

DISEÑO METODOLÓGICO DE LA MUESTRA V4.0





ENCUESTA CONTINUA DE EMPLEO

UNIDAD DE MUESTREO Y LOGÍSTICA

Bolivia

OFICINA CENTRAL LA PAZ

Avenida José Carrasco N° 1391 - Miraflores, Telf.: (591-2) 2222333 • Fax (591-2) 2222885

 www.ine.gob.bo  info@ine.gob.bo  [ineboliviaoficial](https://www.facebook.com/ineboliviaoficial)  [@INE_Bolivia](https://twitter.com/INE_Bolivia)

ÍNDICE

1.	Introducción	5
2.	Antecedentes	5
3.	Marco Conceptual Estadístico	6
	3.1 Tipo de operación estadística.....	6
	3.2 Período de referencia	6
	3.3 Período de recolección	6
4.	Diseño muestral.....	6
	4.1 Objetivos del diseño de la muestra	6
	4.2 Universo.....	7
	4.3 Población objetivo	7
	4.4 Tipo de Muestreo.....	7
	4.5 Unidades básicas de investigación.....	8
	4.6 Unidad de análisis	8
	4.7 Unidad de muestreo	8
	4.8 Cobertura.....	8
	4.9 Marco Muestral.....	8
	4.10 Tamaño de la muestra	9
	4.11 Selección de la muestra	10
	4.12 Traslape de la Muestra	10
	4.13 Asignación de la Muestra en los Paneles	11
	4.14 Distribución de la Muestra en el Tiempo.....	12
	4.15 Limitaciones	13
	4.16 Representatividad.....	13
5.	Factores de Expansión.....	13
	5.1 Ajustes a los Factores Expansión	15
	5.1.1 Por Incidencias Finales.....	15
	5.1.2 Método post estratificación (4t-2015 al 4t-2018).	15
	5.1.4 Métodos para la corrección del sesgo	18
	5.1.5 Factores de expansión para un periodo "t".	18
	5.1.6 Ajuste por probabilidad de respuesta (Propensity score)	19
	5.2 Calibración de factores de expansión (wkcal).....	23
6.	Estimadores y Errores de Muestreo.....	24

7. Anonimización de códigos de UPM	24
8. Bibliografía	25



ÍNDICE DE TABLAS

Tabla 1: Estratificación.....	9
Tabla 2: Distribución de la muestra de UPM y viviendas por departamento y área geográfica (trimestral)	9
Tabla 3: Rotación de Panel ECE (urbano).....	10
Tabla 4: Tamaño de muestra (UPM) distribuido por departamento y semana.....	12
Tabla 5: Tamaño de muestra (viviendas) distribuido por departamento y semana.....	12

ÍNDICE DE ILUSTRACIONES

Ilustración 1: Traslape de los paneles entre trimestres	11
Ilustración 2: Distribución de los estratos por panel	11
Ilustración 3: Esquema de ajuste de factores de expansión (corrección de sesgo)	19
Ilustración 4: Esquema de los respondientes y no respondientes entre un tiempo "t-1" y "t"	20
Ilustración 5: Esquema de los Pesos ajustados por el Propensity score y el ajuste de los registros Nuevos	23



ENCUESTA CONTINUA DE EMPLEO

1. Introducción

El Instituto Nacional de Estadística (INE) en su política de fortalecimiento y actualización de las estadísticas sectoriales y en particular de los indicadores del mercado de trabajo, ha diseñado y formulado la propuesta para llevar a cabo una Encuesta Continua de Empleo (ECE). La misma se viene desarrollando desde el 4to. Trimestre del año 2015, con cobertura nacional (urbana y rural).

El presente documento fundamenta la metodología del diseño muestral empleados en la Encuesta Continua de Empleo de panel y permitirá conocer la dinámica de la desocupación urbana.

El documento incluye los antecedentes, objetivos, marco conceptual, diseño muestral empleado en la Encuesta Continua de Empleo – ECE y factores de expansión.

2. Antecedentes

De acuerdo a lo estipulado en el Decreto Ley 14100, el Instituto Nacional de Estadística (INE) es el órgano ejecutivo y técnico del Sistema Nacional de Información Estadística (SNIE); el INE es una entidad descentralizada con autonomía administrativa y de gestión que tiene por responsabilidad la dirección, planificación, ejecución, control y coordinación de las actividades estadísticas del Sistema.

El Instituto Nacional de Estadística viene recabando información de empleo desde la gestión 1978. Las encuestas específicas de empleo relacionadas al comportamiento de la oferta de trabajo y sus características, se vienen desarrollando en el INE desde el año 2007, inicialmente con la "Encuesta Nacional de Empleo Urbano" ENEU; encuesta de hogares que abarcó únicamente las ciudades capitales, más la ciudad de El Alto, cuyos dominios de estudio dieron resultados trimestrales: Eje central (La Paz, Cochabamba, Santa Cruz) y Nacional, también proporcionaron resultados anuales: Chuquisaca, La Paz, Cochabamba, Oruro, Potosí, Tarija, Santa Cruz, Beni-Pando (agregado).

Posteriormente, se desarrolló la Encuesta Trimestral de Empleo (ETE), que se efectuó a partir del 1er. Trimestre del 2009 al 1er. Trimestre del 2011, encuesta de diseño panel, los principales indicadores que se obtienen del mercado de trabajo son: la medición de la desocupación y las características de los desocupados por edad y nivel de educación, en la cual hacen mediciones sucesivas sobre un mismo grupo de observación en distintos momentos de tiempo, que sirven para conocer los cambios que experimentan los individuos, con relación a distintas variables o fenómenos de interés.

La Encuesta Continua de Empleo genera información estadística sobre el comportamiento de la desocupación urbana, la oferta y movilidad del trabajo urbano y rural de la población boliviana, para la toma de decisiones de políticas públicas, nacionales y regionales.

3. Marco Conceptual Estadístico

3.1 Tipo de operación estadística

La Encuesta Continua de Empleo – ECE es una investigación que se realiza a través de encuestas considerando la temporalidad.

- Las que se realizan en un solo momento
- Las que recogen información periódicamente

Las que se realizan en un solo momento temporal (investigaciones puntuales) son efectuadas en el área rural y resto urbano (centros poblados) y las que se realizan periódicamente (investigaciones longitudinales) efectuadas en el área urbana y el tipo de diseño es de panel.

