,63	3,59	0,0003	1,23
Carencia adscripción al sistema de salud	-1,16	0,22	-5,24	<0,0001	1,58
Carencia en jubilaciones	-0,91	0,27	-3,36	0,0008	1,24
Carencia en habitabilidad	0,59	0,14	4,33	<0,0001	1,73
Carencia en accesibilidad	-0,53	0,16	-3,25	0,0012	1,27
Carencia en apoyo y participación	1,43	0,22	6,43	<0,0001	1,22
Carencia en seguridad	-0,38	0,09	-4,36	<0,0001	1,78

Fuente. Elaborado por el Instituto Nacional de Estadísticas.

En cuanto a los supuestos asociados al modelo, se comenta que:

- Los valores para la inflación de la varianza se encuentran por debajo de 5, indicando falta de colinealidad entre las variables explicativas, lo que garantiza el supuesto de independencia.
- Las estimaciones de los coeficientes se realizan haciendo el ajuste de White, lo que permite tener estimaciones robustas garantizando el supuesto de homocedasticidad.

Así, una vez configurado el modelo de regresión lineal y, en base a las predicciones de la variable dependiente, es decir, la probabilidad de respuesta a nivel de manzana, se obtienen deciles de ajuste dentro de los cuales se redistribuye el peso de las unidades que no respondieron entre las que sí lo hicieron.

En consecuencia, para cada decil se calcula la siguiente razón de ajuste:

$$R_d = \frac{\sum_{j \in \varphi_{ij}} w_{rgij}^{t*}}{\sum_{j \in \varphi'_{ij}} w_{rgij}^{t*}}, \quad \forall d$$

Donde:

- $\varphi_{ij}$  : Conjunto de viviendas seleccionadas.
- $\varphi'_{ij}$  : Conjunto de viviendas que respondieron.
- $R_d$  : Razón de ajuste para el decil  $d$ .

Las razones de ajuste obtenidas para cada uno de los deciles, se pueden observar en la tabla 4.

**Tabla 4. Razón de ajuste según decil.**

Deciles	Razón Ajuste
1	2,63
2	1,70
3	1,66
4	1,54
5	1,45
6	1,35
7	1,29
8	1,29
9	1,26
10	1,18

Fuente. Elaborado por el Instituto Nacional de Estadísticas.

Así, el factor ajustado por no respuesta, queda definido por la siguiente expresión:

$$w_{rgij}^* = w_{rgij}^{t*} \cdot R_d$$

## I.7. Ponderador de Selección de Unidades de Tercera Etapa (Personas)

---

Obtenido el factor de viviendas ajustado por no respuesta, se procede a obtener el ponderador para la tercera etapa de selección, es decir, para las personas al interior de cada vivienda. Dicha selección se realiza por medio de un algoritmo programado en el dispositivo electrónico de levantamiento.

El procedimiento corresponde a la selección aleatoria de un integrante de la vivienda cuyas características obedecen a la definición de la población objetivo (personas de 15 años o más). Por lo tanto, se construye un listado dentro de cada vivienda, con todas las personas que cumplen con dicha característica, siendo seleccionada una de ellas en forma aleatoria simple. Así, la probabilidad de selección de la persona es:

$$p_q = \frac{1}{n_q}$$

Donde:

$p_q$  : Probabilidad de seleccionar a la persona  $q$ .  
 $n_q$  : Total de personas seleccionables en la vivienda.

Obtenido el ponderado de tercera etapa, el factor de expansión queda constituido a través de la siguiente fórmula:

$$w_{rgij}^{**} = w_{rgij}^* \cdot n_q$$

Donde:

$w_{rgij}^{**}$  : Ponderador correspondiente a la selección de tercera etapa.

## 1.8. Suavizamiento ponderador de selección de personas.

---

Al igual que en el factor de expansión ajustado por no elegibilidad, se realiza un suavizamiento al ponderador de selección de tercera etapa, con el objetivo de controlar la dispersión producida en algunas regiones.

La metodología utilizada es la misma que para el suavizamiento anterior, es decir, se busca truncar los factores de expansión que se ubican sobre el umbral de suavizamiento, distribuyendo la diferencia entre el resto de los factores.

