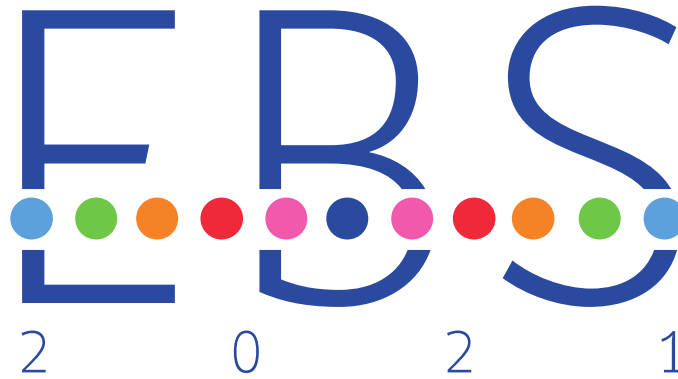




Encuesta de Bienestar Social



METODOLOGÍA DE DISEÑO MUESTRAL

Observatorio Social

Octubre 2021

Tabla de contenido

PRESENTACIÓN	3
1 ANTECEDENTES GENERALES DEL DISEÑO MUESTRAL	4
1.1 Objetivo	4
1.2 Población objetivo	4
1.3 Unidad de información	4
1.4 Marco muestral	4
1.5 Dominios de estudio	5
1.6 Objetivos de precisión y tamaño de muestra	5
1.7 Periodo de referencia	5
2 DISEÑO MUESTRAL	6
2.1 Marco muestral	6
2.1.1 Cobertura del Marco Muestral	7
2.1.2 Estratificación geográfica del Marco Muestral	7
2.1.3 Construcción del marco de selección	8
2.2 Estrategia muestral	9
2.3 Unidad de muestreo	9
2.4 Método de selección de la muestra	9
2.5 Tamaño muestral	9
2.5.1 Estimación del tamaño muestral	10
2.5.2 Objetivos de precisión	11
2.6 Selección de unidades muestrales	11
2.7 Errores efectivos	13
3 FACTORES DE EXPANSIÓN DE BIENESTAR SOCIAL	15
3.1 Factor de expansión de primera fase	15
3.2 Factor de expansión de segunda fase	18
3.3 Factor de expansión de primera y segunda fase	18
3.4 Ajustes al factor de expansión	18
3.4.1 Antecedentes de la disposición final de casos	18
3.4.2 Ajuste por no elegibilidad	19
3.4.3 Ajuste por no respuesta	19
3.4.4 Suavizamiento del factor de expansión (método mixto)	21
3.5 Calibración de los factores de expansión	22
4 Anexos	25

PRESENTACIÓN

El Ministerio de Desarrollo Social y Familia es responsable de evaluar la situación socioeconómica, así como el impacto de los programas sociales en las condiciones de vida de la población, con el propósito de contribuir a mejorar la eficiencia y eficacia de las políticas sociales.

Durante el año 2019 donde la población expuso distintas necesidades y demandas sociales, el Ministerio de Desarrollo Social y Familia (MDSF) determinó que el diagnóstico de la situación socioeconómica del país estaba incompleto, habiéndose centrado fuertemente en las condiciones materiales de las personas y sus hogares, pero con carencias de información respecto a la percepción y oportunidades que enfrenta la ciudadanía, así como de aspectos de calidad de vida que van más allá de lo meramente material.

A fin de disponer de la información necesaria para ello, el MDSF diseña y realiza la primera Encuesta de Bienestar Social (EBS). Esta encuesta permite complementar la información obtenida en la encuesta Casen en Pandemia 2020, y así disponer de una medición del bienestar para la población, aportando al diseño y evaluación de la política social del país.

El marco de medición¹ de la EBS se basa en el enfoque de capacidades (Sen 1982) y sigue los lineamientos de la serie de documentos de la OCDE “How’s Life? Measuring Well-Being” (2011 – 2020)² y el enfoque de capacidades de Sen (1982). De esta forma, se busca comprender el bienestar de manera comparable e integral a través de un conjunto de indicadores de bienestar que den cuenta de las oportunidades y resultados que tienen las personas en su vida.

De esta manera, la EBS se centra en la medición de 11 dimensiones: 3 dimensiones centradas en las condiciones materiales y 8 dimensiones que representan la calidad de vida de las personas.

Este documento tiene por objeto presentar la metodología definida para el diseño muestral de la EBS, teniendo en consideración las siguientes secciones: antecedentes generales del diseño muestral, especificaciones del diseño muestral y finalmente la metodología de construcción de los factores de expansión.

¹ Marco teórico de la EBS disponible en: http://observatorio.ministeriodesarrollosocial.gob.cl/storage/docs/bienestar-social/EBS_2021_Marco_teorico.pdf

² OCDE 2011, 2013, 2015, 2017, 2020. “How’s Life? Measuring Well-Being”

1 ANTECEDENTES GENERALES DEL DISEÑO MUESTRAL

La Encuesta de Bienestar Social (EBS), está dirigida a personas de 18 años y más³ pertenecientes a viviendas particulares dentro del territorio nacional, cuya aplicación, fue realizada durante los meses de abril y mayo del año 2021, considerando de esta manera 8 semanas de levantamiento.

A continuación, se presentan los antecedentes generales referentes al diseño muestral de la encuesta.

1.1 Objetivo

- Disponer de información sobre cómo las personas evalúan su propia vida respecto a diversas dimensiones.
- Generar una batería de indicadores del bienestar, que complementados con la información de la encuesta Casen en Pandemia 2020, contribuyan a la generación y evaluación de políticas sociales en el país.

1.2 Población objetivo

La población objetivo de la encuesta son las personas de 18 años o más que residen en viviendas particulares en el territorio nacional, excluyendo las áreas de difícil acceso definidas por el INE.

1.3 Unidad de información

La unidad de información de la EBS corresponde a las personas de 18 años o más que han sido seleccionadas para el levantamiento de esta encuesta.

1.4 Marco muestral

El marco muestral está conformado, en una primera fase, por el Marco muestral de la encuesta Casen en Pandemia 2020, que está conformado por dos marcos mutuamente excluyentes, uno es el Marco Muestral de Manzanas (MMM 2016) para el área urbana y el Marco Muestral de Secciones (MMS 2016) para el área rural, ambos marcos construidos en base al Censo 2002 y actualizados con información de Precenso 2016. Luego, en una segunda fase, el marco muestral fue construido en base a 143.673 personas de 18 años y más, que respondieron la encuesta Casen en Pandemia 2020.