3.2 Periodo de referencia

La información investigada a partir de la Encuesta Continua de Empleo es realizada bajo los siguientes periodos de referencia: semana pasada, últimas cuatro semanas y últimos doce meses.

3.3 Periodo de recolección

El periodo de recolección de la información para la Encuesta Continua de Empleo es de acuerdo al cronograma semanal, la misma se lleva a cabo mediante entrevistas directas a los miembros del hogar en cada vivienda seleccionada, a lo largo de todo el año a nivel nacional según plan de recorridos por departamento, urbano y rural.

4. Diseño muestral

4.1 Objetivos del diseño de la muestra

El objetivo principal del diseño de la muestra para la ECE está enmarcado en los objetivos de la encuesta; proporcionar información estadística mensual y trimestral, que refleje el comportamiento de la desocupación del área urbana y rural.

Los objetivos específicos del diseño muestral son:

- Obtener estimaciones de coyuntura sobre los niveles de empleo, desocupación urbana de la población en edad de trabajar.
- Determinar el tamaño de la muestra por departamento (Chuquisaca, La Paz, Cochabamba, Oruro, Potosí, Tarija, Santa Cruz, Beni y Pando, estos

dos últimos departamentos de manera conjunta), los cuales permitan efectuar estimaciones de las variables de interés.

Realizar desagregaciones mayores a los planteados en el diseño, estará sujeto a la valoración de los errores muestrales.

4.2 Universo

La investigación está dirigida al conjunto de hogares constituidos en viviendas ocupadas particulares de los nueve departamentos del país del área urbana y rural, donde residen personas de 10 años o más, excluyendo así a las personas que habitan en viviendas colectivas, como hospitales, cárceles, conventos, cuarteles y otros, pero incluye a las personas que residen en viviendas particulares dentro de dichos centros, como porteros, conserjes y otros.

4.3 Población objetivo

Todos los hogares y miembros de un hogar mayores a 10 años, que residen en viviendas particulares seleccionadas.

4.4 Tipo de Muestreo

El tipo de muestreo es probabilístico, estratificado, por conglomerado y bietápico.

Probabilístico, La unidad de selección, es decir las viviendas, tienen una probabilidad conocida y distinta de cero de ser seleccionadas para formar parte de la muestra.

Estratificado, previo a la selección de conglomerados, estos, son clasificados con el propósito de mejorar la precisión de los estimadores y permite un mejor control en la distribución de la muestra.

Bietápico, porque los elementos pertenecientes a la muestra se seleccionan en dos etapas:

Etapas 1: Selección de conglomerados (UPM), la selección se la realiza proporcional al número de viviendas en la UPM.

Etapas 2: Selección de viviendas dentro la UPM elegida, la selección en esta segunda etapa es de manera sistemática con arranque aleatorio.

Por conglomerados:

Debido a que las unidades objeto de estudio (viviendas/hogares) se encuentran agrupadas en áreas conformadas, con fines de muestreo, en base a la información y cartografía del Censo de Población y Vivienda 2012, denominadas conglomerados (UPM).

En general la formación de conglomerados tiende a incrementar la varianza muestral cuando no cumple la heterogeneidad dentro del conglomerado. Sin embargo, su utilización en el diseño muestral permite flexibilidad en el número de etapas de selección y reduce substancialmente los costos asociados a la construcción del marco y a la realización del trabajo de campo.

4.5 Unidades básicas de investigación

Las unidades básicas de la investigación son los hogares dentro las viviendas particulares ocupadas.

4.6 Unidad de análisis

Es la población residente de 10 años y más de edad en las viviendas seleccionadas.

4.7 Unidad de muestreo

En su última etapa, es la vivienda.

4.8 Cobertura

La Encuesta Continua de Empleo considera una cobertura geográfica sobre los nueve departamentos y el área urbana y rural.

4.9 Marco Muestral

Características del Marco Muestral MM-2012:

- Está basada en el Censo de Población y Vivienda de 2012 (CNPV-2012), la Actualización Cartográfica Multipropósito (ACM-2010-2012) y el Censo Nacional Agropecuario de 2013 (CNA-2013). Por tanto, el MM-2012 se describe como un marco de áreas y listas.
- Cubre todo el territorio nacional incluyendo las áreas de difícil acceso.
- Está constituido por unidades primarias de muestreo-UPM que son áreas geográficas codificadas alfa numéricas a nivel nacional (urbana y rural), con 102.593 conglomerados.
- El MM-2012 incluye la estratificación de áreas geográficas y socioeconómicas

Para la construcción de la estratificación socioeconómica se consideró las variables estructurales del Censo de Población y Vivienda 2012: Características de la vivienda; Acceso a servicios básicos; Hacinamiento; Capital Humano; Educación; Patrimonio de los hogares; y Capacidad para la generación de ingresos. Conformando cuatro (4) estratos socioeconómicos que se describen en la tabla siguiente:

Tabla 1: Estratificación

Estrato Estadístico	Descripción	Reagrupación Estadístico
1	Bajo	1. Bajo
2	Medio Bajo	
3	Medio Alto	2. Alto
4	Alto	

Fuente: Instituto Nacional de Estadística

4.10 Tamaño de la muestra

El tamaño de la muestra para la ECE es de 17.940 viviendas para el trimestre. Los tamaños de muestra fueron calculados en base a la variable: Tasa de desocupación abierta de la Encuesta Trimestral de Empleo 2009.

El diseño permite:

- Medir los cambios entre-trimestres (área urbana) – Bajo un esquema de rotación. Donde el tamaño de muestra alcanza a 14.028 viviendas por trimestre (UPM de Panel).
- Ampliar la cobertura geográfica en el área rural – Bajo un diseño puntual, con una muestra de 3.912 viviendas por trimestre (UPM Sin Panel).

Distribución de la Muestra por Departamento y área geográfica

El número de UPM planificadas para la ECE es de 1.495 UPM por cada trimestre, distribuidas en 13 semanas, la distribución a nivel departamental (urbana – rural) es la siguiente:

En cada una de las UPM se selecciona 12 viviendas.

