

Documentos de trabajo

Análisis del efecto del COVID-19
sobre el diseño muestral y las
estimaciones de la ENE en EFM 2020

Autores:

Ivan Touron

Jonathan Paredes

Marly Olivares

Miguel Guerrero

Pedro Aravena

Mayo 2020



INSTITUTO NACIONAL DE ESTADÍSTICAS

Morandé 801, Santiago de Chile

Teléfono: 56 232461010

Correo: ine@ine.cl

Facebook: @ChileINE

Twitter: @INE_Chile

Departamento de Metodologías e Innovación Estadística

Denisse López

Subdepartamento de Diseño de Marcos y Muestras

Ivan Touron

Jonathan Paredes

Marly Olivares

Miguel Guerrero

Pedro Aravena

Ximena Espinosa

Los autores agradecen el valioso aporte a los expertos de la Comisión Económica para América Latina y el Caribe (Cepal) y de la Organización Internacional del Trabajo (OIT), cuyos comentarios y orientaciones permitieron desarrollar y enriquecer el presente trabajo. De igual manera, extendemos un agradecimiento especial al equipo del Departamento de Estadísticas del Trabajo, específicamente a Eduardo Toro y José Aranguiz quienes se encargaron de facilitar los insumos necesarios y a Álvaro Krause cuyos análisis ayudaron a fortalecer los resultados de este trabajo.

Los Documentos de Trabajo del INE están dirigidos a investigadores, académicos, estudiantes y público en general, y tienen como objetivo proporcionar un análisis exhaustivo sobre aspectos conceptuales, analíticos y metodológicos claves de los productos estadísticos que elabora la institución y, de esta forma, contribuir al intercambio de ideas entre los distintos componentes del Sistema Estadístico Nacional.

Las interpretaciones y opiniones que se expresan en los Documentos de Trabajo pertenecen en forma exclusiva a los autores y colaboradores y no reflejan necesariamente el punto de vista oficial del INE ni de la institución a la que pertenecen los colaboradores de los documentos.

El uso de un lenguaje que no discrimine ni marque diferencias entre hombres y mujeres ha sido una preocupación en la elaboración de este documento. Sin embargo, y con el fin de evitar la sobrecarga gráfica que supondría utilizar en castellano “o/a” para marcar la existencia de ambos sexos, se ha optado por utilizar -en la mayor parte de los casos- el masculino genérico, en el entendido de que todas las menciones en tal género representan siempre a hombres y mujeres, abarcando claramente ambos sexos.

Índice

Resumen	6
Abstract.....	6
1. Contexto	7
2. Primer análisis: Ejercicio sobre el trimestre DEF 2020 con base en las tasas de logro de EFM 2020.....	10
3. Segundo análisis: Tratamiento sobre los factores de expansión	20
3.1. Método de ajuste para la no respuesta de unidades provenientes del MMM2016 y MMS2016.....	21
3.2. Método evaluado de ajuste para la no respuesta.....	23
3.3. Método de suavizamiento de los factores de expansión.....	24
3.3.1. Método de contracción a la media (CM)	25
3.3.2. Método R-K	26
3.3.3. Método mixto	28
3.3.4. Parametrización del suavizamiento para la ENE.	29
4. Método a evaluar y resultados	30
5. Disposiciones sobre la calibración de los factores de expansión	35
6. Conclusiones	36
7. Referencias.....	37
Anexos	38

Índice de tablas

Tabla 1. Tasas de logro para el mes de marzo de la muestra del MMV2017	8
Tabla 2. Tasas de logro para el mes de marzo de la muestra del MMM2016 y MMS2016	9
Tabla 3. Comparación de tasas de logro DEF 2020 y DEF' 2020	13
Tabla 4. Comparación de estimaciones y errores DEF 2020 y DEF' 2020.....	14
Tabla 5. Estadísticas descriptivas de los factores de expansión por método de ajuste a nivel nacional	31

Índice de gráficos

Gráfico 1. Estimación de la tasa de desocupación a nivel nacional.....	15
Gráfico 2. Estimación de la tasa de desocupación a nivel regional (zona norte)	16
Gráfico 3. Estimación de la tasa de desocupación a nivel regional (zona centro 1).....	17
Gráfico 4. Estimación de la tasa de desocupación a nivel regional (zona centro 2)	18
Gráfico 5. Estimación de la tasa de desocupación a nivel regional (zona sur).....	19
Gráfico 6. Distribución de los factores de expansión según tipo de ajuste, a nivel nacional	31
Gráfico 7. Estimación de la tasa de desocupación según tipo de ajuste del FE, a nivel nacional	33
Gráfico 8. Estimación de la tasa de desocupación según tipo de ajuste del FE, a nivel regional (zona norte)	33
Gráfico 9. Estimación de la tasa de desocupación según tipo de ajuste del FE, a nivel regional (zona centro 1)	34
Gráfico 10. Estimación de la tasa de desocupación según tipo de ajuste del FE, a nivel regional (zona centro 2)	34
Gráfico 11. Estimación de la tasa de desocupación según tipo de ajuste del FE, a nivel regional (zona sur).....	35

Índice de figuras

Figura 1. Recomendaciones para tres escenarios de recolección de información en marzo de 2020.....	10
Figura 2. Esquema de simulación	11
Figura 3. Esquema de cómo se configuran las sumas a repartir	27
Figura 4. Aplicación de método mixto CM y R-K.....	29

Análisis del efecto del COVID-19 sobre el diseño muestral y las estimaciones de la ENE en EFM 2020

Resumen

El presente documento presenta los análisis realizados por SDMM para cuantificar los efectos que ha tenido la pandemia de Coronavirus COVID-19 sobre el diseño muestral y las estimaciones de la Encuesta Nacional de Empleo (ENE), ya que, actuando en concordancia con las medidas de distanciamiento social decretadas por las autoridades de Gobierno para afrontar la emergencia sanitaria, a contar de la segunda quincena de marzo 2020 se ha suspendido el habitual levantamiento con modalidad presencial de la encuesta, lo que ha acrecentado la pérdida en los tamaños muestrales respecto a lo esperado.

En particular, este documento presenta simulaciones de los efectos de la pérdida muestral con base en lo ocurrido con la no respuesta en marzo 2020, cuyos resultados permiten aproximar el sesgo a incluir en las estimaciones, así como determinar, bajo este nuevo escenario, qué niveles de desagregación de las estimaciones cumplen con el estándar de calidad mínimo requerido. Por otro lado, para mitigar dichos efectos, se proponen distintos tipos de ajuste al factor de expansión, específicamente referidos al ajuste por no respuesta y al suavizamiento de éstos.

Abstract

This document presents analyses carried out by SDMM to quantify the effects that the Coronavirus COVID-19 pandemic has had on the sample design and estimates of the National Employment Survey (ENE). Acting in accordance with the social distancing measures decreed by the Government authorities to face the health emergency, since the second half of March 2020, the usual data collection with face-to-face survey mode has been suspended, which has increased the loss in the sample sizes compared to what was expected.

In particular, this document presents simulations of the effects of the sample size loss based on what happened with the non-response in March 2020, whose results allow us to approximate the bias to be included in the estimates, as well as to determine, under this new scenario, what levels of breakdown in the estimates meet the required minimum quality standard. In addition, different types of adjustment to the sampling weight are proposed to mitigate these effects, specifically referring to the adjustment for non-response and their smoothing.

Palabras clave: COVID-19, diseño muestral, estimaciones, pérdida muestral, sesgo, factor de expansión, suavizamiento.

1. Contexto

Ante la confirmación de la fase 4 de la pandemia de COVID-19 en nuestro país, y en conformidad a las direcciones entregadas por las autoridades de Gobierno, el Instituto Nacional de Estadísticas de Chile (INE) ha tomado, desde marzo de 2020, una serie de medidas que permiten dar continuidad a la producción estadística, en particular de la Encuesta Nacional de Empleo (ENE). Entre las medidas se encuentran la suspensión temporal, a contar de la segunda quincena de marzo de 2020, de la recolección presencial de viviendas, establecimientos u otras fuentes informantes, aplicando en su lugar, la metodología de recolección telefónica y/o web.

Por su parte, durante el 2019 se definió un proceso de renovación de las unidades muestrales de la ENE, planificado entre los meses de enero a junio de 2020. En esa línea, y con apoyo de un grupo asesor experto, se definió hacer una incorporación progresiva de la muestra actualizada, proveniente del Marco Muestral de Viviendas (MMV2017) y la desincorporación de las viviendas provenientes del Marco Muestral basado en información de PreCenso 2016 (MMM2016 y MMS2016) y Censo 2002 (MMS2002)¹.

Sin embargo, tal proceso de actualización solo pudo ser realizado de forma efectiva hasta el mes de febrero de 2020 pues, en el escenario antes expuesto, durante marzo no fue posible llevar a cabo el levantamiento de la muestra ENE en la forma que había sido planificada, pues si bien la muestra enviada al trabajo de campo responde a lo definido en el diseño, a partir de las medidas tomadas a causa del COVID-19 la suspensión del levantamiento presencial conlleva a que las tasas de logro se encuentren por debajo de lo esperado, afectando los resultados del trimestre móvil EFM 2020². Por tanto, la estrategia de levantamiento realizada en marzo trae consigo la suspensión del levantamiento presencial a partir de la segunda quincena, a cambio del levantamiento telefónico o web en aquellas viviendas pertenecientes a los marcos antiguos (MMM2016 y MMS2016), donde pudo utilizarse la información de contacto que se disponía de éstas para ubicar los informantes. Sin embargo, la incorporación de las unidades que provenían del marco nuevo (MMV2017) tuvo que ser detenida dado que no se dispone de información de contacto de éstas al ser la primera entrevista que se realizaría.

¹Para mayor información consultar el documento metodológico disponible en: https://www.ine.cl/docs/default-source/ocupacion-y-desocupacion/metodologia/espanol/metodolog%C3%ADa-encuesta-nacional-de-empleo-ene-2020.pdf?sfvrsn=793380e9_8.

² EFM 2020 corresponde al trimestre móvil conformado por los meses de enero, febrero y marzo de 2020.

En este sentido, la evaluación de los efectos del COVID-19 sobre las estimaciones de la encuesta, parte por una revisión de las tasas de logro de la submuestra del mes de marzo 2020, que como se comentó anteriormente ha sufrido una pérdida muestral sustancial, debido a los cambios en la forma de levantamiento de la encuesta.

La Tabla 1 contiene las tasas de logro para el mes de marzo de la muestra del MMV2017, donde puede observarse que, tanto a nivel de unidades primarias de muestreo (UPM) como a nivel de viviendas, alcanzan 37,9% y 24,0% a nivel nacional respectivamente. En lo que respecta a las regiones la mayor tasa de logro tanto a nivel de UPM como de viviendas se observó en la región del Maule siendo estas 51,6% y 33,9% respectivamente. Este comportamiento se debe a que la muestra del MMV2017 corresponde a aperturas, es decir, a viviendas que participan por primera vez en la encuesta, y por experiencia, aún en situaciones normales de levantamiento, la tasa de logro en aperturas suele ser baja con respecto a la de las visitas sucesivas, comportamiento acentuado en gran medida, por las modificaciones que fue necesario implementar para el levantamiento de la muestra durante el mes de marzo, pues al ser aperturas no se disponía de información de contacto de las viviendas del marco nuevo.

Tabla 1. Tasas de logro para el mes de marzo de la muestra del MMV2017

Nivel	MMV2017			
	UPM a levantar	Tasa logro UPM	Viviendas a levantar	Tasa logro viviendas
Nacional	919	37,9%	9.028	24,0%
Arica y Parinacota	25	36,0%	250	21,2%
Tarapacá	17	47,1%	170	31,8%
Antofagasta	28	32,1%	280	22,5%
Atacama	39	43,6%	320	24,4%
Coquimbo	55	38,2%	550	24,7%
Valparaíso	102	40,2%	1.020	25,3%
Metropolitana	178	37,6%	1.780	20,3%
O´Higgins	52	38,5%	520	29,0%
Maule	64	51,6%	640	33,9%
Ñuble	42	38,1%	420	25,5%
Biobío	102	33,3%	1.020	23,8%
La Araucanía	51	29,4%	510	21,4%
Los Ríos	44	29,5%	440	17,3%
Los Lagos	68	41,2%	680	25,9%
Aysén	27	33,3%	224	16,1%
Magallanes	25	32,0%	204	21,6%

Fuente: Instituto Nacional de Estadísticas.

Por su parte, la Tabla 2 presenta las tasas de logro para el mes de marzo de la muestra del MMM2016 y MMS2016, donde puede apreciarse que las tasas tanto de UPM (manzanas en el caso del MMM2016 y secciones en el MMS 2016) como de viviendas, se vieron afectadas en menor medida por las modificaciones implementadas en el levantamiento, ya que, al no

ser viviendas de apertura, se disponía de información de contacto. A nivel nacional se logró 98,7% de las UPM y, a 73,3% de las viviendas. A nivel regional, solo en las regiones de Atacama, Valparaíso, Metropolitana y Los Lagos la tasa de logro a nivel de UPM fue menor a 100%, no obstante, dichas tasas fueron superiores a 90%. Con relación al logro a nivel de viviendas, en todas las regiones se logró más de 70%, con excepción de la región Metropolitana, donde la tasa de logro solo alcanzó 64,6%.

Tabla 2. Tasas de logro para el mes de marzo de la muestra del MMM2016 y MMS2016

Nivel	MMM2016 y MMS2016			
	UPM a levantar	Tasa logro UPM	Viviendas a levantar	Tasa logro viviendas
Nacional	745	98,7%	6.235	73,3%
Arica y Parinacota	34	100,0%	192	72,4%
Tarapacá	17	100,0%	170	76,5%
Antofagasta	23	100,0%	174	75,9%
Atacama	23	91,3%	157	70,1%
Coquimbo	35	100,0%	351	74,9%
Valparaíso	101	99,0%	835	77,2%
Metropolitana	148	95,9%	1.304	64,6%
O´Higgins	45	100,0%	380	76,1%
Maule	46	100,0%	401	78,6%
Ñuble	18	100,0%	175	83,4%
Biobío	88	100,0%	709	77,3%
La Araucanía	39	100,0%	338	71,9%
Los Ríos	22	100,0%	228	81,6%
Los Lagos	51	98,0%	498	70,1%
Aysén	27	100,0%	175	72,0%
Magallanes	28	100,0%	148	74,3%

Fuente: Instituto Nacional de Estadísticas.

Una vez obtenidas las tasa de logro y en el marco de las recomendaciones emanadas por la División de Estadística de la Comisión Económica para América Latina y el Caribe (Cepal)³, donde se proponen los pasos a seguir en función a la cobertura lograda durante la recolección de marzo (Figura 1), se procede a realizar los análisis y ajustes inherentes al nivel de cobertura alcanzado.

³ https://repositorio.cepal.org/bitstream/handle/11362/45372/4/2000274_es.pdf

Figura 1. Recomendaciones para tres escenarios de recolección de información en marzo de 2020



Fuente: https://repositorio.cepal.org/bitstream/handle/11362/45372/4/2000274_es.pdf

Dada la pérdida de unidades muestrales de la submuestra de marzo, la estimación del trimestre EFM 2020, debe ser abordada bajo el escenario de cobertura media, por ende, previo a la publicación, se realizan dos análisis: el primero, se basa en una simulación que permite cuantificar el sesgo que se puede generar debido a la ausencia de respuestas y busca proveer información para definir el nivel de desagregación hasta el cual es posible entregar cifras robustas; mientras que, el segundo análisis consiste en la revisión de los métodos de tratamiento que se realizan a los factores de expansión ante la pérdida de respuesta de las unidades y un eventual aumento en la magnitud de éstos, buscando determinar los ajustes necesarios para mitigar este efecto.

2. Primer análisis: Ejercicio sobre el trimestre DEF 2020 con base en las tasas de logro de EFM 2020

Con el objetivo de estimar los efectos de la pérdida muestral ocurrida en el mes de marzo, y tomando en consideración las recomendaciones emanadas por la División de Estadística de la Cepal, se obtuvieron muestras aleatorias del trimestre DEF 2020⁴, tomando como estructura la tasa de logro de marzo.

Específicamente, el ejercicio consistió en replicar en la submuestra del mes de febrero 2020, las tasas de logro obtenidas en la submuestra del mes de marzo, y a partir de esto, reconstruir

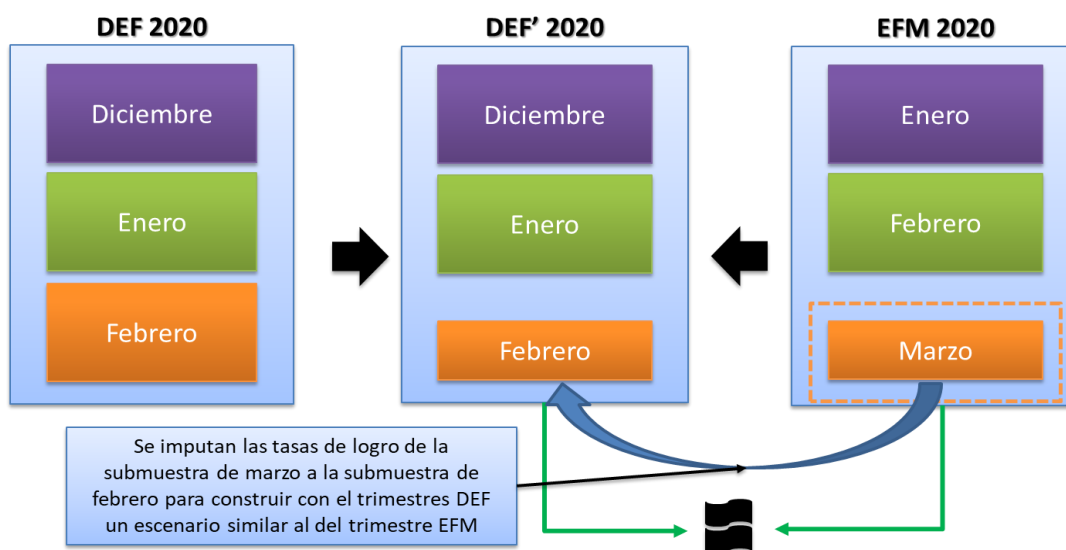
⁴ DEF 2020 corresponde al trimestre móvil conformado por los meses de diciembre 2019, enero y febrero de 2020.

un trimestre móvil denominado DEF' 2020, cuya tasa de logro sea similar a la del trimestre móvil EFM 2020.