El factor suavizado, se describe mediante la siguiente expresión<sup>3</sup>:

$$w_{rgij}^{t**} = \begin{cases} w_{rgij}^{**} \cdot \frac{(\sum_{g \in \Psi_r} \sum_{i \in \Theta_g} \sum_{j \in \Omega_i} w_{rgij}^{**} - \sum_{g \in \Psi_r} \sum_{i \in \Theta_g} \sum_{w_{rgij}^{**} > M_{ra}} M_{ra})}{\sum_{g \in \Psi_r} \sum_{i \in \Theta_g} \sum_{w_{rgij}^{**} \leq M_{ra}} w_{rgij}^{**}} & , w_{rgij}^{**} \leq M_{ra} \\ M_{ra} & , w_{rgij}^{**} > M_{ra} \end{cases}$$

Donde:

$w_{rgij}^{t**}$  : Ponderador de selección de tercera etapa suavizado.

$a$  : Punto de corte en el cual se minimiza el error cuadrático medio, del índice de participación cultural. En esta caso se prueba con los valores entre 4 y 10, resultando como corte más favorable  $a=5$ .

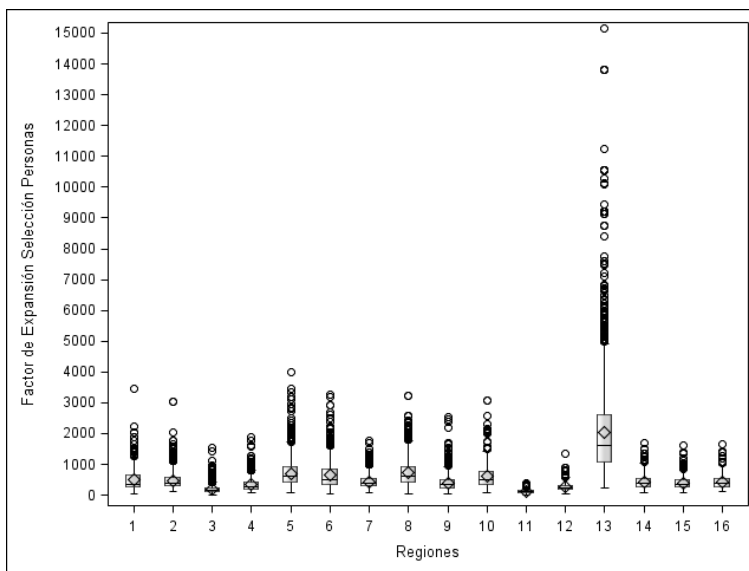
$M_{ra}$  : Producto entre el punto de corte y la media del factor de expansión a nivel de región.

Los gráficos 3 y 4 muestran la distribución de los valores del factor de expansión antes y después del suavizamiento.

---

<sup>3</sup> Más detalles en Anexo 1: "Metodología de suavizamiento de factores de expansión".

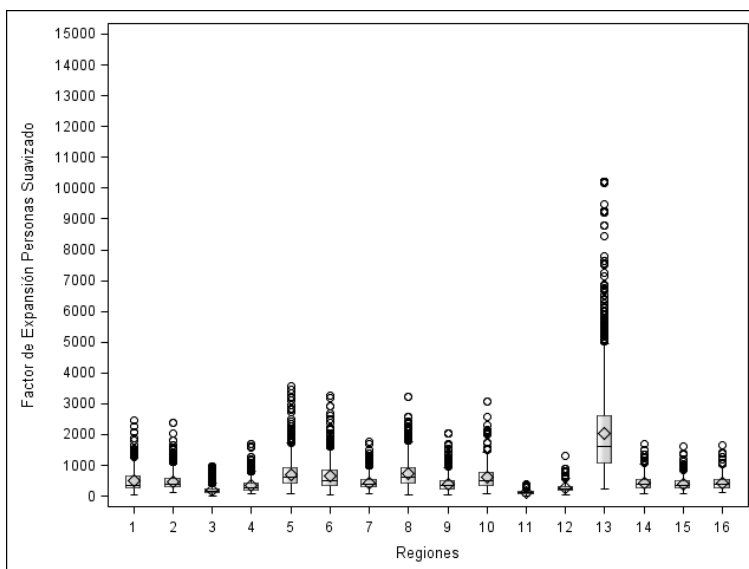
**Gráfico 3. Distribución del factor antes del suavizamiento.**



Fuente: Elaborado por el Instituto Nacional de Estadísticas.

