



**MANUAL CONCEPTUAL Y
METODOLOGICO
DISEÑO MUESTRAL
ENCUESTA NACIONAL DE EMPLEO
(ENE)**

INSTITUTO NACIONAL DE ESTADÍSTICAS (INE)

31 de enero de 2018



MANUAL CONCEPTUAL Y METODOLOGICO

DISEÑO MUESTRAL

ENCUESTA NACIONAL DE EMPLEO (ENE)

DEPARTAMENTO DE ESTUDIOS LABORALES

SUBDIRECCIÓN TÉCNICA

INSTITUTO NACIONAL DE ESTADÍSTICAS

VERSIÓN 2: 31 DE ENERO DE 2018

La presente versión del documento es una actualización del manual publicado en enero del 2010. Incorpora actualizaciones únicamente en el Capítulo II: Manual de Diseño Muestral.

ÍNDICE

CAPITULO I: MANUAL CONCEPTUAL	5
1. Introducción.....	5
2. Clasificación de la Población en Edad de Trabajar	6
2.1 Conceptos y criterios de medición.....	7
Anexo 1: Diagramas de Flujo de Preguntas de.....	17
Nuevo Encuesta Nacional de Empleo	17
(NENE).....	17
Anexo 2: Glosario.....	21
Anexo 3: Principales Indicadores de Razón	23
Anexo 4: Indicadores Analíticos de Razón	24
Anexo 5: Otras Definiciones	25
Anexo 6: Nomenclaturas de Clasificación Utilizadas	26
CAPITULO II: MANUAL DE DISEÑO MUESTRAL	28
1 Introducción.....	28
2 Objetivo de la Encuesta	29
3 Población Objetivo	29
4 Marco Muestral y estratificación	30
4.1 Marco Muestral: Conformación, partición y actualización del Marco Muestral	30
4.2 Estratificación del marco muestral.....	31
5 Descripción del Diseño Muestral	33
6 Cobertura Geográfica y áreas de estimación	34
7 Unidades de muestreo y unidades de información	34
8 Determinación del tamaño muestral	35
9 Distribución de la muestra.....	36
10 Rotación de la muestra	36
11 Selección de Unidades de muestreo y Factores de Expansión	37
11.1 Selección de unidades de primera y segunda etapa	37
11.2 Factores de Expansión	42
12 Estimadores de Totales y Razones	44
12.1 Estimadores del Total de la variable Y	45
12.3 Estimadores de razón	46
12.3.1 Estimador de razón entre las variables X e Y.....	46
13 Varianza de los Estimadores.....	47
13.1 Varianza del Total de la variable X	47
13.2 Varianza de la razón del Total de la variable Y respecto al Total de la variable X.....	47
14 Coeficiente de Variación y Error asociado al estimador	48
Coeficiente de Variación y Error relativo.....	49
Anexo 1: Tamaño y Errores Muestrales según Nivel de Estimación.	53
Anexo 2: Programación para cálculo de Error Estándar y Coeficientes de Variación.....	57

Cálculo usando STATA	57
Cálculo usando SPSS	58
Cálculo usando SAS	59
15 Bibliografía.....	62

CAPITULO I: MANUAL CONCEPTUAL

1. Introducción

El presente documento tiene por objeto introducir a los usuarios en el detalle conceptual que define a la Nueva Encuesta Nacional de Empleo (NENE), la que reemplaza a partir de abril de 2010 a la antigua Encuesta Nacional de Empleo (ENE) vigente desde 1966.

Las principales definiciones que se describen en este Manual, se refieren a las variables que componen el Código Sumario de Empleo (CSE). Las definiciones recogen la actualización de los criterios de clasificación emanados de las resoluciones de las Conferencias Internacionales de Estadísticos del Trabajo (CIET), de la OIT, en particular de la XII CIET realizada en 1982.

Como resultado de la incorporación de los estándares internacionales de medición, el diseño conceptual y metodológico de la Nueva Encuesta Nacional de Empleo incluye, entre otros aspectos, la actualización de los conceptos de población ocupada, de población desocupada y la distinción de los ocupados entre los ocupados "tradicionales" y los "no tradicionales".

Las actualizaciones conceptuales se complementan con el detalle de los principales indicadores de razón calculados a partir de las categorías de condición de actividad laboral de la población y con un resumen de los principales indicadores analíticos, consistentes en un conjunto de cocientes que facilitan el análisis de la coyuntura y que están asociados a la población ocupada.

2. Clasificación de la Población en Edad de Trabajar

En la Nueva Encuesta Nacional de Empleo (NENE), la clasificación de la población en edad de trabajar (PET) según condición de actividad económica o estatus en relación a la fuerza de trabajo, se sintetiza en una variable denominada "Código Sumario de Empleo" (CSE).

El objetivo general del CSE es clasificar a todas las personas de 15 años y más, según su situación laboral, aplicando un conjunto de reglas de prioridad, en las que se basa, a nivel internacional, el enfoque de medición del empleo y el desempleo a partir de encuestas a hogares. Estas reglas de prioridad, se traducen en una serie de preguntas y secuencias que constituyen el nuevo cuestionario de la encuesta.

Los fundamentos conceptuales del CSE se basan en la actualización de los criterios de clasificación, emanados de las resoluciones de las Conferencias Internacionales de Estadísticos del Trabajo (CIET), de la OIT, en particular de la XII CIET realizada en 1982¹.

Como resultado de la incorporación de los estándares internacionales de medición, el diseño conceptual y metodológico de la Nueva Encuesta Nacional de Empleo se traduce en:

- Actualizar la conceptualización de la población ocupada, incorporando los criterios de haber trabajado al menos una hora en la semana de referencia y recibir un ingreso por la actividad realizada.
- Distinguir, entre los ocupados, a los ocupados "tradicionales" de los "no tradicionales", en consideración de si el entrevistado reconoce o identifica en primera instancia, su actividad como trabajo.
- Actualizar la conceptualización de población desocupada, incorporando los criterios de haber buscado empleo en las últimas cuatro semanas y de estar disponible para trabajar en, al menos, las dos semanas siguientes a las de la entrevista.
- Clasificar las personas fuera de la fuerza de trabajo, según razones de inactividad, así como también, según su potencialidad de participar en el mercado del trabajo. Esto marca diferencias notables respecto del cuestionario anterior en el que, basado

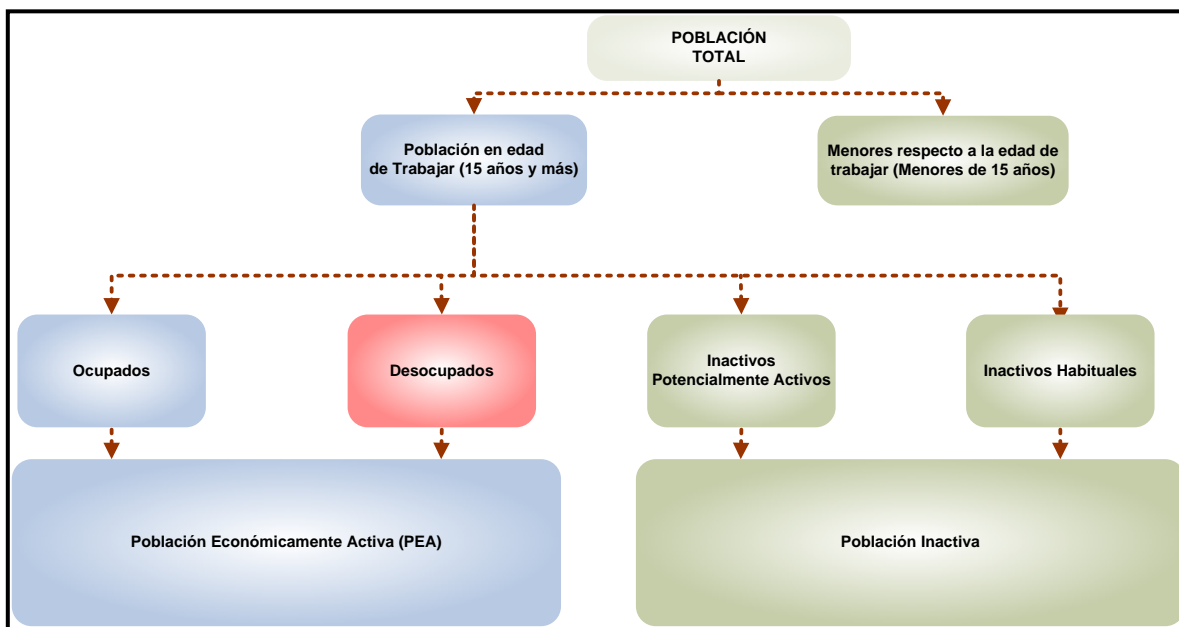
1 "Encuestas de Población Económicamente Activa. Empleo, Desempleo y Subempleo". Vol. 1. En especial: Parte 1. Cap. 2 (puntos 2, 3 y 4), Cap. 3 (puntos 2, 3 y 4) y Cap.6 (punto 3).

en la autoclasificación, la inactividad estaba asociada a la habitualidad² de la condición. El nuevo instrumento clasifica a los inactivos a través de una batería de preguntas, que se acogen a la regla de prioridad en la clasificación, la cual parte con la condición de ocupado y continúa con la de desocupado. Para ser inactivo, una persona tiene que, no haber cumplido con los criterios de clasificación de dichas condiciones.

2.1 Conceptos y criterios de medición

El siguiente esquema conceptual, muestra las grandes categorías en que se clasifica la población, para fines del análisis de la condición laboral. La Nueva Encuesta Nacional de Empleo se aplica a todas las personas de 15 años o más, las cuales constituyen la población en edad de trabajar (PET) en Chile. Las personas a quienes se aplica el cuestionario, pueden quedar clasificadas como ocupadas, desocupadas o inactivas³, categorías mutuamente excluyentes.

Esquema 1:
CLASIFICACIÓN DE LA POBLACIÓN ACTIVA E INACTIVA EN LA NUEVA ENCUESTA NACIONAL DE EMPLEO (NENE)



Los criterios de clasificación y los periodos de referencia asociados a dichas categorías se describen a continuación.

² La condición de actividad de los mayores de 15 años en la antigua encuesta se derivaba de lo que las personas declaraban haber hecho "la mayor parte de la semana pasada" (semana de referencia).

³ En el Anexo 1 se puede revisar las secuencias a partir de las que se construyen, todos y cada uno, de los stocks del CSE de la NENE.

2.1.1 Ocupados

Los ocupados son personas que deben cumplir los siguientes criterios:

- Haber trabajado al menos una hora durante la semana de referencia
- Haber devengado (recibió o recibirá) un pago, ya sea en dinero o en especie, por ese trabajo

Esquema 2:
CRITERIOS DE CLASIFICACIÓN: POBLACIÓN OCUPADA

CRITERIOS	PERÍODO DE REFERENCIA
A. Realizar alguna actividad económica	Durante la semana anterior (de lunes a domingo)
B. Trabajo realizado por lo menos 1 hora	Durante la semana anterior (de lunes a domingo)
C. Recibir o recibirá un pago en dinero o en especie	El pago por el trabajo puede ser a futuro

El trabajo se refiere a cualquier actividad que esté dentro de los límites de producción del Sistema de Cuentas Nacionales (SCN) y cubre toda la producción de mercado, así como algunos tipos de producción de no mercado (productos primarios o procesados para autoconsumo).

i. Ocupados Tradicionales y No Tradicionales

La diferenciación entre ocupados “tradicionales” y “no tradicionales” dice relación con la concepción que cada persona tiene de la labor que desempeña. Al formular las preguntas de la sección A del cuestionario, quienes perciban su trabajo como tal, quedarán clasificados, de acuerdo a los criterios arriba especificados, como “ocupados tradicionales”. Mientras que, aquéllos informantes que no identifiquen desde un principio su actividad como un trabajo⁴, se clasificarán, como resultado de una mayor indagación respecto de la actividad realizada, como “ocupados no tradicionales”.

La mayor precisión en la clasificación de los ocupados, permite incluir personas que desarrollan trabajos que, por sus características, ya sea en horas trabajadas, frecuencia con que lo realizan o lugar de trabajo, no habrían sido identificadas, en principio, como una actividad económica. Esta precisión, es particularmente pertinente, en los países en vía de

⁴ Se exige que por dicha actividad se haya devengado un ingreso (a menos que se trate del familiar no remunerado de un miembro de la familia).

desarrollo, donde por razones culturales, ciertas actividades no son declaradas, pero hacen parte, en muchos casos, de la sobrevivencia y sostenibilidad económica de los hogares, ya sea en la coyuntura, la estacionalidad o el ciclo económico.

i.i Ocupados Ausentes

La definición de ocupados incluye a las personas con empleo, pero que temporalmente no trabajaron en la semana de referencia por diversas razones, mantienen su vínculo formal con el empleo, y se espera que volverán al trabajo después de su periodo de ausencia.

La definición internacional especifica ciertos principios para determinar la ausencia temporal del trabajo (“ocupados ausentes”), según se trate de trabajadores Asalariados o Independientes y de acuerdo a tres posibilidades en línea con los estándares de la OIT .

- Ocupados ausentes con vínculo efectivo: incluye como razones de no haber trabajado la semana de referencia: vacaciones o permisos, licencia médica, horario o jornada variable o flexible, huelga, conflicto laboral, asistencia a cursos de capacitación.
- Ocupados ausentes con pronto retorno: si declaran que volverán al trabajo en menos de cuatro semanas.
- Ocupados ausentes con sueldo o ganancia: si declaran que durante su ausencia siguen percibiendo ingresos por su trabajo.

La relevancia de esta clasificación, responde a que la ausencia laboral sigue patrones temporales diferenciados. Dependiendo del ciclo económico, las razones técnicas o económicas pueden determinar la duración de la ausencia y el retorno al trabajo. Nuevas formas contractuales o la temporalidad de algunas actividades pueden tener asociados periodos de ausencia, que resultan de interés analítico identificar. La ausencia temporal también exhibe un sesgo estacional. Durante el invierno por ejemplo, pueden aumentar las licencias médicas, en tanto que en verano, las “vacaciones o permisos”.

La pregunta A6 del cuestionario, precisa las razones de ausencia de los ocupados en la semana de referencia:

A6 ¿Por qué razón no trabajó la semana pasada?