Los resultados obtenidos a partir de estas simulaciones contribuyen a evaluar el sesgo como consecuencia de la pérdida de unidades muestrales, a la vez que permite definir los niveles de desagregación para los cuales la ENE puede generar indicadores robustos para el trimestre EFM 2020.

La Figura 2 ilustra el esquema de la simulación, mediante el cual se generaron 100 réplicas aleatorias denominadas DEF' 2020.

Figura 2. Esquema de simulación



Fuente: Instituto Nacional de Estadísticas.

Dado que el diseño de la ENE, es un diseño complejo, la imputación de las tasas de logro en febrero se realizó dentro de cada estrato en dos etapas. La primera consistió en imputar, al interior de cada estrato de muestreo, las tasas de logro de las UPM. En el caso de las UPM del MMV2017 y MMS2016, la imputación se realizó considerando que las UPM son homogéneas en cuanto a la cantidad de viviendas a seleccionar. En el caso de las UPM del MMM2016, la imputación de la tasa de logro fue realizada a nivel de estrato de muestreo, bajo el diseño anterior, y grupo de tamaño, permitiendo así, replicar de mejor forma la cobertura de marzo.

En algunos casos, la tasa de logro a nivel de UPM fue mayor en marzo que en febrero, por ende, en estos casos particulares, si con la tasa de logro imputada se obtenía como resultado un número de UPM superior al logrado en febrero, se seleccionaron para la simulación todas las UPM logradas en febrero.

La segunda etapa, consistió en replicar las tasas de logro de viviendas, para ello al interior de cada estrato de muestreo, primero se determinó el número de viviendas a seleccionar en el estrato de muestreo, en el caso de UPM del MMV2017 y MMS2016, o estrato de muestreo y grupo de tamaño, en el caso de las UPM del MMM2016, luego se dividió ese número entre el número de UPM definidas según la tasa de imputación, obteniendo así el número de viviendas a seleccionar dentro de cada UPM seleccionada en la primera etapa.

Dado que este procedimiento asigna un mismo número de viviendas a seleccionar dentro de cada UPM en un determinado estrato, sin considerar la cantidad de viviendas disponibles para seleccionar⁵ en la submuestra de febrero 2020 oficial dentro de esa UPM, era posible que en algunos casos el número de viviendas a seleccionar superara la cantidad de viviendas disponibles para seleccionar, en dichos casos, si el número de viviendas a seleccionar era superior al logrado en febrero dentro de la UPM, se seleccionan para la simulación todas las viviendas logradas en febrero dentro de la UPM.

Finalmente, en el ejercicio se generaron 100 réplicas de la submuestra de febrero 2020, con las tasas de logro imputadas, y posteriormente, en función a éstas, se construyen 100 muestras de DEF' 2020, manteniendo en su estado original las submuestras de diciembre y enero.

Por su parte, para evaluar la magnitud de los posibles sesgos introducidos en las estimaciones como consecuencia de la pérdida muestral, se analiza el comportamiento de las estimaciones para la tasa de desocupación de estas 100 simulaciones generadas del trimestre DEF' 2020 respecto a las del trimestre DEF 2020 oficial. Hay que considerar que adicionalmente se realizaron análisis de otros indicadores como los niveles de ocupados y ocupados informales.

Las Tabla 3 presenta las tasas de logro a nivel nacional y regional de DEF 2020 y los resultados obtenidos con las simulaciones de DEF' 2020. El porcentaje de logro tanto de UPM como de viviendas para DEF' 2020, se calcula en función al promedio de los tamaños de muestra simulados en las 100 réplicas. Como se puede observar, la tasa de logro simulada a nivel nacional fue en promedio 88,0% a nivel UPM y 66,8% a nivel de viviendas. A nivel

⁵ El número de viviendas disponibles para seleccionar y poder conformar la muestra de febrero simulada con las tasas de logro de marzo, depende del número de viviendas que hayan sido logradas en la submuestra de febrero 2020 oficial, ya que, para generar las estimaciones, necesitamos la información a nivel de personas.

regional, resalta que, todas las tasas de logro a nivel de UPM resultaron menores a 95,0% y a nivel de viviendas no llega a 80,0%.

Tabla 3. Comparación de tasas de logro DEF 2020 y DEF' 2020

Nivel	UPM teórica	Viviendas teóricas	DEF 2020		DEF' 2020	
			UPM % logro	Viviendas % logro	UPM % logro ⁶	Viviendas % logro ⁷
Nacional	4.876	42.982	99,5%	80,0%	88,0%	66,8%
Arica y Parinacota	169	1.253	100,0%	84,0%	90,5%	68,3%
Tarapacá	113	1.013	100,0%	87,5%	92,9%	75,9%
Antofagasta	157	1.238	100,0%	84,2%	87,9%	68,5%
Atacama	171	1.294	100,0%	78,2%	86,0%	63,4%
Coquimbo	264	2.490	100,0%	78,6%	87,9%	65,9%
Valparaíso	630	5.393	99,8%	84,0%	90,6%	72,1%
Metropolitana	953	8.805	97,9%	68,2%	86,6%	57,4%
O' Higgins	260	2.564	100,0%	83,8%	88,1%	71,1%
Maule	295	2.869	100,0%	83,1%	90,5%	70,9%
Ñuble	165	1.520	100,0%	87,6%	84,2%	69,5%
Biobío	555	4.834	99,8%	86,8%	87,6%	71,4%
La Araucanía	275	2.378	100,0%	78,7%	86,9%	64,0%
Los Ríos	195	1.857	100,0%	80,3%	84,1%	65,8%
Los Lagos	350	3.308	100,0%	82,3%	88,0%	68,4%
Aysén	171	1.177	99,4%	81,1%	88,9%	67,4%
Magallanes	153	989	99,3%	81,2%	88,2%	66,5%

Fuente: Instituto Nacional de Estadísticas.

La Tabla 4 por otra parte, presenta las estimaciones de la tasa de desocupación y sus correspondientes errores a nivel nacional y regional. El valor de la tasa de desocupación y del error absoluto que se muestran corresponden al promedio de lo obtenido en las 100 réplicas. El promedio de las estimaciones de la tasa de desocupación calculadas con DEF' 2020, arrojó resultados muy cercanos a la estimación oficial de DEF 2020, no obstante, como consecuencia de la pérdida muestral, se observa que en promedio el error absoluto en DEF' 2020 es mayor que en DEF 2020. Adicionalmente, se muestra la diferencia entre las estimaciones de la tasa de desocupación obtenidas en DEF 2020 y el promedio de las estimaciones de las 100 simulaciones de DEF' 2020, observándose que la mayor diferencia entre las tasas de desocupación estimadas es de 0,3% en las regiones de Antofagasta y Maule.

⁶ Corresponde al promedio de la tasa de logro de UPM obtenida en las 100 simulaciones.

⁷ Corresponde al promedio de la tasa de logro de viviendas obtenida en las 100 simulaciones.

Tabla 4. Comparación de estimaciones y errores DEF 2020 y DEF' 2020

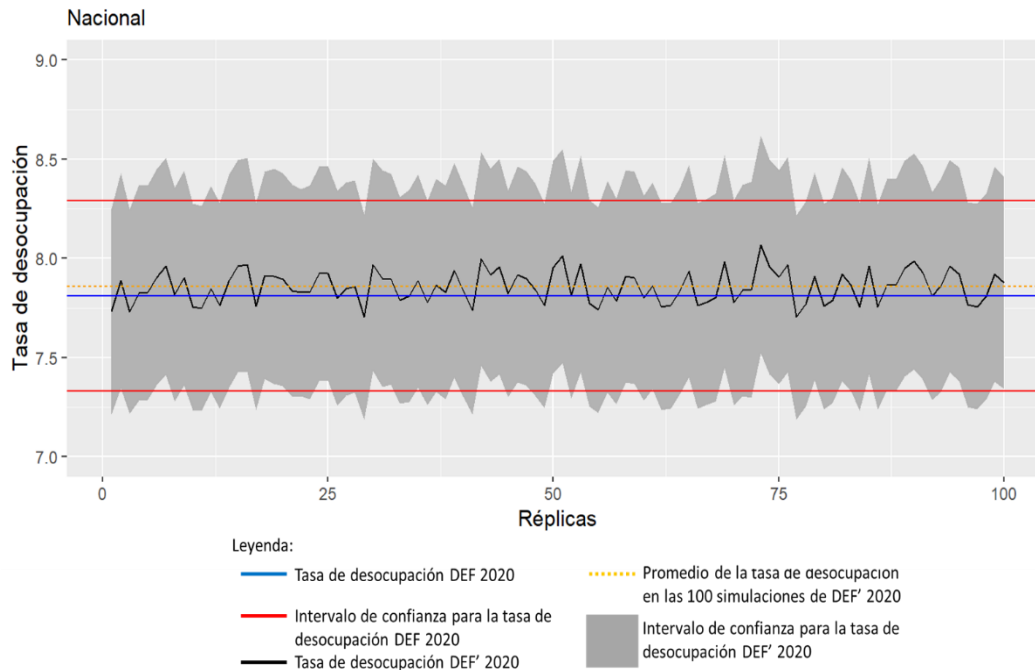
Nivel	DEF 2020		DEF' 2020		Dif. de tasa
	Tasa de desoc.	Error absoluto	Tasa de desoc. ⁸	Error absoluto ⁹	
Nacional	7,81	0,48	7,86	0,53	0,05
Arica y Parinacota	6,99	1,74	7,21	2,06	0,22
Tarapacá	7,67	1,9	7,64	2,12	-0,03
Antofagasta	7,64	1,66	7,94	1,85	0,3
Atacama	7,63	1,67	7,46	1,93	-0,17
Coquimbo	8,77	2,03	9,02	2,27	0,24
Valparaíso	8,46	1,03	8,44	1,12	-0,02
Metropolitana	8,57	0,98	8,65	1,09	0,08
O' Higgins	8,08	1,74	8,15	1,93	0,07
Maule	6,05	1,12	6,35	1,26	0,3
Ñuble	7,03	1,34	6,93	1,45	-0,11
Biobío	8,31	0,95	8,29	1,06	-0,02
La Araucanía	6,36	1,23	6,17	1,41	-0,19
Los Ríos	5,58	1,34	5,63	1,47	0,05
Los Lagos	3,84	0,88	3,64	0,97	-0,21
Aysén	3,55	1,23	3,3	1,34	-0,25
Magallanes	4,58	1,44	4,49	1,56	-0,09

Fuente: Instituto Nacional de Estadísticas.

El Gráfico 1, muestra los resultados obtenidos para la tasa de desocupación a nivel nacional. Como se puede observar el promedio de la tasa de desocupación de las 100 simulaciones, se encuentra por encima de la estimación oficial, reflejándose una sobreestimación de la tasa de desocupación de 0,05 puntos porcentuales, lo que indica que el sesgo introducido es pequeño. Adicionalmente, se observa que con el escenario simulado DEF' 2020 se obtienen estimaciones con los mismos niveles de precisión que la oficial. Por su parte, todos los intervalos de confianza obtenidos con DEF' 2020 se solapan con el oficial.

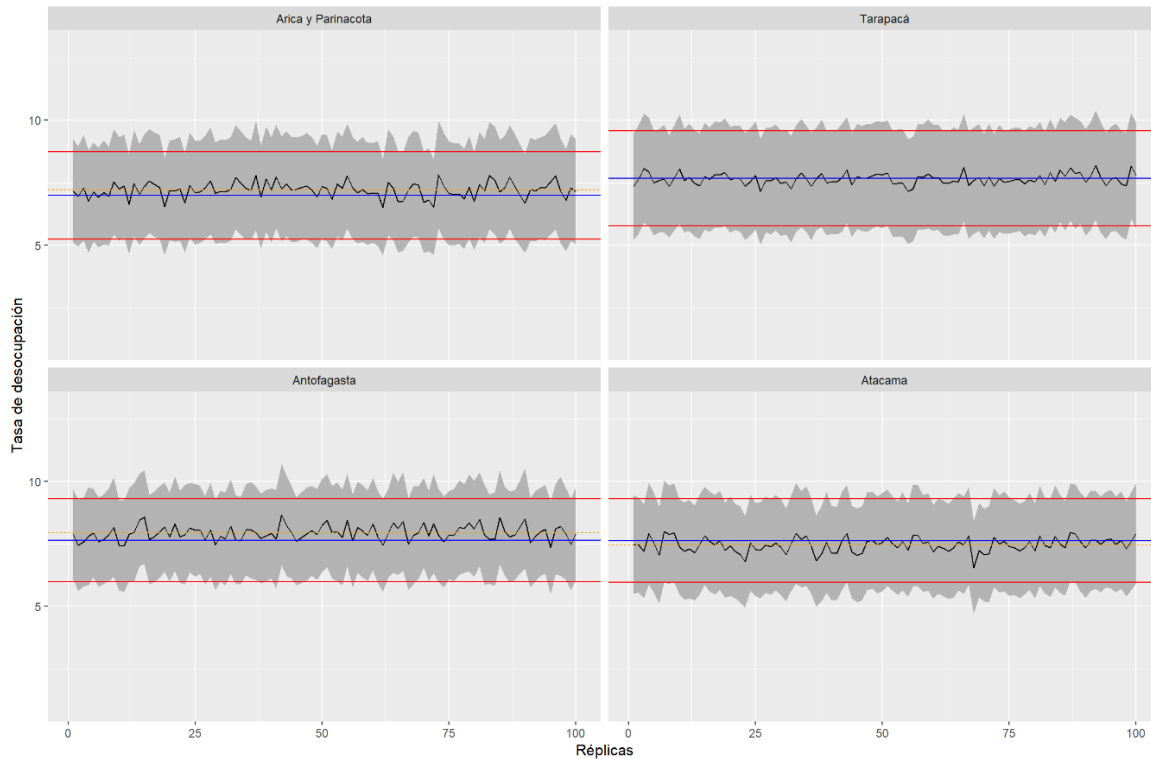
⁸ Corresponde al promedio de la tasa de desocupación obtenida en las 100 simulaciones.

⁹ Corresponde al promedio del error absoluto de la tasa de desocupación obtenido en las 100 simulaciones.

Gráfico 1. Estimación de la tasa de desocupación a nivel nacional

Fuente: Instituto Nacional de Estadísticas.

El Gráfico 2 muestra los resultados obtenidos para las regiones de la zona norte. En Arica y Parinacota y Antofagasta, se observa una sobrestimación de la tasa de desocupación, y los límites superiores de los intervalos de confianza, tienden a superar el límite superior del intervalo de confianza oficial, en donde cabe mencionar que, de las cuatro regiones éstas tienen las menores tasas de logro. Por el contrario, las tasas de desocupación de las regiones de Tarapacá y Atacama presentan una leve subestimación, y los límites inferiores de los intervalos de confianza tienden a ser menores que el límite inferior del intervalo de confianza oficial. No obstante, todos los intervalos de confianza se solapan en gran medida con el intervalo de confianza oficial.

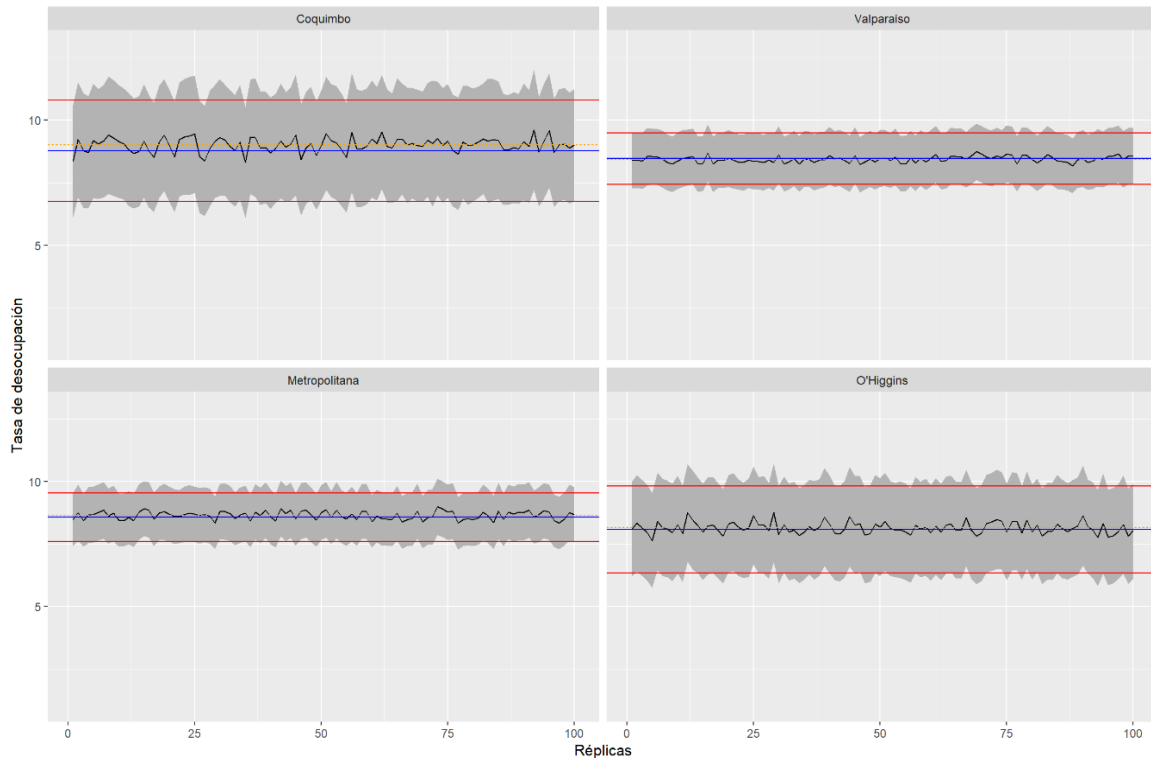
Gráfico 2. Estimación de la tasa de desocupación a nivel regional (zona norte)

Fuente: Instituto Nacional de Estadísticas.