Tabla 2: Distribución de la muestra de UPM y viviendas por departamento y área geográfica (trimestral)

Departamento	UPM			VIVIENDAS		
	Urbano	Rural	Total	Urbano	Rural	Total
Total	1238	257	1495	14856	3084	17940
Chuquisaca	95	10	105	1140	120	1260
La Paz	288	79	367	3456	948	4404
Cochabamba	218	43	261	2616	516	3132
Oruro	67	12	79	804	144	948
Potosí	108	24	132	1296	288	1584
Tarija	64	17	81	768	204	972
Santa Cruz	237	50	287	2844	600	3444
Beni	95	10	105	1140	120	1260
Pando	66	12	78	792	144	936

Fuente: Instituto Nacional de Estadística

4.11 Selección de la muestra

La selección de la muestra es bietápico, porque los elementos pertenecientes a la muestra se seleccionan en dos etapas descritas en una sección anterior

Esquema de Rotación de la Muestra por panel

Tabla 3: Rotación de Panel ECE (urbano)

Rotación de Panel	<i>Gestión i</i>
	<i>Trimestre j</i>
	<i>Ronda z</i>
z	4ta Visita
$z + 1$	3ra Visita
$z + 2$	2da Visita
$z + 3$	1ra Visita

Fuente: Instituto Nacional de Estadística

Donde¹:

i – esima: Gestión de referencia de la encuesta (IV trimestre 2015, 2016, 2017, ...)

j - esima: Trimestre de la gestión (I, II, III, IV)

z - esima: Número de ronda de panel dentro de la gestión y trimestre correspondientes (donde $z = 1, 2, 3, \dots$)

Los hogares de las UPM de Panel, son visitados durante 4 trimestres continuos y posteriormente salen de la muestra permanentemente. En cada trimestre tiene conformado 4 paneles de rotación (grupos de rotación). Donde estos grupos se encuentran equilibrados en los estratos a nivel temporal, a lo largo del trimestre.

4.12 Traslape de la Muestra

El Traslape de las muestras entre dos trimestres consecutivos (muestra en común) corresponde al 75% y favorece a medir los cambios existentes de un trimestre a otro, por otro lado, el 25% es renovado (la forma en que se produce la renovación es denominada "Esquema de rotación").

Este traslape disminuirá su precisión para una agregación a lo largo de varios periodos bajo el siguiente esquema:

- Entre un trimestre y el siguiente existe un 75% de muestra en común, encuestada en los mismos meses y semanas de esos trimestres.
- Entre un trimestre y otro, separados por un trimestre intermedio, existe un 50% de muestra en común.
- Entre un trimestre y otro separado por dos trimestres intermedios existe un 25% de muestra en común.

¹ $i=2025$; $j=III$; $z=39$ (trigésima novena ronda)

- Entre un trimestre y el mismo del año siguiente, se carece de una muestra en común.

Por el contrario, un bajo Traslape mejora la precisión cuando se agrega muestra, pero disminuye la precisión de la estimación del cambio entre periodos sucesivos. A continuación, se ilustra el Traslape de los paneles entre trimestres durante un año:

Ilustración 1: Traslape de los paneles entre trimestres

1° trimestre	2° trimestre	3° trimestre	4° trimestre
Sin Panel			
P1	Sin Panel		
P2	P2	Sin Panel	
P3	P3	P3	Sin Panel
P4	P4	P4	P4
	P5	P5	P5
		P6	P6
			P7
	75% de Traslape con el trimestre anterior.	50% de Traslape entre un trimestre y otro, separados por un trimestre intermedio.	25% de Traslape entre un trimestre y otro, separados por dos trimestres intermedios.

Fuente: Instituto Nacional de Estadística

4.13 Asignación de la Muestra en los Paneles

Cada uno de los 4 paneles de rotación (o grupos de rotación), representa a una cuarta parte de la muestra que se tiene por trimestre. Estos paneles se encuentran equilibrados a nivel estrato socio-económico y a nivel temporal en las semanas a lo largo del trimestre.

A continuación, se expone la asignación por panel según estrato:

Ilustración 2: Distribución de los estratos por panel

	ROTACION						
	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇
1 Estrato Bajo	a ₁	b ₁	c ₁	d ₁	a ₁	b ₁	c ₁
2 Estrato Medio Bajo	a ₂	b ₂	c ₂	d ₂	a ₂	b ₂	c ₂
3 Estrato Medio Alto	a ₃	b ₃	c ₃	d ₃	a ₃	b ₃	c ₃
4 Estrato Alto	a ₄	b ₄	c ₄	d ₄	a ₄	b ₄	c ₄
	Trimestre 1						
		Trimestre 2					
			Trimestre 3				
				Trimestre 4			

Fuente: Instituto Nacional de Estadística

$a_i, b_i, c_i, d_i =$ Numero de UPM por estrato $\forall i$ de 1 a 4

Es relevante remarcar que existe parte de la muestra, aproximadamente 22% de la muestra independiente del total en el trimestre, correspondiente al área rural y resto

urbano (centros poblados). Similarmente se encuentran equilibrados a nivel estrato socio-económico y a nivel temporal en las semanas a lo largo del trimestre. En estos casos la visita es una sola vez y se retiran de la muestra.

4.14 Distribución de la Muestra en el Tiempo

- La muestra se distribuye a lo largo de 13 semanas en un Trimestre de manera que el número de UPM por departamento y semana sean constantes.
- Cada UPM seleccionada tiene asignada una semana de referencia dentro del trimestre. Esta semana es la misma para todos los trimestres, no cambia de trimestre a trimestre.
- Las muestras de diseño puntual (UPM Sin Panel), fueron distribuidas a lo largo de las 13 semanas de referencia.