En el gráfico 3, se aprecia presencia de factores de expansión atípicos que sobrepasan 10.000 unidades, específicamente en la Región Metropolitana, donde en la mayoría de los casos, estos valores no sobrepasan 5.000 unidades.

**Gráfico 4. Distribución del factor después del suavizamiento.**



Fuente: Elaborado por el Instituto Nacional de Estadísticas.

Luego del suavizamiento, se trunca el 0,2% de los factores de expansión, ocasionando que no superen las 11.000 unidades.

### I.9. Ajuste para calibrar al total poblacional por región, sexo y tramo etario.

Este ajuste al factor de expansión tiene por objetivo calibrar las estimaciones de la muestra levantada a los stocks poblacionales para el área urbana por región, sexo y tramo etario. Para esto, es necesario contar con las proyecciones de población correspondientes al mes central del periodo de levantamiento de la encuesta, que son obtenidas en base a la información del Censo 2002 con actualización del componente de crecimiento demográfico (natalidad, mortalidad y migración) mediante información de registros administrativos.

La tabla 5 muestra la población proyectada al 30 de octubre de 2017 por sexo según región para la población de 15 años y más.

**Tabla 5. Proyecciones de Población por región y sexo para personas de 15 años y más.**

Región	Total	Hombres	Mujeres
<b>Total</b>	<b>15.843.567</b>	<b>7.756.281</b>	<b>8.087.286</b>
I Tarapacá	337.800	169.269	168.531
II de Antofagasta	604.113	310.257	293.856
III de Atacama	266.057	134.062	131.995
IV de Coquimbo	647.988	316.925	331.063
V de Valparaíso	1.728.917	846.235	882.682
VI de O'Higgins	672.628	335.514	337.114
VII del Maule	720.596	351.887	368.709
VIII del Biobío	1.467.572	715.540	752.032
IX de La Araucanía	698.632	334.612	364.020
X de Los Lagos	650.234	325.614	324.620
XI de Aisén	96.818	49.375	47.443
XII de Magallanes y La Antártica	151.236	76.437	74.799
XIII Metropolitana	7.060.881	3.437.169	3.623.712
XIV de Los Ríos	264.472	127.658	136.814
XV de Arica y Parinacota	149.850	68.685	81.165
XVI de Ñuble	325.773	157.042	168.731

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

Para este ajuste, se estima, a través de la muestra, el total de personas por región, sexo y tramo etario, utilizando el factor suavizado ( $w_{rgij}^{t**}$ ).

Luego, se calcula la razón entre las proyecciones y la estimación a partir del levantamiento, según la siguiente fórmula:

$$\hat{R}_{rst} = \frac{P_{rst}}{\hat{P}_{rst}}$$

Donde:

$P_{rst}$  : Proyección de población por región, sexo y tramo etario<sup>4</sup>, de personas de 15 años y más.  
 $\hat{P}_{rst}$  : Estimación de población por región, sexo y tramo etario, de personas de 15 años y más.

con

$$\hat{P}_{rst} = \sum_{rgij} w_{rgij}^{t**} \cdot p_{rgij}$$

Siendo:

$p_{rgij}$  : Total de personas en la vivienda  $j$  en la manzana  $i$  de grupo tamaño  $g$  dentro de la región  $r$ , por sexo y tramo etario.

Luego del cálculo de la razón, el factor de expansión calibrado a las proyecciones de población se expresa cómo:

$$w_{rgij}^c = w_{rgij}^{t**} \cdot \hat{R}_{rst}$$

---

<sup>4</sup> Se consideraron cuatro tramos etarios para las 15 regiones existentes al momento del levantamiento de la encuesta (15-29, 30-44, 45-59, 60 y más); y un tramo etario para la nueva Región de Ñuble (15 años y más).