En el Gráfico 3, se presentan los resultados obtenidos para un primer grupo de las regiones de la zona central del país, denotado como zona centro 1 y conformado por las regiones de Coquimbo, Valparaíso, Metropolitana y O'Higgins. De este grupo, la región con mayor diferencia entre la tasa de desocupación oficial y el promedio de dicha tasa obtenido con las 100 réplicas, es Coquimbo, siendo esta la región con la segunda menor tasa de logro y el menor tamaño muestral objetivo del grupo, lo que la hace más susceptible de presentar sesgo debido a la pérdida muestral. Por otro lado, Valparaíso y la región Metropolitana, aunque también tienen tasas de logro muy bajas, presenta estimaciones en el escenario simulado muy cercanas a la estimación oficial, junto con intervalos de confianza más pequeños, debido a que son algunas de las regiones con mayor tamaño de muestra objetivo.

Al igual que en el caso anterior, se observa que ningún intervalo del escenario simulado se escapa completamente del intervalo oficial.

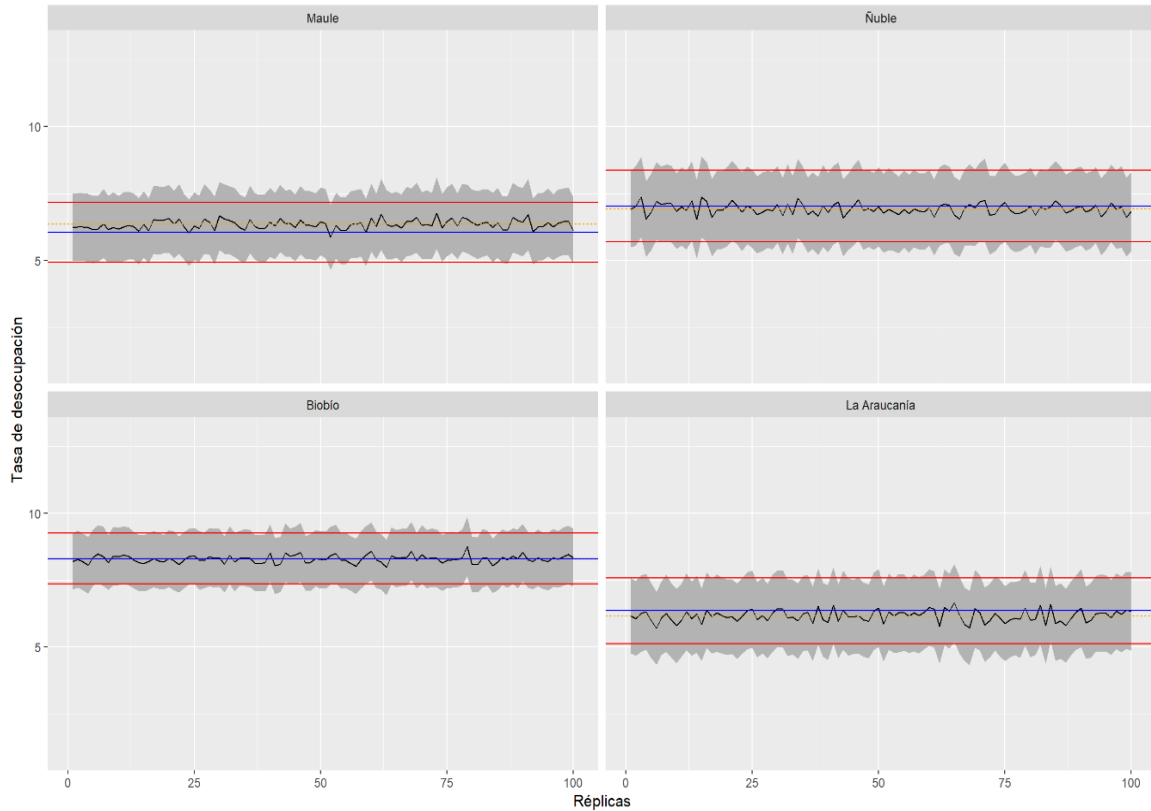
Gráfico 3. Estimación de la tasa de desocupación a nivel regional (zona centro 1)



Fuente: Instituto Nacional de Estadísticas.

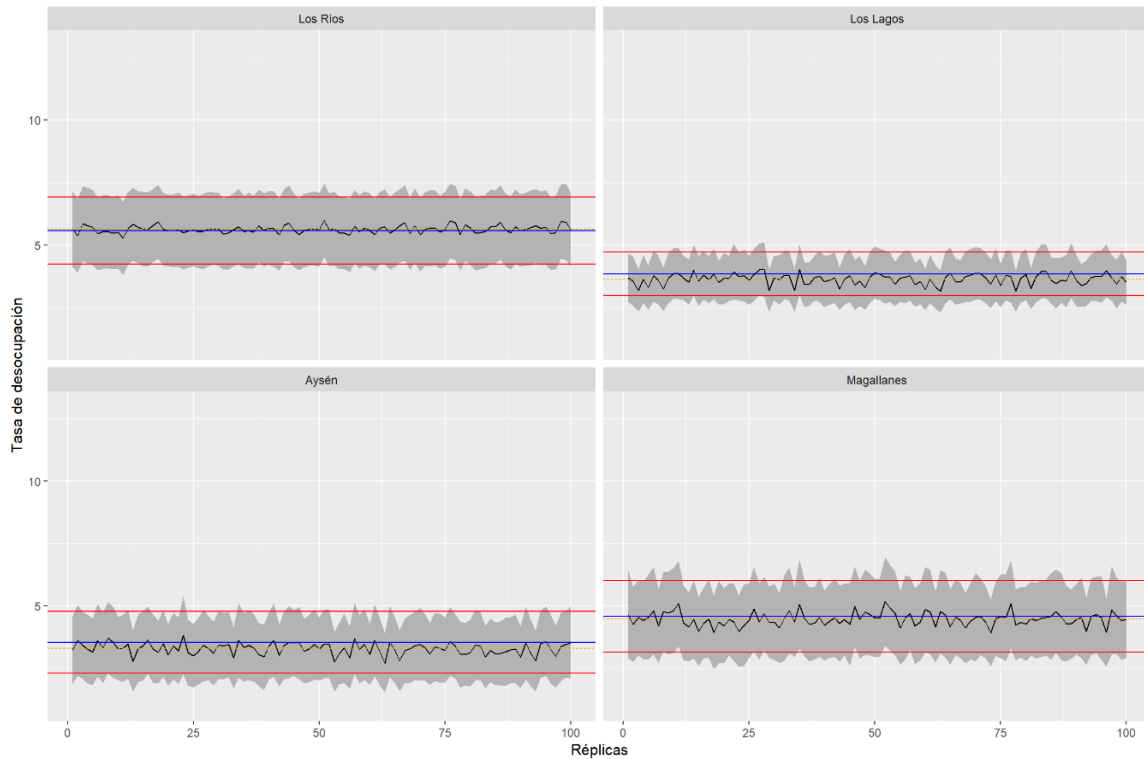
Los resultados del segundo grupo de las regiones centrales, denotado zona central 2, se presentan en el Gráfico 4, donde se observa que Maule es la región que presenta mayor diferencia entre las estimaciones simuladas y la estimación oficial de la tasa de desocupación, observándose sobre-estimación a pesar de ser una de las regiones con mayor tasa de logro. Por su parte, en las regiones de Ñuble y La Araucanía se observa una ligera subestimación. En Biobío el promedio de las simulaciones es prácticamente igual a la estimación oficial, pues también es una de las regiones con mayor tamaño de muestra objetivo, lo que la hace menos susceptible a los efectos de la pérdida muestral. Al igual que los casos anteriores, los intervalos simulados se solapan en gran medida con el oficial.

Gráfico 4. Estimación de la tasa de desocupación a nivel regional (zona centro 2)



Fuente: Instituto Nacional de Estadísticas.

El Gráfico 5 presenta los resultados de las regiones de la zona sur. Los Ríos, a pesar de ser una de las regiones con la menor tasa de logro simulada, produjo estimaciones e intervalos de confianza muy cercanos a los resultados oficiales de DEF 2020. Las regiones de Los Lagos, Aysén y Magallanes presentaron, en promedio, una subestimación de la tasa de desocupación y un comportamiento menos estable en lo que respecta a los intervalos de confianza. Sin embargo, ninguno de los intervalos simulados se escapa del intervalo oficial.

Gráfico 5. Estimación de la tasa de desocupación a nivel regional (zona sur)

Fuente: Instituto Nacional de Estadísticas.

En general, el promedio de las estimaciones para la tasa de desocupación de las 100 réplicas del escenario simulado DEF' 2020, resultó ser muy similar a la estimación oficial, tanto a nivel nacional como regional.

Al comparar los valores puntuales de la tasa de desocupación obtenidas a nivel regional para las 100 réplicas respecto de la estimación oficial, las mayores diferencias se observaron en las regiones de Atacama, Antofagasta, Aysén y Arica y Parinacota. Mientras que las mayores diferencias medidas en función del promedio de la tasa de desocupación de las 100 réplicas a nivel regional, se observaron en las regiones de Antofagasta, Maule, Aysén y Coquimbo.

No obstante, aunque en algunas regiones, con el escenario simulado, se obtuvo una mayor diferencia entre las tasas oficiales versus simuladas, en relación con otras, dichas diferencias siguen siendo bajas.

En cuanto a la precisión de las estimaciones, los intervalos de confianza obtenidos de las 100 simulaciones también estuvieron muy cercanos al intervalo oficial, aunque se haya observado mayor variabilidad de éstos en algunas regiones.

3. Segundo análisis: Tratamiento sobre los factores de expansión

En vista de la pérdida de unidades muestrales ya revisada anteriormente, se revisan los mecanismos que usa la encuesta para la construcción de los factores de expansión, específicamente, el ajuste que se realiza para tratar la pérdida de información por motivo de la falta de respuesta, además de la inclusión de un suavizamiento a los factores de expansión para el tratamiento de valores atípicos.

Como se ha comentado, la muestra del trimestre móvil EFM 2020 está compuesta por unidades que corresponden a dos diseños muestrales distintos. El nuevo diseño, elaborado sobre el MMV2017, que incorpora el ajuste para la no respuesta de viviendas en la construcción de los factores de expansión dentro de cada estrato de muestreo, mientras que el diseño antiguo, considera las dos etapas de muestreo, ajustando inicialmente para las unidades primarias de muestreo seguido del ajuste por pérdida de viviendas, siempre dentro de cada estrato de muestreo.

Dada la falta de respuesta a nivel de UPM observada en marzo, resulta conveniente analizar cómo se comportan los factores de expansión, por ejemplo, aplicando a la muestra antigua el ajuste para la no respuesta de manera similar a la muestra proveniente del MMV2017, además de evaluar la pertinencia de emplear alguna estrategia de suavizamiento de los factores de expansión.

Para ello, en los apartados 3.1 y 3.2, se presenta el método por el cual se realiza actualmente el ajuste para la no respuesta en la muestra antigua y el método que se prueba, respectivamente; con relación al ajuste que se realiza para la muestra nueva, éste puede revisarse con mayor detalle en el documento metodológico de la encuesta¹⁰, ya que para efectos de esta prueba permanecerá invariante. Asimismo, en el apartado 3.3 se detalla la estrategia de suavizamiento a emplear también para la muestra antigua, dado que fue en esta muestra donde se observa la presencia de factores de expansión atípicos. Para finalmente en el apartado 4, presentar los cuatro métodos a evaluar junto con los resultados obtenidos.

¹⁰ Documento disponible en el siguiente enlace: https://www.ine.cl/docs/default-source/ocupacion-y-desocupacion/metodologia/espanol/metodolog%C3%ADa-encuesta-nacional-de-empleo-ene-2020.pdf?sfvrsn=793380e9_8.

3.1. Método de ajuste para la no respuesta de unidades provenientes del MMM2016 y MMS2016

La ENE tiene un diseño muestral probabilístico estratificado y bietápico. En la primera etapa se seleccionan manzanas o secciones y en la segunda etapa, viviendas. Por tanto, el factor de expansión da cuenta de estas dos etapas de selección.

Los componentes del factor de expansión se obtienen como el producto de los inversos de las probabilidades de selección de las unidades de primera etapa (UPM definidas como manzanas o secciones) y de las probabilidades de selección de las unidades de segunda etapa (viviendas) seleccionadas desde las UPM ya muestreados, por lo que depende de:

- 1) El número de viviendas en el estrato y en la UPM seleccionada, según información reportada por el marco muestral.
- 2) El número de viviendas seleccionadas y existentes en la UPM de acuerdo con el proceso de enumeración o de actualización del marco muestral y el número de UPM seleccionadas dentro de cada estrato.

Los componentes del factor de expansión son constantes por unidad de primera etapa y, por consiguiente, se aplican a todas las viviendas que pertenecen a la unidad de primera etapa i del estrato h .

Por tanto, el cálculo del factor de expansión de selección, asociado a la vivienda j de la UPM i del estrato h de la muestra antigua de la ENE se calcula de acuerdo con la siguiente ecuación:

$$FO_{hij} = \frac{M_h}{n_h^{sel} \cdot M_{hi}} \cdot \frac{M'_{hi}}{m_{hi}^{sel}} \quad (1)$$

Donde:

FO_{hij} : Factor de expansión de selección de la vivienda j dentro de la unidad de primera etapa i en el estrato h .

M_h : Número de viviendas particulares en el estrato h al Censo 2002 o al año de actualización del marco.

M'_{hi} : Número de viviendas particulares ocupadas, reportado en el proceso de enumeración, en la unidad de primera etapa i en el estrato h .

M_{hi} : Número de viviendas particulares en la unidad de primera etapa i en el estrato h al Censo 2002 o al año de actualización del marco.

n_h^{sel} : Número de unidades de primera etapa seleccionadas en el estrato h .

m_{hi}^{sel} : Número de viviendas seleccionadas en la unidad de primera etapa i en el estrato h .

Luego, para dar cuenta de la calidad del levantamiento en terreno y ajustar el factor de expansión a la respuesta obtenida del trabajo de campo, se ajusta por la no respuesta de las UPM y de las viviendas logradas, por lo que se calculan los siguientes dos ponderadores de ajuste:

-Ponderador para unidades de primera etapa: Razón entre el número de UPM seleccionadas y el número de UPM efectivamente logradas.

$$Ponderador\ conglomerado = \frac{n_h^{sel}}{n_h} \quad (2)$$

Donde:

n_h^{sel} : Número de unidades de primera etapa seleccionadas en el estrato h .

n_h : Número de unidades de primera etapa logradas en el estrato h .

-Ponderador para viviendas: razón entre el número de viviendas seleccionadas y el número de viviendas efectivamente logradas.

$$Ponderador\ vivienda = \frac{m_{hi}^{sel}}{m_{hi}} \quad (3)$$

Donde:

m_{hi}^{sel} : Número de viviendas seleccionadas en la unidad de primera etapa i en el estrato h .

m_{hi} : Número de viviendas logradas en la unidad de primera etapa i en el estrato h .

Así, el factor de expansión es ajustado por no respuesta mediante la siguiente ecuación:

$$F1_{hij} = F0_{hi} \cdot \frac{n_h^{sel}}{n_h} \cdot \frac{m_{hi}^{sel}}{m_{hi}} = \frac{M_h}{n_h^{sel} \cdot M_{hi}} \cdot \frac{M'_{hi}}{m_{hi}^{sel}} \cdot \frac{n_h^{sel}}{n_h} \cdot \frac{m_{hi}^{sel}}{m_{hi}} = \frac{M_h}{n_h \cdot M_{hi}} \cdot \frac{M'_{hi}}{m_{hi}} \quad (4)$$

Donde:

$F1_{hij}$: Factor de expansión de selección de la vivienda j dentro de la unidad de primera etapa i en el estrato h , ajustado por no respuesta.

M_h : Número de viviendas particulares en el estrato h al Censo 2002 o al año de actualización del marco.

M'_{hi} : Número de viviendas particulares ocupadas, reportado en el proceso de enumeración, en la unidad de primera etapa i en el estrato h .

M_{hi} : Número de viviendas particulares en la unidad de primera etapa i en el estrato h al Censo 2002 o al año de actualización del marco.

n_h^{sel} : Número de unidades de primera etapa seleccionadas en el estrato h .

m_{hi}^{sel} : Número de viviendas seleccionadas en la unidad de primera etapa i en el estrato h .

n_h : Número de unidades de primera etapa logradas en el estrato h .

m_{hi} : Número de viviendas logradas en la unidad de primera etapa i en el estrato h .