▼ No se leen las opciones

<input type="checkbox"/>	1. Vacaciones o permisos	
<input type="checkbox"/>	2. Licencia médica	
<input type="checkbox"/>	3. Por horario o jornada variable o flexible	Pasa a B1
<input type="checkbox"/>	4. Huelga, conflicto laboral	
<input type="checkbox"/>	5. Asistencia a cursos de capacitación	
<input type="checkbox"/>	6. Problemas de salud	
<input type="checkbox"/>	7. Suspensión temporal del trabajo	
<input type="checkbox"/>	8. No tuvo pedidos (clientes)	
<input type="checkbox"/>	9. Razones climáticas o catástrofes naturales	
<input type="checkbox"/>	10. Su trabajo es ocasional	
<input type="checkbox"/>	11. Razones económicas o técnicas de la empresa o negocio (rotura de equipos, falta de materias primas o capital para comprar)	
<input type="checkbox"/>	12. Su trabajo es estacional	
<input type="checkbox"/>	13. Clausura de negocio o de empresa	
<input type="checkbox"/>	14. Otras razones (especificar) _____	

2.1.2. Desocupados

Los desocupados son personas que deben cumplir los siguientes criterios:

- No haber trabajado durante la semana de referencia.
- Haber buscado un empleo en las últimas cuatro semanas (incluyendo la de referencia).
- Estar disponible para trabajar en las próximas dos semanas, posteriores a la de referencia.

Esquema 3:
CRITERIOS DE CLASIFICACIÓN: POBLACIÓN DESOCUPADA

CRITERIOS	PERÍODO DE REFERENCIA
A. No haber trabajado 1 hora	Durante la semana anterior (de lunes a domingo)
B. Haber buscado empleo activamente	Durante las últimas cuatro semanas
C. Estar disponible para trabajar	Desde el fin de la semana de referencia y hasta 15 días después.

Las dos últimos criterios permiten precisar las “presiones de oferta” en el mercado de trabajo. Buscar trabajo, pero no estar disponible implica que, en los hechos, no se está compitiendo por un puesto de trabajo. Estar disponible, pero no buscar, supone que no se concurrió al mercado como oferente.

Dentro de los desocupados se distingue entre cesantes y buscan trabajo por primera vez, a partir de si el desocupado tuvo o no, anteriormente, un trabajo.

2.1.3 Fuera de la Fuerza de Trabajo o Inactivos

La población inactiva está conformada por personas que no tuvieron un empleo y tampoco lo buscaron y/o estuvieron disponibles para trabajar. Es decir, comprende a todas las personas no clasificadas como ocupadas o desocupadas.

i. Inactivos según razones y nuevas subcategorías

Con la Nueva Encuesta Nacional de Empleo se ordenan los inactivos según razones de inactividad (no búsqueda o no disponibilidad, o ambas al mismo tiempo). Este es un cambio conceptual que apela a las motivaciones de las personas a no entrar al mercado laboral, que permite entender la inactividad no como una condición permanente sino como una condición que puede cambiar por determinadas razones, en periodos cortos de tiempo⁵. Las razones corresponden a razones familiares, razones de estudio, razones de jubilación, razones de pensión o montepío y razones de salud permanente. Adicionalmente, se distinguen los inactivos por razones de “desaliento”, por razones “estacionales”, “razones personales temporales” y los “iniciadores”.

⁵ Esta precisión está asociada a la consideración de “la semana de referencia” como un periodo corto dentro del cual se indaga por la situación laboral, de acuerdo al marco conceptual de la fuerza de trabajo en el que se basa la medición del empleo y el desempleo.

E9 ¿Por qué razón no buscó un empleo o no ha hecho preparativos para iniciar o reanudar una actividad por cuenta propia durante las últimas cuatro semanas?

▼ No se leen las opciones

1. Iniciará pronto una actividad por cuenta propia → Pasa a E10a

2. Encontró un empleo que empezará pronto → Pasa a E10b

3. Por responsabilidades familiares permanentes
(cuidado de hijos o personas dependientes)

4. Está estudiando o preparando estudios

5. Es jubilado/a

6. Es rentista

7. Es pensionado/a o montepiada

8. Por motivos de salud PERMANENTES

9. Espera la estación de mayor actividad

10. Por motivos de salud TEMPORALES → Pasan a E11

11. Por responsabilidades familiares de carácter temporal

12. Está embarazada

13. Espera los resultados de un proceso de selección o que lo llamen

14. Algún miembro del hogar no se lo permite

15. Cree que por su edad no le darán empleo

16. Cree que no lo encontrará

17. Se cansó de buscar

18. Cree que ningún empleo o actividad se adapte a su calificación

19. Le piden demasiados trámites para iniciar
una actividad por cuenta propia

20. No sabe donde dirigirse ni a quien acudir

21. No quiere, no necesita trabajar → Pasa a E13

ii. Inactivos potencialmente activos

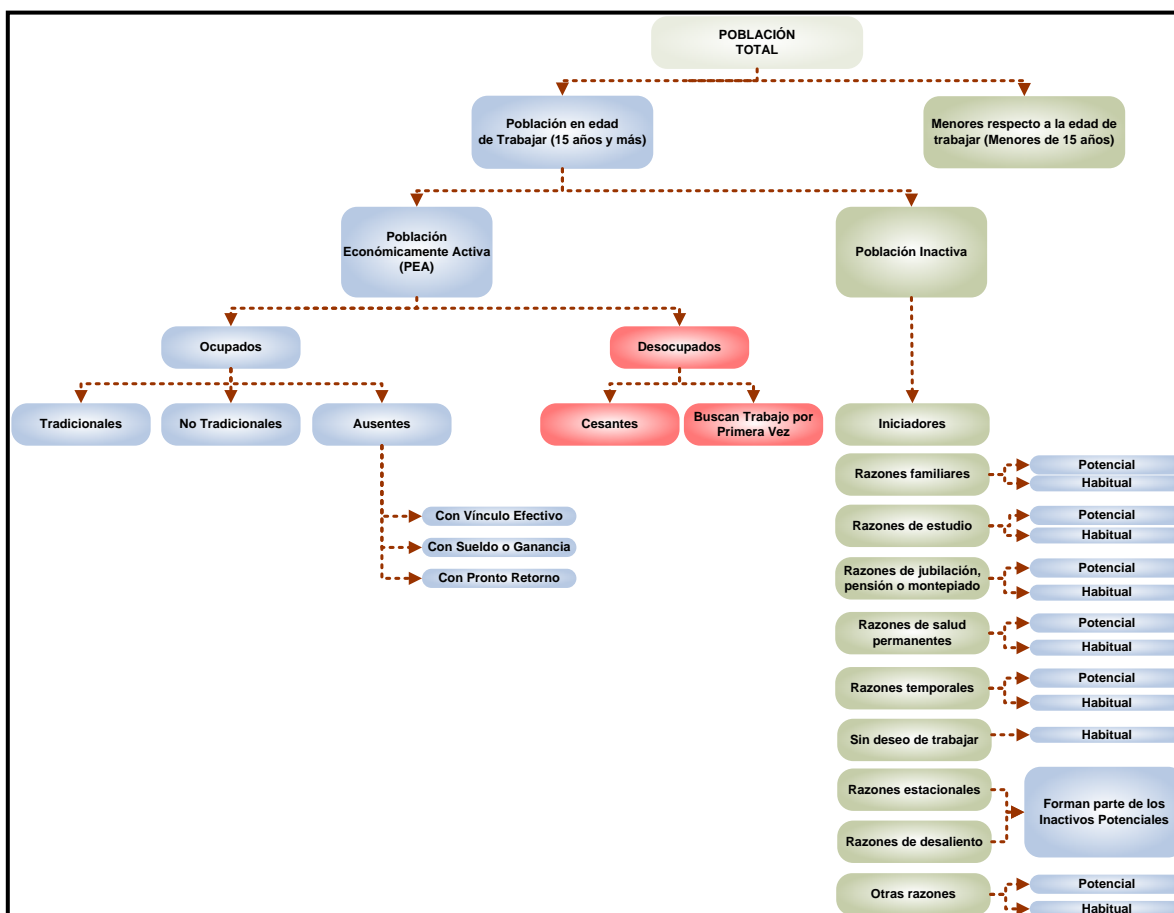
Los inactivos potencialmente activos son aquellos inactivos que, independiente de la razón a la que está asociada su condición de inactivo, o buscaron empleo, pero no estarían disponibles para trabajar o no buscaron, pero estarían disponibles. La potencialidad entre los inactivos tiene que ver con la declaración de una u otra actividad o disposición, respectivamente.

El término “potencial” hace referencia a una presunción de la factibilidad de que una persona que no está presionando el mercado laboral, efectivamente lo haga en algún momento.

Los inactivos que ni buscaron ni estarían disponibles, se identifican como los inactivos más típicos o habituales.

Mención especial merecen las nuevas categorías de inactivos por razones de “desaliento”, y razones “estacionales”, que han sido de creciente importancia en el análisis del comportamiento de la fuerza de trabajo a nivel internacional, por cuanto corresponden a grupos de población que, por sus características, tendrían un mayor vínculo con la fuerza de trabajo respecto del resto de los inactivos. La denominación a nivel internacional es de población “marginamente vinculada a la fuerza de trabajo”. Bajo la Nueva Encuesta Nacional de Empleo, estas categorías se identifican dentro de los inactivos potencialmente activos.

Esquema 4:
CLASIFICACIÓN DE LA POBLACIÓN ACTIVA E INACTIVA EN LA NUEVA ENCUESTA NACIONAL DE EMPLEO (NENE)



ii.1 Iniciadores

Son personas que no han buscado empleo en las últimas cuatro semanas debido a que “iniciarán pronto” un trabajo y que pueden o no estar disponibles para trabajar en las dos semanas siguientes a la semana de referencia⁶.

De estar disponibles, en los países de la OCDE, se agregan a los desocupados. En caso contrario, se les considera inactivos pues no buscaron ni estarían disponibles. En general, la disponibilidad supone una mayor disposición -en el corto plazo- de participación aunque, en esencia, ser iniciador (disponible o no) es una etapa intermedia entre la inactividad y formar parte de la fuerza de trabajo⁷. De acuerdo a la gradación de los inactivos de acuerdo

⁶ También conocidos como personas con “incorporación futura”. OIT, 1990.

⁷ Los iniciadores que estuvieron disponibles, bajo consideraciones especiales de la OIT y adaptaciones de la OCDE, pueden ser considerados como desocupados. No existe una normativa estándar a este respecto, por lo

a su potencialidad de hacer parte de la fuerza de trabajo, los iniciadores son el subconjunto con mayor potencialidad.

ii. 2 Razones Estacionales

Los inactivos “por razones estacionales” están constituidos por personas que declaran esperar “la estación de mayor actividad” y que, al mismo tiempo, señalan estar disponibles.

Su grado de potencialidad es alto y, por lo mismo, en algunos países este subconjunto se suma a la población “marginallymente vinculada” que conforma al grupo que tiene mayor posibilidad de ingresar a la fuerza de trabajo en un breve plazo de tiempo o en cuanto cambien las condiciones en el mercado laboral.

Aunque no se trata de una población muy numerosa, pueden llegar a tener un impacto en las medidas de presión de oferta laboral.

ii.3 Razones de Desaliento

Los desalentados, son personas que no buscaron empleo en las últimas cuatro semanas, se cansaron de buscar un empleo porque creen que no lo encontrarán, pero estarían disponibles para iniciar un trabajo en las dos semanas siguientes a la semana de referencia. Al igual que los inactivos estacionales, en algunos países, esta población se suma a los marginalmente vinculados siendo, de hecho, el grupo más numeroso entre éstos.

ii.4 Otras Razones

Es una categoría de inactivos compleja desde el punto de vista de su composición, ya que considera a todos aquéllos individuos que no califican en un cien por ciento en otra categoría de inactividad. Sin embargo, y a diferencia de los grupos que se mencionarán más abajo, entre esta población predomina la potencialidad ya sea porque se buscó empleo, o bien, debido a la declaración de disponibilidad. De hecho, hay personas que señalan que iniciarán pronto un trabajo/actividad por cuenta propia, pero que no son considerados “iniciadores” porque -paralelamente- indican que no han realizado gestiones/formalizado un acuerdo laboral que garantice, efectivamente, una condición de iniciador.

ii.5 Razones Personales Temporales

Dicen relación con motivos de no búsqueda de empleo por razones de salud, de embarazo, o de responsabilidades familiares de corta duración. En caso de declarar disponibilidad se clasifican como “potenciales”.

que serán considerados bajo la Nueva Encuesta Nacional de Empleo como un grupo entre los inactivos y para fines de comparabilidad con la OCDE, como desocupados (tasa de desempleo estandarizada).

ii. 6 Razones Familiares Permanentes

Esta es una categoría de inactivos importante, en la que predomina la “habitualidad”⁸ y en la que, mayormente se encuentran mujeres.

Se trata de un subconjunto de inactivos “típicos”, en términos de exhibir una baja propensión a la participación laboral. Sin embargo, si se declaran “disponibles para trabajar”, se clasifican como inactivos potencialmente activos.

ii. 7 Razones de Estudio

Como la anterior, también es una categoría importante entre los inactivos, asociada a la condición de Estudiantes. Quienes dentro de esta población señalan tener “disponibilidad” representan una escasa proporción del total, lo que hace a este segmento, exhibir una baja potencialidad.

ii. 8 Razones de Pensión o Montepiado

Son las personas que gozan de un ingreso permanente como consecuencia del trabajo realizado en su vida laboral activa, o bien por viudez. La potencialidad de este grupo es muy reducida, predominando la habitualidad en su condición.

ii. 9 Razones de Jubilación

Se trata de una población con características muy similares a la anterior. En esencia, son personas que perciben un flujo de ingresos asociado a su desempeño laboral previo.

ii.10 Razones de Salud Permanentes

Se trata de personas que señalan que o no buscaron empleo, o no estuvieron disponibles para trabajar, debido a que su salud no se los permite. Esta categoría de inactivos es, en la práctica, un grupo poblacional con prácticamente nula potencialidad (existen casos de disponibilidad, pero minoritaria).

ii. 11 Sin Deseo de Trabajar

Son personas que declaran no querer trabajar por lo que su potencial de participación es igual a cero. En general, se trata de individuos que no necesitan trabajar, ya que tienen garantizado un flujo de ingresos.