Tabla 4: Tamaño de muestra (UPM) distribuido por departamento y semana

DEPARTAMENTO	Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Semana 6	Semana 7	Semana 8	Semana 9	Semana 10	Semana 11	Semana 12	Semana 13	Total
Nacional	115	115	115	115	115	115	115	115	115	115	115	115	115	1495
Chuquisaca	8	8	8	8	9	8	8	8	8	8	8	8	8	105
La Paz	28	28	28	29	28	28	28	29	28	28	28	28	29	367
Cochabamba	20	21	20	20	20	20	20	20	20	20	20	20	20	261
Oruro	7	6	6	6	6	6	6	6	6	6	6	6	6	79
Potosí	10	10	10	10	10	11	10	10	10	10	11	10	10	132
Tarija	6	6	7	6	6	6	7	6	6	6	6	7	6	81
Santa Cruz	22	22	22	22	22	22	22	22	22	23	22	22	22	287
Beni	8	8	8	8	8	8	8	8	9	8	8	8	8	105
Pando	6	6	6	6	6	6	6	6	6	6	6	6	6	78

Fuente: Instituto Nacional de Estadística

Tabla 5: Tamaño de muestra (viviendas) distribuido por departamento y semana

DEPARTAMENTO	Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Semana 6	Semana 7	Semana 8	Semana 9	Semana 10	Semana 11	Semana 12	Semana 13	Total
Total	1380	1380	1380	1380	1380	1380	1380	1380	1380	1380	1380	1380	1380	17940
Chuquisaca	96	96	96	96	108	96	96	96	96	96	96	96	96	1260
La Paz	336	336	336	348	336	336	336	348	336	336	336	336	348	4404
Cochabamba	240	252	240	240	240	240	240	240	240	240	240	240	240	3132
Oruro	84	72	72	72	72	72	72	72	72	72	72	72	72	948
Potosí	120	120	120	120	120	132	120	120	120	120	132	120	120	1584
Tarija	72	72	84	72	72	72	84	72	72	72	72	84	72	972
Santa Cruz	264	264	264	264	264	264	264	264	264	276	264	264	264	3444
Beni	96	96	96	96	96	96	96	96	108	96	96	96	96	1260
Pando	72	72	72	72	72	72	72	72	72	72	72	72	72	936

Fuente: Instituto Nacional de Estadística

4.15 Limitaciones

- Perdida de Muestra

Se presenta registros de viviendas en los que se tiene ausencia de información en alguno de los periodos que se pretende analizar (ausencia de personas o la vivienda familiar). Estos registros de "NO RESPUESTA" son debidos a las salidas de las personas en un grupo familiar o bien al cambio del grupo familiar que habita la vivienda, cambio de domicilio, hogares seleccionados rechazan seguir participando.

A esto también se denomina movimiento de un hogar completo y causa un desgaste de la muestra del panel, la dimensión será las personas que dejen de formar parte del panel.

En los casos de traslados (cambio de domicilio de todo el hogar), donde se verifica que hay otro hogar habitando la vivienda, se realiza la entrevista al nuevo hogar, de tal forma no perder información para el análisis mensual.

Si una o varias personas de un hogar se alejan, pero no la totalidad. En este caso, la o las que siguen en esa vivienda siguen formando parte del panel. A esto se denomina movimiento de personas dentro de un hogar y se considera que la persona fue una pérdida de muestra del panel. Sin embargo, es necesario aclarar que si la persona regresa puede volver a formar parte del panel en el siguiente trimestre.

- Inconsistencias de respuesta

Analizando dos trimestres consecutivos, por ejemplo, es posible encontrar personas que han trabajado en un trimestre y que posteriormente declaran que están buscando su primer empleo, o que tienen un nivel de estudios inferior al que declaro anteriormente. Teniendo en cuenta que la ECE admite informantes calificados distintos en los diferentes periodos en el tiempo, estos cambios de informante pueden dar lugar a inconsistencias de respuesta.

4.16 Representatividad

Tomando en cuenta los objetivos de la Encuesta Continua de Empleo mensuales y trimestrales, la representatividad de la ECE con periodicidad mensual es Nacional, en el caso de la periodicidad trimestral la representatividad es Total Nacional, Urbano y conurbaciones de La Paz, Santa Cruz y Cochabamba.

5. Factores de Expansión

El factor de expansión se interpreta como la cantidad de personas en la población, que representa una persona en la muestra.

El uso de los factores de expansión garantiza que:

- La estimación sea insesgada y consistente.
- El error de muestreo sea pequeño condicionado al diseño muestral y al tamaño de la muestra.
- Las deficiencias de cobertura sean corregidas.

La construcción del factor de expansión se realiza en base al diseño muestral expuesto en primera instancia, se calculan los factores básicos las cuales resultan hallando la inversa de la probabilidad de selección de la vivienda, posteriormente se realiza un ajuste de incidencias finales, y por último se realiza la calibración por proyecciones de población.

De acuerdo al diseño de la muestra, el muestreo es bi-etápico, por lo cual la probabilidad de selección de una vivienda está dividida en dos etapas, en la primera etapa se calcula la probabilidad de seleccionar la j – ésima UPM del estrato h . En la segunda etapa se calcula la probabilidad de seleccionar la i – ésima vivienda de la j – ésima UPM del estrato h . De la siguiente manera,

$$P(Viv_{ijh}) = \left(\frac{A_h N_{jh}}{N_h} \right) * \left(\frac{VE_{jh}}{VOE_{jh}} \right) = \left(\frac{A_h N_{jh}}{N_h} \right) * \left(\frac{c}{VOE_{jh}} \right)$$

Dónde:

- $P(Viv_{ijh})$: Probabilidad de seleccionar la i – ésima vivienda de la j – ésima UPM, del estrato h de una UPM del área amanzanada o dispersa.
- A_h : Número de UPM seleccionadas del estrato h .
- N_h : Número de viviendas del estrato h .
- N_{jh} : Número de viviendas en la j – ésima UPM del estrato h .
- VE_{jh} : Total de viviendas elegidas (viviendas elegidas ocupadas) en la j – ésima UPM, del estrato h , es decir, $VE_{jh} = c$
- VOE_{jh} : Total de viviendas objeto de estudio (viviendas ocupadas) listadas en la j – ésima UPM, del estrato h .
- c : Número fijo de viviendas seleccionadas en la última etapa, $c = 12$.

Los factores de expansión básico mensual y trimestral, están dados por:

$$d_k = F_{ijh} = \frac{1}{P(Viv_{ijh})}$$

5.1 Ajustes a los Factores Expansión

5.1.1 Por Incidencias Finales

Los factores de expansión se corrigen por incidencias de campo, es decir, se utiliza los valores de la muestra efectiva por UPM.

$$IF = \frac{c}{V_{jh}}$$

Dónde:

V_{jh} : Es la muestra efectiva, es el número de viviendas con encuestas válidas en la j –ésima UPM del estrato h .

c : Es el número de viviendas planificadas para ser encuestadas = 12

Los factores de expansión básico, ajustado con las incidencias finales, están dados por:

$$w_k = d_k * IF$$

5.1.2 Método post estratificación (4t-2015 al 4t-2018).

Por Proyecciones de Población.