3.2. Método evaluado de ajuste para la no respuesta

El método que se propone introduce un cambio para el cálculo de $F1_{hij}$ (Ecuación (4)), de tal manera que no hace uso de los ponderadores a nivel de UPM y vivienda (Ecuaciones (2) y (3)), en su defecto realiza el ajuste a nivel de viviendas dentro del estrato de selección, de manera que se obtiene el ponderador $\hat{R}_{h,R}$ definido como la razón entre la estimación del total de viviendas elegibles en la población, calculada a partir del factor de expansión $F0_{hij}$ y la estimación del total de viviendas de las que se obtiene respuesta de sus informantes. Así, el ajuste por no respuesta queda determinado por la siguiente expresión:

$$\hat{R}_{h,R} = \begin{cases} \frac{\sum_{i \in \Omega_h} \sum_{j \in \Theta_i} F0_{hij}}{\sum_{i \in \Omega_h} \sum_{j \in \Theta_{i,R}} F0_{hij}} & , \quad \text{Si la vivienda } j \text{ responde} \\ . & , \quad \text{Si la vivienda } j \text{ no responde} \end{cases} \quad (5)$$

Donde,

Θ_i : Conjunto de viviendas j seleccionadas en la UPM i .

$\Theta_{i,R}$: Conjunto de viviendas j logradas en la UPM i .

Ω_h : Conjunto de UPM i seleccionados en el estrato h .

$F0_{hij}$: Factor de expansión de selección de la vivienda j dentro de la UPM i en el estrato h .

Luego, se calcula el ponderador de selección de vivienda ajustado por no respuesta ($F1_{hij,R}$), como:

$$F1_{hij,R} = \hat{R}_{h,R} \cdot F0_{hij} \quad (6)$$

Este factor de expansión es aplicado a las k personas dentro de la vivienda j en la UPM i del estrato h , escribiéndose como $F1_{hijk,R}$, para ser aplicado el ajuste de calibración por el método multiplicativo o *raking*.

Dado que para la muestra proveniente del MMM2016 y MMS2016 no se aplicaba una hoja de ruta, a partir del cual clasificar el estado de las viviendas, este método supone que todas las unidades provenientes del marco antiguo son elegibles, lo que puede justificarse por el hecho de que corresponden a viviendas que pertenecen a la muestra de la ENE desde trimestres anteriores.

Por último, los resultados que se obtendrán permitirán comparar las estimaciones provenientes de ambos métodos de construcción del factor de expansión y brindarán los antecedentes que permitan decidir cuál método es el apropiado.

3.3. Método de suavizamiento de los factores de expansión

El método propuesto se denota método mixto (MM) en alusión a que mezcla dos métodos, a saber, método de contracción a la media (CM) y el método R-K, los cuales se vienen estudiando desde 2019 dentro del instituto¹¹, no obstante, dichos métodos ya han mostrado características favorables que mejoran la dispersión de las estimaciones, además que permiten mantener el orden y la distribución inicial de los factores de expansión¹². A continuación, se detallan los métodos CM, R-K y mixto.

¹¹ Como precedente se puede mencionar la aplicación de un método similar en la Quinta Encuesta Longitudinal de Empresas (ELE5), a partir de lo cual nace el interés de estudiar y desarrollar estos métodos de suavizamiento.

¹² Para mayor información consultar (Guerrero, 2019).

3.3.1. Método de contracción a la media (CM)

Este método contrae los factores de expansión acercándolos en distancia a la media. De esta forma, la distribución de los factores se hace más homogénea.

Sea w_i el factor i -ésimo y \bar{w} el promedio de los factores al nivel al cual se requiera aplicar el suavizamiento. El factor suavizado o contraído a la media w_i^S viene dado por:

$$w_i^S = \beta \cdot w_i + (1 - \beta) \cdot \bar{w} ; \quad 0 \leq \beta \leq 1 \quad (7)$$

Donde β es un parámetro real que permite controlar la intensidad de la contracción. Si se desea contraer muy poco la distribución original de los factores, entonces se le debe dar más peso a los valores originales (w_i) asignándole a β valores cercanos a 1. Por otra parte, si se desea maximizar la homogeneidad de los factores, es decir, que todos los factores sean lo más parecido a la media, se debe ponderar más el valor de la media (\bar{w}), asignándole a β valores cercanos a 0.

Una propiedad interesante de este método es que mantiene la suma total de los factores originales, ya que independientemente del valor que tome β se garantiza llegar al mismo *stock* poblacional. Esta afirmación queda demostrada en la siguiente expresión, en donde al suponer que hay n factores de expansión cuya suma es $t = \sum_i w_i = n \cdot \bar{w}$, y desarrollar la expresión de la suma de los factores contraídos a la media se obtiene el mismo valor, como sigue:

$$\sum_i w_i^S = \beta \cdot \sum_i w_i + (1 - \beta) \cdot \sum_i \bar{w} = \beta \cdot n \cdot \bar{w} + (1 - \beta) \cdot n \cdot \bar{w} = n \cdot \bar{w} = t \quad (8)$$

Adicionalmente, la contracción a la media constituye una transformación lineal de la variable w_i que es el factor original antes de la contracción, por lo que, considerando las propiedades de la media (el operador M de la media, en el denominador de la ecuación (9)) y varianza (el operador σ^2 de la varianza, en el numerador) se obtiene siempre un coeficiente de variación menor al suavizar los factores de expansión. En efecto:

$$Cv^2[w^S] = \frac{\sigma^2[\beta \cdot w + (1 - \beta) \cdot \bar{w}]}{M^2[\beta \cdot w + (1 - \beta) \cdot \bar{w}]} = \frac{\beta^2 \cdot \sigma^2[w]}{(\beta \cdot \bar{w} + (1 - \beta) \cdot \bar{w})^2} = \frac{\beta^2 \cdot \sigma^2[w]}{\bar{w}^2} \quad (9)$$

Sacando raíz cuadrada a ambos lados de la igualdad de la ecuación (9), se llega a:

$$Cv[w^S] = \frac{\beta \cdot \sigma[w]}{\bar{w}} = \beta \cdot Cv[w] \leq Cv[w], \quad \text{por ser } 0 \leq \beta \leq 1 \quad (10)$$

Por lo que queda demostrado que, al aplicar el método CM, por estar β entre 0 y 1, siempre se cumple que $Cv[w^s] < Cv[w]$.

3.3.2. Método R-K

Este método se caracteriza por dar un tratamiento diferenciado a los conjuntos de datos atípicos y no atípicos, ya que, genera un aumento proporcional en los factores no atípicos, mientras que, por el otro lado, origina una disminución proporcional de los factores atípicos.

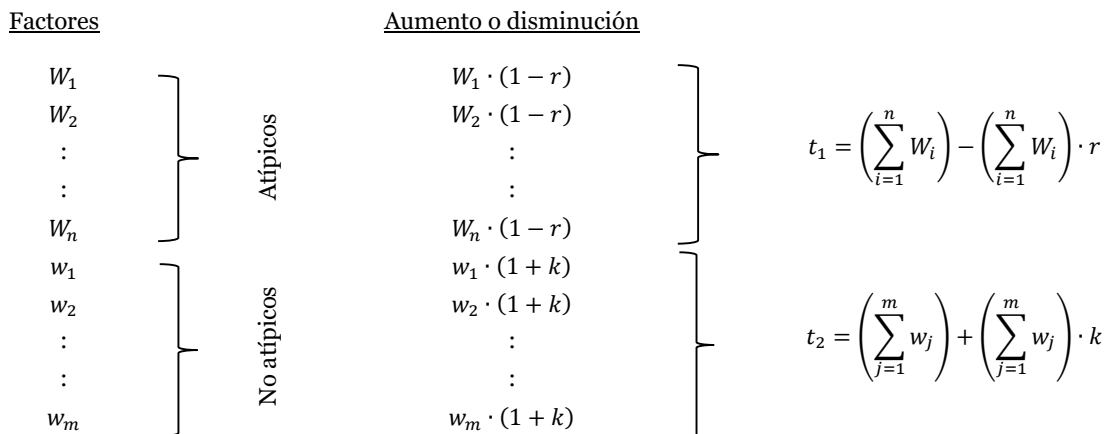
El nombre del método se debe a que la disminución de los factores de expansión atípicos se realiza en una proporción r , mientras que el aumento de los factores de expansión no atípicos se realiza en una proporción k .

Para lograr esto, se plantea un sistema de ecuaciones para distribuir los pesos, con la restricción de que el factor mayor de los no atípicos suavizados, sea inferior al factor menor de los atípicos suavizados, lo que se traduce en hallar los valores de r y k que cumplan estas condiciones.

Sean $W_1 \geq W_2 \geq \dots \geq W_n \geq w_1 \geq w_2 \geq \dots \geq w_m$ los factores de expansión ordenados en forma decreciente, distinguiendo con letra mayúscula a aquellos que son considerados como atípicos y a los cuales se les va disminuir el valor en r por ciento, mientras que, con letra minúscula se distingue a aquellos considerados no atípicos y a los cuales se les va a aumentar el valor en k por ciento. En otras palabras, los n factores considerados como atípicos se disminuyen en r por ciento, mientras que los m factores considerados como no atípicos se aumentan en k por ciento.

Más en detalle, si W_i es un factor atípico, la disminución de r por ciento se escribe como $W_i \cdot (1 - r)$, mientras que si w_j es un factor no atípico, el aumento de k por ciento se denota como $w_j \cdot (1 + k)$. En la Figura 3 se muestran las sumas a repartirse entre los factores que aumentan y los que disminuyen:

Figura 3. Esquema de cómo se configuran las sumas a repartir



Fuente: Instituto Nacional de Estadísticas.

La siguiente expresión muestra la primera condición, dada por la ecuación (11) que se emplea para encontrar los valores de r y k :

$$\left(\sum_{i=1}^n W_i \right) \cdot r = \left(\sum_{j=1}^m w_j \right) \cdot k \quad \Leftrightarrow \quad k = r \cdot \left(\sum_{i=1}^n W_i \right) / \left(\sum_{j=1}^m w_j \right) \quad (11)$$

Lo que se descuenta de los factores atípicos debe coincidir con lo que se agrega a los factores no atípicos. Es decir, el minuendo de t_1 debe coincidir con lo que se agrega a la suma de factores no atípicos (el segundo sumando de t_2 (Figura 3)).

La expresión siguiente muestra la segunda condición (ecuación (12)): que el menor de los factores atípicos disminuidos, $W_n \cdot (1 - r)$, sea mayor o igual que el mayor de los no atípicos aumentados, $w_1 \cdot (1 + k)$, es decir:

$$W_n(1 - r) > w_1(1 + k) \quad \Leftrightarrow \quad r < [(W_n - w_1) \cdot \sum_{j=1}^m w_j] / [w_1 \sum_{i=1}^n W_i + W_n \sum_{j=1}^m w_j] \quad (12)$$

Notar que el porcentaje de disminución r es estrictamente menor que la expresión encontrada. Con la igualdad, el menor de los atípicos tiene el mismo valor que el mayor de los no atípicos. A medida que r es menor que este valor, se va incrementando la distancia y es recomendado conceptualmente que exista una diferencia para que el valor que era atípico no llegue a igualarse a alguno de los no atípicos, en este caso, al mayor de los no atípicos.

Por otro lado, el caso extremo es cuando $r = 0$, es decir, no hacer disminución alguna de los factores atípicos y que va en contra del principio de suavizamiento, es decir, disminuir el valor de los factores atípicos y distribuir el total de las diferencias producidas, entre los factores no atípicos.

Tomando como valor inicial $r_0 = [(W_n - w_1) \cdot \sum_{j=1}^m w_j] / [f_1 \sum_{i=1}^n W_i + F_n \sum_{j=1}^m w_j]$, se pueden tomar distintos valores de r , parametrizándolo de la forma $r = \lambda \cdot r_0$, con $0 \leq \lambda \leq 1$.

3.3.3. Método mixto

Para el método mixto se separan los factores de expansión en dos subconjuntos: Φ_1 siendo el conjunto de los factores no atípicos y Φ_2 siendo el conjunto de los factores atípicos (Figura 4 (a)). Posteriormente se procede de la siguiente manera.

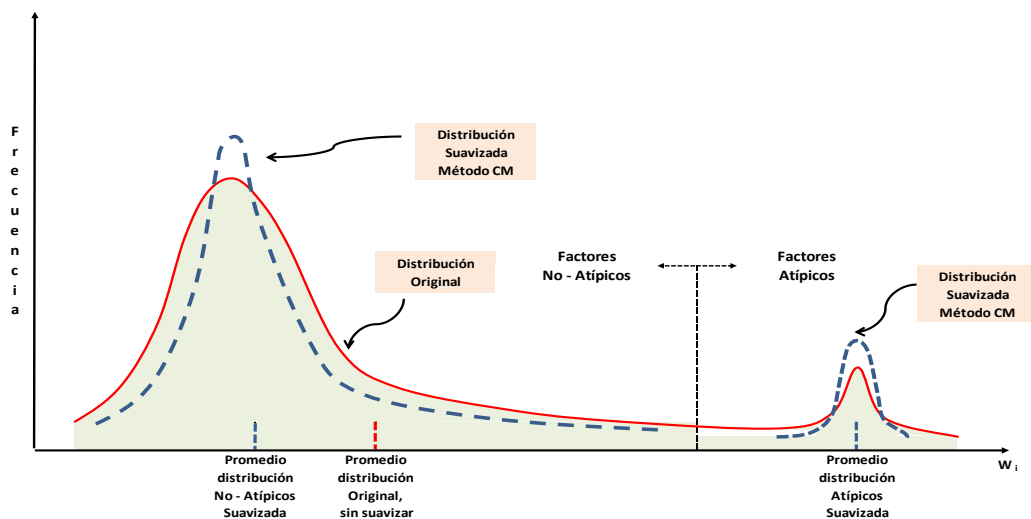
1. Para el conjunto Φ_1 , de factores no atípicos, se aplica el método de contracción a la media con un valor de beta relativamente cercano a la unidad, de modo que la distribución se mantenga casi intacta, es decir, que no varíe mucho respecto a la original, por ejemplo, aplicar el método CM con $\beta = 0,8$.
2. Para el conjunto Φ_2 de factores atípicos, se aplica el mismo método CM, pero con un valor de beta cercano relativamente a cero, de modo de potenciar la homogeneidad de ellos en este conjunto, es decir, de modo que se parezcan lo más posible entre ellos, por ejemplo aplicar el método CM con $\beta = 0,2$. En el caso extremo que se encuentre tan solo un valor atípico, el método CM no tiene efecto sobre este valor, es decir, el factor suavizado coincide con el valor atípico.

Cabe resaltar que en la aplicación separada del método CM a ambos subconjuntos (factores atípicos y no-atípicos), se debe contraer levemente los factores no atípicos y fuertemente los atípicos, por lo que se recomienda usar intensidades de contracciones complementarias a la unidad, es decir, si para los no atípicos se aplica β , para los atípicos se aplique $(1 - \beta)$.

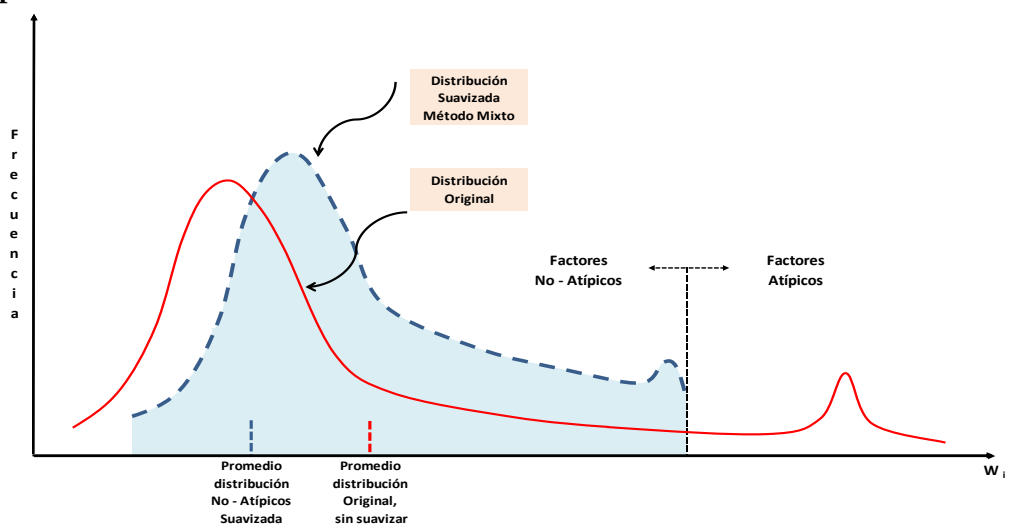
3. Habiendo aplicado en ambos conjuntos por separado el método CM, se aplica entonces el método R-K, juntando ambas distribuciones de factores (no atípicos con atípicos) en un continuo que se puede lograr con valores de λ cercanos a la unidad (Figura 4 (b)). Específicamente, se recomienda un valor de λ entre 0,95 a 0,99, para conservar una pequeña distancia entre el mayor de los factores no atípicos y el menor de los factores atípicos, después del suavizamiento.

Figura 4. Aplicación de método mixto CM y R-K

(a) Aplicación de método contracción a la media CM, en ambos conjuntos de factores; no atípicos y atípicos



(b) Aplicación de método R-K en ambos conjuntos de factores atípicos y no atípicos suavizados por método CM



Fuente: Instituto Nacional de Estadísticas.

3.3.4. Parametrización del suavizamiento para la ENE.

El método mixto en el contexto de la ENE, se aplica a nivel de región-área, y solo para la muestra antigua, que es donde se identifica la presencia de factores anómalos. Además, se toman como parámetros para la implementación de este método $\beta=0,8$ y $\lambda=0,99$, con la intención de intervenir poco los factores no atípicos y mucho los factores atípicos. También vale recordar que este método conserva la suma y la media de los factores a los niveles requeridos.