⁸ Por no haber buscado empleo durante las últimas cuatro semanas y no estar disponible para trabajar en las dos semanas siguientes a la semana de referencia.

Anexo 1: Diagramas de Flujo de Preguntas de Nuevo Encuesta Nacional de Empleo (NENE)

En este Anexo se muestra los diagramas de flujo que permiten dar seguimiento a la secuencia de preguntas de las que se derivan las distintas categorías del **Código Sumario de Empleo**. Como resultado, se genera una clasificación nominal y excluyente de la Población en Edad de Trabajar (véase Esquema 4, del cuerpo principal).

Diagramas de Flujo de Preguntas Cuestionario Nueva Encuesta Nacional De Empleo

El **Código Sumario de Empleo Específico (CSE_E)**, es la versión más desagregada del CSE, y se determina a partir de los módulos A y E del nuevo cuestionario (NENE). Su elaboración es compleja, por cuanto requiere el filtrado de ambos módulos simultáneamente.

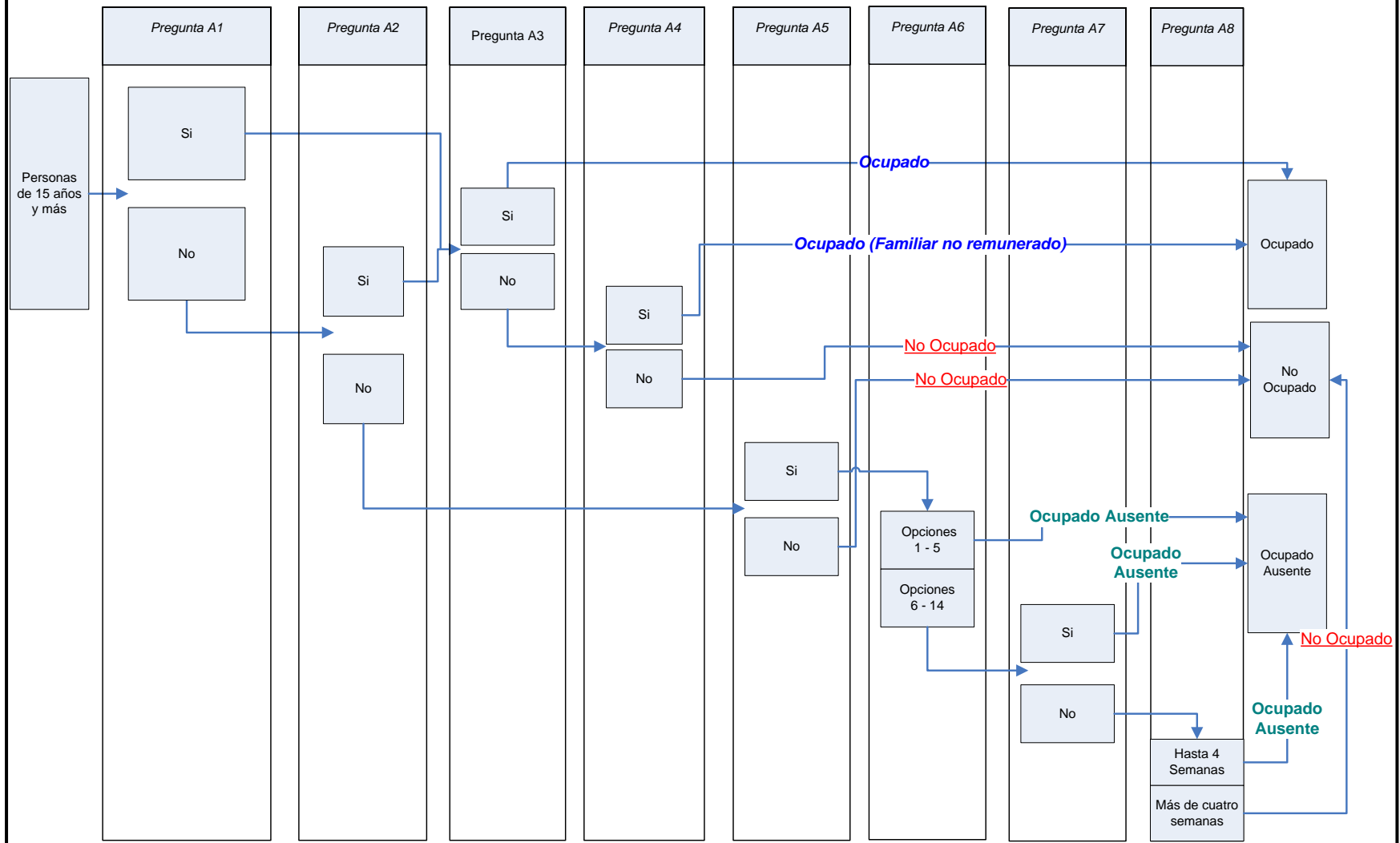
En el Módulo A, se establece el stock de ocupados entre la población mayor de 15 años en el país.

Las personas “no ocupadas”, deben pasar al Módulo E para determinar a cuál de los distintos stocks de la población mayor de 15 años pertenecen.

En primer lugar, se muestra la secuencia de preguntas del Módulo A; posteriormente, lo relativo al Módulo E.

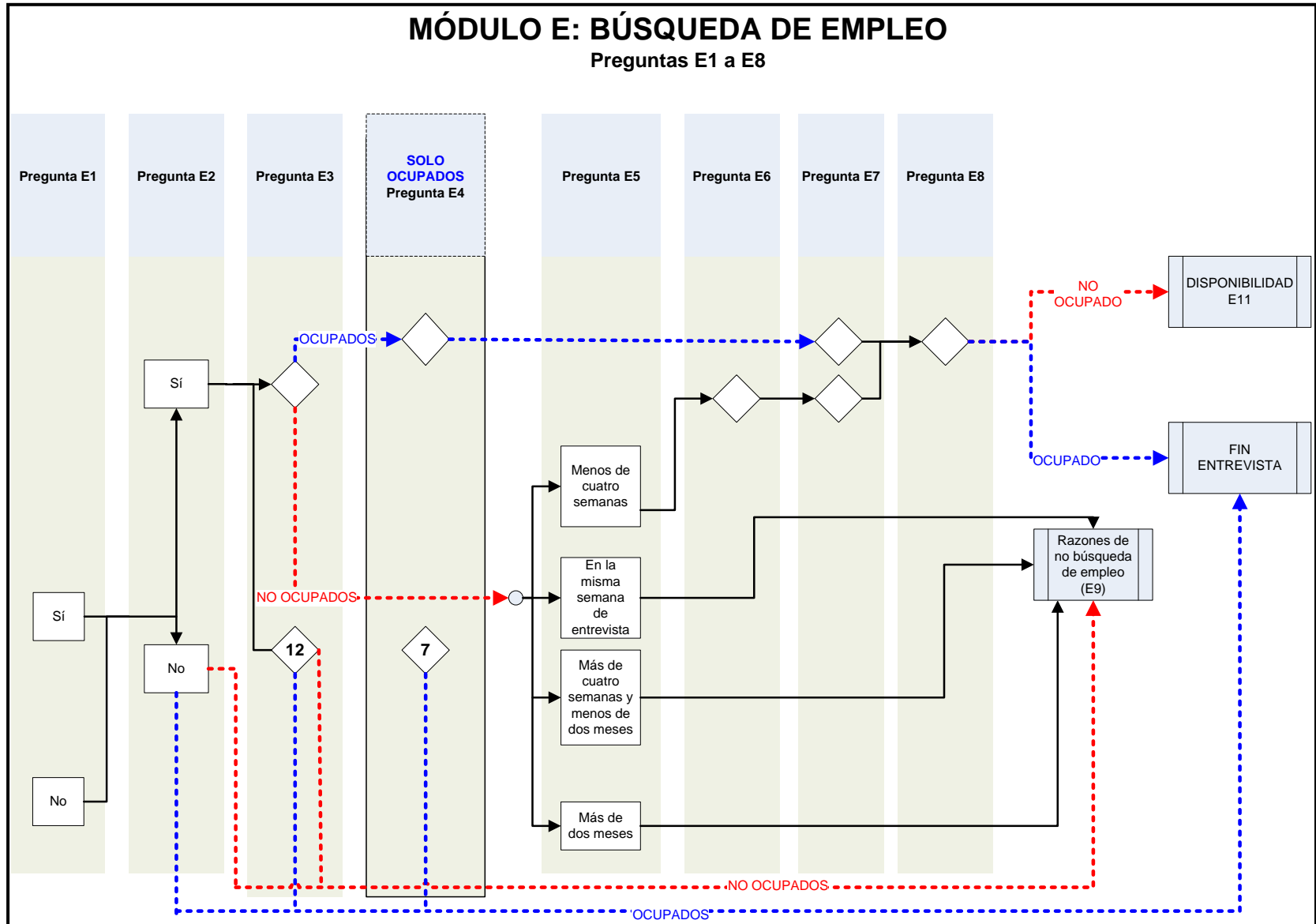
MÓDULO A: SITUACIÓN LABORAL EN LA SEMANA DE REFERENCIA

Preguntas A1 a A8



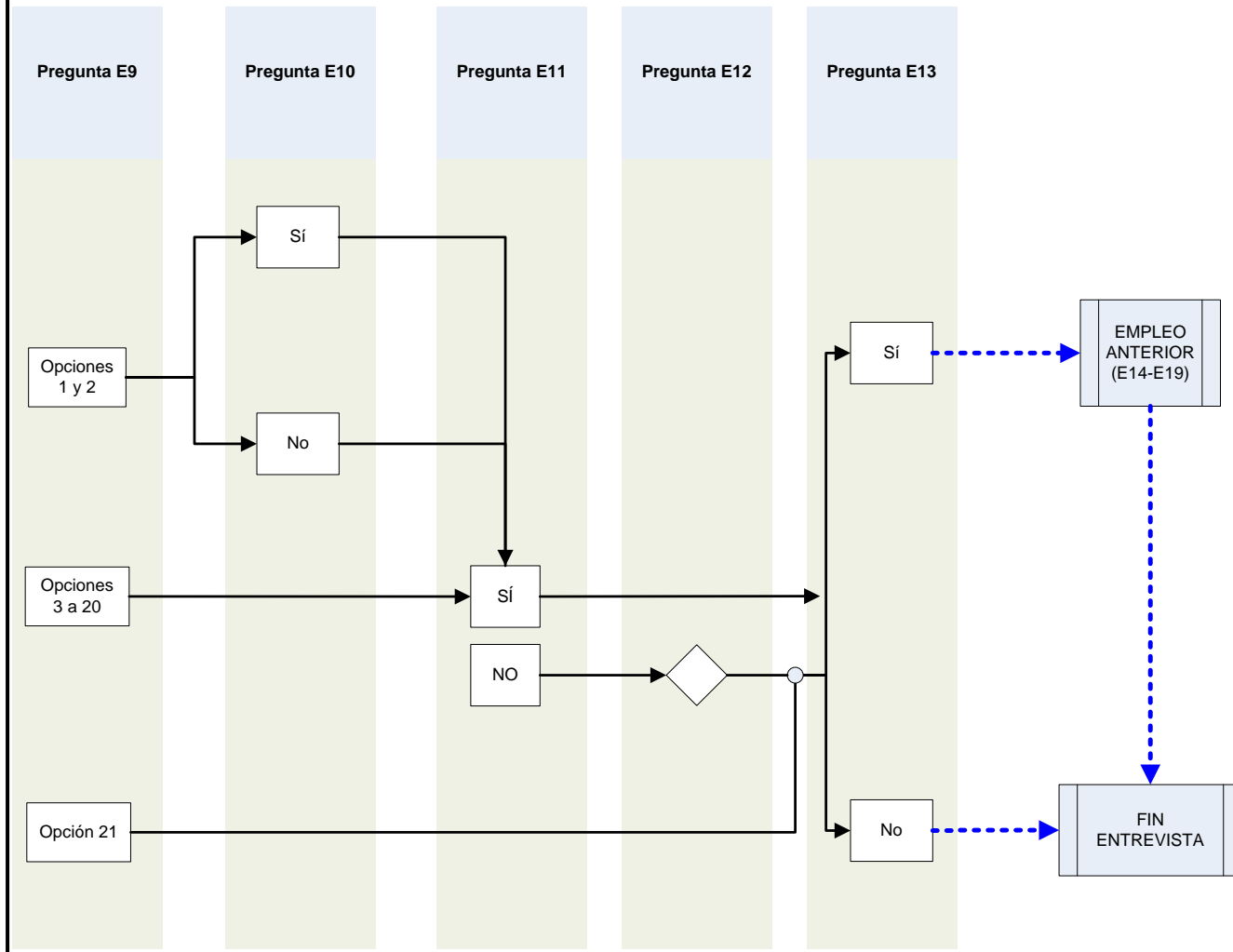
MÓDULO E: BÚSQUEDA DE EMPLEO

Preguntas E1 a E8



MÓDULO E: BÚSQUEDA DE EMPLEO

Preguntas E9 a E13



Anexo 2: Glosario

Población en Edad de Trabajar (PET): Población actualmente residente en el país de 15 años y más.

Población Económicamente Activa (PEA) O Fuerza de Trabajo: Personas en edad de trabajar, que durante la semana de referencia, cumplen los requisitos para ser incluidas en la categoría de ocupados o desocupados.

Ocupados: Todas las personas en edad de trabajar que durante la semana de referencia, trabajaron al menos una hora, recibiendo un pago en dinero o en especie, o un beneficio de empleado/empleador o cuenta propia.

- I. **Ocupados Tradicionales:** Ocupados que respondieron afirmativamente desde un inicio que trabajaron la semana pasada.
- II. **Ocupados no Tradicionales:** ocupados que al consultar inicialmente si trabajaron la semana pasada, respondieron que no lo hicieron, y luego en una serie de preguntas más inclusivas y abiertas responden que sí trabajaron.
- III. **Ocupados Ausentes:** ocupados que durante la semana de referencia, no trabajaron por diversas razones, pero que mantuvieron un vínculo estrecho con su empleo.