³ Se excluyen 98 personas que en Casen en Pandemia 2020 corresponden a Servicio Doméstico Puertas Adentro (SDPA).

1.5 Dominios de estudio

Los niveles de estimación o dominios de estudio de la EBS 2021 para los cuales la muestra fue diseñada para tener representatividad son: nacional, nacional urbano, nacional rural y regional, los cuales coinciden con los dominios de estudio de Casen en Pandemia 2020. Así mismo, la muestra continúa teniendo presencia en las mismas 324 comunas.

1.6 Objetivos de precisión y tamaño de muestra

Los criterios de precisión se establecieron definiendo la variable de interés como una proporción para la cual se asume varianza máxima. Así, los errores muestrales esperados se fijaron en: nivel nacional 1 punto porcentual (pp.) en el caso del error absoluto y 2% el error relativo. A nivel regional los valores máximos son de 4 pp. y el error relativo de 8%.

1.7 Periodo de referencia

La EBS fue aplicada entre el 5 de abril y el 30 de mayo de 2021, es decir consideró 8 semanas en el proceso de recolección de datos.

2 DISEÑO MUESTRAL

El diseño muestral de la EBS corresponde a un diseño muestral bifásico donde la primera fase está supeditada al diseño de la encuesta Casen en Pandemia 2020, la que corresponde a una selección probabilística, estratificada y bietápica, siendo los estratos conformados por la dupla comuna-áreas⁴. Adicionalmente, en el área urbana, proveniente del Marco Muestral de Manzanas (MMM2016), las manzanas se clasificaron según su tamaño (en número de viviendas), creando 5 estratos de tamaño, en los cuales la muestra fue seleccionada de forma independiente, quedando los estratos del área urbana conformados por la terna comuna-área-grupo de tamaño⁵. De esta manera, las personas de 18 años y más que conformaron los hogares entrevistados en la encuesta Casen en Pandemia 2020 se utilizan como marco de selección para la muestra de la EBS 2021.

Luego, la segunda fase, se define como un muestreo aleatorio estratificado por región, realizando una selección (directa de personas) en forma sistemática con asignación proporcional al tamaño de la población en cada estrato o región.

2.1 Marco muestral

El marco muestral de primera fase, corresponde al marco muestral utilizado para la muestra de la encuesta Casen en Pandemia 2020 (levantada entre noviembre de 2020 y enero de 2021). Éste es un marco de viviendas y comprende dos marcos de muestreo mutuamente excluyentes:

- Marco Muestral de Manzanas (MMM 2016): Este marco se genera a partir de la actualización con información de Precenso 2016 del marco de manzanas con base en el Censo 2002. Éste contiene las áreas geográficas urbanas y resto de área urbana (RAU). Las unidades que lo componen se denominan “manzanas” y corresponden a delimitaciones geográficas fijas.
- Marco Muestral de Secciones (MMS 2016): Este marco de muestreo corresponde al generado a partir del Censo 2002, y actualizado con información de Precenso 2016. En él se incluye el área rural y parte del área RAU, donde las unidades que lo componen se denominan “secciones”.

Luego, la segunda fase, considera un marco de personas que comprende a 143.673 personas de 18 años y más⁶ entrevistadas en el levantamiento de la encuesta Casen en Pandemia 2020.

⁴ Las áreas clasificadas como urbana (Ciudad, Resto de Área Urbana, Urbana) y Rural.

⁵ Para más detalles respecto al Diseño Muestral de Casen en Pandemia 2020, ver: http://observatorio.ministeriodesarrollosocial.gob.cl/storage/docs/casen/2020/Diseno_Muestral_Casen_en_Pandemia_2020.pdf

⁶ Se excluyen 98 personas que en Casen en Pandemia 2020 corresponden a Servicio Doméstico Puertas Adentro (SDPA).

2.1.1 Cobertura del Marco Muestral

Dado que la EBS corresponde a un estudio de carácter bifásico, comparte las propiedades de cobertura del marco de selección de primera fase, obtenido a partir del levantamiento de la encuesta Casen en Pandemia 2020. Por lo anterior, la cobertura geográfica contempla la población residente en viviendas particulares en todo el país, con exclusión de las áreas de difícil acceso (ADA⁷) y las manzanas de 7 o menos viviendas y las manzanas y secciones incluidas en otras muestras seleccionadas por el INE para el periodo de recolección de datos de la encuesta Casen en Pandemia 2020.

2.1.2 Estratificación geográfica del Marco Muestral

La estratificación geográfica del marco de muestral de primera fase (Casen en Pandemia 2020), se clasifican de acuerdo con una ordenación división político-administrativa y las unidades corresponden, en orden jerárquico, a región, provincia y comuna. Al interior de cada comuna se conforma la división censal que da origen a las áreas geográficas denominadas urbana y rural. Estas áreas se encuentran definidas por la actividad económica preponderante y dan origen a las siguientes entidades:

1. Ciudad (CD): Es un gran centro urbano conformado por uno o un conjunto de centros urbanos adyacentes con 40.000 o más habitantes.
2. Resto de Área Urbana (RAU): Conformado por un conjunto de centros urbanos que totalizan menos de 40.000 y más de 2.000 habitantes. Esta clasificación se da cuando en una comuna existe una ciudad (CD) y entonces todos los centros urbanos restantes, si es que existen, se denominan resto de área urbana (RAU).
3. Urbano (U): Es el centro urbano con menos de 40.000 y más de 2.000 habitantes. Esta clasificación se da cuando en la comuna no existe una ciudad (CD), por lo que cada uno de sus centros urbanos se denominan simplemente como urbanos (U).
4. Rural (R): Conformado por el conjunto de entidades clasificadas como rurales de acuerdo con un tamaño poblacional menor a 1.000 habitantes o entre 1.001 y 2.000 habitantes con predominio de población económicamente activa (PEA) dedicada a actividades primarias.

Además, y para fines de selección muestral, en el interior de cada comuna del marco muestral de la encuesta Casen, las manzanas correspondientes al área urbana fueron clasificadas según su tamaño en cinco grupos, con el objetivo de poder distribuir la muestra de viviendas, en cada uno de estos

⁷ Se excluyen las comunas de: General Lagos, Colchane, Ollagüe, Juan Fernández, Isla de Pascua, Cochamó, Chaitén, Futaleufú, Hualaihué, Palena, Lago Verde, Guaitecas, O'Higgins, Tortel, Laguna Blanca, Río Verde, San Gregorio, Cabo de Hornos (Ex - Navarino), Antártica, Primavera, Timaukel, Torres del Paine.

grupos de manzanas y de esta forma, tener representadas en la muestra los diversos grupos de tamaño⁸.