4. Método a evaluar y resultados

De los puntos desarrollados durante este análisis surgen cuatro métodos para su evaluación, que se diferencian por la forma como se realiza el ajuste por no respuesta y/o suavizamiento, a saber:

Método	Ajuste por no respuesta	Suavizamiento
M1	Cálculo tradicional	No
M2	Ajuste para no respuesta a nivel de estrato	No
M3	Cálculo tradicional	Método mixto
M4	Ajuste para no respuesta a nivel de estrato	Método mixto

Fuente: Instituto Nacional de Estadísticas.

Para tomar una decisión sobre el método más idóneo, es pertinente revisar la distribución de los factores de expansión en cada uno de ellos. Adicionalmente, también es importante observar la estimación puntual y la precisión de ésta, obtenida para el nivel nacional y regiones.

La Tabla 5 presenta las estadísticas descriptivas de los factores de expansión obtenidos en cada método evaluado, a nivel nacional. Al observar los valores máximos de cada método, es evidente la depuración de los factores de expansión más atípicos, sin que los valores de la media se vean afectados. En concordancia con lo anterior, los valores de los coeficientes de variación (CV) indican que al aplicar los ajustes por no respuesta y/o suavizamiento se genera efectivamente una reducción en la dispersión de los factores de expansión, además, se aprecia que la reducción más importante se genera al emplear suavizamiento. Cabe resaltar, que al observar la suma de los factores de expansión se evidencia que en todos los métodos probados se apunta al mismo *stock* poblacional.

Tabla 5. Estadísticas descriptivas de los factores de expansión por método de ajuste a nivel nacional

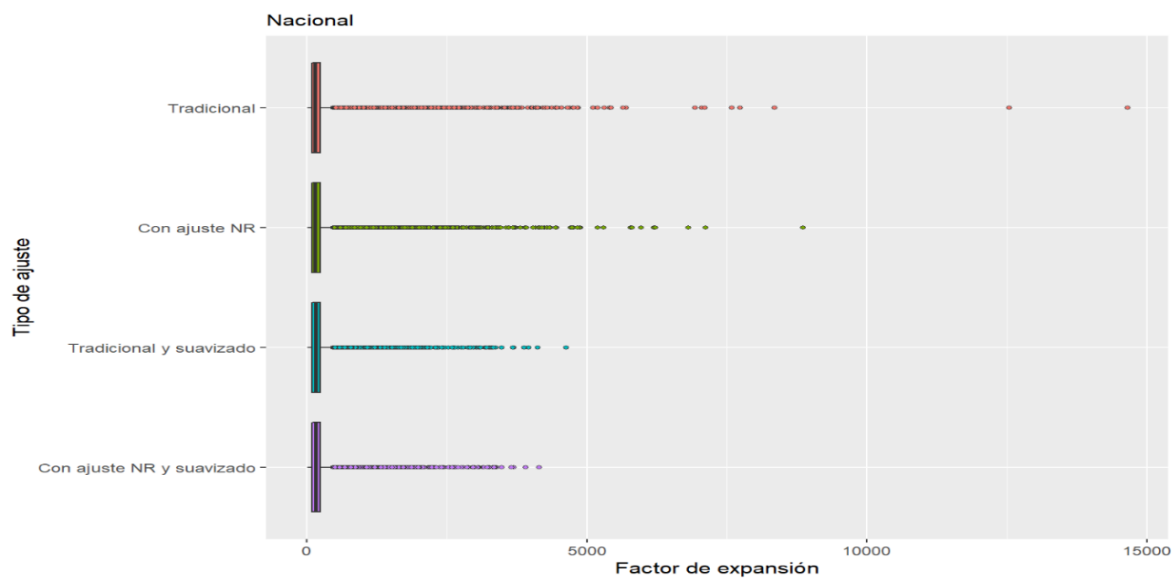
Estadístico	Método			
	M1	M2	M3	M4
N	87.842	87.842	87.842	87.842
Mínimo	9,2	9,8	12,0	13,2
Q1	88,2	88,9	96,2	96,8
Mediana	147,5	148,4	155,0	155,7
Q3	238,7	239,1	247,1	246,6
Máximo	14.648,2	8.865,1	4.634,9	4.147,9
Media	220,0	220,0	220,0	220,0
CV	130,7	123,6	104,1	103,0
Suma	19.326.617	19.326.617	19.326.617	19.326.617

Fuente: Instituto Nacional de Estadísticas.

El Gráfico 6 presenta la distribución de los factores de expansión calibrados, a nivel nacional según el método de ajuste que les fue aplicado, evidenciando cambios importantes entre cada uno de los métodos, sobre todo al realizar el ajuste por no respuesta o el suavizamiento. No obstante, la diferencia entre el método tradicional suavizado (M3) y el ajustado por no respuesta suavizado (M4) no es tan significativa. Es de notar que el método M4 reduce a poco menos de un tercio el rango que tienen los factores con el método tradicional M1.

En cuanto a la comparación de los factores obtenidos en los métodos M2, M3, y M4 respecto a los factores del método tradicional (M1), cualquiera de los ajustes propuestos en dichos métodos, representan cambios significativos, siendo el método M2 el que sigue un comportamiento más cercano.

Gráfico 6. Distribución de los factores de expansión según tipo de ajuste, a nivel nacional



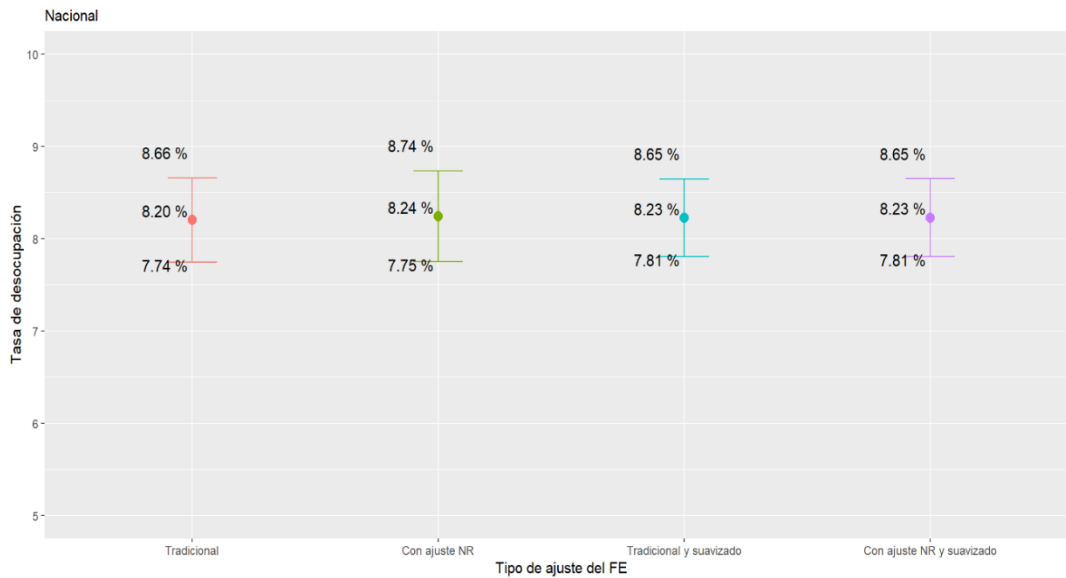
Fuente: Instituto Nacional de Estadísticas.

Las estadísticas descriptivas y los gráficos de caja de los factores de expansión según método de ajuste a nivel regional, se presentan en el Anexo 1 y el Anexo 2, respectivamente. En general, es la aplicación de los métodos que incluyen suavizamiento, lo que conlleva a una mejora importante en lo que respecta a la dispersión de los factores de expansión. No obstante, se presentan algunos patrones que resaltan:

- Al observar las diferencias entre los coeficientes de variación de cada método, Antofagasta y Valparaíso son las regiones donde la implementación de los ajustes con suavizamiento (M2, M3 o M4) impactan en mayor medida sobre la distribución de los factores de expansión.
- En algunas regiones como Tarapacá, Valparaíso y Biobío, se observan factores de expansión aún más grandes luego de aplicar el ajuste por no respuesta (M2).
- En la región de O'Higgins, aunque se observa una depuración de los factores de expansión atípicos, sobre todo en los métodos M3 y M4, persiste la presencia de valores atípicos en el factor de expansión, lo que se debe al ajuste de calibración, específicamente al ajustar al *stock* de población extranjera.

Finalmente, a continuación, se presentan los resultados de la estimación de la tasa de desocupación a nivel nacional para cada uno de los tipos de ajuste del factor de expansión. Como se puede observar en el Gráfico 7, no se aprecian diferencias significativas en cuanto al valor estimado para la tasa de desocupación. Sin embargo, se puede notar que bajo el ajuste tradicional con suavizamiento (M3) y el ajuste por no respuesta con suavizamiento (M4), se obtienen estimaciones levemente más precisas, dado que la amplitud de los intervalos de confianza en cada caso es menor a los observados cuando no se realiza el suavizamiento.

Gráfico 7. Estimación de la tasa de desocupación según tipo de ajuste del FE, a nivel nacional



Fuente: Instituto Nacional de Estadísticas.

Los Gráficos 8, 9, 10 y 11, presentan los resultados obtenidos a nivel regional. A este nivel también se observa que las estimaciones de la tasa de desocupación son bastante similares para cada uno de los métodos de ajuste, aunque como es de esperarse, varían un poco más que lo observado a nivel nacional. En cuanto a la precisión de las estimaciones la tendencia son intervalos de confianza más cerrados cuando se realiza suavizamiento.

Gráfico 8. Estimación de la tasa de desocupación según tipo de ajuste del FE, a nivel regional (zona norte)



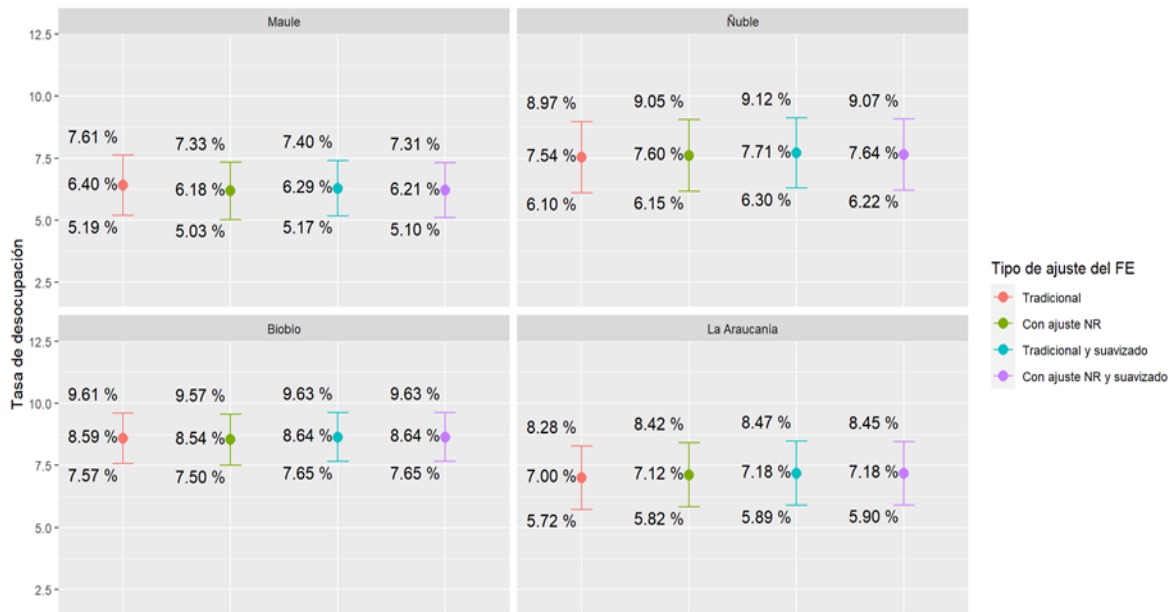
Fuente: Instituto Nacional de Estadísticas.

Gráfico 9. Estimación de la tasa de desocupación según tipo de ajuste del FE, a nivel regional (zona centro 1)



Fuente: Instituto Nacional de Estadísticas.

Gráfico 10. Estimación de la tasa de desocupación según tipo de ajuste del FE, a nivel regional (zona centro 2)



Fuente: Instituto Nacional de Estadísticas.

Gráfico 11. Estimación de la tasa de desocupación según tipo de ajuste del FE, a nivel regional (zona sur)



Fuente: Instituto Nacional de Estadísticas.

5. Disposiciones sobre la calibración de los factores de expansión

En lo que respecta al proceso de calibración de los factores de expansión, se revisaron los coeficientes de variación de cada marginal, los cuales resultaron menores al 10% para todas las categorías con el factor de expansión ajustado por no respuesta. Teniendo en cuenta estos resultados, se determina que no es necesarios generar cambios en el proceso de calibración definido en el diseño muestral¹³.

¹³Para mayor información, consultar el Manual Metodológico de la ENE disponible en el siguiente enlace: https://www.ine.cl/docs/default-source/ocupacion-y-desocupacion/metodologia/espanol/metodolog%C3%ADa-encuesta-nacional-de-empleo-ene-2020.pdf?sfvrsn=793380e9_8.

6. Conclusiones

Los resultados de la simulación del sesgo generado en la estimación debido a la pérdida muestral, indican que es posible concluir que el tamaño muestral obtenido para el trimestre móvil EFM 2020 permite obtener estimaciones para el nivel nacional y regional, para los que no hay evidencia de que la pérdida muestral genere sesgos significativos.

Por su parte los resultados obtenidos en el segundo análisis para determinar el método más idóneo de ajuste de los factores de expansión de las unidades provenientes de los marcos MMM2016 y MMS 2016, evidencian que los métodos de suavizamiento, excepto casos puntuales como el de O'Higgins, reducen significativamente la dispersión de los factores de expansión, lo que además se traduce en una leve mejora de la precisión. Se puede concluir entonces, que el método M4 puede ser implementado en la coyuntura EFM 2020, sin necesidad de modificar las marginales de calibración.

7. Referencias

Comisión Económica para América Latina y el Caribe (CEPAL). (Abril de 2020). *Recomendaciones para la publicación de estadísticas oficiales a partir de encuestas de hogares frente a la coyuntura de la enfermedad por coronavirus (COVID-19)*.

Elliott-Little. (2000, 2008, 2009). Model-based Alternatives to Trimming Survey Weights; Bayesian Weight Trimming for Generalized Linear Regression Models; Model Averaging Methods for Weight Trimming. *Journal of Official Statistics*.

Guerrero, M. (2019). *Impacto de Expansores Atípicos en la Medición de la Pobreza. Métodos de suavización de factores de expansión atípicos (Tesis de Magíster)*. Chile: Univeridad Diego Portales.

INE. (Enero de 2019). *Quinta Encuesta Longitudinal de Empresas. Informe de diseño muestral*. Obtenido de <https://www.economia.gob.cl/wp-content/uploads/2019/03/Informe-Dise%C3%B1o-Muestral-ELE-5.pdf>

INE. (2020). *Manual Metodológico de la Encuesta Nacional de Empleo (ENE)*. Obtenido de https://www.ine.cl/docs/default-source/ocupacion-y-desocupacion/metodologia/espanol/metodolog%C3%ADa-encuesta-nacional-de-empleo-ene-2020.pdf?sfvrsn=793380e9_8

Instituto Nacional de Estadísticas, Chile. (Enero de 2018). *Manual conceptual y metodológico diseño muestral Encuesta Nacional de Empleo*. Obtenido de sitio web INE: <http://www.ine.cl/docs/default-source/laborales/ene/antecedentes-metodologicos/manual-conceptual-y-metodol%C3%B3gico-dise%C3%B1o-muestral.pdf?sfvrsn=6>

ONU. (2009). *Diseño de muestras para encuestas de hogares. Directrices prácticas*. Recuperado el 15 de Noviembre de 2017, de https://unstats.un.org/unsd/publication/seriesf/Seriesf_98s.pdf

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

Anexos

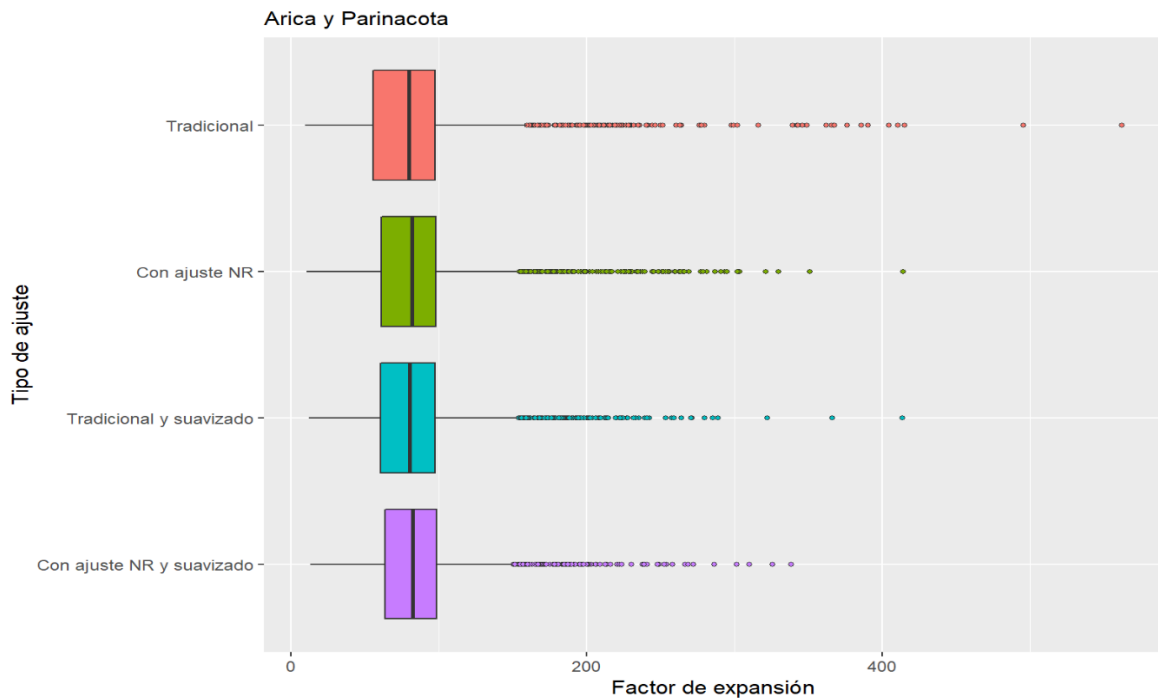
Anexo 1. Estadísticas descriptivas de los factores de expansión por método de ajuste a nivel regional