Para estimar el Total de hogares y el Total de población, se corrige el factor de expansión con las proyecciones de población que genere el INE, esta información por cada dominio de estudio.

La corrección se realiza multiplicando por: $\frac{P_h}{\hat{P}_h}$

Dónde:

P_h : Población en el h -ésimo dominio, según la proyección.

\hat{P}_h : Población en el h -ésimo dominio, a la que expande la encuesta.

Los factores de expansión finales estarán dados por:

$$F_{jh} = \frac{1}{P(Viv_{ijh})} \times \frac{c}{V_{jh}} \times \frac{P_h}{\hat{P}_h} = d_k \times IF \times \frac{P_h}{\hat{P}_h}$$

Ajuste post-estratificación.

El objetivo de este ajuste es reducir los errores de estimación, estrechar los intervalos de confianza y mejorar las precisiones esperadas, donde se genera post-estratos según departamento, área (urbana, rural) y grupo etario (0 a 13 años, 14 y más

años), ajustando el factor de expansión de acuerdo a las proyecciones de población generadas para los años 2015-2016-2017 y 2018.

Para ello se calcula el valor delta:

$$\delta = \frac{U}{U_{hat}}$$

Dónde:

U = Proyección de población mensual y trimestral, en fechas específicas.
 U_{hat} = Expansión de la población por medio de la encuesta por post-estrato.

El factor final se obtiene multiplicando el valor encontrado δ con el factor ajustado por proyección de población, de acuerdo a los estratos generados (post-estratificación).

Sin embargo, ante nuevos retos para contar con análisis en sub grupos de interés desagregados (sexo y población entre 16 a 28 años) hacen necesario contar con una metodología de calibración con mayor flexibilidad para introducir restricciones en la calibración de los factores de expansión y controlar variables auxiliares, la cual se implementó a partir de la gestión 2019.

5.1.3 Calibración de factores - Método lineal con distancia Ji-cuadrado (a partir del 1^{er} trimestre 2019)

Este método permite agrupar múltiples variables de control, en base a sus marginales. El objetivo de este ajuste (Calibración) es reducir los errores de estimación, estrechar los intervalos de confianza y mejorar las precisiones esperadas. Los ajustes al igual que la post estratificación se realizan con el propósito de que las estimaciones poblacionales de las **variables de control (variables auxiliares)** reproduzcan los totales poblacionales (Proyecciones).

Para utilizar esta metodología, deben cumplir las siguientes condiciones:

- El número de variables de control no debe ser elevado.
- Los coeficientes de variación de los estimadores de Horvitz y Thompson de los totales poblacionales de las variables de control estimados, con los factores básicos ajustados por incidencias finales, deben ser menor al 10%.

Estas condiciones deben de cumplirse, para que las inferencias sobre los estimadores sean insesgados.

(Deville J.C., Särndal C.E. y Sautory O., 1993) La calibración de los factores de expansión es un ajuste que se realiza a los factores de expansión con el propósito de que las estimaciones de algunas variables de control reproduzcan con exactitud los totales poblacionales de dichas variables.

El objetivo de la calibración es obtener un nuevo sistema de factores de expansión w_k^{cal} que se encuentren cerca de w_k , de tal forma que cuando los factores sean usados para estimar los totales de las variables auxiliares, dichos totales sean reproducidos con exactitud, de manera que los nuevos factores conserven cualquier propiedad buena de estimación de los factores básicos.

Para la construcción estos nuevos factores w_k^{cal} , se debe minimizar una pseudo-distancia $G(w_k^{cal}/w_k)$ entre w_k^{cal} y w_k en toda la muestra y está dada por:

$$\sum_{k \in S} w_k G(w_k^{cal}/w_k)$$

Los factores de calibración deben satisfacer la siguiente restricción:

$$\sum_S w_k^{cal} X^{demográficas} = \sum_U X^{demográficas} = t_X \quad (*)$$

Dónde: X es un vector de valores auxiliares (Variables de control) conocidos (proyecciones de población) para todos los elementos de la muestra

El método utilizado en la calibración de los factores de expansión por el INE es el Método lineal con distancia Ji cuadrado

Este método calcula la distancia, en toda la muestra, de los nuevos factores w_k^{cal} y los factores w_k como:

$$\sum_{k \in S} w_k G(w_k^{cal}/w_k) = \frac{1}{2} \sum_S \frac{(w_k^{cal} - w_k)^2}{w_k}$$

Por lo tanto, la idea es definir factores de calibración que minimicen $G(w_k^{cal}/w_k)$ sujeto a la restricción de la ecuación (*), el cual se resuelve mediante multiplicadores de Lagrange y se obtiene la ecuación de los valores de los nuevos factores de expansión calibrados.

$$w_k^{cal} = w_k F(q_k \lambda' X_k)$$

El vector λ se obtiene al resolver el siguiente sistema de ecuaciones

$$\sum_{k \in S} \underbrace{w_k F(q_k \lambda' X_k)}_{w_k^{cal}} X'_k = t'_k$$

Luego los estimadores de calibración están dados por:

$$\hat{t}_{y,cal} = \sum_S w_k^{cal} y_k$$

Para las calibraciones mensuales y trimestrales se han utilizado variables de control descritas en la sección anterior (**4.16 Representatividad**) en ambos casos,

adicionalmente se han incluido otras variables de control, siempre y cuando cumplan con la evaluación de los coeficientes de variación calculados (sean menores al 10%), previa a la calibración.

Para realizar la calibración se considera la información auxiliar disponible de las proyecciones poblacionales de las variables de control.

Para el cálculo de los factores de expansión calibrados se utilizan los siguientes paquetes que están implementados en el software estadístico “R”:

- svyr
- survey

Donde el primero se encarga de la creación del diseño muestral mediante la función **as_survey_design** tomando como punto de partida los factores ajustados por incidencias finales y el segundo realiza la calibración de los factores de expansión mediante la función **calibrate**.

5.1.4 Métodos para la corrección del sesgo

Se debe hacer notar, que durante la ejecución de la encuesta puede traer consigo sesgos por ausencia de respuesta, puesto que las personas que no pudieron ser contactadas pueden tener altas probabilidades de tener alguna característica en particular, el cual puede incidir sobre los indicadores laborales.