Desocupados: todas las personas en edad de trabajar, que no tuvieron un empleo durante la semana de referencia, buscaron uno durante las últimas cuatro semanas (incluyendo la de referencia) y están disponibles para trabajar en las próximas dos semanas (posteriores a la de referencia).

- I. **Cesantes:** es toda aquella persona que habiendo cumplido con los criterios de desocupado, tuvo anteriormente un empleo que duró por lo menos 1 mes.
- II. **Buscan trabajo por primera vez:** es toda aquella persona, que habiendo cumplido con los criterios de desocupado, que no hayan tenido un empleo que durara por lo menos 1 mes.

Iniciadores Disponibles: Todas las personas de la población en edad de trabajar, que esgrimen no haber buscado trabajo en las últimas cuatro semanas por la razón de que iniciarán pronto una actividad laboral y que, al mismo tiempo, declaran disponibilidad.

Este grupo poblacional, se incluye dentro de los inactivos pero, para ciertos efectos (de armonización estadística con la OCDE) se suma a la fuerza de trabajo.

Población No Económicamente Activa: Todas las personas de la población en edad de trabajar, no ocupados ni desocupados. Caen en esta categoría son personas con las siguientes razones de inactividad:

- Iniciadores⁹
- Razones estacionales
- Razones de desaliento
- Razones temporales
- Razones familiares permanentes
- Razones de estudio
- Razones de pensión o montepiado
- Razones de jubilación
- Razones de salud permanentes
- Sin deseos de trabajar

⁹ Individuos que lograron un acuerdo laboral o hicieron gestiones para empezar como cuenta propia, en las últimas cuatro semanas y que, al mismo tiempo pueden o no estar disponibles. De estar disponibles, son considerados en la PEA para fines de tasas estandarizadas OCDE.

Anexo 3: Principales Indicadores de Razón

A partir de las categorías de condición de actividad se construye los principales indicadores de razón:

Tasa de desempleo: número de personas desocupadas expresado como porcentaje de la fuerza de trabajo.

Tasa de desocupación estandarizada: número de personas desocupadas más los iniciadores disponibles, expresado como porcentaje de la fuerza de trabajo (incluyendo iniciadores disponibles). Es decir, se agrega tanto en el numerador como en el denominador a los iniciadores disponibles.

Tasa de participación: número de personas en la fuerza de trabajo expresado como porcentaje de la población en edad de trabajar.

Tasa de ocupación: número de personas ocupadas como porcentaje de la población en edad de trabajar.

Tasa de presión general: número de personas ocupadas que buscan un empleo, más los desocupados, más los iniciadores disponibles, expresado como porcentaje de la fuerza de trabajo (incluyendo iniciadores disponibles). Es decir, se considera en el numerador a los ocupados que buscaron un empleo y se agrega tanto en el numerador como en el denominador a los iniciadores disponibles.

Anexo 4: Indicadores Analíticos de Razón

Se trata de una batería de cuocientes que facilitan el análisis de la coyuntura laboral y, que están asociados a la población ocupada. Con los mismos, se pretende lograr una mejor caracterización de las condiciones ocupacionales de la población ya empleada en términos de: horas trabajadas y protección laboral.

% de ocupados con jornada a tiempo parcial involuntario: es el número de personas que trabajan, involuntariamente¹⁰, a tiempo parcial¹¹, expresado como porcentaje del total de personas que trabajan a tiempo parcial.

% de ocupados con jornada a tiempo parcial voluntario: es el número de personas que trabajan, voluntariamente, a tiempo parcial¹², expresado como porcentaje del total de personas que trabajan a tiempo parcial. En otras palabras, es el complemento del indicador inmediato anterior.

% de ocupados que trabajan más de 45 hrs. a la semana: es el número de personas que trabajan, habitualmente, más de 45 hrs. a la semana, expresado como porcentaje del total de personas ocupadas.

% de asalariados con contrato escrito: es el número de asalariados que tienen contrato escrito, expresado como porcentaje del total de personas asalariadas.

% de asalariados con contrato definido: es el número de asalariados que tienen contrato definido, expresado como porcentaje del total de personas asalariadas.

% de asalariados con un alto grado de protección: es el número de asalariados que tienen contrato escrito, cotización previsional, de salud, y seguro de desempleo, expresado como porcentaje del total de personas asalariadas.

% de asalariados con un bajo grado de protección: es el número de asalariados que NO tienen contrato escrito, NI cotización previsional, NI de salud, NI seguro de desempleo, expresado como porcentaje del total de personas asalariadas.

¹⁰ Ya que expresan su deseo y disponibilidad de trabajar más horas a la semana.

¹¹ Es decir, no más de 30 hrs. semanales, habitualmente.

¹² Es decir, no más de 30 hrs. semanales, habitualmente.

Anexo 5: Otras Definiciones

Horas Habituales: Las horas habituales corresponden a las horas comúnmente trabajadas, en un período de referencia determinado.

Horas Efectivas: Las horas efectivamente trabajadas, cuantifican el tiempo dedicado por las personas, al desempeño de actividades que contribuyen a la producción de bienes y servicios, durante un período de referencia determinado.

Trabajo a Tiempo Parcial Involuntario: Se consideran trabajadores a tiempo parcial involuntario a los ocupados cuya jornada de trabajo que declaran es de tipo parcial, que desean trabajar más horas y están disponibles para trabajarlas (inmediatamente o dentro de las dos semanas siguientes a la de referencia).

Anexo 6: Nomenclaturas de Clasificación Utilizadas¹³

Clasificación de Situación en el Empleo

Para fines de clasificación de la población económicamente activa, se toma en cuenta la Clasificación Internacional de la Situación en el Empleo (CISE), adoptada por la Organización Internacional del Trabajo en la decimotercera CIET de 1982. Este sistema clasifica la población económicamente activa en categorías tales como Empleador, Trabajador por cuenta propia, Empleado a sueldo o salario, y Trabajador familiar no remunerado¹⁴.

Clasificación Industrial de Actividades Económicas

La NENE utiliza la Clasificación Industrial Internacional Uniforme de todas las Actividades Económicas CIIU Revisión 3. Esta es una clasificación estandarizada de las actividades económicas, cuyo propósito es clasificar las entidades según su actividad económica principal. Este clasificador fue aprobado por la Comisión de Estadística de las Naciones Unidas de 1998, quienes fijaron como objetivo la implementación de una clasificación para la comparabilidad de la información a nivel Internacional.

La estructura general de la CIIU Rev.3 se compone de 17 clasificaciones representadas por letras que van desde la A hasta la Q, 60 divisiones, 159 grupos y 292 clases. Los dos primeros dígitos indican la división, tres dígitos identifican al grupo y cuatro dígitos tomados conjuntamente indican la clase. Para fines de clasificación, la representatividad de las estimaciones entregada por la NENE, es a nivel de las 17 clasificaciones.

Clasificación Internacional de Ocupaciones

Para la clasificación de ocupaciones se toma en cuenta la Clasificación Internacional Uniforme de Ocupaciones CIUO-88 publicada por la Organización Internacional de Trabajo en 1988 la cual reemplaza la CIUO-68 (o “Cota 70”, como se le conoce comúnmente). Este es un sistema clasificador de datos e información sobre las ocupaciones, concebida para establecer una clasificación nacional y facilitar comparaciones internacionales.

La estructura general de la CIUO 88 esta compuesta por 10 grandes grupos, 28 subgrupos principales, 116 subgrupos y 360 grupos primarios. La desagregación a un dígito indica el gran grupo, a 2 dígitos subgrupos principales, 3 dígitos subgrupos y 1 dígito grupo primario. Para fines de clasificación, la representatividad de las estimaciones entregada por la NENE, es a nivel de los grandes grupos.

¹³ Para información más detallada de los 3 primeros clasificadores ingresar a Buscador Asistido de Nomenclaturas BAN en http://www.ine.cl/pop_up/nomenclaturas.php

¹⁴ Para información más detallada ver Manual Conceptual.

Clasificación de Educación

Para la clasificación de la población según nivel educativo, se utiliza la Clasificación Internacional Normalizada de la Educación (CINE), concebida por UNESCO en 1970. Este es un instrumento que sirve de compilación y presentación de estadísticas de educación en los distintos países y permite comparabilidad a nivel internacional. Este clasificador cubre básicamente dos variables de clasificación: niveles y campos de educación.

* Para información más detallada de estos 3 últimos clasificadores en **Buscador Asistido de Nomenclaturas BAN**.

http://www.ine.cl/pop_up/nomenclaturas.php

CAPITULO II: MANUAL DE DISEÑO MUESTRAL¹⁵

1 Introducción

En el marco del mejoramiento continuo de las estadísticas generadas por el Instituto Nacional de Estadísticas (INE) y del interés por el acercamiento a las mejores prácticas utilizadas en países miembros de la OCDE, el Departamento de Investigación y Desarrollo del INE elaboró en 2010 un diseño muestral para una nueva forma de medir el empleo, con el objetivo de comparabilidad de sus estimaciones con aquellas de países pertenecientes a dicha organización.

Esta nueva forma de medir el empleo y sus características consideró la utilización de dos marcos muestrales según área geográfica (Urbana, RAU¹⁶ y Rural). Para el área urbana del país se utiliza el Marco Muestral de Manzanas. Mientras que en las áreas RAU y Rural el marco muestral utilizado es el tradicionalmente empleado por la antigua Encuesta Nacional de Empleo, un Marco Muestral de Secciones construido a partir del CENSO 2002.

Los tamaños muestrales fijados en esta encuesta fueron calculados de manera que, para los mismos niveles de estimación de la Encuesta Nacional de Empleo anterior, se obtuvieran errores similares, en relación a lo registrado históricamente en la encuesta.

¹⁵ Este capítulo contiene actualizaciones respecto de la primera versión de enero de 2010.

¹⁶ RAU: Resto de Área Urbana. Conglomeraciones de centros urbanos con menos de 40.000 habitantes.

2 Objetivo de la Encuesta

El objetivo de la Encuesta Nacional de Empleo ENE es caracterizar y cuantificar la población de 15 años y más respecto de su situación en el mercado laboral.

Esta caracterización permite capturar de una forma similar a las encuestas de los países miembros de la OCDE, las transformaciones y/o evolución en la Fuerza de Trabajo, específicamente de la Tasa de Desocupación, tanto a nivel nacional como en los correspondientes Niveles de Estimación¹⁷.

3 Población Objetivo

La población objetivo la constituye toda la población del país residente en viviendas particulares ocupadas. Esta definición excluye a la población que habita 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 son los porteros, conserjes y otros residentes. Por otro lado, la cobertura geográfica excluye a la población residente en las áreas de difícil acceso, denominadas ADA's.

¹⁷ Para mayor información de los Niveles de estimación revisar Cuadro n° 3.

4 Marco Muestral y estratificación

4.1 Marco Muestral: Conformación, partición y actualización del Marco Muestral

La Encuesta Nacional de Empleo, considera a contar de 2010 la utilización de un Marco Muestral de manzanas para las áreas urbanas del país, que en un principio se generaron a partir de cartografías digitales de alta validez métrica, de información anexa como lo son Mapcity y Dmapas y de la información proporcionada por las municipalidades referente a los permisos de edificación, además del material cartográfico obtenido a partir del precenso del año 2001 y el Censo de Población y Vivienda del año 2002, compuestos por los mapas regionales, provinciales, comunales y distritales que contienen las áreas urbanas.

En cuanto a las áreas RAU y Rural, el marco muestral que se utiliza es aquel generado a partir de la información proporcionada por el Censo de Población y Vivienda levantado el año 2002.

El marco muestral correspondiente al área urbana, involucra un nuevo concepto de unidad muestral, la manzana, cuya configuración corresponde a la geografía natural que las delimita. Este se distingue del marco de secciones de las áreas RAU y Rural, cuya configuración obedece a una conglomeración de un número determinado de viviendas (ver Cuadro n°1), con tamaños relativamente estables en el tiempo y configurados a partir de la información del CENSO año 2002, con fácil identificación en terreno ya que respeta los límites comunales.

Cuadro 1. Tamaño de secciones y número de viviendas a encuestar según área RAU y Rural.

Área	Número de viviendas	
	por sección	a encuestar por sección
RAU	50 - 150	12
Rural	40 - 120	15

4.2 Estratificación del marco muestral

En ambos casos el marco muestral se estratifica según su condición geográfica (División Político Administrativa) y según el número de viviendas y población que contenían al CENSO 2002, además de una segregación dependiendo de la actividad económica preponderante en el área.

Con esta información es posible clasificar los estratos según:

- Ciudades o grandes Centros Urbanos (CD): Conformadas por ciudades o conjuntos de ciudades adyacentes con 40.000 o más habitantes.
- Resto de Área Urbana (RAU): Conformadas por conjuntos de Centros Urbanos con menos de 40.000 habitantes.
- Área Rural (R): Conformado por el conjunto de entidades clasificadas como rurales de acuerdo a 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 primarias¹⁸.

El marco muestral de manzanas respeta la estratificación geográfica y segregación de acuerdo a la actividad económica de los marcos muestrales anteriormente usados por el INE¹⁹, y por otro lado, incorpora un nuevo concepto de estratificación que permite obtener muestras de manzanas de distintos tamaños. Es así como en la configuración del marco, se incorporó una variable que clasifica las manzanas de acuerdo a su tamaño en número de viviendas, formándose 5 grupos denominados “Estratos de Tamaño”, y 30 subgrupos. Esto permite, que estén representadas en la muestra, manzanas de todos los tamaños en forma proporcional, seleccionando un número fijo de viviendas, según corresponda el grupo y subgrupo (ver cuadro 2).

¹⁸ Se entiende por Actividad Primaria a toda aquella actividad relacionada con la extracción de recursos naturales (agricultura, caza, pesca, minería, etc.).

¹⁹ Para mayor información ver “Metodología de la Encuesta Nacional de Empleo”, de 2006 en la Web del INE.

Cuadro 2. Clasificación de las manzanas en estratos de tamaño por número de viviendas en el marco y a encuestar según tamaño muestral asociado al subgrupo.