## II. Estimadores de precisión

### II.1. Estimación de la varianza, cálculo de estratos y conglomerados de varianza.

---

La Encuesta Nacional de Participación Cultural 2017, como ya fue señalado, contempla un diseño muestral complejo probabilístico, estratificado y trietápico. En general, cuanto más complejo es el diseño muestral bajo el cual se implementa una encuesta, más compleja se vuelve la forma de determinar los errores muestrales. Tanto así, que no existen fórmulas exactas y/o explícitas para esto. Sin embargo, paquetes estadísticos en software especializados, facilitan los cálculos a través de aproximaciones realizadas mediante distintos modelos o métodos de estimación, para lo cual se debe identificar las variables que definen el diseño muestral (estratos, conglomerados) y el factor de expansión apropiado (considerando todos los ajustes pertinentes).

En ocasiones pueden existir algunas dificultades en la implementación de la estimación de los errores mediante un paquete estadístico, originadas por las características del diseño muestral, por ejemplo: más de una fase de muestreo; muestreo multietápico de las unidades muestrales, selección de unidades sin reemplazo, estratos de muestreo con sólo una unidad primaria con unidades elegibles, variabilidad de los tamaños de los conglomerados, etc.

En esta encuesta se observan principalmente tres dificultades:

- Diseño muestral complejo.
- Existencia de estratos de muestreo que poseen sólo una manzana (unidad de primera etapa o conglomerado).
- El número de viviendas (unidades de segunda etapa) que responde en cada conglomerado es desigual y muy variable.

A fin de minimizar los problemas señalados anteriormente y, siguiendo las recomendaciones internacionales<sup>5</sup>, los errores son estimados a partir de modelos que buscan dar cuenta, lo más fielmente posible del diseño muestral. Para ello se agruparon estratos y conglomerados a fin de que estos nuevos pseudo-estratos y pseudo-conglomerados, garanticen la estimación de varianzas en cada nuevo estrato

---

<sup>5</sup> Ver Capítulo 15.5 en Valliant *et al.* (2013).



y, de ésta forma no subestimar los errores. A continuación, se detallan los procedimientos y criterios utilizados en la creación de dichas variables.

### **II.1.1. Creación de pseudos-estratos.**

Los pseudo estratos o varstrat, son reagrupaciones de los estratos de muestreo originales cuyo objetivo es garantizar que contengan al menos dos conglomerados. Estos varstrat son agrupados de acuerdo a reglas de ordenamiento jerárquico de variables geográficas o de tamaño.

Los pseudo-estratos o varstrat son construidos con el objetivo de corregir los problemas generados por la existencia de estratos con un solo conglomerado (estratos unitarios), que trae como consecuencia subestimar la varianza de cualquier variable de interés.

### **II.1.2. Creación de pseudo-conglomerados.**

Por su parte, los pseudo conglomerados o varunit buscan estabilizar las estimaciones de varianza al interior de los conglomerados, ya que suele presentar mayor dispersión cuando se cuenta con conglomerados con baja cantidad de unidades muestrales.

Por consiguiente, se reagrupan los conglomerados originales en varunit, considerando que contengan entre 30 y 50 viviendas. Sin embargo, durante el proceso de reagrupación se dieron casos en que se tuvo que flexibilizar el mínimo de unidades, quedando este en 20.

Con esto, se crean 68 varstrat y 321 varunit, los cuales se distribuyen a nivel regional como se muestra en la tabla 6.

**Tabla 6. Total de Estratos, conglomerados, varstrat y varunit según región.**

Región	Estratos Nacional de Participación Cultural	Varstrat	Conglomerados (Manzanas)	Varunit
<b>Total</b>	<b>79</b>	<b>68</b>	<b>1.941</b>	<b>321</b>
I Tarapacá	5	4	51	10
II de Antofagasta	5	5	78	16
III de Atacama	5	4	152	23
IV de Coquimbo	5	4	123	21
V de Valparaíso	5	5	239	38
VI de O'Higgins	5	5	86	14
VII del Maule	5	5	123	22
VIII del Biobío	5	5	166	29
IX de La Araucanía	5	4	160	23
X de Los Lagos	5	4	75	15
XI de Aisén	4	3	78	11
XII de Magallanes y La Antártica	5	3	87	9
XIII Metropolitana	5	5	352	60
XIV de Los Ríos	5	4	54	10
XV de Arica y Parinacota	5	4	61	11
XVI de Ñuble	5	4	56	9

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

## II.2. Representatividad de las estimaciones.

Para garantizar la representatividad de las estimaciones se pueden tomar como referencia dos indicadores: el coeficiente de variación y el número de pseudo conglomerados con información para la variable de interés.