Asimismo, CEPAL propone técnicas que permiten corregir el sesgo, entre ellas: ajuste por probabilidad de respuesta (Propensity Score), calibración de los factores de expansión en dos etapas y post estratificación basada en modelos multinivel.

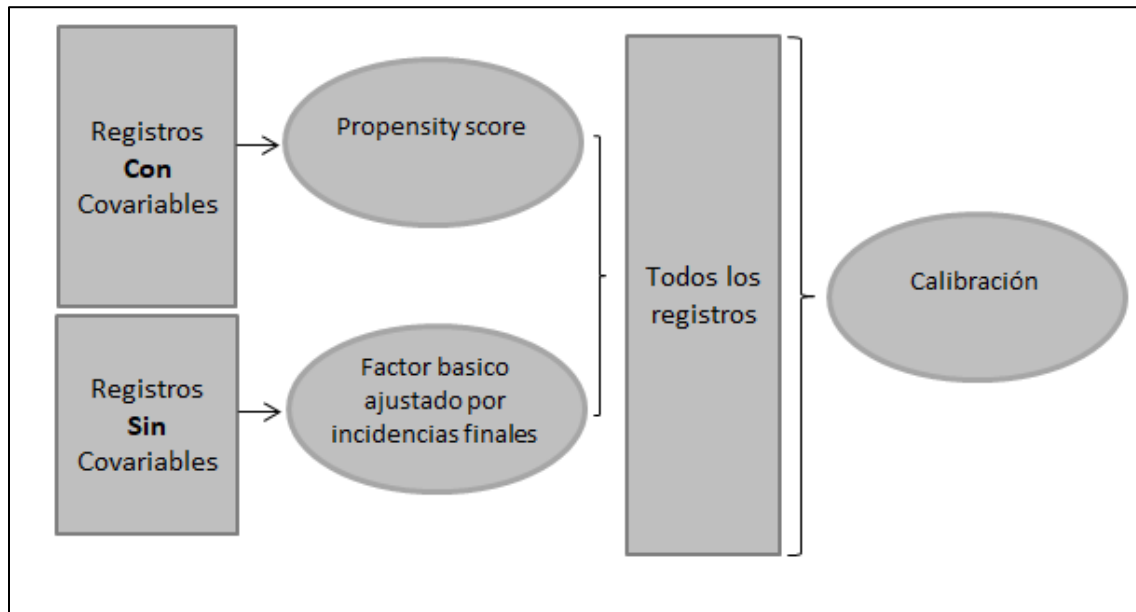
La metodología adoptada por el INE para la corrección de sesgo, es la probabilidad de respuesta (**Propensity Score**), que consiste en clasificar las unidades como respondientes o no respondientes entre dos periodos y aplicar un modelo logístico en función de las covariables disponibles en un periodo “t-1”. Con la finalidad de obtener la probabilidad estimada de respuesta o ausencia de respuesta.

5.1.5 Factores de expansión para un periodo “t”.

En el proceso del cálculo de los factores de expansión, se empleó ajustes y técnicas que permiten reducir el sesgo generado por la ausencia de respuesta, con el objetivo de lograr que las unidades que no responden sean representadas por las que sí, con información de los respondientes.

(CEPAL, 2020) Utilizar el factor de expansión ajustado por probabilidad de respuesta en el cálculo de los estimadores deseados, minimizaría el sesgo de selección que se generó por el cambio de modo de recolección de la información.

Ilustración 3: Esquema de ajuste de factores de expansión (corrección de sesgo)



Fuente: Reuniones de coordinación CEPAL – INE

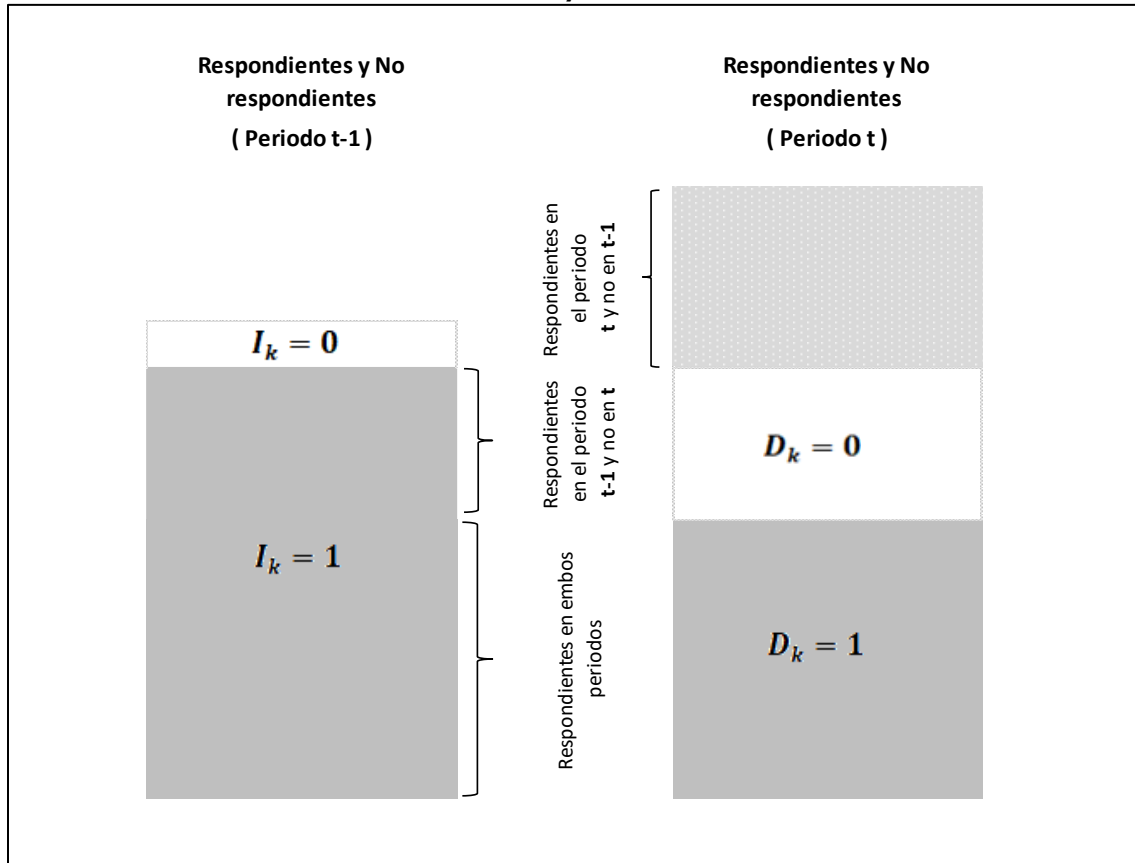
5.1.6 Ajuste por probabilidad de respuesta (Propensity score)