Grupo	Subgrupo	N° Viviendas por Manzana		Viviendas a Encuestar *
		Mínimo	Máximo	
1	1	8	9	2
	2	10	13	3
	3	14	17	4
	4	18	23	5
2	5	24	25	6
	6	26	29	7
	7	30	33	8
	8	34	37	9
	9	38	44	10
3	10	45	45	11
	11	46	49	12
	12	50	53	13
	13	54	57	14
	14	58	61	15
	15	62	65	16
	16	66	69	17
	17	70	73	18
	18	74	77	19
	19	78	81	20
4	20	82	85	21
	21	86	89	22
	22	90	93	23
	23	94	97	24
	24	98	102	25
	25	103	105	26
	26	106	109	27
	27	110	113	28

Grupo	Subgrupo	N° Viviendas por Manzana		Viviendas a Encuestar *
		Mínimo	Máximo	
	28	114	154	29
5	29	155	303	30
	30	304	1043	31

* Se considera una fracción de muestreo de 25% aproximadamente

5 Descripción del Diseño Muestral

La confección del diseño muestral contempla el estudio de las variables que se desea investigar, en este caso, la evolución de la Fuerza de Trabajo, específicamente la Tasa de Desocupación.

El diseño muestral utilizado se define como probabilístico, bi-etápico, estratificado. Concordantemente las probabilidades de selección difieren según se seleccione en áreas urbanas, resto de área urbana o rurales:

Área Urbana: La probabilidad de selección de las Manzanas o unidades de primera etapa (UPE) es proporcional a su tamaño expresado por el número de viviendas o unidades de segunda etapa (USE) que tiene el estrato (geográfico y de tamaño) y del número de viviendas que tiene cada manzana. La selección de viviendas (o unidades de segunda etapa) es con igual probabilidad y con método de selección sistemático, a partir de una vivienda de arranque dada en forma aleatoria y una constante fija de selección.

Área RAU y Rural: La probabilidad de selección de las Secciones o unidades de primera etapa (UPE) es proporcional a su tamaño expresado en número de viviendas o unidades de segunda etapa (USE) que contiene cada Sección. La selección de viviendas (unidades de segunda etapa) es con igual probabilidad y con método de selección sistemática ejecutada en gabinete. Hecho esto, el encuestador va a terreno con un directorio predeterminado de las viviendas que entrevistará.

6 Cobertura Geográfica y áreas de estimación

Como se mencionó anteriormente, la cobertura geográfica contempla la población residente en todo Chile con exclusión de las ADA's y de las manzanas con 7 o menos viviendas, debido a potenciales problemas operativos y los costos que implica recolectar información en unidades de muy poca densidad poblacional.

Las áreas de estimación cubiertas por la muestra corresponden, en términos generales a:

Cuadro 3. Niveles de estimación según área geográfica..

Nivel Nacional	Total : Urbano y Rural
	U : Urbano ²⁰
	CD : Grandes Ciudades
	R : Rural
Nivel Regional	Total : Urbano y Rural
	U : Urbano
	CD : Grandes Ciudades
	GC : Grandes Centros ²¹
	R : Rural
Nivel Provincial	Total : Urbano y Rural

7 Unidades de muestreo y unidades de información

Las unidades de muestreo dependen del área en cuestión. Considerando la metodología de selección de la muestra, existen 2 unidades de muestreo. La unidad de muestreo de primera etapa (UPE), en el caso del área urbana (ciudad) corresponde a “manzanas”, y en el caso del área RAU y Rural, corresponde a “secciones”. En cuanto a la unidad de muestreo de segunda etapa (USE), en ambos casos corresponde a “viviendas particulares ocupadas” dentro de las manzanas o secciones seleccionadas.

²⁰ El nivel de estimación Urbano incluye a Grandes Ciudades y a Resto de Área Urbana (RAU)

²¹ El nivel de estimación Grandes Centros corresponde a una aglomeración de Grandes Ciudades dentro de una región, como Gran Valparaíso, Gran Santiago y Gran Concepción.

Las unidades de información o análisis corresponden a todas las personas que componen las viviendas, siendo encuestadas sólo las personas mayores de 15 años, aun cuando se registran algunas características (Tarjeta de registro del hogar) de cada una de las personas pertenecientes a la vivienda seleccionada.

8 Determinación del tamaño muestral

Considerando el propósito de la encuesta y en concordancia con los niveles para los cuales se desea obtener estimaciones representativas y confiables de la variable de interés, se determinó el tamaño muestral a partir de la tasa de desocupación media calculada a partir del promedio de las estimaciones de las tasas de 5 trimestres móviles de un estudio piloto levantado el año 2008, y de la obtención de errores muestrales similares a los reportados por la antigua encuesta de empleo en los mismos niveles de estimación (ver Anexo).

Así, el tamaño por nivel de estimación fue calculado según la siguiente fórmula:

$$n \cdot \bar{m} = \frac{S^2 \cdot Deff}{((Cv/100) \cdot \bar{Y})^2}$$

Donde:

$n \cdot \bar{m}$: Tamaño muestral o número total de viviendas en el nivel de estimación.

n : Número de secciones o manzanas por estrato.

\bar{m} : Número de viviendas a encuestar por sección o manzana.

S^2 : Cuasi varianza poblacional promedio del número de desocupados⁶ obtenido a partir de 5 trimestres móviles del estudio piloto.

$Deff$: Efecto del diseño, obtenido en el estudio piloto.

\bar{Y} : Tasa de Desocupación promedio de 5 trimestres móviles.

La afijación de la muestra en los distintos estratos, consideró además de la estructura piramidal²² de los errores de muestreo, los siguientes factores relativos a cada estrato:

²² Por estructura piramidal se entiende que a medida que se va desagregando el nivel de estimación, se obtienen mayores errores muestrales.

- a) Tasa de Desocupación dentro de cada estrato.
- b) El coeficiente de variación del número de Desocupados²³

Existe una mayor variabilidad de las estimaciones obtenidas de las unidades primarias en el nuevo marco debido, en parte, a la heterogeneidad en sus tamaños, por lo que se requiere un tamaño de muestra mayor que el empleado en la antigua encuesta de empleo para obtener los mismos errores muestrales en los mismos niveles de estimación. Debido a restricciones presupuestarias, se optó por disminuir el nivel de confiabilidad de 95% a 90% para mantener tanto tamaños como errores muestrales similares.

9 Distribución de la muestra

La muestra total se distribuye en un periodo trimestral, dividiéndose en tres submuestras de un tamaño aproximadamente similar, cada una de las cuales se asigna en forma sistemática a uno de los tres meses del trimestre móvil. De esta forma se generan tres submuestras, una para cada mes, las cuales cada una por sí sola no presentan la precisión suficiente para todos los niveles de estimación.

Con esta distribución, cada vivienda encuestada en el mes “t” vuelve a ser encuestada en el mes “t+3” y a su vez, cada vivienda de la muestra es encuestada una vez por trimestre. Así, las estimaciones de un trimestre “t” se calculan con la información correspondiente a los meses “t-1”, “t” y “t+1”.

10 Rotación de la muestra

La rotación de la muestra es un procedimiento que tiene como finalidad mantener actualizada la muestra y evitar el cansancio de los informantes.

Para llevar a cabo la rotación, la muestra de secciones y manzanas se divide en turnos de rotación, los que se estructuran considerando 1/6 de la muestra Urbana (manzanas), 1/9 de la muestra RAU y 1/12 de la muestra Rural (secciones). Esto permite rotar la totalidad de la muestra de viviendas en un periodo de 18 meses en el caso Urbano, 27 meses en el área RAU y 36 meses en el área Rural.

²³ Como una aproximación del coeficiente de variación de la Tasa de Desocupación.

La rotación consiste en reemplazar las viviendas dentro de la manzana o sección; en caso que se agoten los reemplazos, se sustituye la manzana o sección completa, por otra con similares características.

11 Selección de Unidades de muestreo y Factores de Expansión

11.1 Selección de unidades de primera y segunda etapa

11.1.1 Selección de manzanas desde el MM2008

La selección de las manzanas se realiza en cada estrato de la muestra (en el área urbana de la comuna) en dos etapas: se elige primero un grupo de tamaño de forma independiente, aleatoria y proporcional al tamaño en cuanto al total de viviendas y manzanas que posee y enseguida se elige una manzana del grupo seleccionado en forma aleatoria con igual probabilidad de selección. La selección se implementa en base a un algoritmo de selección que considera las dos etapas, que se describen a continuación. El procedimiento descrito permite seleccionar en la muestra manzanas grandes y pequeñas.

Etapas 1 Selección de un grupo de tamaño

Las manzanas se agrupan en Z grupos de tamaño a nivel del marco de manzanas (ver Cuadro 4 siguiente). Esta información se cruza con las comunas, de modo que en cada comuna puede existir a lo más 30 grupos ($l = 1, 2, \dots, z$ donde $z \leq 30$). El procedimiento aplicado es un proceso iterativo en que se escoge en forma aleatoria un grupo de tamaño²⁴ l en forma proporcional a su tamaño. Este último medido en términos de W_l que corresponde al producto entre el número total de manzanas y el número total de viviendas de cada grupo l , ($W_l = N_l \cdot M_l$).

²⁴ Los grupos de tamaño son conglomerados de manzanas censales. Los grupos de tamaño son las unidades primarias de muestreo que se utilizan en el procedimiento de selección implementado en el marco de manzanas de INE.

Cuadro 4. Método de selección del grupo de tamaño y manzanas.

Grupo de Tamaño G_l	Número de Manzanas N_l	Número de Viviendas M_l	Peso del Grupo de Tamaño $W_l = N_l \cdot M_l$	Ponderación $P_l = \frac{W_l}{\sum_{l=1}^Z W_l}$
1	N_1	M_1	$N_1 * M_1$	P_1
2	N_2	M_2	$N_2 * M_2$	P_2
3	N_3	M_3	$N_3 * M_3$	P_3
:	:	:		
:	:	:		
:	:	:		
Z	N_Z	M_Z	$N_Z * M_Z$	P_Z
Total	N_h	M_h	$\sum W_l$	1

Donde,

$$N_h = \sum_{l=1}^Z N_l \quad ; \quad M_h = \sum_{l=1}^Z M_l \quad ; \quad W_l = N_l \cdot M_l \quad ; \quad P_l = \frac{W_l}{\sum_{l=1}^Z W_l} \quad ; \quad \sum_{l=1}^Z P_l = 1$$

Etapa 2 Selección de una manzana.

Una vez escogido el grupo de tamaño, la manzana se elige con igual probabilidad.

En el cálculo de la probabilidad de selección, se descuenta la manzana (disminuye en uno el número de manzanas en el grupo y en el total de la comuna y/o estrato) y se descuenta el número de viviendas²⁵ (disminuye el número de viviendas en el grupo seleccionado y en el total de la comuna y/o estrato), para finalmente actualizar tanto el número de manzanas y viviendas en cada grupo de tamaño, iterando el procedimiento hasta seleccionar todas las manzanas que completan el número de viviendas requeridas.

Como se ha mencionado, este método asegura que la muestra quede conformada tanto por manzanas grandes, medianas y pequeñas, asignando mayor ponderación a grupos de tamaño que concentran mayor cantidad de manzanas y viviendas. Como las probabilidades de selección se obtienen de forma iterativa y su cálculo se puede formular mediante la siguiente aproximación:

²⁵ De acuerdo al número total de viviendas que registra el marco, cada manzana tiene predeterminado el total de viviendas que pueden ser seleccionadas para una muestra. Este total es consistente con una fracción de muestreo de aproximadamente 25% (ver Cuadro 2). Para efectos de selección, sin embargo, se descuenta el número de viviendas registradas en el marco con el fin de recalcular las probabilidades de selección P_l .

$$\pi_{hi} = n_h \cdot \frac{M_{hi}}{M_h}$$

donde:

h : Corresponde al estrato h .

i : Corresponde a la manzana “ i ”.

n_h : Número de manzanas en la muestra en el estrato “ h ”.

M_{hi} : Número de viviendas en la manzana “ i ” del estrato “ h ”.

M_h : Número de viviendas en el estrato “ h ”.

De esta manera se asume que la probabilidad de selección de una manzana es proporcional al tamaño que tiene esta manzana dentro del estrato del que fue seleccionada.

11.1.2 Selección de secciones desde el MM2002

La selección preliminar de las secciones se realiza al interior de cada comuna y área en forma proporcional al total de viviendas ocupadas que éstas contienen, según información del CENSO 2002. Esta selección se realiza en forma sistemática proporcional al tamaño, mediante un algoritmo de acumulación de viviendas en el estrato, en el que se elige un período y arranque aleatorio, como sigue:

Paso 1

Se construyen Q intervalos (equivalente al número de secciones del estrato o comuna) de la manera como se muestra en el Cuadro 5:

Cuadro 5: Método de selección de secciones.

Sección	N° Viviendas	Acumulado del número de viviendas	Límite inferior	Límite superior
i	M_i	ΣM_i	L_{inf}	L_{sup}
1	M_1	$\Sigma_1 = M_1$	1	Σ_1
2	M_2	$\Sigma_2 = M_1 + M_2$	$\Sigma_1 + 1$	Σ_2
3	M_3	$\Sigma_3 = M_1 + M_2 + M_3$	$\Sigma_2 + 1$	Σ_3
⋮	⋮	⋮	⋮	⋮
Q-1	M_{Q-1}	$\Sigma_{Q-1} = M_1 + M_2 + \dots + M_{Q-1}$	$\Sigma_{Q-2} + 1$	Σ_{Q-1}
Q	M_Q	$\Sigma_Q = M_1 + M_2 + \dots + M_{Q-1} + M_Q$	$\Sigma_{Q-1} + 1$	Σ_Q

Paso 2

Se calcula el período dado por: $K = \Sigma Q/n$, con ΣQ siendo el total de viviendas de la comuna o estrato y n siendo el número de secciones a seleccionar (se supone $n \leq Q$, es decir, que el número de secciones a seleccionar es menor o igual al número de secciones de la comuna o estrato).

Paso 3

Se genera un número aleatorio entero A entre 1 y el período K .