Nivel	Método	n	Mín	Q1	Mediana	Q3	Máx	Media	CV	Suma
Arica y Parinacota	Tradicional	3.024	9,5	55,6	79,9	97,3	562,1	82,7	58,4	250.207
Arica y Parinacota	Con ajuste NR	3.024	10,8	60,9	81,6	98,0	414,2	82,7	52,6	250.207
Arica y Parinacota	Tradicional suavizado	3.024	12,0	60,5	80,4	97,6	413,8	82,7	48,5	250.207
Arica y Parinacota	Con ajuste NR suavizado	3.024	13,2	63,7	82,4	98,5	338,5	82,7	46,0	250.207
Tarapacá	Tradicional	2.475	13,7	88,3	130,7	178,4	1.028,8	152,6	79,4	377.576
Tarapacá	Con ajuste NR	2.475	14,0	88,5	134,3	180,5	1.210,3	152,6	80,5	377.576
Tarapacá	Tradicional suavizado	2.475	14,0	100,1	142,4	186,0	827,9	152,6	63,3	377.576
Tarapacá	Con ajuste NR suavizado	2.475	14,2	98,9	143,7	188,5	839,4	152,6	63,6	377.576
Antofagasta	Tradicional	3.137	24,2	130,3	182,1	233,9	4.299,7	217,8	85,0	683.109
Antofagasta	Con ajuste NR	3.137	24,2	131,4	183,4	228,4	2.030,7	217,8	80,7	683.109
Antofagasta	Tradicional suavizado	3.137	26,1	152,5	188,7	241,6	1.417,8	217,8	53,2	683.109
Antofagasta	Con ajuste NR suavizado	3.137	26,0	153,3	191,2	245,3	1.066,3	217,8	51,5	683.109
Atacama	Tradicional	2.933	13,9	82,1	100,6	118,6	670,5	106,9	51,9	313.452
Atacama	Con ajuste NR	2.933	13,9	84,2	102,8	119,7	436,9	106,9	43,7	313.452
Atacama	Tradicional suavizado	2.933	13,7	86,5	103,3	120,4	447,2	106,9	36,4	313.452
Atacama	Con ajuste NR suavizado	2.933	13,7	87,2	104,4	121,1	396,8	106,9	35,0	313.452
Coquimbo	Tradicional	5.160	23,3	65,7	121,2	213,7	1.885,2	161,0	88,4	830.691
Coquimbo	Con ajuste NR	5.160	26,5	66,1	119,3	213,3	1.442,2	161,0	87,9	830.691
Coquimbo	Tradicional suavizado	5.160	32,2	70,1	125,9	217,9	996,8	161,0	75,2	830.691
Coquimbo	Con ajuste NR suavizado	5.160	35,0	70,5	125,4	215,8	1.060,8	161,0	75,2	830.691
Valparaíso	Tradicional	11.125	17,2	89,0	136,6	200,2	2.208,0	175,4	94,6	1.950.853
Valparaíso	Con ajuste NR	11.125	17,2	88,6	136,9	198,1	2.559,9	175,4	96,4	1.950.853
Valparaíso	Tradicional suavizado	11.125	29,8	103,8	146,9	203,1	1.540,9	175,4	65,5	1.950.853
Valparaíso	Con ajuste NR suavizado	11.125	28,2	104,7	148,5	204,2	1.492,8	175,4	63,6	1.950.853
Metropolitana	Tradicional	15.301	30,1	242,2	404,4	630,2	14.648,2	525,9	100,3	8.046.427
Metropolitana	Con ajuste NR	15.301	38,3	248,1	412,9	632,0	8.865,1	525,9	91,7	8.046.427
Metropolitana	Tradicional suavizado	15.301	42,5	283,6	428,9	641,3	4.634,9	525,9	70,9	8.046.427
Metropolitana	Con ajuste NR suavizado	15.301	42,6	286,6	433,3	643,3	4.147,9	525,9	69,3	8.046.427
O´ Higgins	Tradicional	5.417	20,7	82,0	149,2	238,7	1.682,1	182,1	83,2	986.526
O´ Higgins	Con ajuste NR	5.417	22,8	81,6	148,3	238,2	1.650,5	182,1	81,5	986.526
O´ Higgins	Tradicional suavizado	5.417	31,1	91,9	157,6	238,8	1.698,1	182,1	68,7	986.526
O´ Higgins	Con ajuste NR suavizado	5.417	31,0	91,8	157,9	237,6	1.702,2	182,1	68,8	986.526

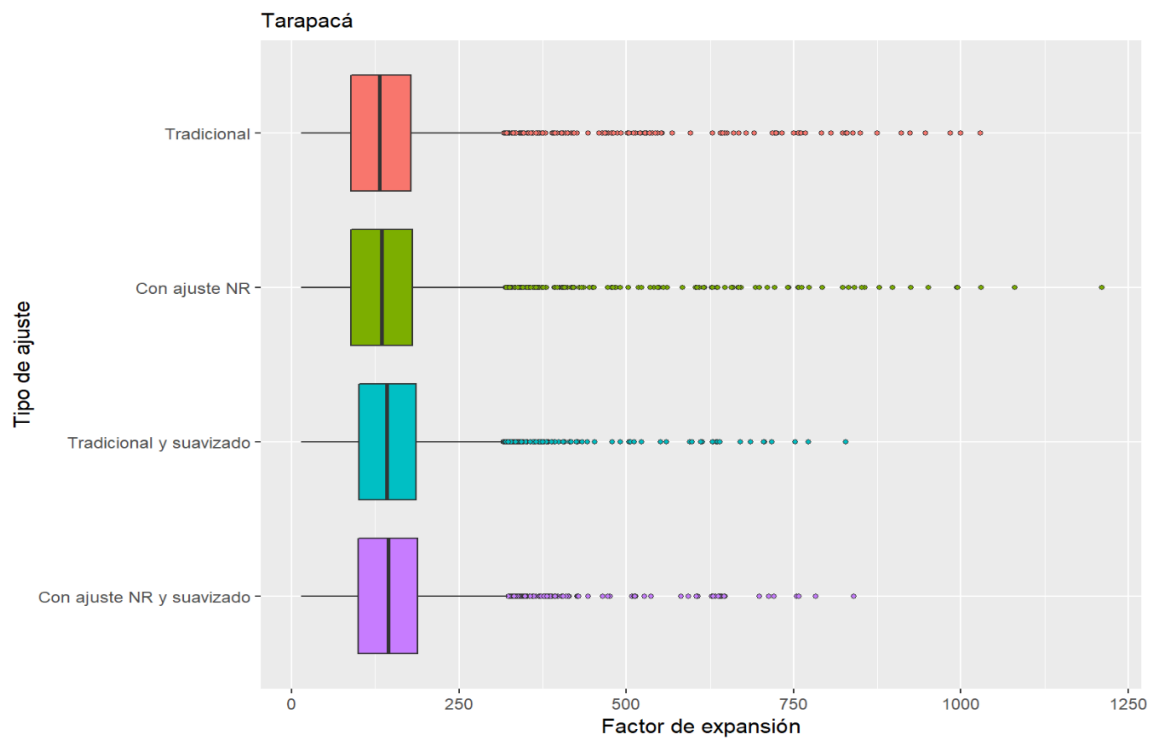
Nivel	Método	n	Mín	Q1	Mediana	Q3	Máx	Media	CV	Suma
Maule	Tradicional	6.434	20,0	102,7	155,7	225,1	2.051,2	175,2	67,4	1.127.032
Maule	Con ajuste NR	6.434	23,2	104,6	158,3	224,8	1.105,4	175,2	62,7	1.127.032
Maule	Tradicional suavizado	6.434	36,7	112,6	160,1	225,5	862,7	175,2	48,3	1.127.032
Maule	Con ajuste NR suavizado	6.434	41,4	113,0	162,1	224,9	715,9	175,2	47,6	1.127.032
Ñuble	Tradicional	3.571	26,1	97,6	132,6	171,4	1.065,4	142,9	55,0	510.175
Ñuble	Con ajuste NR	3.571	26,6	99,2	132,8	172,9	866,1	142,9	52,4	510.175
Ñuble	Tradicional suavizado	3.571	39,6	101,1	135,4	172,8	592,4	142,9	42,5	510.175
Ñuble	Con ajuste NR suavizado	3.571	40,1	101,7	135,6	172,4	783,8	142,9	45,9	510.175
Biobío	Tradicional	10.717	9,2	99,6	142,1	184,9	2.193,7	154,9	69,5	1.660.367
Biobío	Con ajuste NR	10.717	9,8	100,2	142,8	185,3	2.472,9	154,9	69,9	1.660.367
Biobío	Tradicional suavizado	10.717	28,1	105,9	144,1	185,4	833,3	154,9	49,1	1.660.367
Biobío	Con ajuste NR suavizado	10.717	27,6	106,6	145,2	185,4	903,6	154,9	49,0	1.660.367
La Araucanía	Tradicional	4.624	12,7	138,8	198,9	260,9	2.732,2	218,9	82,3	1.011.984
La Araucanía	Con ajuste NR	4.624	11,0	139,9	198,3	262,6	1.737,6	218,9	71,1	1.011.984
La Araucanía	Tradicional suavizado	4.624	30,2	143,1	201,6	265,9	1.020,7	218,9	58,2	1.011.984
La Araucanía	Con ajuste NR suavizado	4.624	29,3	144,4	199,9	262,2	1.464,6	218,9	62,5	1.011.984
Los Ríos	Tradicional	3.691	21,6	65,6	97,0	138,9	1.160,3	109,7	59,6	404.909
Los Ríos	Con ajuste NR	3.691	18,3	66,5	97,3	138,6	1.011,6	109,7	55,6	404.909
Los Ríos	Tradicional suavizado	3.691	31,8	69,6	99,2	137,8	618,5	109,7	48,8	404.909
Los Ríos	Con ajuste NR suavizado	3.691	28,7	69,8	98,3	138,5	604,4	109,7	48,6	404.909
Los Lagos	Tradicional	6.354	21,1	70,2	117,7	186,8	1.261,6	139,9	73,3	888.789
Los Lagos	Con ajuste NR	6.354	19,5	70,6	118,4	185,5	1.130,9	139,9	72,4	888.789
Los Lagos	Tradicional suavizado	6.354	29,5	73,2	120,2	186,6	847,4	139,9	64,0	888.789
Los Lagos	Con ajuste NR suavizado	6.354	31,3	74,5	123,0	182,3	834,3	139,9	62,4	888.789
Aysén	Tradicional	2.017	11,3	38,7	47,1	56,6	561,0	53,1	63,7	107.047
Aysén	Con ajuste NR	2.017	12,6	39,2	47,1	58,6	367,1	53,1	51,8	107.047
Aysén	Tradicional suavizado	2.017	17,7	41,0	48,0	58,1	262,6	53,1	42,5	107.047
Aysén	Con ajuste NR suavizado	2.017	18,7	41,0	48,1	59,1	215,4	53,1	38,6	107.047
Magallanes	Tradicional	1.862	20,2	66,7	84,3	105,1	654,8	95,3	57,9	177.473
Magallanes	Con ajuste NR	1.862	35,6	68,3	86,8	108,8	488,0	95,3	45,9	177.473
Magallanes	Tradicional suavizado	1.862	37,1	70,6	86,0	108,4	364,6	95,3	41,4	177.473
Magallanes	Con ajuste NR suavizado	1.862	36,8	70,9	87,7	108,6	362,2	95,3	39,8	177.473

Fuente: Instituto Nacional de Estadísticas.

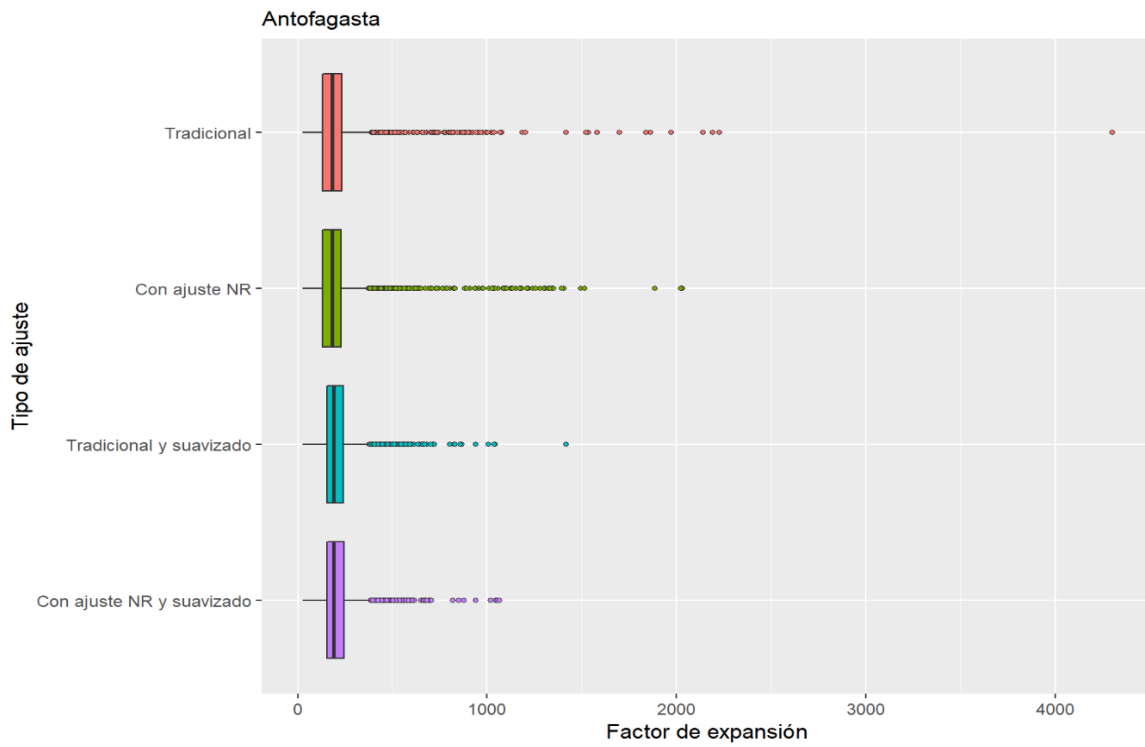
Anexo 2. Distribución de los factores de expansión según tipo de ajuste, a nivel regional



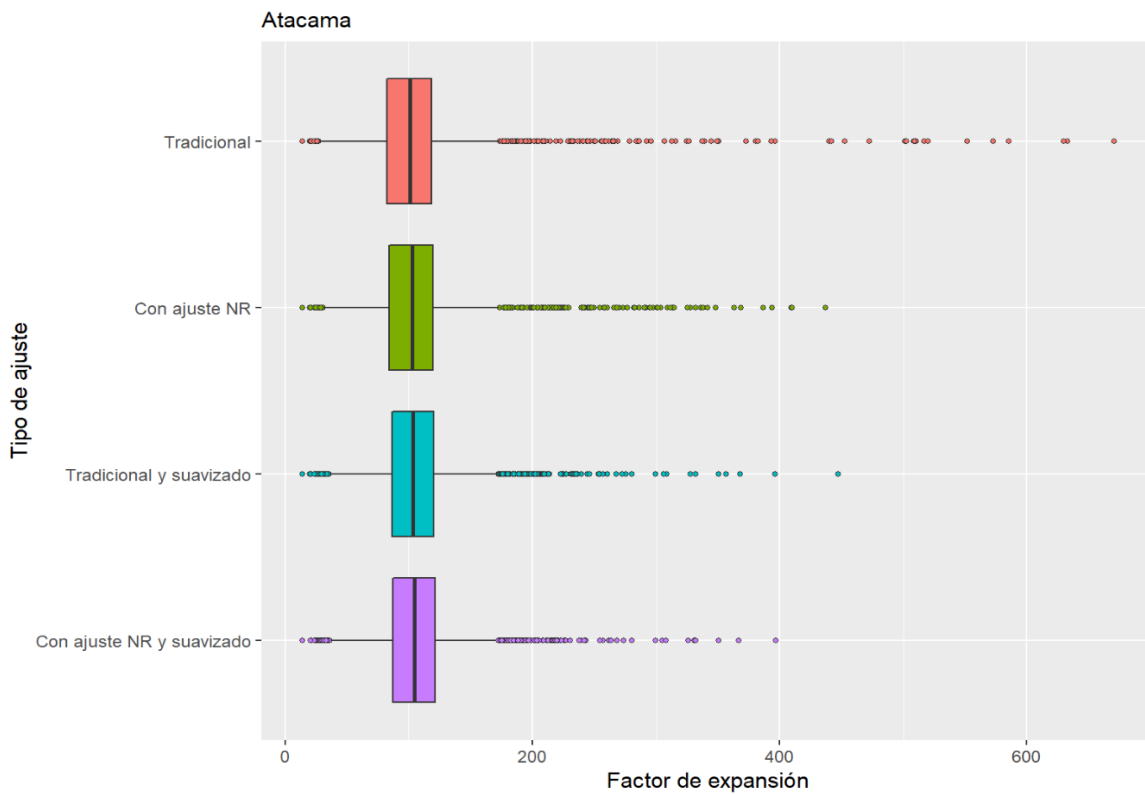
Fuente: Instituto Nacional de Estadísticas.