En cuanto al coeficiente de variación se recomiendan los siguientes rangos:

- hasta un 5% se considera que la estimación es precisa.
- entre un 5% y un 15% se considera una buena estimación.
- entre un 15% y un 20% se considera una estimación aceptable.
- sobre el 20% la estimación se presenta como poco precisa y por ende no recomendable.

En cuanto al número de conglomerado, según la literatura de muestreo, se recomienda entre 30 y 60 conglomerados como mínimo, para con esto garantizar consistente de los estimadores de precisión.

Por consiguiente, el cruce entre el número de conglomerados con información de la variable de interés y el coeficiente de variación efectivo, al nivel de desagregación evaluado, definirá el uso que se le puede dar a las estimaciones, el cual podrá ser el siguiente:

- **Estimador confiable:** el estimador se puede utilizar según los objetivos propuestos.
- **Uso descriptivo:** se recomienda que el estimador se utilice para estadística descriptiva. Si se desea utilizar para otro tipo de estadística, se recomienda agruparlo a un nivel superior de la clasificación de consumo.
- **Estimador poco confiable:** no se recomienda utilizar el estimador, sino que debe agruparse a un nivel superior de la clasificación.

Los cruces que se proponen son los que se observan en la tabla siguiente:

**Tabla 7. Criterio de representatividad**

Nivel del coeficiente de variación efectivo	Cantidad de conglomerados con información en la variable de interés	
	Baja	Alta
Mayor a 20%	Agrupar	Agrupar
Entre 15% y 20%	Agrupar	Descriptiva o nota metodológica
Entre 5% y 15%	Descriptiva o nota metodológica	Estimador confiable
Entre 0% y 5%	Estimador confiable	Estimador confiable

Elaborado por el Instituto Nacional de Estadísticas.

### II.3. Estimación de índice de participación cultural y errores efectivos.

A efectos del cálculo de los estimadores de precisión es importante destacar que todas las estimaciones deben realizarse bajo módulos de muestras complejas, que se caracterizan por considerar el diseño muestral asociado al estudio<sup>6</sup>, de lo contrario se estaría realizando cálculos bajo muestreo aleatorio simple, por lo que se podría incurrir en subestimaciones de las métricas de precisión.

En la tabla 8, se observa la estimación de la variable de interés (índice de participación cultural) y sus principales indicadores de precisión.

**Tabla 8. Estimación índice de participación cultural y errores efectivos.**

Región	índice de participación cultural	Error estándar	Coefficiente de variación
<b>Total</b>	<b>23,78%</b>	<b>0,01</b>	<b>3,87%</b>
Arica y Parinacota	30,77%	0,04	13,47%
Tarapacá	11,16%	0,02	18,14%
Antofagasta	7,25%	0,02	25,57%
Atacama	28,01%	0,02	7,59%
Coquimbo	16,04%	0,03	17,05%
Valparaíso	23,13%	0,02	7,69%
Metropolitana	28,54%	0,02	6,09%
O'Higgins	26,69%	0,02	8,64%
Maule	14,51%	0,02	15,26%
Biobío	25,48%	0,02	9,32%
La Araucanía	15,92%	0,02	13,39%
Los Ríos	28,45%	0,04	15,09%
Los Lagos	13,54%	0,04	26,98%
Aisén	20,33%	0,05	23,84%
Magallanes	27,39%	0,05	19,70%
Ñuble	13,86%	0,03	21,37%

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

<sup>6</sup> En el caso de la Encuesta Nacional de Participación Cultural 2017 para definir el diseño complejo se deben tomar en cuentas los estratos (varstrat), conglomerados (varunit) y el factor de expansión.

### III. ANEXOS

#### III.1. Metodología de suavizamiento de factores de expansión.

---

Debido a las etapas de selección, los ponderadores iniciales de selección a medida que se van ajustando, aumentan significativamente su variabilidad. Como medida de mitigación se realiza un procedimiento que suaviza dichos ponderadores. Para ello, se requiere como insumo la principal variable de interés a medir, que permita identificar el umbral de suavizamiento a través de la minimización de su error cuadrático medio.