Paso 4

Enseguida se va sumando sucesivamente el período K para obtener distintos valores que pertenecerán a los distintos intervalos que indican qué sección es la seleccionada. De acuerdo a este algoritmo, las secciones seleccionadas son aquellas cuyo intervalo asociado contiene a los valores dados por: $A, A + K, A + 2K, A + 3K, \dots, A + (n - 1)K$

El procedimiento anterior, para secciones homogéneas en cuanto al tamaño (de una cantidad de viviendas similar), no permite repeticiones. Por el contrario, cuando las secciones son de tamaño muy distinto, llegando a ser algunas el doble y más del doble de las demás, el procedimiento o algoritmo selecciona menos secciones de las requeridas, ya que es posible seleccionar dos o más veces una misma sección, si ésta es demasiado grande respecto a las demás. Se puede demostrar que para secciones homogéneas en cuanto al tamaño, la probabilidad de inclusión de una sección con M_i viviendas está dada por: $n \cdot M_i / \Sigma_Q M_i$.

11.1.3 Selección de viviendas

Una vez determinadas las manzanas y secciones definitivas de la muestra, se procede a la selección de las viviendas al interior de los respectivos conglomerados.

En el marco de manzanas la fracción de muestreo que se aplica al interior de cada manzana es de 25% aproximadamente, con un mínimo de 2 viviendas a encuestar. Para no concentrar la muestra en manzanas grandes (en número de viviendas), se establece una cota superior de 31 viviendas a encuestar. Esto permite que manzanas de todos los tamaños estén representadas en la muestra en forma proporcional, a través de un número fijo de viviendas, según el grupo al que pertenezca.

En el marco de secciones la selección de viviendas se realiza al interior de cada sección, en forma sistemática, con igual probabilidad. Específicamente, el algoritmo de selección sistemático de viviendas, que es aplicado también en otras encuestas, se estandariza para que independientemente de quien aplique el mismo algoritmo, bajo las mismas condiciones de ordenamiento previo²⁶, seleccione las mismas viviendas. Este procedimiento es análogo a introducir una semilla de aleatorización para selecciones aleatorias, las cuales arrojan también las mismas selecciones.

Sean “ M_i ” el número de viviendas que posee la manzana o sección y m_i el número de viviendas a seleccionar. Para la selección de “ m_i ” viviendas se siguen los siguientes pasos:

Paso 1

Se calcula el período $K = M_i/m_i$.

Paso 2

Se calcula el arranque A o primera selección como la mitad entre el número uno y el período K . Al sumarle 0,5 al período K , se evita que el arranque empiece con cero:

$$A = \text{Redondear}([k + 0,5]/2; 0)$$

Paso 3

Luego se va sumando el período K al arranque A para obtener distintos valores que una vez redondearlos van generando las sucesivas selecciones:

$$A, A + K, A + 2K, A + 3K, \dots, A + (m_i - 1)K.$$

²⁶ Tanto en manzanas como en secciones las viviendas son ordenadas según la variable “orden vivienda”, mediante el cual se asegura el recorrido del conglomerado.

La primera selección A ya es un número redondeado a entero, la segunda es el redondeo de $A + K$, la tercera es el redondeo de $A + 2K$, y así sucesivamente hasta la m_i selección dada por el redondeo de $A + (m_i - 1)K$.

Podría pensarse que se resta aleatoriedad al partir la selección siempre con la mitad entre uno y el período, pero el arranque es impredecible debido a que se pre-seleccionan sólo las elegibles, generando que la primera vivienda elegible no corresponda a la primera vivienda de acuerdo al orden de la vivienda según recorrido de la enumeración. Finalmente, la probabilidad de selección de una vivienda dentro de un conglomerado está dada por m_i/M_i .

11.2 Factores de Expansión

El Factor de Expansión se interpreta como la cantidad de viviendas en la población que representa una vivienda en la muestra. Lo que se transmite a las personas que viven en la vivienda.

Este depende muestralmente del número de viviendas particulares ocupadas que tienen las unidades de primera etapa (manzanas o secciones) al momento de la enumeración, y de la cantidad de manzanas/secciones en la muestra al interior de cada estrato. Mientras que según la información reportada en el respectivo marco muestral, depende del número de viviendas en el estrato y en la sección o manzana seleccionada.

El Factor de expansión Teórico $F1_{hi}$, es calculado como el inverso de la probabilidad de selección de cada vivienda. Este factor es constante por sección o manzana y se aplica sobre todas las viviendas y por lo tanto a todas las personas que pertenecen a la sección o manzana i del estrato h .

El cálculo corresponde a:
$$F1_{hi} = \frac{M_h}{M_{hi} \cdot n_h} \cdot \frac{M'_{hi}}{m_{hi}}$$

donde:

M_h : Número de viviendas particulares en el estrato h al Censo 2002 (en el caso de secciones) o al año de actualización del marco en el caso de manzanas.

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 Urbano.

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

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

El factor de expansión incluye un ajuste de población de personas por estrato, donde cada estrato tiene sus propias estimaciones de proyecciones de población, por sexo y tramo de edad (menores de 15 años y de 15 años y más) con el objeto de aumentar la precisión de las estimaciones y ser consistente con los stocks poblacionales de las proyecciones de población. De esta manera, después de este ajuste, por cada sección o manzana del estrato h , se tienen 4 ajustes adicionales a los factores de expansión que se aplican respectivamente a todas las personas dentro de la sección, esto es: hombres menores (menores de 15 años), hombres mayores (de 15 años o más), mujeres menores (menores de 15 años) y mujeres mayores (de 15 años o más).

El cálculo del Factor Ajustado para la persona k , de sexo s y tramo de edad e corresponde a:

$$F2_{hik} = \frac{M_h}{M_{hi} \cdot n_h} \cdot \frac{M'_{hi}}{m_{hi}} \cdot \frac{P_{hse}}{\hat{P}_{hse}} = F1_{hi} \cdot \frac{P_{hse}}{\hat{P}_{hse}}$$

donde:

P_{hse} : Proyección de población total de personas de sexo s y tramo de edad e en el estrato h , considerado como un parámetro para el ajuste.

\hat{P}_{hse} : Número total de personas de sexo s y tramo de edad e estimadas con el factor teórico para el estrato h .

$$\text{con } \hat{P}_{hse} = \sum_i \sum_j \frac{M_h}{M_{hi} \cdot n_h} \cdot \frac{M'_{hi}}{m_{hi}} \cdot P_{hijse} = \sum_i \sum_j F1_{hi} \cdot P_{hijse}$$

P_{hijse} : Número total de personas de sexo s y tramo de edad e en la vivienda j , de la unidad de primera etapa (sección o manzana) i en el estrato h .

El resultado de este ajuste implica que la sumatoria sobre todas las personas en el estrato h , expandidas con el factor de expansión ajustado $F2_{hik}$, es equivalente a la proyección de población para ese estrato, tal como se esperaba, de acuerdo a la construcción de este factor ajustado.

12 Estimadores de Totales y Razones

Tanto para los estimadores como sus varianzas, denotamos estrato de muestreo h al área geográfica básica correspondiente en forma general a la intersección o unión de intersecciones de los estratos Urbano y Rural con las comunas. Estos fueron definidos para la selección aleatoria e independiente de las unidades de primera etapa (secciones o manzanas como conglomerados de viviendas). En ellos donde se obtienen las primeras estimaciones desagregadas. A partir de estos estratos, se construyen los niveles de estimación como su agregación en dichos niveles.

Para facilitar la lectura de los estimadores, varianzas y sumas, el subíndice h se referirá al estrato geográfico, el subíndice i se referirá a la sección o manzana dentro del estrato h , el subíndice j se referirá a la vivienda dentro de la sección o manzana i del estrato h y finalmente el subíndice k a la persona dentro de la vivienda j de la sección o manzana i del estrato h .

La nomenclatura utilizada para los estimadores es:

$F2_{hik}$: Factor de Expansión para la persona k ajustado por proyecciones de población y asociado a la sección o manzana i en el estrato h .

$P_{hijk} = 1$, para toda persona k de la vivienda j de la sección o manzana i en el estrato h .

P_{hijk_c} : Persona k de la vivienda j de la sección o manzana i en el estrato h , que cumplen con la característica C .

Es decir:
$$P_{hijk_c} = \begin{cases} 1 & , \text{ si la persona cumple con característica } C \\ 0 & , \text{ en otro caso} \end{cases}$$

Y_{hijk} : Valor de la variable Y medida en la persona k de la vivienda j de la sección o manzana i en el estrato h .

Y_{hijk_c} : Valor de la variable Y medida en la persona k de la vivienda j de la sección o manzana i en el estrato h que cumple con la característica C .

Es decir:
$$Y_{hijk_c} = \begin{cases} Y_{hijk} & , \text{ si la persona cumple con característica } C \\ 0 & , \text{ en otro caso} \end{cases}$$

También podemos expresar:

$$Y_{hijk_c} = Y_{hijk} \cdot P_{hijk_c} = \begin{cases} Y_{hijk} \cdot 1 = Y_{hijk} & , \text{ Si la persona cumple} \\ & \text{ con característica } C \\ Y_{hijk} \cdot 0 = 0 & , \text{ en otro caso} \end{cases}$$

12.1 Estimadores del Total de la variable Y

$$\hat{Y}_h = \sum_i \sum_j \sum_k F2_{hik} \cdot Y_{hijk}, \text{ Total Estimado de la Variable Y en el estrato h.}$$

$$\hat{Y} = \sum_h \sum_i \sum_j \sum_k F2_{hik} \cdot Y_{hijk} = \sum_h \hat{Y}_h, \text{ Total estimado de la variable Y en el nivel de estimación requerido.}$$

La Estimación del Total en el nivel de estimación requerido se obtiene como la suma de los Totales sobre todos los estratos que componen dicho nivel.

$$\hat{Y}_{h_c} = \sum_i \sum_j \sum_k F2_{hik} \cdot Y_{hijk_c}, \text{ Total estimado de la variable Y en personas con característica C, en el estrato h.}$$

$$\hat{Y}_c = \sum_h \sum_i \sum_j \sum_k F2_{hik} \cdot Y_{hijk_c} = \sum_h \hat{Y}_{h_c}, \text{ Total estimado de la variable Y en personas con característica C, en el nivel de estimación requerido.}$$

Nuevamente la Estimación del Total para el nivel de estimación requerido se obtiene como la suma de los Totales de los estratos que componen dicho nivel.

12.3 Estimadores de razón

Se denomina razón al cociente entre dos variables que son estimadas. Tanto el numerador como denominador de este cociente pueden ser variables cuantitativas o cualitativas.

Nuevamente el índice **h** se referirá al estrato, el **i** a la sección o manzana, el **j** a la vivienda y finalmente el índice **k** a la persona.

12.3.1 Estimador de razón entre las variables X e Y

La razón de la variable Y respecto a la variable X, en el estrato h, está dada por:

$$\hat{R}_{h_{y/x}} = \frac{\hat{Y}_h}{\hat{X}_h} = \frac{\sum_i \sum_j \sum_k F2_{hik} \cdot Y_{hijk}}{\sum_i \sum_j \sum_k F2_{hik} \cdot X_{hijk}} = \frac{\text{Total estimado de variable Y en el estrato h}}{\text{Total estimado de variable X en el estrato h}}$$

De forma similar obtenemos el estimador de razón en el nivel de estimación requerido, como una razón combinada.

$$\hat{R}_{y/x} = \frac{\hat{Y}}{\hat{X}} = \frac{\sum_h \sum_i \sum_j \sum_k F2_{hik} \cdot Y_{hijk}}{\sum_h \sum_i \sum_j \sum_k F2_{hik} \cdot X_{hijk}} = \frac{\text{Total estimado de variable Y en nivel requerido}}{\text{Total estimado de variable X en nivel requerido}}$$

Nota: recordar que $\hat{P}_h = P_h$, es decir, por construcción, la estimación de personas dentro de cada estrato y nivel, coincide con la proyección de población.

13 Varianza de los Estimadores

La precisión de las principales estimaciones se estima a partir del Método del Conglomerado Último²⁷, considerando que la mayor contribución a la varianza de un estimador, en un diseño multi-etápico, es la que se origina a partir de las Unidades de Primera etapa de Muestreo (UPE). El Método del Conglomerado Último permite simplificar un diseño complejo multi-etápico en un caso especial del muestreo de conglomerados sin submuestreo, es decir en una etapa. Se denomina *conglomerado último* al conjunto de unidades muestrales de última etapa que pertenecen a una unidad primaria de muestreo, independiente del número de etapas de selección que se efectúan al interior de la primera etapa de selección.

Por otro lado, al igual que en encuestas de gran tamaño, se asume que, independientemente del número de etapas de selección, en la primera etapa de selección los conglomerados (UPM) han sido seleccionados con remplazo aun cuando el actual esquema de selección pueda realizarse sin remplazo.

13.1 Varianza del Total de la variable X

La estimación de la varianza del estimador de un total, se calcula mediante la siguiente expresión:

$$V(\hat{X}) = \sum_{h=1}^L \left[\frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{X}_{hi} - \frac{1}{n_h} \hat{X}_h \right)^2 \right]$$

13.2 Varianza de la razón del Total de la variable Y respecto al Total de la variable X

Para medir la precisión de los estimadores de razón, conjuntamente al método del Conglomerado Último se aplica la aproximación lineal de Taylor sobre el estimador no lineal de la razón²⁸. Entonces, tenemos que la varianza de una razón \hat{R} definida por:

$$\hat{R} = \frac{\hat{Y}}{\hat{X}}$$

²⁷ Hansen, Hurwitz y Madow (1953),

²⁸ En el documento de Franca y Soares (2007) se encuentra la derivación de la expresión.

viene dada por:

$$V(\hat{R}) = \frac{1}{\hat{X}^2} \{V(\hat{Y}) + \hat{R}^2 V(\hat{X}) - 2\hat{R} COV(\hat{X}, \hat{Y})\}$$

Donde, $V(\hat{Y})$ y $V(\hat{X})$, son varianzas de los estimadores de totales, los cuales se calculan como se especificaron en el punto anterior, mientras que la covarianza entre \hat{Y} y \hat{X} , $COV(\hat{X}, \hat{Y})$, viene dada por:

$$COV(\hat{X}, \hat{Y}) = \sum_{h=1}^L \left[\frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{X}_{hi} - \frac{1}{n_h} \hat{X}_h \right) \left(\hat{Y}_{hi} - \frac{1}{n_h} \hat{Y}_h \right) \right]$$

14 Coeficiente de Variación y Error asociado al estimador

Los errores de muestreo existen como consecuencia de la utilización de una parte de la población para estimar características de todo el conjunto poblacional. Sin embargo, existen distintas formas de medir estos errores para las variables de interés como por ejemplo, la varianza, la desviación estándar o algunas transformaciones de ellas, como el error absoluto, el coeficiente de variación, error relativo, entre otros. En la Encuesta Nacional de Empleo (ENE) se utilizan, principalmente, las tres últimas medidas de error, que ayudan a determinar la precisión estadística de las estimaciones.