Por otro lado, la estratificación del marco muestral de la EBS, correspondiente a la segunda fase, está conformada a partir de las 16 regiones del país.

2.1.3 Construcción del marco de selección

El marco de selección de la EBS fue construido desde un listado de personas de 18 años y más⁹ que fueron entrevistadas en la encuesta Casen en Pandemia 2020, durante los meses de noviembre de 2020 a enero de 2021, de esta manera se cuenta con los datos de contacto del hogar, en este caso el teléfono del informante de la encuesta Casen en Pandemia 2020.

A continuación, en la Tabla N°1 se presenta la composición del marco de selección a nivel de región y zona:

Tabla N°1: Número de personas del marco de selección por región y zona, EBS 2021

Región / Zona	N (personas)	%
Nacional	143.673	100,0%
Urbano	120.766	84,1%
Rural	22.907	15,9%
Arica y Parinacota	5.446	3,8%
Tarapacá	6.163	4,3%
Antofagasta	5.656	3,9%
Atacama	5.561	3,9%
Coquimbo	6.293	4,4%
Valparaíso	14.466	10,1%
Metropolitana	30.821	21,5%
O'Higgins	9.641	6,7%
Maule	9.770	6,8%
Ñuble	5.043	3,5%
Biobío	13.907	9,7%
Araucanía	9.119	6,3%
Los Ríos	6.222	4,3%
Los Lagos	8.160	5,7%
Aysén	3.415	2,4%
Magallanes	3.990	2,8%

Fuente: Observatorio Social, Ministerio de Desarrollo Social y Familia.

⁸ Es posible que en algunas comunas no existan los cinco grupos de tamaño porque no existen manzanas en el rango del grupo faltante.

⁹ Se excluyen 98 personas que en Casen en Pandemia 2020 corresponden a Servicio Doméstico Puertas Adentro (SDPA).

2.2 Estrategia muestral

Como ya mencionamos, la EBS posee un diseño probabilístico, bifásico y estratificado. Para efectos de este estudio y a partir de la segunda fase del diseño, se trabajará con 16 estratos explícitos de muestreo conformados por todas las regiones del país. Luego, al interior de cada uno de estos estratos, se realiza la selección de las personas en forma sistemática y con igual probabilidad al interior de cada estrato de muestreo. Además, se definen estratos implícitos (para efectos de ordenamiento), definidos por: comuna, unidad primaria de muestreo (UPM), hogar y orden de la persona dentro del hogar en la encuesta Casen en Pandemia 2020.

2.3 Unidad de muestreo

Respecto a las unidades de muestreo de cada fase estas son:

Primera fase: manzanas y secciones al interior de cada estrato¹⁰ (unidades primarias de muestreo) y viviendas al interior de cada manzana o sección (unidades secundarias de muestreo).

Segunda fase: personas de 18 años y más al interior de cada estrato de muestreo (región).

2.4 Método de selección de la muestra

Una vez definido el marco de selección de personas de 18 años y más logrados en la encuesta Casen en Pandemia 2020, el procedimiento es el siguiente:

- Se definen los estratos que corresponden a cada región.
- Dentro de cada estrato se definen estratos implícitos (para efectos de orden) conformados por información geográfica recogida en la encuesta Casen en Pandemia 2020, correspondiente a: comuna, unidad primaria de muestreo (UPM), hogar y orden de la persona dentro del hogar en la encuesta Casen en Pandemia 2020.
- Una vez ordenado el estrato, con la información geográfica, se procede a realizar la selección en forma sistemática y con igual probabilidad dentro de cada estrato de muestreo definido.

2.5 Tamaño muestral

Se definió una muestra objetivo de 10.400 personas, sin embargo, con la finalidad de anticipar pérdidas debido a problemas del marco muestral, problemas de no-contacto y posibles problemas de rechazo por parte de los seleccionados, es que se seleccionaron 20.800 personas (muestra con sobremuestreo). Ante este aumento de muestra objetivo cabe hacer notar que, dado el cambio del método de recolección de información desde una modalidad presencial a una modalidad telefónica,

¹⁰ Los estratos en Casen en Pandemia 2020 están conformados por la terna Comuna-Área-Grupo de tamaño para el área urbana y Comuna-Área para el área rural.

fue necesario ajustar la tasa de no respuesta que se considera en la estimación del sobremuestreo de la muestra objetivo. De acuerdo con la experiencia internacional y nacional (OIT, JPAL), la tasa de personas que no responden las encuestas realizadas a través de teléfono, u otros medios no presenciales, es mayor que en encuestas con levantamiento presencial. De hecho, la experiencia actual de la Encuesta Nacional de Empleo (ENE) aplicada por el Instituto Nacional de Estadísticas (INE), que cambió su forma de levantamiento por uno no presencial a mediados de marzo 2020, muestra tasas de logro en torno a 60%, mientras que durante un periodo normal (previo a octubre 2019) se encontraban alrededor del 90%, y en el caso de la encuesta Covid-19 el 50% de la muestra presentó los mismos problemas, debido a esto se optó por considerar una tasa del 50% de sobremuestreo.

2.5.1 Estimación del tamaño muestral

El tamaño muestral se determinó bajo un muestreo aleatorio simple, se calcula por región considerando el número de personas en viviendas particulares del marco de muestreo del INE, donde se fija el error absoluto en un 4% inicialmente. Luego se realizan algunos ajustes de la muestra según la distribución de la población por región del Censo 2017 para llegar a un tamaño de muestra país con la suma de los tamaños regionales. Con el objetivo de abarcar con exactitud los indicadores planteados en los objetivos de la encuesta, y suponiendo que su distribución es disímil, de manera general, la variable de diseño se define como una proporción para la cual se asume una varianza máxima ($p=q=0,5$).