En este contexto, se calculó una variable de interés denominada “índice de participación cultural”, la cual considera la asistencia o participación a una serie de prácticas culturales consultadas en la Encuesta Nacional de Participación Cultural 2017 (éstas prácticas se describen en la tabla 6).

El indicador construido es de carácter aditivo y se obtiene contabilizando el total de prácticas culturales a las que asistió cada ciudadano. Se miden 26 actividades del ámbito cultural a las que pueden asistir o participar los entrevistados.

**Tabla 9. Dimensiones incluidas en el índice de participación cultural.**

Variable	Código BBDD	Pregunta
Artes visuales	c9a	9.a. Durante los últimos 12 meses, ¿ha visitado alguna exposición de arte?
Museos	c13a	13.a. Durante los últimos 12 meses, ¿ha ido a algún museo?
Compra de artesanía	c10b	10.b. En su vida, ¿ha comprado algún objeto elaborado por un artesano?
Teatro	c1a	1.a. Durante los últimos 12 meses, ¿ha ido a ver alguna obra de teatro?
Danza	c2a	2.a. Durante los últimos 12 meses, ¿ha asistido a un espectáculo de danza?
Circo	c11a	11.a. Durante los últimos 12 meses, ¿ha ido al circo?
Espectáculos en vivo en el espacio público	c12a1	12.a.1. Malabarismo, payasos u otras artes circenses.
	c12a2	12.a.2. Mimos
	c12a3	12.a.3. Espectáculos de teatro
	c12a4	12.a.4. Músicos o cantantes
	c12a5	12.a.5. Bailes o coreografías (K-pop, Tinku, comparsa, tango, etc.)
	c12a6	12.a.6. Títeres o marionetas
	c12a7	12.a.7. Cuenta cuentos
	c12a8	12.a.8. Proyecciones en edificios (mapping)
	c12a9	12.a.9. Graffitis
	c12a10	12.a.10. Murales
Fiestas religiosas	c17w	17.w. Durante los últimos 12 meses, ¿ha participado en fiestas de tipo religioso?
Lectura	c16c	16.c. Sin contar los libros que ha leído por estudio o trabajo, ¿cuántos libros ha leído en los últimos 12 meses?
Bibliotecas	c15a	15.a. Durante los últimos 12 meses, ¿ha ido a alguna biblioteca?
Sitios naturales	c17d	17.d. Durante los últimos 12 meses, ¿ha visitado parques nacionales, santuarios...?
Cine	c7a	7.a. Durante los últimos 12 meses, ¿ha ido a ver una película en una sala de cine?
Películas	c8d	8.d. Durante los últimos 12 meses, ¿ha visto películas chilenas?
Internet	c20a	20.a. ¿Con qué frecuencia utiliza Internet?
Videojuegos	c21a	21.a. ¿Con qué frecuencia usted juega videojuegos?
Música	C6a	6.a. ¿Con qué frecuencia escucha música?
Conciertos	c5a	5.a. Durante los últimos 12 meses, ¿ha ido a ver algún recital, concierto o espectáculo?

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

La tabla 10 muestra la cantidad de prácticas a las cuales asistió cada individuo y la frecuencia de cada una.

**Tabla 10. Distribución de frecuencia del Índice de participación cultural.**

Número actividades en que participa	Frecuencia	Porcentaje	Frecuencia acumulada	Porcentaje acumulado
0	153	1,26	153	1,26
1	525	4,32	678	5,58
2	649	5,34	1.327	10,92
3	638	5,25	1.965	16,17
4	750	6,17	2.715	22,34
5	803	6,61	3.518	28,95
6	829	6,82	4.347	35,77
7	887	7,3	5.234	43,07
8	870	7,16	6.104	50,23
9	858	7,06	6.962	57,3
10	853	7,02	7.815	64,32
11	810	6,67	8.625	70,98
12	788	6,49	9.413	77,47
13	662	5,45	10.075	82,91
14	575	4,73	10.650	87,65
15	482	3,97	11.132	91,61
16	368	3,03	11.500	94,64
17	249	2,05	11.749	96,69
18	172	1,42	11.921	98,11
19	102	0,84	12.023	98,95
20	67	0,55	12.090	99,5
21	32	0,26	12.122	99,76
22	24	0,2	12.146	99,96
23	5	0,04	12.151	100

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

Los resultados de la encuesta muestran que el máximo puntaje obtenido es de 23 actividades, siendo el mínimo cero. La media de los puntajes obtenidos es 9,4 y la mediana de 9,0 puntos.