En primer lugar, el error absoluto asociado a una estimación de un 95% de confiabilidad se puede interpretar como la diferencia máxima que se esperaría observar entre el valor estimado y el verdadero valor poblacional, en una estimación con una confiabilidad del 95%. En tanto, el error relativo y el coeficiente de variación son medidas que indican la precisión relativa de la estimación. Este último, se obtiene del cociente entre la desviación estándar (raíz cuadrada de la varianza) y el valor medio de la estimación.

Error absoluto e intervalo de confianza

Para el cálculo del intervalo de confianza de un 95% para la estimación de la variable Y, se utiliza el error absoluto (error de estimación que está expresado en las mismas unidades de la variable en estudio²⁹).

²⁹ Cuando la variable Y es un porcentaje o tasa, el error absoluto se expresa en puntos porcentuales.

El error absoluto nos permite observar cuanto es la diferencia máxima esperada entre el valor estimado y el verdadero parámetro poblacional, donde dicha estimación se puede realizar a distintos niveles de confianza, siendo la más común la del 95%. A partir de este error se construye el intervalo de confianza de la variable de interés Y.

El error absoluto se obtiene del producto entre la desviación estándar ($S(\hat{y})$) de la estimación y el valor crítico asociado al percentil de la distribución T-student con v grados de libertad ($t_{1-\frac{\alpha}{2},v}$), donde el primer argumento hace referencia al nivel de confiabilidad con el cual se desea hacer inferencia (95%), mientras que el segundo argumento corresponde a los v grados de libertad con los que se calcula el valor crítico³⁰. Su expresión algebraica está dada por:

$$E_{\alpha}(\hat{y}) = t_{1-\frac{\alpha}{2},v} \cdot S(\hat{y})$$

Alternativamente, en términos probabilísticos, se puede expresar como:

$$\Pr(|\hat{Y} - Y| \leq E_{\alpha}(\hat{y})) = 1 - \alpha$$

Con \hat{Y} e Y siendo los valores estimado y verdadero en la población, respectivamente.

A través de este error se puede construir un intervalo de confianza del 95% con el verdadero valor de la variable en la población, considerando como el valor mínimo del parámetro poblacional el parámetro estimado menos el error absoluto; y como valor máximo el parámetro estimado más el error absoluto. Matemáticamente, el intervalo de confianza es igual a:

$$IC(\hat{y})_{95\%} = (\hat{y} - E_{\alpha}(\hat{y}); \hat{y} + E_{\alpha}(\hat{y})) = \left(\hat{y} - t_{1-\frac{\alpha}{2},v} \cdot S(\hat{y}); \hat{y} + t_{1-\frac{\alpha}{2},v} \cdot S(\hat{y}) \right)$$

Coefficiente de Variación y Error relativo

Complementariamente, para cuantificar en términos relativos los errores asociados al muestreo, se utiliza el error relativo y el coeficiente de variación. El error relativo representa, en términos

³⁰ Los grados de libertad corresponden a la diferencia entre el número de conglomerados no vacíos (con presencia de la variable de interés) y el número de estratos no vacíos. Para un ejemplo práctico, ver Anexo 2 dentro de este documento.

porcentuales, la diferencia máxima que existe entre la estimación puntual y el verdadero valor de la variable \hat{y} en la población, a un nivel de confiabilidad definido generalmente en 95%, denotado por el percentil de distribución T-student ($t_{1-\frac{\alpha}{2},v}$).

La expresión matemática del error relativo es igual a:

$$E_r(\hat{y}) = t_{1-\frac{\alpha}{2},v} \cdot \frac{\sqrt{S(\hat{y})}}{\hat{y}}$$

Donde $\frac{\sqrt{S(\hat{y})}}{\hat{y}}$ representa el coeficiente de variación (CV) asociado a la estimación de la variable \hat{y} .

Ambos, el error relativo y el coeficiente de variación, pueden ser expresados en porcentajes, lo que los distingue de otras mediciones del error, al permitir comparar entre diferentes estimaciones, soslayando las diferencias entre unidades de medida.

Por ejemplo, considerando un error relativo de un 20% asociado a un nivel de confianza del 95% con 120 grados de libertad (valor t crítico = 1,9799), se puede interpretar que el verdadero valor poblacional de la variable, está a más o menos 20% del valor estimado, con una confiabilidad del 95%. La interpretación del coeficiente de variación es idéntica, al ser el error relativo una transformación monótona del coeficiente de variación, ya que este se multiplica por el escalar $t_{1-\frac{\alpha}{2},v}$. Por eso, con un 95% de confiabilidad, el coeficiente de variación es aproximadamente la mitad del error relativo en muestras suficientemente grandes e idéntico cuando el número de observaciones (y en consecuencia los grados de libertad) tienden a infinito.

El Cuadro 6, muestra distintos rangos en niveles del error relativo y coeficiente de variación, y la confiabilidad de su estimación en términos estadísticos. Cabe señalar que esta clasificación es un primer paso, ya que la decisión final depende del tamaño de la muestra, y será finalmente responsabilidad del usuario dar buen uso e interpretación de las cifras apoyándose en estos dos antecedentes.

Cuadro 6: Clasificación del error relativo (al 95%) y coeficiente de variación.

Clasificación	Rango del Error Relativo	Rango del Coeficiente de Variación
Muy Bueno	(0 , 10)	(0 , 5)
Bueno	[10 , 30)	[5 , 15)
Regular	[30 , 40)	[15 , 20)
No Recomendable	[40 , 100]	[20 , 100]

Para realizar un análisis adecuado de la precisión estadística de las estimaciones se debe no sólo considerar el error de muestreo asociado al estimador (error relativo o coeficiente de variación), sino que también si es suficiente el tamaño de la muestra asociado a dicha estimación.

Entonces, para evaluar la confiabilidad de un estimador se propone analizar de forma conjunta tanto el coeficiente de variación como el tamaño de la muestra. Se recomienda, a modo de referencia, utilizar los siguientes criterios:

- I. Hasta un 5% se considera que la estimación es precisa.
- II. Entre un 5% y un 15% se considera una buena estimación, sin embargo, hay que tomar en cuenta si el número de observaciones es suficiente o no.
- III. Entre un 15% y un 20% se podría considerar una estimación poco confiable, dependiendo si el número de observaciones es suficiente o no.
- IV. Sobre el 20% la estimación se presenta como poco precisa y por ende no recomendable, independiente el número de observaciones que se tenga disponible.

Para el tamaño de la muestra no existe una recomendación oficial al respecto, debido a que los totales poblacionales de las variables de interés difieren entre sí, lo que hace imposible definir un único umbral que sea transversal a todas las categorías o desagregaciones de la encuesta. Por lo anterior, se deja a criterio de los analistas la responsabilidad de definir el límite mínimo de observaciones que aseguren representatividad y precisión del estimador. A su vez, es importante que al realizar alguna estimación se considere que las observaciones sean distribuidas en distintos conglomerados y estratos y que además exista varianza al interior de ellos, dicho de otro modo, que exista un número suficiente de observaciones para el cálculo de la varianza.

El cruce entre ambos criterios, que se explicita en el Cuadro 7, definirá el trato que se deba realizar a cada estimador basándose en los siguientes 3 niveles:

- I. Estimador no confiable: No se recomienda utilizar este estimador para predecir características poblacionales (inferencia estadística). Una alternativa es agrupar categorías. Por ejemplo, se pueden agrupar los sectores de actividad económica según el sector productivo al que pertenezcan (primario, secundario y terciario)³¹. Luego, al agrupar se debe analizar nuevamente su precisión, ya que se obtendrán coeficientes de variación distintos.
- II. Uso descriptivo (Nota metodológica): En los casos que la estimación realizada por el investigador no pueda ser agrupada es necesario incluir una nota metodológica que advierta que de los errores que conlleva dicha estimación y que su uso es sólo descriptivo. Esto en general se pueda usar para sacar conclusiones de tendencias, pero no de niveles.
- III. Estimador confiable: Se puede utilizar el estimador según los objetivos propuestos.

Cuadro 7: Criterio de confiabilidad del estimador relacionado a su coeficiente de variación y número de observaciones.

Coeficiente de variación (%)	Número de Observaciones	
	Bajo	Alto
[20 , 100]	Estimador no confiable	Estimador no confiable
[15 , 20)	Estimador no confiable	Descriptivo (Nota metodológica)
[5 , 15)	Descriptivo (Nota metodológica)	Estimador confiable
(0 , 5)	Estimador confiable	Estimador confiable

³¹ Por ejemplo, se podría agrupar a los Ocupados que pertenezcan a ramas de actividad similares en un sector más amplio, definido por el investigador. Considérese a los Ocupados clasificados según rama de actividad económica utilizando CIIU4.CL 2012 según CAENES (Clasificador de Actividad Económica Nacional para Encuestas Sociodemográficas). Eventualmente, se podría observar para una región particular, que las categorías Agricultura, Ganadería, Silvicultura y Pesca, y Minería, cuenten individualmente con una participación inferior al 1%, pero que conjuntamente, al categorizarlas en un grupo más amplio, que podría denominarse sector primario, tengan una participación relativa superior al 1%, lo que mejoraría la estimación en términos de variabilidad, permitiendo hacer una mejor inferencia.

Anexo 1: Tamaño y Errores Muestrales según Nivel de Estimación.

REGION	NIVEL DE ESTIMACIÓN	TASA DESOCUPACIÓN PROMEDIO 5 TRIMESTRES	TAMAÑO TEÓRICO CALCULADO	ERROR RELATIVO	ERROR ABSOLUTO
PAIS	NACIONAL	9,1%	35.719	3,3%	0,30%
PAIS	URBANO	9,9%	28.717	3,6%	0,4%
PAIS	GRANDES CENTROS	9,9%	23.573	4,1%	0,4%
PAIS	RURAL	5,6%	7.002	8,5%	0,5%
1	I REGION DE TARAPACÁ	7,6%	1.026	18,1%	1,4%
1	IQUIQUE CD	6,9%	564	26,4%	1,8%
1	ALTO HOSPICIO CD	9,3%	222	31,1%	2,9%
2	II REGION DE ANTOFAGASTA	6,8%	1.050	20,5%	1,4%
2	CALAMA CD	6,9%	384	29,1%	2,0%
2	ANTOFAGASTA CD	6,6%	504	33,5%	2,2%
3	III REGION DE ATACAMA	8,6%	999	15,9%	1,4%
3	COPIAPO CD	8,6%	438	25,1%	2,2%
3	VALLENAR CD	9,5%	180	30,5%	2,9%
4	IV REGION DE COQUIMBO	6,8%	2.193	13,3%	0,9%
4	URBANO	8,0%	1.233	15,8%	1,3%
4	COQUIMBO CD	8,2%	279	31,1%	2,5%
4	LA SERENA CD	8,5%	306	31,8%	2,7%
4	OVALLE CD	4,8%	297	36,3%	1,7%
4	PR ELQUI	5,8%	1.071	19,9%	1,2%
4	PR LIMARI	3,7%	726	27,5%	1,0%
4	PR CHOAPA	15,2%	396	17,2%	2,6%
4	RURAL	5,3%	960	23,1%	1,2%
5	REGION DE VALPARAÍSO	11,3%	5.055	7,5%	0,8%
5	URBANO	11,6%	4.575	7,9%	0,9%
5	RURAL	7,7%	480	24,7%	1,9%
5	PR PETORCA	12,4%	216	24,0%	3,0%
5	PR LOS ANDES	5,6%	348	33,8%	1,9%
5	LOS ANDES CD	13,7%	462	21,9%	3,0%
5	PR SAN FELIPE	8,7%	396	24,9%	2,2%
5	PR QUILLOTA	16,0%	531	21,6%	3,4%
5	PR VALPARAISO	9,5%	1.209	18,0%	1,7%
5	PR MARGA MARGA	11,3%	879	17,6%	2,0%

REGION	NIVEL DE ESTIMACIÓN	TASA DESOCUPACIÓN PROMEDIO 5 TRIMESTRES	TAMAÑO TEÓRICO CALCULADO	ERROR RELATIVO	ERROR ABSOLUTO
5	VALPARAISO CD	9,7%	639	24,2%	2,3%
5	VIÑA DEL MAR CD	8,4%	648	26,2%	2,2%
5	QUILPUE CD	10,0%	384	30,1%	3,0%
5	VILLA ALEMANA CD	12,6%	399	23,8%	3,0%
5	PR SAN ANTONIO	13,1%	600	20,3%	2,7%
5	GRAN VALPARAISO	10,3%	1.941	13,5%	1,4%
5	SAN ANTONIO CD	14,0%	537	21,4%	3,0%
5	QUILLOTA	16,4%	702	18,3%	3,0%
6	VI REGION DEL LIBERTADOR GENERAL BERNARDO O'HIGGINS	7,2%	2.259	12,6%	0,9%
6	URBANO	7,7%	1.389	16,1%	1,2%
6	RURAL	6,3%	870	20,0%	1,2%
6	RANCAGUA CD	8,1%	609	27,1%	2,2%
6	SAN FERNANDO CD	7,3%	288	40,0%	2,9%
6	PR CACHAPOAL	7,6%	1.095	18,0%	1,4%
6	PR COLCHAGUA	6,4%	681	23,0%	1,5%
6	PR CARDENAL CARO	7,1%	483	26,8%	1,9%
7	VII REGION DEL MAULE	9,6%	2.456	10,5%	1,0%
7	URBANO	11,3%	1.616	11,9%	1,4%
7	RURAL	6,3%	840	20,9%	1,3%
7	CURICO CD	8,4%	240	35,7%	3,0%
7	TALCA CD	14,2%	408	21,1%	3,0%
7	LINARES CD	12,8%	375	23,5%	3,0%
7	PR TALCA	14,0%	663	17,5%	2,5%
7	PR LINARES	10,6%	765	17,3%	1,8%
7	PR CAUQUENES	6,9%	344	30,4%	2,1%
7	PR CURICO	5,7%	684	23,7%	1,3%
8	VIII REGION DEL BÍO BÍO	12,7%	4.920	7,5%	1,0%
8	URBANO	13,5%	4.365	7,8%	1,0%
8	RURAL	6,6%	555	27,7%	1,8%
8	CHILLAN CD	14,5%	441	20,6%	3,0%
8	LOTA CD	17,6%	522	17,0%	3,0%
8	CORONEL CD	19,8%	567	22,7%	4,5%
8	LOS ANGELES CD	8,5%	369	30,5%	2,6%
8	PR ÑUBLE	13,9%	618	18,2%	2,5%
8	PR CONCEPCION	13,2%	2.458	11,3%	1,5%
8	CONCEPCION CD	8,4%	1.045	20,9%	1,8%
8	TALCAHUANO CD	11,3%	393	23,6%	2,7%