Asumiendo muestreo aleatorio simple para determinar el tamaño muestral se tiene:

$$n = \frac{N * Z_{(1-\alpha/2)}^2 * S^2}{(N - 1) * e^2 + Z_{(1-\alpha/2)}^2 S^2} \quad (1)$$

Donde:

- N : Número de personas en viviendas particulares en el marco muestral del INE.
- $Z_{(1-\alpha/2)}^2$: Percentil de nivel $(1 - \alpha/2)$ de la distribución Normal, correspondiente a la estimación del intervalo de $(1 - \alpha)$ de confianza.
- S^2 : Cuasi varianza poblacional ($p*q$), usando varianza máxima.
- e^2 : Error absoluto esperado

Luego, la muestra estimada para esta encuesta es de 10.400 personas. A continuación, en la Tabla N°2, se describen los tamaños muestrales originales propuestos por región y zona:

Tabla N°2: Distribución del tamaño muestral por región y zona, EBS 2021

Región / Zona	Muestra objetivo	Distribución %	Error absoluto estimado	Error relativo estimado	Tamaño con sobremuestreo
Nacional	10.400	100,0	1,0%	2,0%	20.800
Urbano	9.078	87,3	1,0%	2,1%	17.367
Rural	1.322	12,7	2,7%	5,4%	3.433
Arica y Parinacota	600	5,7	4,0%	8,0%	1.200
Tarapacá	600	5,8	4,0%	8,0%	1.200
Antofagasta	600	5,8	4,0%	8,0%	1.200
Atacama	600	5,7	4,0%	8,0%	1.200
Coquimbo	600	5,8	4,0%	8,0%	1.200
Valparaíso	800	7,7	3,5%	6,9%	1.600
Metropolitana	1.100	10,6	3,0%	5,9%	2.200
O'Higgins	600	5,8	4,0%	8,0%	1.200
Maule	600	5,7	4,0%	8,0%	1.200
Ñuble	600	5,8	4,0%	8,0%	1.200
Biobío	700	6,7	3,7%	7,4%	1.400
La Araucanía	600	5,8	4,0%	8,0%	1.200
Los Ríos	600	5,8	4,0%	8,0%	1.200
Los Lagos	600	5,7	4,0%	8,0%	1.200
Aysén	600	5,8	4,0%	8,0%	1.200
Magallanes	600	5,8	4,0%	8,0%	1.200

Fuente: Observatorio Social, Ministerio de Desarrollo Social y Familia.

2.5.2 Objetivos de precisión

Se espera lograr a nivel Nacional un margen de error absoluto de 1 punto porcentual (pp.) y un margen de error relativo de un 2%. Por otro lado, a nivel regional se espera un margen de error absoluto máximo de 4 pp. y un margen de error relativo máximo de 8%.

2.6 Selección de unidades muestrales

Conformado el marco de selección, a partir del levantamiento de la encuesta Casen en Pandemia 2020, se procede a la selección de personas de 18 años y más para la EBS. Dicha selección se realiza por estrato de muestreo, que corresponde a región, en forma sistemática y con igual probabilidad.

La selección es implementada fijando una semilla aleatoria, a fin de que pueda ser replicable en cualquier momento.

Sean M_r el número de personas de 18 años y más encontrado en el estrato (región) y $m_{r\ sel}$ el número de personas de 18 años y más a seleccionar. Para la selección de $m_{r\ sel}$ personas, se ejecutan los siguientes pasos:

Paso 1

En primera instancia se ordenan de forma geográfica ascendente todas las personas; según región, zona, comuna, conglomerado (Cod_UPM), la identificación de la vivienda (id_vivienda), identificación del hogar (folio) e identificación de la persona (id_persona).

Paso 2

Al interior de cada estrato se calcula el período (k) que corresponde a:

$$k = \frac{M_r}{m_{r\ sel}} \quad (2)$$

Notar que " k " puede ser un número real, no entero (puede tener decimales).

Paso 3

Luego se determina el arranque " A " o primera selección, que corresponde a una semilla aleatoria propia para la encuesta.

Paso 4

Luego, se suma sucesivamente el período " k " al arranque " A " para obtener distintos valores, los que dan origen a la selección de unidades de la siguiente forma: " A ", " $A+k$ ", " $A+2k$ ", " $A+3k$ ", ..., " $A+(m_{r\ sel} - 1)k$ ".

La primera persona seleccionada es " A " y es un número entero, la segunda es el redondeo de " $A+k$ ", la tercera es el redondeo de " $A+2k$ " y así hasta la $m_{r\ sel}$ selección, dada por el redondeo de " $A+(m_{r\ sel} - 1)k$ ".

Luego, la probabilidad de seleccionar la j -ésima persona dentro del estrato r está dada por:

$$P_r(j) = \frac{m_{r\ sel}}{M_r} \quad (3)$$

2.7 Errores efectivos

A partir del levantamiento de la encuesta EBS, se calcularon los estadísticos asociados a parámetros de interés definidos por el Ministerio, en las siguientes variables:

- Insatisfacción con la vida (a1)¹¹ con la categoría de respuesta “Totalmente insatisfecho o Insatisfecho” con su vida, ver Tabla N°3.
- Balance afectivo negativo (a2) correspondiente al porcentaje (%) de personas que presentan un balance negativo, es decir, las emociones negativas (enojado(a), triste) superan las emociones positivas (feliz, tranquilo(a)), ver Tabla N°4.

Tabla N°3: Errores muestrales por región y zona para Insatisfacción con la vida, EBS 2021

Región / Zona	Insatisfecho con la vida		
	Estimación	Error absoluto	Error relativo
Nacional	12,6%	1,0%	7,8%
Urbano	12,4%	1,1%	8,5%
Rural	14,1%	2,6%	18,2%
Arica y Parinacota	15,5%	3,1%	20,1%
Tarapacá	13,6%	3,0%	22,2%
Antofagasta	12,0%	2,7%	22,1%
Atacama	10,6%	2,5%	23,5%
Coquimbo	12,2%	3,0%	24,4%
Valparaíso	12,0%	2,3%	19,0%
Metropolitana	13,3%	2,0%	15,3%
O'Higgins	10,9%	2,6%	23,9%
Maule	12,7%	2,9%	22,7%
Ñuble	10,6%	2,5%	23,3%
Biobío	12,3%	2,5%	20,4%
La Araucanía	12,0%	2,7%	22,2%
Los Ríos	12,4%	2,8%	22,3%
Los Lagos	13,1%	3,1%	23,6%
Aysén	10,0%	2,7%	26,9%
Magallanes	9,0%	2,3%	25,1%

Fuente: Observatorio Social, Ministerio de Desarrollo Social y Familia.