Una forma mejor de comprender la **falta de respuesta** es introducir el concepto de PROBABILIDAD DE RESPUESTA (**Propensity score**), para ello se debe ajustar un modelo que sea capaz de explicar las probabilidades de respuesta a partir de un conjunto de variables auxiliares, el modelo más utilizado es el de regresión logística (MODELO LOGIT)

Propuesto por Rosenbaum y Rubin (1983), este enfoque es útil para dilucidar la estructura de la ausencia de respuesta y, por consiguiente, corregir el sesgo de cobertura y el sesgo por ausencia de respuesta (Lensvelt-Mulders, Lugtig y Hubregtse, 2009).

Para el manejo efectivo de la ausencia de respuesta se consideran las variables dicotómicas: I_k y D_k , que indican si el individuo pertenece a la muestra en un periodo anterior ("t-1") y si ha respondido la encuesta en ambos periodos ("t-1" y "t"), respectivamente.

Ilustración 4: Esquema de los respondientes y no respondientes entre un tiempo “t-1” y “t”



Suponiendo que la distribución de las respuestas efectivas puede ser estimada, la probabilidad de respuesta (Propensity Score) de un individuo en la muestra está dada por:

$$\phi_k = \Pr(D_k = 1 | I_k = 1)$$

Esta probabilidad es distinta para cada individuo y puede ser estimada usando los datos del panel (de un periodo t-1).

(CEPAL–2020) Contar con la muestra en un periodo “t-1”, para la cual se obtuvo toda la información del cuestionario en un período anterior, constituye un excelente punto de partida para tratar de eliminar el sesgo, puesto que se tendrá acceso a un conjunto de covariables x para determinar el mejor modelo a fin de estimar el patrón de ausencia de respuesta en la muestra de respondientes efectivos.

Con la finalidad de obtener la probabilidad estimada de respuesta o no respuesta, en los cuales la variable dependiente es dicotómica e identifica si el individuo respondió a la encuesta (1) o no (0) para un periodo “t” tomando como punto de partida un periodo “t-1” (panel), consideremos el siguiente modelo:

$$\text{logit}(\hat{\theta}_k) = x_k \hat{\beta}$$

$$\hat{\theta}_k = \frac{1}{1 + e^{-(x_k \hat{\beta})}}$$

Dónde:

$\hat{\beta}$: Vector de coeficientes estimado de la regresión logística.

$\hat{\theta}$: Probabilidades estimadas (Propensity score) para respondientes y no respondientes.

x_k : Vector de covariables (Predictores).

Las variables Predictores o independientes utilizadas en el modelo son:

- Panel (paneles que se traslapan)
- Estrato socioeconómico
 - 1: Estrato bajo
 - 2: Estrato medio bajo
 - 3: Estrato medio alto
 - 4: Estrato alto
- Semana (Semana de referencia de la que se preguntó la encuesta)
- Departamento (9 departamentos)
- Sexo (periodo t-1)
 - 1: Hombre
 - 2: Mujer
- Número de personas por hogar (periodo t-1)
- Condición de actividad (periodo t-1)
 - 1: Pent (población menor a 14 años)
 - 2: Ocupado
 - 3: Cesante → Desocupado
 - 4: Aspirante → Desocupado
 - 5: Temporal → Inactivo
 - 6: Permanente → Inactivo
- Nivel de educación (periodo t-1)
 - 0: Ninguno
 - 1: Primaria incompleta
 - 2: Primaria completa
 - 3: Secundaria incompleta
 - 4: Secundaria Completa
 - 5: Superior
 - 7: Otros
 - 9: Sin especificar
- Jefe del hogar (periodo t-1)
 - 1: Es Jefe del hogar

2: No es jefe del Hogar.

Teniendo en cuenta que el factor de expansión de la muestra del periodo "t-1" se representan como d_k^* , y habiendo estimado $\hat{\varnothing}_k$ para respondientes y no respondientes de la muestra en el periodo "t", entonces el factor de expansión ajustado tomaría la siguiente forma:

$$w_k = \frac{d_k^*}{\hat{\varnothing}_k}$$

Estos nuevos pesos se transfieren a la muestra del periodo "t" para $D_k = 1$ (respondientes en ambos periodos)