La tabla 11 muestra estadísticas descriptivas del índice de participación cultural.

**Tabla 11. Estadísticas descriptivas para el índice de participación cultural.**

Media	Mediana	Moda	Rango
9,45	9,00	7,00	23

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

Por otro lado, al realizar un test de fiabilidad del indicador de participación cultural, éste muestra un Alfa de Cronbach<sup>7</sup> de 0,82 lo que muestra que el indicador tiene buena fiabilidad.

**Tabla 12. Test fiabilidad construcción de indicador.**

Coeficiente Alfa de Cronbach	
Variables	Alfa
Sin tratar	0,82
Estandarizada	0,82

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

A partir de este análisis, se establecen 3 niveles de participación (bajo, medio y alto), donde se distribuyen las frecuencias de los entrevistados en función de su nivel de participación en actividades de tipo cultural, lo cual se muestra en la Tabla 13.

**Tabla 13. Niveles de participación en actividades culturales.**

Índice de participación cultural	Frecuencia	Porcentaje	Frecuencia acumulada	Porcentaje acumulado
Bajo	5.234	43,07	5.234	43,07
Medio	2.581	21,24	7.815	64,32
Alto	4.336	35,68	12.151	100,00

Fuente: Elaborado por el Instituto Nacional de Estadísticas.

Luego, para construir el indicador que servirá como variable de interés en el suavizamiento, se agrupa la variable de participación cultural en dos categorías, según la variable indicatriz " $I_{pc}$ " calculándose la prevalencia de dicha variable para cada informante, donde  $I_{pc}$  se define como:

$$I_{pc} = \begin{cases} 1, & \text{si participación es alta} \\ 0, & \text{en otro caso} \end{cases}$$

<sup>7</sup> El Alfa de Cronbach es un coeficiente que busca medir la fiabilidad de una escala de medida para indicadores aditivos y se considera un valor confiable cuando este se encuentra sobre 0,70, siendo el valor máximo 1.



Una vez obtenida la estimación, se procedió a realizar el suavizamiento del factor a partir de la implementación de las siguientes etapas:

1. Se inspeccionó la existencia de valores extremos de éste, al interior de cada una de las regiones.
2. Se determinaron puntos de corte a partir de los cuales realizar el suavizamiento.
3. Se suavizaron los valores extremos identificados.
4. Se estimó el error cuadrático medio (ECM) para el indicador de participación cultural para los distintos puntos de corte.
5. Se eligió la opción de corte que minimizaba dicho error.

Considerando lo anterior, se analizaron 7 puntos de cortes distintos, definidos como sigue:

$$M_{ra} = a * \bar{F}_r, \text{ para } a = 4, 5, 6, 7, 8, 9, 10,$$

Donde:

$M_{ra}$  : Producto entre el punto de corte y la media del factor de expansión a nivel de región

$a$  : Punto de corte en el cual se prueba el suavizamiento.

$\bar{F}_r$  : Media del factor de expansión en la región  $r$ .

$$\text{siendo } \bar{F}_r = \frac{\sum_{g \in \Psi_r} \sum_{i \in \Theta_g} \sum_{j \in \Omega_i} w_{rgij}^{**}}{n_r}$$

Donde:

$\Omega_i$  : Es el conjunto de viviendas seleccionadas en la manzana  $i$ .

$\Theta_g$  : Es el conjunto de manzanas seleccionadas en el grupo de tamaño  $g$ .

$\Psi_r$  : Es el conjunto de grupos de tamaño de la región  $r$ .

$n_r$  : Es el total de personas seleccionadas en la región  $r$ .