¹¹ Cuestionario EBS disponible en:
http://observatorio.ministeriodesarrollosocial.gob.cl/storage/docs/bienestar-social/Cuestionario_Encuesta_Bienestar_Social_2021.pdf

Tabla N°4: Errores muestrales por región y zona para Balance afectivo negativo, EBS 2021

Región / Zona	Balance afectivo negativo		
	Estimación	Error absoluto	Error relativo
Nacional	9,7%	0,9%	8,8%
Urbano	9,8%	0,9%	9,6%
Rural	8,9%	1,7%	19,4%
Arica y Parinacota	11,5%	2,9%	25,1%
Tarapacá	8,0%	2,1%	26,3%
Antofagasta	7,8%	2,3%	29,6%
Atacama	8,8%	2,3%	26,1%
Coquimbo	12,2%	3,0%	24,5%
Valparaíso	8,2%	1,9%	23,6%
Metropolitana	10,3%	1,8%	17,1%
O'Higgins	12,6%	2,8%	22,2%
Maule	10,1%	2,5%	24,5%
Ñuble	8,5%	2,5%	29,8%
Biobío	8,7%	2,1%	24,1%
La Araucanía	8,1%	2,1%	26,3%
Los Ríos	8,5%	2,4%	28,6%
Los Lagos	8,7%	2,5%	28,8%
Aysén	6,3%	1,9%	30,8%
Magallanes	6,1%	1,9%	31,1%

Fuente: Observatorio Social, Ministerio de Desarrollo Social y Familia.

3 FACTORES DE EXPANSIÓN DE BIENESTAR SOCIAL

Los factores de expansión se utilizan para obtener estimaciones que reflejen de manera precisa el comportamiento de la población objetivo. Estos pesos permiten garantizar que las estimaciones sean insesgadas y consistentes, que el error de muestreo sea pequeño condicionado al diseño muestral y al tamaño de la muestra, y que los errores de cobertura sean corregidos.

Como se mencionó antes, la EBS posee un diseño muestral bifásico donde en la primera etapa se tienen a los hogares logrados en la encuesta Casen en Pandemia 2020, y la segunda fase corresponde a un diseño estratificado, donde se realiza la selección de las personas de 18 años y más en forma sistemática y con igual probabilidad al interior de cada estrato de muestreo.

La construcción de los factores de expansión de la EBS contempla el desarrollo de un **factor de expansión de primera fase**, el cual corresponde al ponderador de los hogares logrados en la encuesta Casen en Pandemia 2020; un **factor de expansión de segunda fase**, que refleja cuántas personas en la población representa la persona seleccionada en la muestra de la EBS. Luego, al producto del **factor de primera y segunda fase** se le aplica una serie de **ajustes**: ajuste por no elegibilidad; ajuste por no respuesta, que mitiga el sesgo en las estimaciones producido por la ausencia de respuesta en las personas y un suavizamiento del factor de expansión. Finalmente, se realiza la **calibración del factor de expansión**, lo que permite mantener la consistencia con las proyecciones de población desarrolladas por el Instituto Nacional de Estadísticas (INE) y mejorar las propiedades estadísticas de los estimadores utilizados.

La construcción del factor de expansión de la EBS estuvo a cargo del MDSF, para lo cual se solicitó apoyo al INE para aplicar las metodologías de suavizamiento y calibración. A continuación, se describe la metodología utilizada para la construcción de cada ponderador utilizado en el cálculo final de los factores de expansión.

3.1 Factor de expansión de primera fase

El cálculo del factor de expansión de primera fase de la EBS da cuenta del diseño muestral de la encuesta Casen en Pandemia 2020 cuyas unidades son seleccionadas en dos etapas (selección del conglomerado, UPM, y la selección de la vivienda) y considera un ajuste por elegibilidad y un ajuste por no respuesta. De la encuesta Casen en Pandemia 2020 se tiene la información de los conglomerados y viviendas entrevistados, luego se realiza la corrección de elegibilidad y la corrección por no respuesta, de donde se obtiene la información de todas las personas que pertenecen a las viviendas que respondieron la encuesta Casen en Pandemia 2020. Lo anterior, se define de la siguiente manera:

a) El ponderador de selección de los conglomerados corregido w'_{hi} de Casen en Pandemia 2020, está dado por:

$$w'_{hi} = \hat{R}_h \cdot w_{hi} \quad (4)$$

Siendo

$$\hat{R}_h = \frac{M_h}{\sum_{i \in \Omega_h} w_{hi} \cdot M_{hi}} \quad (5)$$

y,

$$w_{hi} = \begin{cases} \frac{M_h}{n_h \cdot M_{hi}} & , \text{ Si unidad } i \text{ es una sección} \\ \frac{N_{hg}}{n_{hg}} & , \text{ Si unidad } i \text{ es una manzana} \end{cases} \quad (6)$$

Donde,

- \hat{R}_h Es el ajuste por omisión de conglomerados por estrato h
- M_h Es el número de viviendas en el estrato h .
- Ω_h Es el conjunto de conglomerados seleccionados en el estrato h .
- M_{hi} Es el número de viviendas del conglomerado i del estrato h .
- i Es el índice de la unidad primaria de muestreo o conglomerado (manzana o sección).
- h Es el índice del estrato de muestreo, o área (urbana, RAU o rural) de una comuna.
- w_{hi} Es el ponderador de selección base de la unidad primaria de muestreo o conglomerado (sección o manzana).
- n_h Es el número de secciones seleccionadas en el estrato h .
- N_{hg} Es el número de manzanas según el MMM2016, en el grupo de tamaño g del estrato h .
- n_{hg} Es el número de manzanas en la muestra, seleccionadas en el grupo de tamaño g del estrato h .

b) El ponderador de selección de la vivienda de Casen en Pandemia 2020 corregido por elegibilidad desconocida incorporando el ajuste por omisión de conglomerado, al diferenciarlo por el marco desde el cual fue seleccionada la vivienda:

$$w'_{hij} = \hat{R}_{h, know} \cdot w_{hij} \quad (7)$$

Siendo, w'_{hij} es el ponderador de selección de viviendas corregido por elegibilidad y $\hat{R}_{h, know}$ es la proporción de los pesos a redistribuir y se calcula como la razón entre dos estimaciones realizadas en base al ponderador de selección de viviendas.