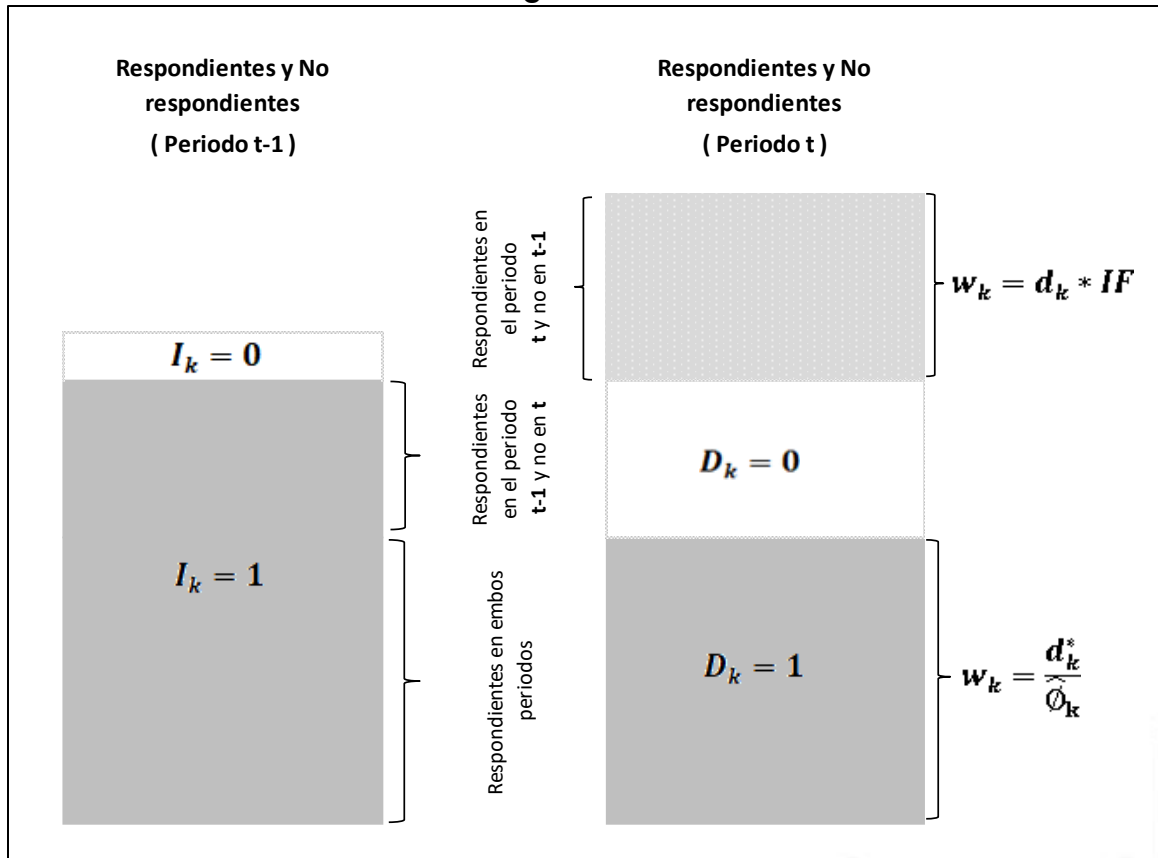
Para los registros que se encuentran fuera del panel (no tienen información auxiliar porque sus paneles rotatorios no se traslapan) y nacimientos (no tienen información auxiliar porque se acaban de incorporar al panel rotatorio) en el periodo "t", es decir que no tienen información anterior de covariables que permitan calcular la probabilidad de ser o no respondiente, en estos casos se calcula los factores básicos normales ajustados por incidencias finales (cobertura de viviendas).

$$w_k = d_k * IF$$

Dónde:

- d_k : es el factor Básico (inversa de la probabilidad de selección en dos etapas) de la muestra en el periodo "t".
- $IF = \frac{c}{V_{jh}} = \frac{\text{Viviendas planificadas por UPM}}{\text{Viviendas ejecutadas por UPM}}$

Ilustración 5: Esquema de los Pesos ajustados por el Propensity score y el ajuste de los registros Nuevos



(CEPAL-2020) Utilizar el factor de expansión w_k en el cálculo de los estimadores deseados minimizaría el sesgo de selección que se generó por el cambio de modo de recolección de la información. Los factores asociados con el sesgo de cobertura pueden no ser los mismos que los factores asociados con el sesgo por ausencia de respuesta, por lo que probablemente sería beneficioso modelar estos problemas por separado y luego usar los dos puntajes de propensión como factores de ajuste independientes.

Calibración de factores de expansión (w_k^{cal})

La calibración se realiza a partir de w_k utilizando lo mencionado en la sección 5.1.3.

Por recomendaciones de la CEPAL en los futuros cálculos de los factores de expansión mensual y trimestral en la Encuesta Continua de Empleo, se debe de considerar la metodología del Propensity Score, con la finalidad de evitar sesgos de no respuesta.

6. Estimadores y Errores de Muestreo

Ya que el diseño es complejo, se requiere utilizar fórmulas especiales que consideran los efectos de la estratificación y la conglomeración al momento de obtener los resultados de la encuesta.

Para calcular los estimadores, es prioritario que la base de datos de la encuesta esté expandida, vale decir, que se empleen los factores de expansión. Para el cálculo de los intervalos de confianza, coeficiente de variación, error estándar, efecto de diseño, entre otros; se debe aplicar el diseño de la muestra antes de utilizar la base de datos en la generación de indicadores.

Se recomienda valerse de algún software estadístico que permita realizar el cálculo de los errores muestrales aplicando el diseño de muestra empleado. Para la ECE, se considera los siguientes estándares para evaluación de los errores muestrales (CV):

Indicador	CV
Optimo	< 10%
Bueno	10% a 15%
Regular	16% a 20%
**No significativo	>20%

**Indicador sujeto a evaluación.

7. Anonimización de códigos de UPM

El término anonimizar es una técnica de tratamiento de datos que elimina o modifica los datos personales identificables para obtener datos que no se puedan asociar a las personas o viviendas, así mismo se realiza la anonimización de los identificadores geográficos de la UPM, con el fin de resguardar la información proporcionada por los informantes y reducir al mínimo el riesgo de identificar a los mismos. Para este fin, se considera la Ley N° 1405 vigente, sobre el Secreto Estadístico, donde la información proporcionada por los informantes, no podrá ser revelada en forma individual y solo podrá ser publicada de forma anonimizada.

Como la encuesta de la ECE tiene la característica de panel, la anonimización de los códigos de UPM se realizará de manera progresiva a partir del segundo trimestre de la gestión 2022.

8. Bibliografía

- Thomas Lumley (2010 John Wiley & Sons, Inc. All rights reserved) "Complex Surveys A Guide to Analysis Using R"
- Leslie Kish "Survey Sampling"
- Hugo Andrés Gutierrez (2009) "Estrategias de muestreo Diseño de encuestas y estimación de parámetros"
- Sharon L. Lohr "Muestreo: Diseño y Análisis"
- Valliant R., Dever J., Kreuter F. Second Edition (2018) "Practical Tools for Designing and Weighting Survey Samples".
- Valliant and Dever - 2017 "Survey Weights A Step-by-step Guide to Calculation"
- Pérez-López C., 2005, "Muestreo Estadístico, Conceptos y Problemas Resueltos". Editorial Pearson Prentice Hall.
- CEPAL y OIT (2020). "Coyuntura Laboral para América Latina y el Caribe - El trabajo en tiempos de pandemia: desafíos frente a la enfermedad por coronavirus (COVID-19)". Santiago-Chile.
- CEPAL. (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). Obtenido de https://rtccea.cepal.org/sites/default/files/rtc_connected/files/recomendaciones-encuestas-hogares-covid-final_1.pdf
- CEPAL. (2020). Recomendaciones para eliminar el sesgo de selección en las encuestas de hogares en la coyuntura de la enfermedad por coronavirus (COVID-19). Obtenido de https://repositorio.cepal.org/bitstream/handle/11362/45552/1/S2000316_es.pdf