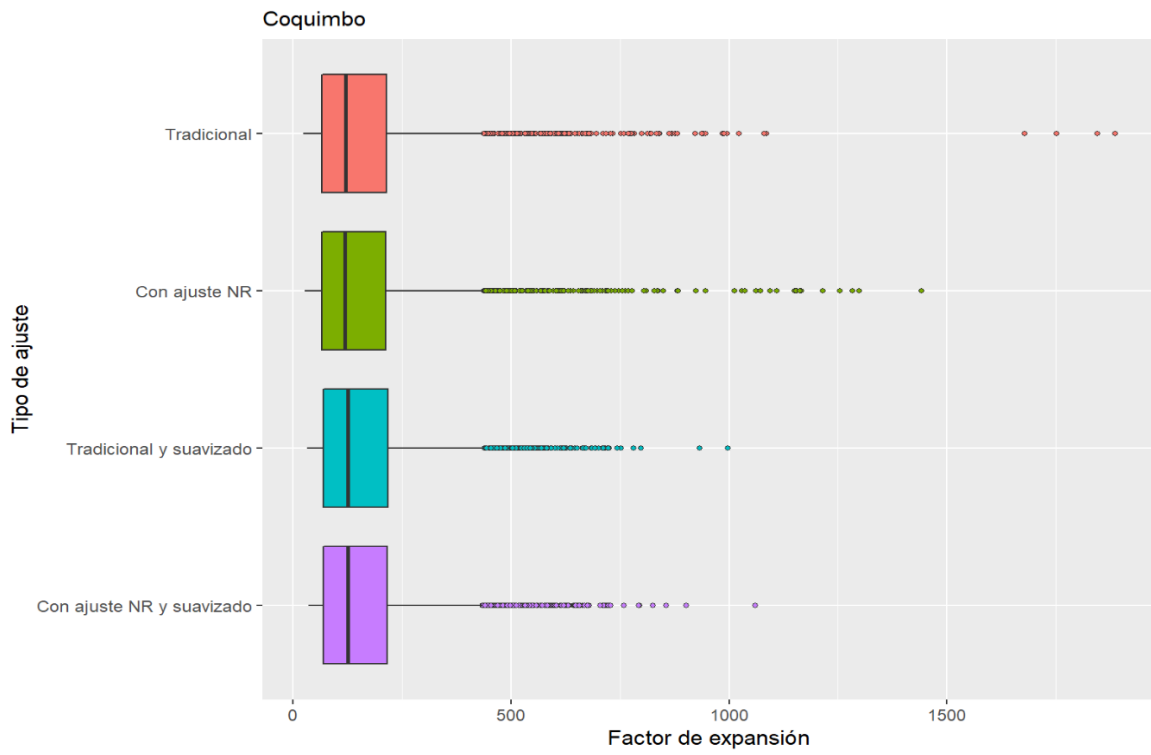
Fuente: Instituto Nacional de Estadísticas.



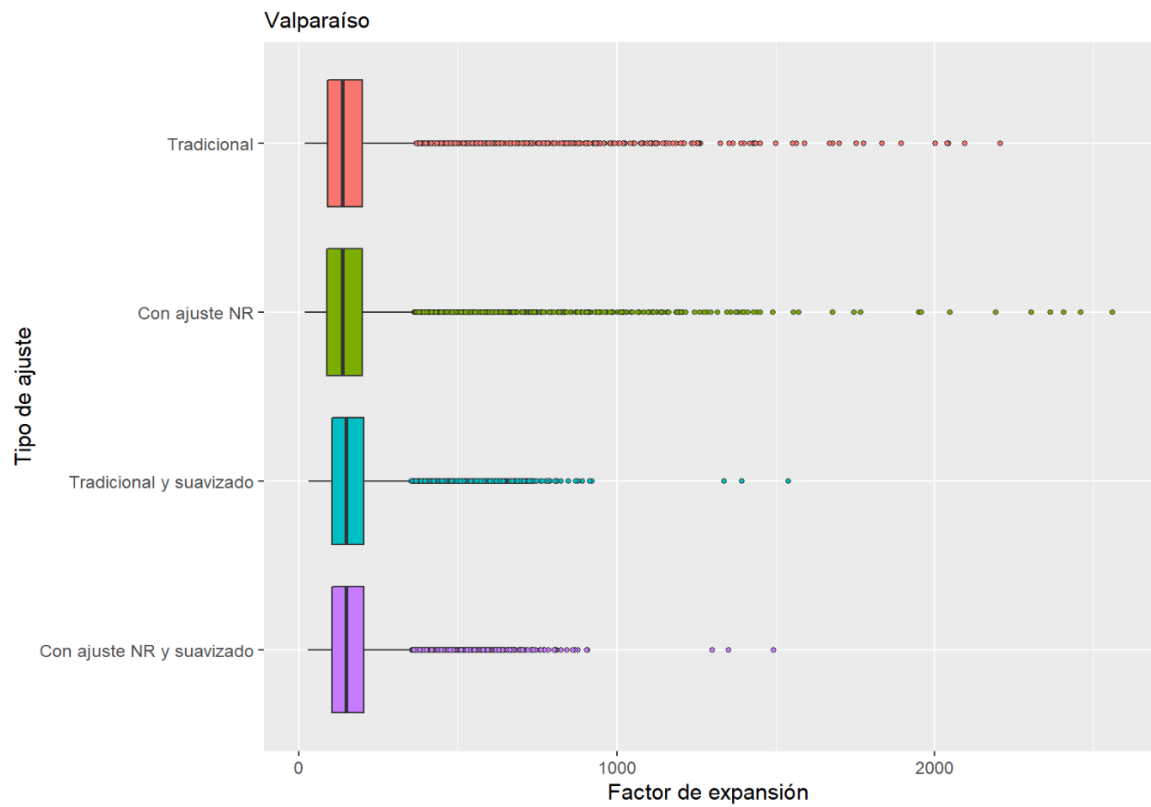
Fuente: Instituto Nacional de Estadísticas.



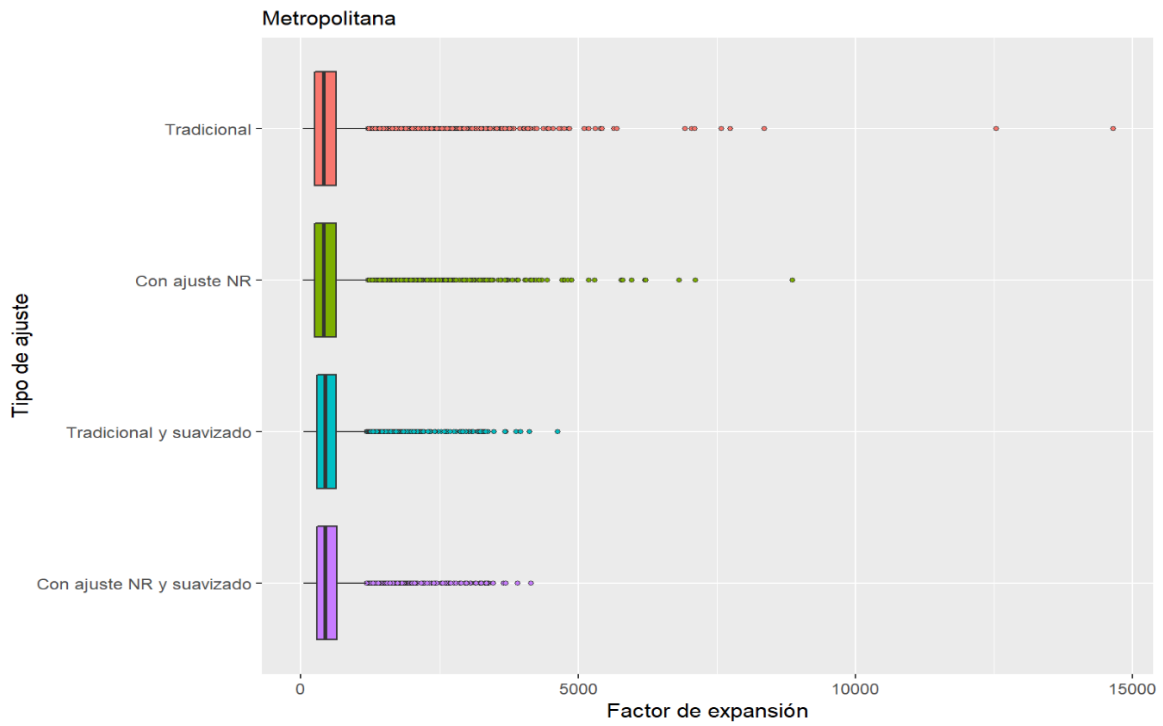
Fuente: Instituto Nacional de Estadísticas.



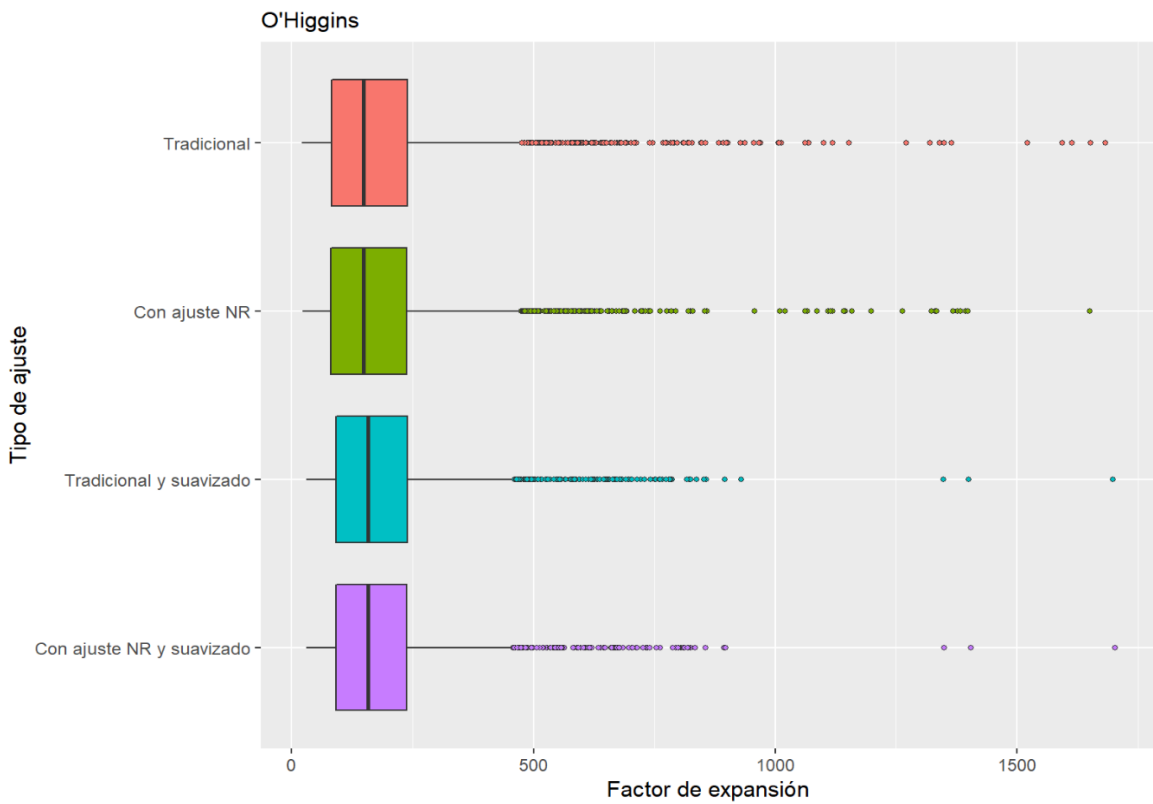
Fuente: Instituto Nacional de Estadísticas.



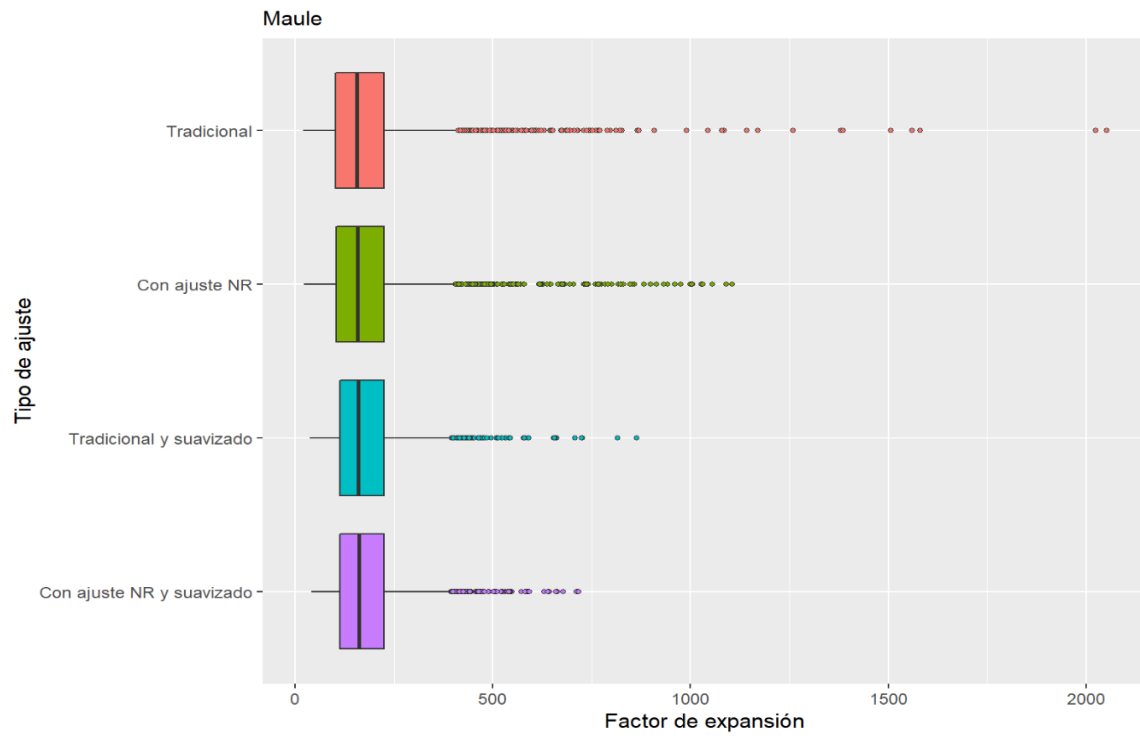
Fuente: Instituto Nacional de Estadísticas.



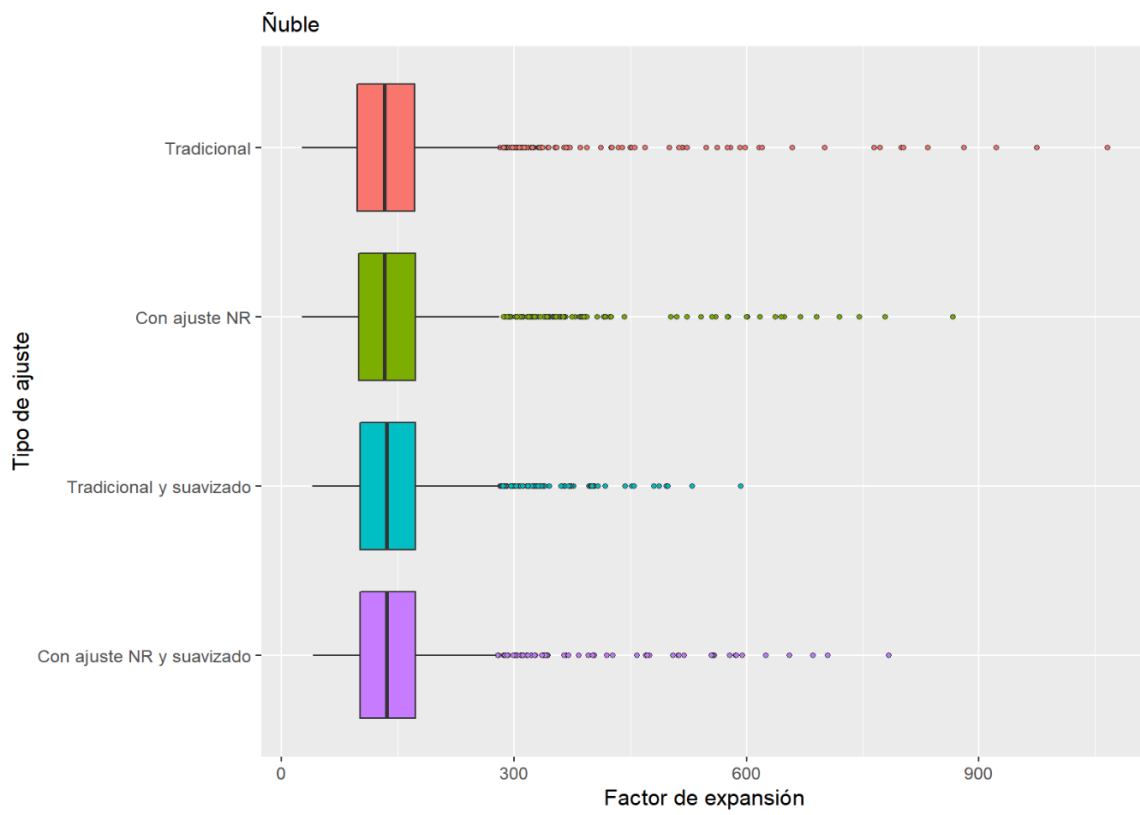
Fuente: Instituto Nacional de Estadísticas.