Donde

$$\hat{R}_{h, know} = \frac{\sum_{i \in \Omega_h} \sum_{j \in \Theta_i} w_{hij}}{\sum_{i \in \Omega_h} \sum_{j \in \Theta_{i, know}} w_{hij}} \quad (8)$$

Θ_i	Es el conjunto de viviendas j seleccionadas en la manzana o sección i .
$\Theta_{i, know}$	Es el conjunto de viviendas j seleccionadas en la en la manzana o sección i y clasificadas como elegibles o no elegibles.
Ω_h	Es el conjunto de estratos h (áreas geográficas urbana, RAU o rural de las comunas).
w_{hij}	Es el ponderador de selección de la vivienda j en el conglomerado i del estrato h .

c) El ponderador de selección corregido por no respuesta w_{hij}^{NR} de Casen en Pandemia 2020 se expresa como,

$$w_{hij}^{NR} = \hat{R}_{g,R} \cdot w'_{hij} \quad (9)$$

Este ponderador se asigna a las viviendas elegibles entrevistadas. A las viviendas elegibles no entrevistadas se asigna un valor blanco (“missing”).

Siendo:

$$\hat{R}_{g,R} = \frac{\sum_{i \in \Phi_g} \sum_{j \in \Theta_{i, eleg}} w_{hij}}{\sum_{i \in \Phi_g} \sum_{j \in \Theta_{i, eleg, R}} w_{hij}} \quad (10)$$

Donde,

El numerador corresponde a la estimación del total de viviendas en la población y el denominador corresponde a la estimación del total de viviendas con elegibilidad conocida (viviendas elegibles y no elegibles). Ambas estimaciones se realizan con el ponderador de selección de viviendas, $\hat{R}_{g,R}$

g	Es el índice del grupo para la corrección de no respuesta.
Φ_g	Es el conjunto de manzanas o secciones asignadas al veintil g .
$\Theta_{g, eleg}$	Es el conjunto de viviendas pertenecientes a la unidad i y catalogadas como elegibles.
$\Theta_{g, eleg, R}$	Es el conjunto de viviendas pertenecientes a la unidad i , catalogadas como elegibles y que responden la encuesta.

Por ende, el ponderador de selección corregido por no respuesta w_{hij}^{NR} puede ser expresado como:

$$F1_{hij} = w_{hij}^{NR} = \hat{R}_{g,R} * w'_{hij} \quad (11)$$

Siendo, $F1_{hij}$ es factor de selección de viviendas logradas en la encuesta Casen en Pandemia 2020.

3.2 Factor de expansión de segunda fase

Una vez obtenido el marco de selección de las personas de 18 años y más que pertenecen a las viviendas logradas en la encuesta Casen en Pandemia 2020, correspondiente a 143.673 personas de 18 años y más por región (estrato EBS), se procede a seleccionar una muestra con sobremuestreo de 20.800 personas, de acuerdo con la siguiente probabilidad de selección:

$$F2_{r,k} = \frac{M_r}{m_{r\text{ sel}}} \quad (12)$$

Donde,

- r, k Es el índice del estrato región r , de la persona k en la EBS 2021.
- $F2_{r,k}$ Es el ponderador de selección de la persona k en el estrato región r .
- $m_{r\text{ sel}}$ Es el total de personas seleccionadas en la EBS 2021 por estrato región r .
- M_r Es el total de personas disponibles en el estrato región r , según levantamiento Casen en Pandemia 2020.

3.3 Factor de expansión de primera y segunda fase

El factor que da cuenta de la selección de personas para la EBS es aquel que considera ambas fases de muestreo, es decir, el factor que da cuenta de la selección de la k -ésima persona para la encuesta EBS, dado que ha sido entrevistada en la encuesta Casen en Pandemia 2020, se obtiene a partir de:

$$w_{r,k} = F1_{hij} * F2_{r,k} \quad (13)$$

Donde,

- $w_{r,k}$ Es el factor de selección de primera y segunda etapa por personas, factor teórico.

3.4 Ajustes al factor de expansión

Los siguientes ajustes aplicados al factor de expansión, se realizan según información proveniente del levantamiento de la EBS e información proveniente de Casen en Pandemia 2020, y tienen como objetivo corregir, entre otras cosas, la falta de respuesta y posibles errores de haber incluido en la muestra a personas que no pertenecen al universo de estudio.

3.4.1 Antecedentes de la disposición final de casos

Desde el cuestionario hoja de ruta obtenida del levantamiento, es posible obtener diferentes estados de las personas seleccionadas en relación con su elegibilidad (elegibles, no elegibles y elegibilidad desconocida). Estos estados se obtienen según el código de disposición final (CDF) registrado al momento del levantamiento. A partir de la información del levantamiento, se pueden identificar por ejemplo, a personas que ya no pertenecen al universo del estudio o bien ya no cumplen con las

restricciones de ser parte de las unidades seleccionadas (ya no forman parte del hogar logrado Casen). La Tabla N°5 se muestra el resultado de elegibilidad del levantamiento de la EBS¹²:

Tabla N°5: Cantidad de personas según relación de elegibilidad, levantamiento EBS 2021

Estado	personas	porcentaje
Personas elegibles que respondieron	10.921	52,5%
Personas elegibles que no respondieron	9.339	44,9%
Personas no elegibles	540	2,6%
Total	20.800	100,0%

Fuente: Observatorio Social, Ministerio de Desarrollo Social y Familia.

3.4.2 Ajuste por no elegibilidad

La muestra de la EBS contempló 20.800 personas incluyendo sobremuestreo. De ese total existen 20.260 personas elegibles y 540 personas consideradas no elegibles, principalmente porque la persona seleccionada ya no formaba parte del hogar en la encuesta Casen en Pandemia 2020, esto puede suceder porque la persona se cambió de vivienda o porque ya no es miembro de ese hogar.

El objetivo de la EBS es producir inferencias hacia la población que reside en viviendas particulares, por lo tanto, a partir de esta etapa no se consideran para fines analíticos aquellas personas que no conforman la población objetivo de la encuesta (no elegibles, tales como la persona seleccionada ya no formaba parte del hogar Casen en Pandemia 2020, teléfono corresponde a empresa, oficina de gobierno u otra organización, entre otras).

Este ajuste consiste en asignar un valor blanco (“missing”), en el factor de expansión de primera y segunda fase, a las personas con clasificación “no elegible”.

3.4.3 Ajuste por no respuesta

Las encuestas a menudo adolecen de no respuesta a la unidad, esto consiste en que una unidad seleccionada en la población objetivo (personas) y, por lo tanto, elegible para responder la encuesta, no provee ningún tipo de información, lo que significa que el cuestionario queda completamente vacío. Una consecuencia de la no respuesta a la unidad es que el tamaño muestral es más pequeño que lo planeado. Si la no respuesta es aleatoria, entonces aún es posible realizar estimaciones válidas, pero aumentará la varianza disminuyendo la precisión.