REGION	NIVEL DE ESTIMACIÓN	TASA DESOCUPACIÓN PROMEDIO 5 TRIMESTRES	TAMAÑO TEÓRICO CALCULADO	ERROR RELATIVO	ERROR ABSOLUTO
8	PR ARAUCO	12,3%	495	22,0%	2,7%
8	PR BIO-BIO	11,1%	996	16,7%	1,9%
8	GRAN CONCEPCION	8,2%	1.120	20,1%	1,7%
9	IX REGION DE LA ARAUCANÍA	9,7%	2.040	12,4%	1,2%
9	URBANO	11,2%	1.365	13,5%	1,5%
9	RURAL	6,8%	675	26,6%	1,8%
9	ANGOL CD	10,6%	423	26,1%	2,8%
9	TEMUCO CD	10,2%	606	19,8%	2,0%
9	PR MALLECO	10,2%	954	17,9%	1,8%
9	PR CAUTIN	9,3%	981	18,6%	1,7%
10	X REGION DE LOS LAGOS	4,2%	2.349	18,8%	0,8%
10	URBANO	4,9%	1.239	22,6%	1,1%
10	RURAL	3,5%	1.110	31,2%	1,1%
10	OSORNO CD	6,5%	519	34,8%	2,3%
10	PUERTO MONTT CD	3,8%	444	39,3%	1,5%
10	PR OSORNO	5,3%	1.020	27,3%	1,4%
10	PR LLANQUIHUE	4,0%	846	28,1%	1,1%
10	PR CHILOE-PALENA	2,4%	483	52,5%	1,3%
11	XI REGION DE AYSÉN DEL GENERAL CARLOS IBAÑEZ DEL CAMPO	5,5%	828	24,7%	1,3%
11	COYHAIQUE CD	5,8%	411	36,4%	2,1%
11	PUERTO AISEN CD	8,8%	183	34,0%	3,0%
12	XII REGION DE MAGALLANES Y LA ANTÁRTICA CHILENA	3,7%	321	44,7%	1,7%
12	PUNTA ARENAS CD	3,3%	207	58,0%	1,9%
13	REGION METROPOLITANA	9,2%	7.991	7,0%	0,6%
13	URBANO	9,3%	7.763	7,1%	0,7%
13	RURAL	5,6%	375	28,0%	1,6%
13	PROVINCIA SANTIAGO	8,4%	5.054	9,76%	0,82%
13	GRAN SANTIAGO	9,0%	6.386	8,3%	0,7%
13	PUENTE ALTO CD	9,4%	582	23,4%	2,2%
13	SAN BERNARDO CD	13,8%	603	21,8%	3,0%
13	MELIPILLA CD	12,6%	333	23,9%	3,0%
13	COLINA CD	10,6%	321	27,0%	2,9%
13	PROV CHACABUCO	8,5%	213	35,0%	3,0%
13	PR CORDILLERA	9,4%	657	22,1%	2,1%
13	PR MAIPO	12,4%	744	19,2%	2,4%
13	PR MELIPILLA	11,1%	465	20,3%	2,2%
13	PR TALAGANTE	9,2%	537	20,2%	1,9%
14	XIV REGION DE LOS RÍOS	8,2%	1.092	17,7%	1,5%
14	URBANO	9,7%	705	20,4%	2,0%

REGION	NIVEL DE ESTIMACIÓN	TASA DESOCUPACIÓN PROMEDIO 5 TRIMESTRES	TAMAÑO TEÓRICO CALCULADO	ERROR RELATIVO	ERROR ABSOLUTO
14	RURAL	5,5%	387	33,8%	1,9%
14	VALDIVIA CD	9,8%	501	24,5%	2,4%
14	PR VALDIVIA	8,2%	1.092	17,7%	1,5%
15	XV REGION DE ARICA Y PARINACOTA	8,1%	1.140	18,0%	1,5%
15	ARICA CD	9,4%	825	19,8%	1,9%

Anexo 2: Programación para cálculo de Error Estándar y Coeficientes de Variación

Cálculo usando STATA³²

***Generamos las variables de interes.**

```
qui gen fdt = (cae_especifico>=1 & cae_especifico <=9)
qui gen ocupado = (cae_especifico>=1 & cae_especifico<=7)
qui gen desocupado = (cae_especifico>=8 & cae_especifico<=9)
```

***Definimos el diseño muestral (plan de muestreo).**

```
qui svyset [pweight=fact], strata(estrato) psu(id_directorio) singleunit(certainty)
```

***Tasa de desocupación Nacional y por Región.**

```
qui svy: ratio desocupado fdt
estat cv
mat td_n = (r(b), r(se), r(cv))
qui svy: ratio desocupado fdt, over(region)
estat cv
qui mat td_r = (r(b)', r(se)', r(cv)')
qui mat td = (td_n\td_r )
mat list td
```

***Número de Ocupados a nivel Nacional y por Región.**

```
qui svy: total ocupado
estat cv
qui mat o_n = (r(b), r(se), r(cv))
qui svy: total ocupado, over(region)
estat cv
qui mat o_r = (r(b)', r(se)', r(cv)')
qui mat o = (o_n\o_r)
mat list o
```

***Ocupados por rama de actividad.**

```
qui replace b14_rev4cl_caenes = 0 if(b14_rev4cl_caenes == .)
qui svy: total ocupado
estat cv
qui mat o_rama_n = (r(b), r(se), r(cv))
qui svy: total ocupado, over( b14_rev4cl_caenes )
estat cv
qui mat est = r(b)'
qui mat se = r(se)'
qui mat cv = r(cv)'
qui mat o_rama_r = (est[2...,1], se[2...,1], cv[2...,1])
qui mat o_rama = (o_rama _ n\ o_rama _ r)
mat list o_rama
```

³² La estimación puntual debe ser multiplicada por 100 para que se exprese como porcentaje. Por otro lado, el coeficiente de variación, bajo esta programación, ya viene expresado como porcentaje.

Cálculo usando SPSS³³

*generamos las variables de interés.
compute fdt = (range(cae_especifico,1,9)).
compute ocupado = (range(cae_especifico,1,7)).
compute desocupado = (range(cae_especifico,8,9)).
execute.

*Definimos el diseño muestral (plan de muestreo).
csplan analysis
/plan file='c:\ENE\bases_spss\plan_ene.csaplan'
/planvars analysisweight=fact
/srsestimator type=wr
/print plan
/design strata=estrato cluster=id_directorio
/estimator type=wr.

*Tasa de desocupación Nacional y por Región.
csdescriptives
/plan file='c:\ENE\bases_spss\plan_ene.csaplan'
/ratio numerator=desocupado denominator=fdt
/statistics se cv
/subpop table=region display=layered
/missing scope=analysis classmissing=exclude.

*Número de Ocupados a nivel Nacional y por Región.
csdescriptives
/plan file='c:\ENE\bases_spss\plan_ene.csaplan'
/summary variables=ocupado
/subpop table=region display=layered
/sum
/statistics se cv
/missing scope=analysis classmissing=exclude.

*Ocupados por rama de actividad.
recode b14_rev4cl_caenes (sysmis=0).
execute.
csdescriptives
/plan file='c:\ENE\bases_spss\plan_ene.csaplan'
/summary variables=ocupado
/subpop table= b14_rev4cl_caenes display=layered
/sum
/statistics se cv
/missing scope=analysis classmissing=exclude.

³³ La estimación del coeficiente de variación en SPSS debe ser multiplicada por 100 para que se exprese como porcentaje.

Cálculo usando SAS³⁴

/* Creación de las variables */

```
data Base_ENE;  
set Base_ENE;
```

```
if CAE_ESPECIFICO >= 1 & CAE_ESPECIFICO <= 9 then FFTT = 1;  
if FFTT = . then FFTT = 0;
```

```
if CAE_ESPECIFICO >= 1 & CAE_ESPECIFICO <= 7 then Ocup = 1;  
if Ocup = . then Ocup = 0;
```

```
if CAE_ESPECIFICO >= 8 & CAE_ESPECIFICO <= 9 then Deso = 1;  
if Deso = . then Deso = 0;
```

```
run;
```

/* Tasa de desocupación Nacional y por Región */

```
ods select ratio;  
proc surveymeans data=Base_ENE alpha=0.05 ncluster nobis ratio stderr;  
cluster ID_DIRECTORIO;  
strata estrato;  
weight FACT;  
var Deso FFTT;  
ratio Deso / FFTT;  
ods output ratio=Ratio_Total;  
run;
```

```
proc sort data=Base_ENE;  
by region;  
run;
```

```
ods select ratio;  
proc surveymeans data=Base_ENE alpha=0.05 ncluster nobis ratio stderr;  
cluster ID_DIRECTORIO;  
strata estrato;  
weight FACT;  
var Deso FFTT;  
ratio Deso / FFTT;  
by region;  
ods output ratio=Ratio_By;  
run;
```

/* Número de Ocupados a nivel Nacional y por Región */

³⁴ La estimación puntual debe ser multiplicada por 100 para que se exprese como porcentaje. Por otro lado, el coeficiente de variación, bajo esta programación, ya viene expresado como porcentaje.

```
ods select Statistics;
proc surveymeans data=Base_ENE alpha=0.05 ncluster nobsum sum std cvsum;
cluster ID_DIRECTORIO;
strata estrato;
weight FACT;
var Ocup;
ods output Statistics=Sum_Total;
run;
```

```
proc sort data=Base_ENE;
by region;
run;
```

```
ods select Statistics;
proc surveymeans data=Base_ENE alpha=0.05 ncluster nobsum sum std cvsum;
cluster ID_DIRECTORIO;
strata estrato;
weight FACT;
var Ocup;
by region;
ods output Statistics=Sum_By;
run;
```

/* Ocupados por rama de actividad */

```
proc format;
value rev4cl
. = 'No Ocupado'
1 = 'AGRICULTURA, GANADERÍA, SILVICULTURA Y PESCA'
2 = 'MINERÍA'
3 = 'INDUSTRIAS MANUFACTURERAS'
4 = 'SUMINISTRO DE ELECTRICIDAD (*)'
5 = 'SUMINISTRO DE AGUA (*)'
6 = 'CONSTRUCCIÓN'
7 = 'COMERCIO'
8 = 'TRANSPORTE Y ALMACENAMIENTO'
9 = 'ACTIVIDADES DE ALOJAMIENTO Y DE SERVICIO DE COMIDAS'
10 = 'INFORMACIÓN Y COMUNICACIONES'
11 = 'ACTIVIDADES FINANCIERAS Y DE SEGUROS'
12 = 'ACTIVIDADES INMOBILIARIAS (*)'
13 = 'ACTIVIDADES PROFESIONALES, CIENTÍFICAS Y TÉCNICAS'
14 = 'ACTIVIDADES DE SERVICIOS ADMINISTRATIVOS Y DE APOYO'
15 = 'ADMINISTRACIÓN PÚBLICA'
16 = 'ENSEÑANZA'
17 = 'ACTIVIDADES DE ATENCIÓN DE LA SALUD HUMANA'
18 = 'ACTIVIDADES ARTÍSTICAS Y ENTRETENIMIENTO (*)'
19 = 'OTRAS ACTIVIDADES DE SERVICIOS'
20 = 'ACTIVIDADES DE LOS HOGARES COMO EMPLEADORES'
21 = 'ACTIVIDADES DE ÓRGANOS EXTRATERRITORIALES (*)';
```

```
value total_rev4cl  
. = 'No Ocupado'  
1 - 21 = 'TOTAL OCUPADOS';  
run;
```

```
proc surveyfreq data=Base_ENE missing;  
cluster ID_DIRECTORIO;  
strata estrato;  
weight FACT;  
tables b14_rev4cl_caenes / alpha=0.05 cvwt varwt nototal;  
format b14_rev4cl_caenes total_rev4cl.;  
ods output OneWay=Total;  
run;
```

```
proc surveyfreq data=Base_ENE missing;  
cluster ID_DIRECTORIO;  
strata estrato;  
weight FACT;  
tables b14_rev4cl_caenes / alpha=0.05 cvwt varwt nototal;  
format b14_rev4cl_caenes rev4cl.;  
ods output OneWay=Frecuencias;  
run;
```

15 Bibliografia

- Cochran W. G. (1977), “**Sampling Techniques**”, John Wiley & Sons, New York.
- Sarndal Carl-Erik, Bengt Swensson, Jan Wretman (1992), “**Model Assisted Survey Sampling**”.
- Des Raj (1968), “**Sampling Theory**”.
- Skinner, C. J., D. Holt y T. M. Smith (1989), “**Analysis of Complex Surveys**”.
- Franca, M. y M. P. Soares (2007), “**Estimação de intervalos de confiança para estimadores de diferenças temporais na Pesquisa Mensal de Emprego**”. Textos para discussão. Diretoria de Pesquisas número 22, IBGE.
- Sharon L. Lohr (2010), “**Sampling: Design and Analysis**”. 2da Edición.
- West, B., P. Berglund y S. Heeringa (2008), “**A closer examination of subpopulation analysis of complex-sample survey data**”. The Stata Journal: 8, Number 4, pp. 520 - 531.