Luego, para realizar el suavizamiento se procedió a trincar aquellos ponderadores identificados como valores extremos de la siguiente forma,

$$w_{rgij}^{**t} = \begin{cases} w_{rgij}^{**} & \text{si } w_{rgij}^{**} \leq M_{ra} \\ M_{ra} & \text{si } w_{rgij}^{**} > M_{ra} \end{cases}$$

Al sumar todos los valores  $w_{rgij}^{**t}$ , se obtiene un total de unidades estimadas inferior que al sumar los ponderadores base, por lo tanto se debe distribuir la diferencia faltante en el resto de los ponderadores que no fueron truncados. Los pesos fueron distribuidos al interior de cada región  $r$  de la siguiente forma:

$$w_{rgij}^{**t} = \begin{cases} w_{rgij}^{**} \cdot \frac{\left(\sum_{g \in \Psi_r} \sum_{i \in \Theta_g} \sum_{j \in \Omega_i} w_{rgij}^{**} - \sum_{g \in \Psi_r} \sum_{i \in \Theta_g} \sum_{w_{rgij}^{**} > M_{ra}} M_{ra}\right)}{\sum_{g \in \Psi_r} \sum_{i \in \Theta_g} \sum_{w_{rgij}^{**} \leq M_{ra}} w_{hij}^{**}} & , w_{rgij}^{**} \leq M_{ra} \\ M_{ra} & , w_{rgij}^{**} > M_{ra} \end{cases}$$

Donde

$w_{rgij}^{**t}$  : es el factor suavizado.

Esto es, aquellos ponderadores identificados como valores extremos son truncados al valor máximo establecido ( $M_{ra}$ ), mientras que el peso “sobrante” de los ponderadores truncados es distribuido sobre el resto de los ponderadores.

Luego, para determinar el punto de corte donde se realiza finalmente el suavizamiento, se calculó un estadígrafo que diera cuenta del sesgo y de la variabilidad. Para esto se obtuvo el Error Cuadrático Medio (ECM) asociado a la variable de interés. Como esta variable tiene relación con el nivel de participación en actividades culturales, se calculó dicho indicador y su desviación típica por región. De esa forma el sesgo y el ECM, respectivamente, se calculan como:

$$sesgo(\hat{P}_{ra}) = P_r - \hat{P}_{ra}$$

$$ECM(\hat{P}_{ra}) = Sesgo^2(\hat{P}_{ra}) + Var(\hat{P}_{ra})$$

Siendo,

$r$  : Subíndice de la región. Toma valores entre 1 y 16 según numeración de las regiones del país.

$P_r$  : La proporción de personas en la región  $r$  con alto porcentaje de participación obtenido con el factor de expansión sin truncar  $w_{rgij}^{**}$

$\hat{P}_{ra}$  : La proporción de personas en la región  $r$  con alto porcentaje de participación con el factor suavizado en el umbral  $a$ .

Por su parte, la tabla 12 muestra la frecuencia de varunit según número de viviendas que la componen.

**Tabla 14. Frecuencia de varunit según número de viviendas que lo componen.**

N° de viviendas que responde	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
<b>Total</b>	<b>12.151</b>	<b>100,00</b>	<b>100,00</b>	
20	40	0,30	0,30	0,30
23	23	0,20	0,20	0,50
25	50	0,40	0,40	0,90
26	52	0,40	0,40	1,40
27	27	0,20	0,20	1,60
28	28	0,20	0,20	1,80
29	58	0,50	0,50	2,30
30	360	3,00	3,00	5,30
31	155	1,30	1,30	6,50
32	672	5,50	5,50	12,10
33	495	4,10	4,10	16,10
34	340	2,80	2,80	18,90
35	1.015	8,40	8,40	27,30
36	936	7,70	7,70	35,00
37	888	7,30	7,30	42,30
38	1.102	9,10	9,10	51,40
39	1.014	8,30	8,30	59,70
40	1.560	12,80	12,80	72,50
41	328	2,70	2,70	75,20
42	378	3,10	3,10	78,40
43	301	2,50	2,50	80,80
44	484	4,00	4,00	84,80
45	315	2,60	2,60	87,40
46	368	3,00	3,00	90,40
47	188	1,50	1,50	92,00
48	480	4,00	4,00	95,90
49	294	2,40	2,40	98,40
50	200	1,60	1,60	100,00

Fuente: Elaborado por el Instituto Nacional de Estadísticas.