En la EBS de las 20.260 personas elegibles seleccionadas en la muestra, se lograron concretar 10.921 personas entrevistadas. Es decir, la tasa de respuesta de la encuesta fue de 53,9%.

El método utilizado para compensar la no respuesta es el método “Propensity Score” que consiste en modelar la probabilidad de respuesta de la EBS con variables obtenidas de la encuesta Casen en Pandemia 2020, las que podrían incidir en la motivación por responder o participar en dicha encuesta.

¹² Para más detalles de los CDF de la EBS 2021 ver Anexo N°1

Dado que la EBS es una encuesta con primera fase en la encuesta Casen en Pandemia 2020, se dispone de la toda información proveniente de ella para las personas que no respondieron la EBS. En la Tabla N°6 se presentan las variables incorporadas en el modelo:

Tabla N°6: Variables de la encuesta Casen en Pandemia 2020 consideradas en modelo de no respuesta de la EBS 2021

Sexo de la persona
Edad al cuadrado de la persona
Si la persona es jefe de hogar
Si el hogar de la persona es unipersonal
Número de personas en el hogar
Si la persona asiste o participa en un establecimiento educacional
Años de educación de la persona
Años de educación de las personas del hogar mayores de 25 años
Ingreso total del hogar corregido por no respuesta
Si la persona ha tenido tratamiento médico por alguna enfermedad o condición de salud
Etnia de la persona
Zona (urbano / rural)
Interacción de zona (urbano/rural) y 4 tramos etarios (18-29; 30-44; 45-59, 60 y más)

Fuente: Observatorio Social, Ministerio de Desarrollo Social y Familia.

A través del modelo “Propensity score” se calcula la probabilidad de responder de cada una de las personas de la encuesta, posteriormente se ordenan las unidades de mayor a menor según las probabilidades estimadas, creando estratos o celdas de ajuste, que serán los niveles sobre los cuales se realizan las correcciones de no respuesta, según la agrupación de individuos que posee características similares asociadas a la respuesta de la encuesta. Se generan 20 celdas de ajustes o veintiles, generando 20 grupos de igual cantidad de personas ordenados por la probabilidad antes mencionada.

Una vez calculadas las 20 celdas de ajuste, se procede a estimar el factor de personas ajustado por la falta de respuesta, el cual está dado por la siguiente ecuación:

$$w_{r,k}^{NR} = \hat{R}_v * w_{r,k} \quad (14)$$

$$\hat{R}_v = \frac{\sum_{k \in \theta_{r,eleg,v}} w_{r,k}}{\sum_{k \in \theta_{r,log,v}} w_{r,k}}, \quad \text{si la persona } k \text{ es elegible y responde} \quad (15)$$

Donde,

$\theta_{r,eleg,v}$ conjunto de personas seleccionadas en el estrato r (región) clasificadas como elegibles dentro del veintil v .

$\theta_{r,log,v}$ conjunto de personas logradas en el estrato r (región) clasificadas como

	elegibles y que responden la encuesta, en el veintil v .
$w_{r,k}$	factor teórico
v	veintil, $v=1, 2, 3, \dots, 20$.
.	si la persona la persona k es elegible y no responde toma el valor perdido o missing.

En el Anexo N°2 se muestra la distribución del factor de corrección de no respuesta según veintil.

3.4.4 Suavizamiento del factor de expansión (método mixto)

Para el suavizamiento de los factores de expansión de la EBS se aplicó una metodología mixta que utiliza el método de contracción a la media (CM) y el método R-K, los cuales se describen a continuación¹³.

Contracción a la Media (CM)

El objetivo del ajuste CM es volver más homogéneos los factores de expansión, acercándolos a la media. La contracción a la media tiene 2 propiedades deseables: (i) Mantiene la suma total de los factores. (ii) Al ser una transformación lineal y aplicando las propiedades de la esperanza, se tiene que el CV de los factores suavizados será siempre menor o igual que el CV de los factores no suavizados (Guerrero, 2019; Touron et al.2020a).

La magnitud de la contracción está dada por el factor β , mediante la siguiente ecuación:

$$w_i^s = \beta w_i + (1 - \beta)\bar{w} ; \text{ con } 0 \leq \beta \leq 1 \quad (16)$$

Donde:

w_i : corresponde al valor puntual del factor de expansión corregido por no respuesta

\bar{w} : corresponde a la media del factor a nivel región-zona. Notar que este valor se calcula a nivel de factores atípicos y no atípicos.

Método R-K

El objetivo del método R-K es disminuir o aumentar los factores en una proporción (r o k), dependiendo de si los valores que toman se consideran atípicos o no atípicos. De esta forma, se busca acercar los valores atípicos y no atípicos, mediante un tratamiento diferenciado.

El procedimiento consiste en encontrar las proporciones r (factor de disminución de factores atípicos) y la proporción k (factor de aumento de los factores no atípicos) de tal forma que se cumplan 2

¹³ Esta metodología permite mejorar la dispersión de los factores de expansión y mantener el orden de la distribución inicial (Touron et al.2020a). Además, esta metodología es la misma que implementa el INE en su encuesta de empleo y que implementa el MDSF en su encuesta Covid-19.

restricciones: (i) que la suma de los factores de expansión siga siendo la misma, (ii) que el mayor factor de los no atípicos suavizados debe ser menor al menor factor de los atípicos suavizados. Lo último con el objetivo de mantener el orden en la distribución de los factores.

En base a las 2 restricciones anteriores se definen las siguientes igualdades:

$$k = r \left(\frac{\sum_1^n W_i}{\sum_j^m w_j} \right) \quad (17)$$

$$r = \lambda r_0 \quad (18)$$

$$r_0 < \frac{[(W_n - w_1) \sum_j^m w_j]}{[w_1 \sum_1^n W_i + W_n \sum_j^m w_j]} \quad (19)$$

Dado que el valor de r_0 se define en base a una desigualdad, el factor r que se aplica se define al resolver el valor de r_0 en igualdad y luego multiplicar ese valor por una fracción menor a 1 (λ), que indica la proximidad que existirá entre el mayor de los factores no atípicos y el menor de los factores atípicos.