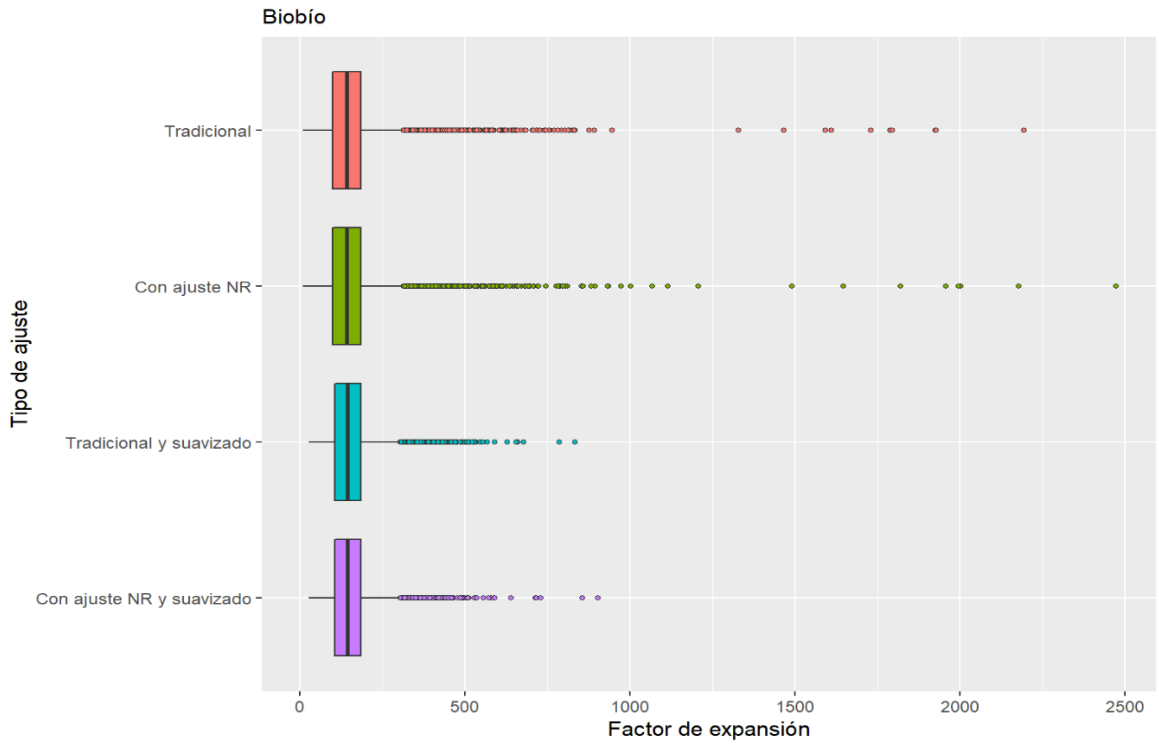
Fuente: Instituto Nacional de Estadísticas.



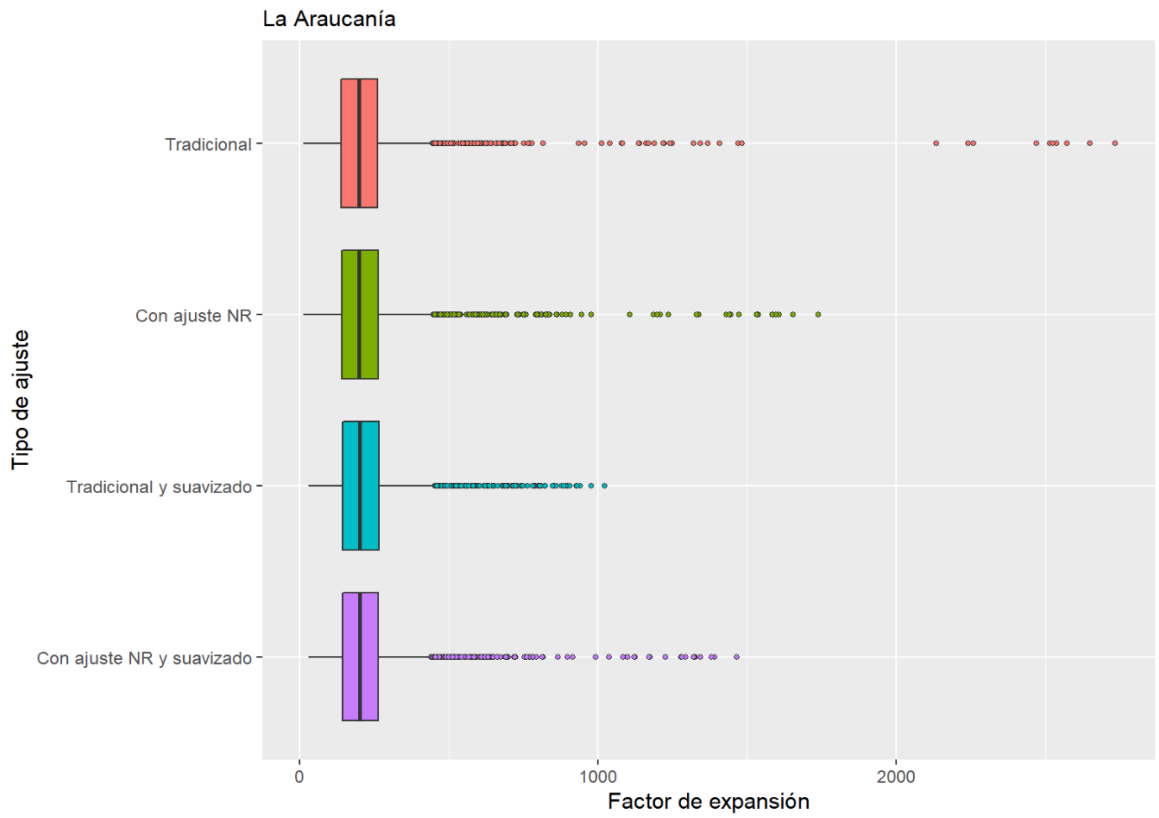
Fuente: Instituto Nacional de Estadísticas.



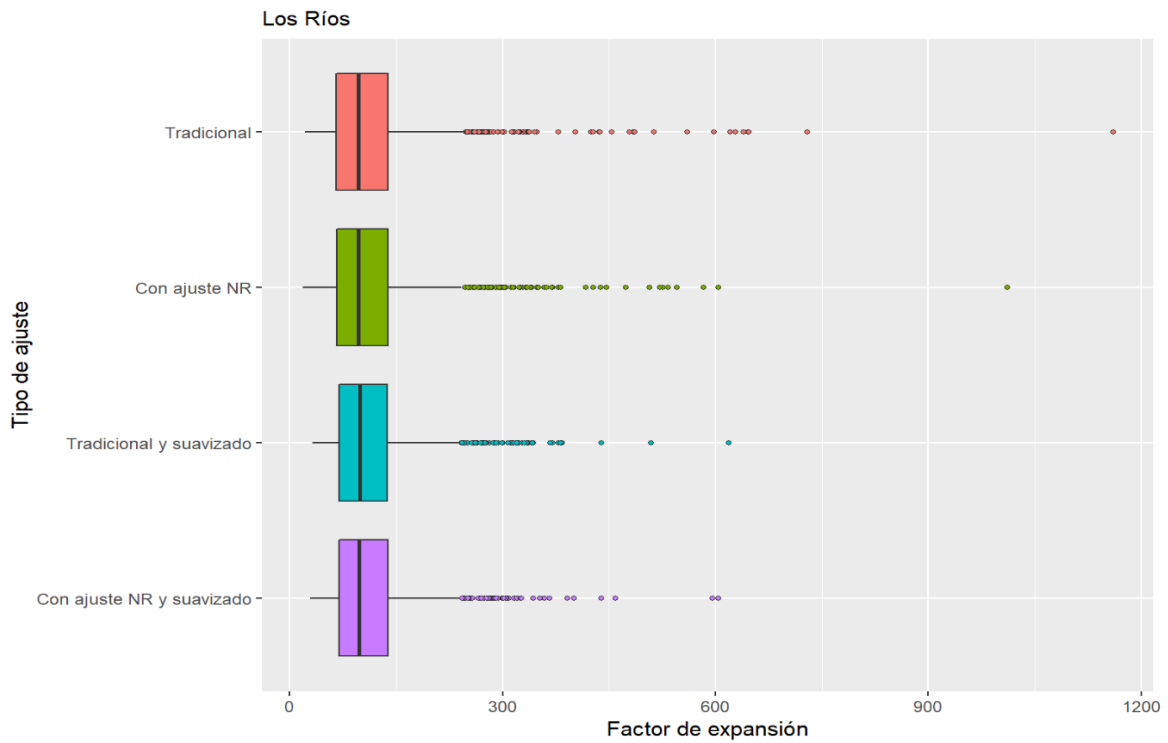
Fuente: Instituto Nacional de Estadísticas.



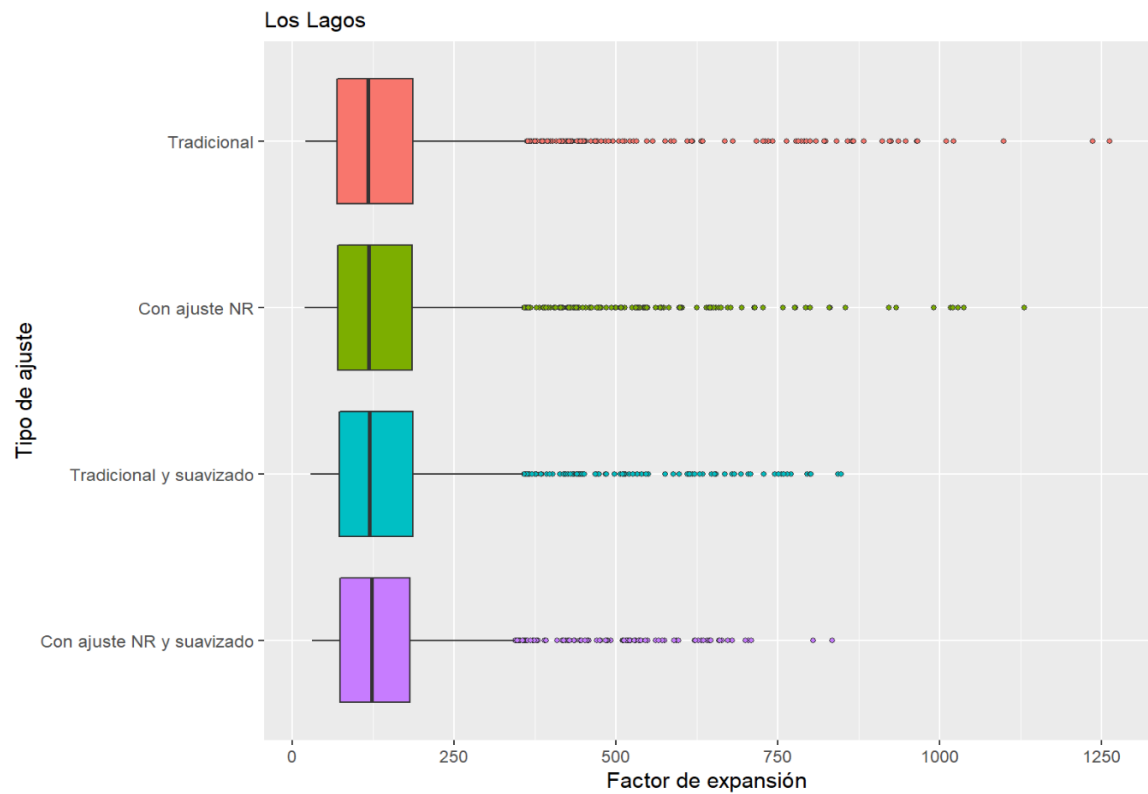
Fuente: Instituto Nacional de Estadísticas.



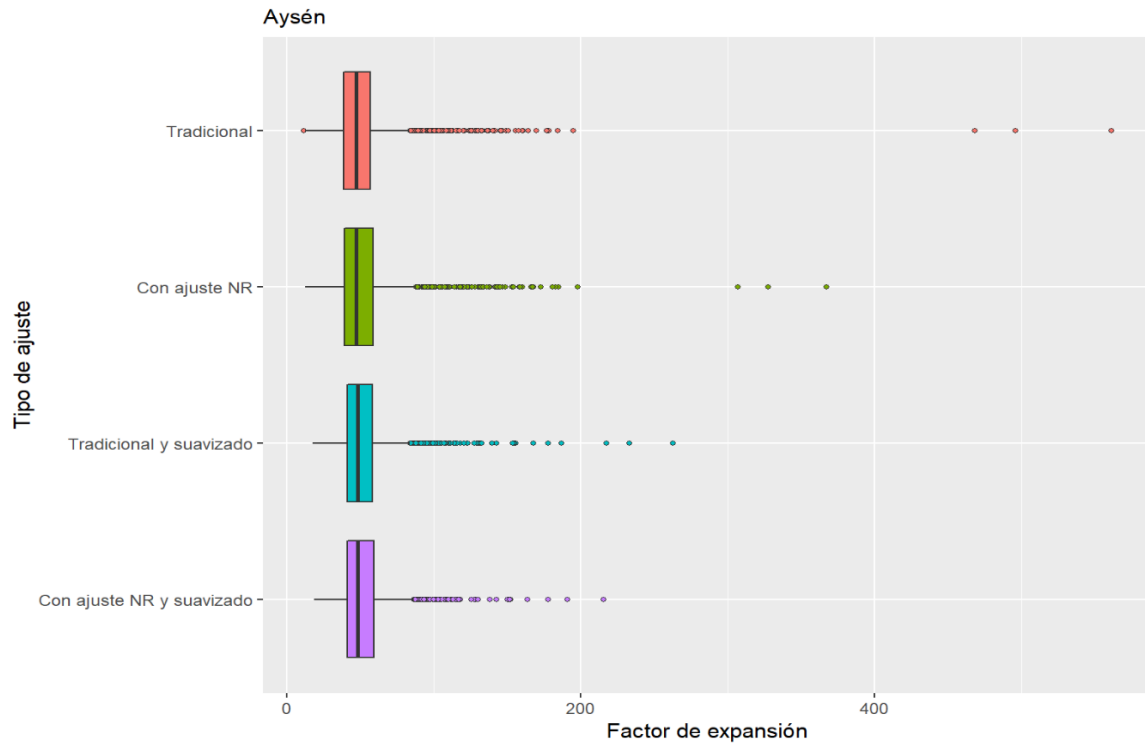
Fuente: Instituto Nacional de Estadísticas.



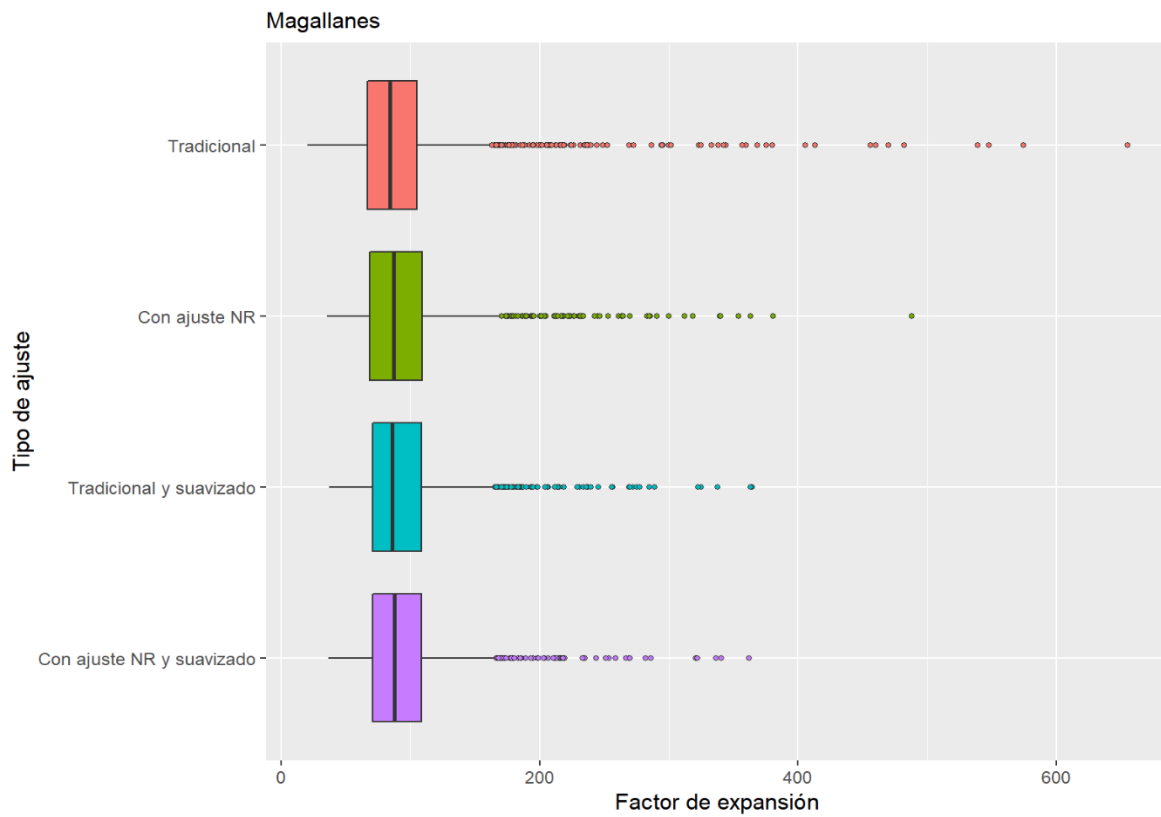
Fuente: Instituto Nacional de Estadísticas.



Fuente: Instituto Nacional de Estadísticas.



Fuente: Instituto Nacional de Estadísticas.



Fuente: Instituto Nacional de Estadísticas.