Método Mixto

El método mixto combina los procedimientos de CM y R-K, notar que los ajustes en la EBS 2021 se realizaron a nivel de región-área. El método mixto se aplicó de la siguiente forma:

- Se separan los factores de expansión en factores típicos y atípicos¹⁴.
- Los factores no atípicos se ajustan con un $\beta = 0.8$
- Los factores atípicos se ajustan con un $\beta = 0.2$
- Luego se aplica método de R-K con un $\lambda = 0.99$

3.5 Calibración de los factores de expansión

La utilización de marcos muestrales desactualizados (o no exhaustivos), excluye conglomerados de viviendas, generando omisión o cobertura insuficiente en las subpoblaciones que residen en las áreas desactualizadas. Si esta exclusión ocurre en subpoblaciones cuyas características sociodemográficas son muy distintas de las características de las personas que no son excluidas del marco muestral, se pueden generar sesgos en la estimación de los parámetros de interés realizados a partir de la encuesta (INE, 2019). Una solución al problema anterior es la calibración de los factores de expansión.

¹⁴ Para identificar los factores atípicos se utilizó como métrica el percentil de distribución del factor de expansión, de esta forma se definió el valor $h = 1.5(p_{75} - p_{25})$. Factores de expansión con un valor inferior a $(p_{25} - h)$ o superior a $(p_{75} + h)$ fueron considerados atípicos.

El enfoque de calibración consiste en el cálculo de los factores de expansión incorporando información auxiliar específica que permite ajustar las estimaciones a ciertos totales poblacionales conocidos. De esta forma, el objetivo es obtener factores de expansión que estén lo más cercano posible a los factores de expansión originales de diseño, pero que a la vez cumplen con ajustarse a ciertos totales poblacionales definidos, los que se conocen como marginales o ecuaciones de calibración (Särndal, 2007; Zang, 2000).

Para el ajuste de la EBS, se contó con información de las proyecciones de población de 18 años y más, al 30 de abril del 2021. Después de un análisis exploratorio se eligieron las siguientes marginales de calibración:

- Población total por región
- Población total por sexo
- Población total por zona
- Población total por tramo de edad: [18-29], [30-44], [45-59], [60+].
- Población total por región y zona
- Población total por región y sexo
- Población total por zona y sexo
- Población total por tramo de edad y sexo

Como resultado de este procedimiento de calibración, los factores de expansión se calibraron a un total de 15.164.390 personas de 18 años y más, tal como se observa en la Tabla N°7.

Tabla N°7: Distribución de la muestra y tamaño poblacional expandido por región y zona, EBS 2021

Región / Zona	Tamaño de muestra	Tamaño expandido a población mayor de 18 años
Nacional	10.921	15.164.390
Urbano	9.307	13.405.487
Rural	1.614	1.758.903
Arica y Parinacota	617	191.704
Tarapacá	623	285.450
Antofagasta	623	527.209
Atacama	621	233.615
Coquimbo	623	641.782
Valparaíso	851	1.541.851
Metropolitana	1.138	6.412.096
O'Higgins	644	766.289
Maule	620	876.150
Ñuble	645	399.028
Biobío	761	1.289.286
La Araucanía	664	776.319
Los Ríos	631	315.343
Los Lagos	633	687.519
Aysén	618	80.017
Magallanes	609	140.732

Fuente: Observatorio Social, Ministerio de Desarrollo Social y Familia.

4 Anexos

Anexo N°1: Códigos de Disposición Final de Casos, levantamiento EBS 2021

	Personas	Porcentaje
Total	20.800	100,0%
Entrevistados		
110. Entrevista completa	10.921	52,5%
No entrevistados, se logra contacto telefónico		
200. Concertó cita con el informante o seleccionado	0	0,0%
211. Se rechazó la entrevista	2.643	12,7%
212. Entrevista interrumpida, sin rechazo	79	0,4%
2250. Seleccionado no disponible, se contacta a otra persona	1.606	7,7%
2251. Seleccionado está fuera del país (período mayor a 6 meses)	46	0,2%
231. Muerte del seleccionado	58	0,3%
232. Seleccionado impedido físico/mental para contestar	470	2,3%
233. Seleccionado no habla idioma español	9	0,0%
234. Calidad de sonido mala o intermitente	16	0,1%
235. Lugar o actividad del informante no permite una entrevista	53	0,3%
236. Teléfono equivocado o no corresponde a la dirección seleccionada	200	1,0%
237. Hogar ya no reside en la vivienda	282	1,4%
238. Otra razón de no entrevista	222	1,1%
No entrevistados, no se logra contactar telefónicamente		
322. Teléfono suena ocupado	286	1,4%
323. Teléfono no contesta (marca tono, pero sin respuesta)	1.590	7,6%
324. Responde buzón de voz	883	4,2%
325. Fuera de área de cobertura, apagado (sin buzón de voz)	445	2,1%
326. Número telefónico no existe	319	1,5%
327. Problemas técnicos	18	0,1%
390. Otra razón (no se logra contacto)	114	0,5%
Teléfono no corresponde a vivienda particular elegible		
410. Seleccionado ya no forma parte del hogar CASEN en Pandemia 2020	539	2,6%
451. Teléfono corresponde a empresa, oficina de gobierno u otra organización	1	0,0%
452. Teléfono corresponde a institución (hospital, cárcel, asilo de ancianos, etc.)	0	0,0%
453. Teléfono corresponde a dormitorio colectivo, militar, de trabajo, internado	0	0,0%
454. Vivienda en demolición, incendiada, destruida o erradicada	0	0,0%
461. Vivienda particular desocupada	0	0,0%
462. Vivienda de veraneo o de uso temporal	0	0,0%

Fuente: Observatorio Social, Ministerio de Desarrollo Social y Familia.

Anexo N°2: Distribución del factor de corrección promedio por no respuesta según veintil, EBS 2021

Celda de ajuste de no respuesta o Veintiles	Componentes del ponderador de corrección por no respuesta			
	Total	Personas que responden	Personas que no responden	\hat{R}_v
Total	20.260	10.921	9.339	0,5
1	1.013	280	733	0,3
2	1.013	404	609	0,4
3	1.013	449	564	0,4
4	1.013	468	545	0,5
5	1.013	474	539	0,5
6	1.013	512	501	0,5
7	1.013	538	475	0,5
8	1.013	532	481	0,5
9	1.013	539	474	0,5
10	1.013	560	453	0,5
11	1.013	581	432	0,5
12	1.013	579	434	0,6
13	1.013	587	426	0,6
14	1.013	608	405	0,6
15	1.013	593	420	0,6
16	1.013	584	429	0,6
17	1.013	626	387	0,6
18	1.013	658	355	0,6
19	1.013	670	343	0,7
20	1.013	679	334	0,7

Fuente: Observatorio Social, Ministerio de Desarrollo Social y Familia.