

ASPECTOS METODOLÓGICOS DEL PROGRAMA INTEGRADO DE ENCUESTAS A LOS HOGARES (PIDEH) DEL INSTITUTO NACIONAL DE ESTADÍSTICAS

Enrique Silva*

EXTRACTO

En este trabajo se entrega una visión panorámica de los principales aspectos metodológicos asociados a la nueva muestra orientada a obtener estadísticas laborales, que el INE puso en operación en 1985, en el marco de un proyecto denominado Programa Integrado de Encuestas de Hogares (PIDEH). Particularmente, se profundiza en los fundamentos estadísticos asociados al diseño de la referida muestra: la estructura de los estimadores y una evaluación estadística de la calidad de la información obtenida; y la organización del trabajo en terreno y otros aspectos operativos.

ABSTRACT

A broad view is presented in this paper of the major methodological aspects associated with a new sample, being used by the National Institute of Statistics (INE), since 1985. The sample's purpose is to obtain labor statistics, under a project named Integrated Program of Household Surveys (PIDEH). Particular attention is given to statistical issues associated to the sample's design: the structure of the estimators, a statistical appraisal of the information's quality, and the organization of the field work as well as other operational aspects.

*Economista y Master en Métodos Estadísticos, Instituto Nacional de Estadísticas (INE). El autor agradece la colaboración de las Divisiones de Estudios y Coordinación, de Estadísticas Sociales y Demográficas Continuas y de Informática del INE, y, muy especialmente, a Víctor Correa y Teresa Varela.

Como es usual, la responsabilidad de esta versión sólo compete al autor.

ASPECTOS METODOLOGICOS DEL PROGRAMA INTEGRADO DE ENCUESTAS A LOS HOGARES (PIDEH) DEL INSTITUTO NACIONAL DE ESTADISTICAS*

Enrique Silva

1. INTRODUCCION

La Encuesta Nacional del Empleo (ENE) del Instituto Nacional de Estadísticas (INE), en su formato actual y salvo algunas modificaciones menores,¹ se ha llevado a cabo en el Gran Santiago, desde 1973; y, a nivel nacional, en forma regular desde 1976.

Entre los años 1976 y 1985, la encuesta se aplicó a una muestra denominada Muestra de Hogares Regionalizada (M.H.R.), cuyo marco se basaba, entre otros antecedentes, en el XIV Censo Nacional de Población y el III Censo de Vivienda levantado en 1970. La M.H.R. permitía entregar estimaciones a los niveles nacional, regional y provincial —área urbana y rural— y del Gran Santiago. Adicionalmente, dentro de cada región o provincia, se distinguían tres áreas de estimación: grandes ciudades de inclusión forzosa, con 40 mil o más habitantes al Censo de 1970; resto del área urbana, entre 2 mil y 40 mil habitantes; y área rural.

Movido por la creciente necesidad de actualizar el diseño muestral, de tal forma de incorporar los cambios ocurridos en las características socio-demográficas y en la distribución espacial de la población, así como las variaciones de la División Político-Administrativa del país, durante los años 1984 y 1985 el INE desarrolló una nueva muestra denominada Programa Integrado de Encuestas a Hogares (PIDEH). Dicha muestra se puso en ope-

*Suplemento, *Estudios de Economía*, publicación del Departamento de Economía de la Facultad de Ciencias Económicas y Administrativas, vol. 15, n° 1, abril de 1988.

¹Una de éstas es que desde el trimestre octubre-diciembre de 1982, el cuestionario se aplica a todos los miembros del hogar de 15 años y más. Anteriormente se encuestaba a la población de 12 años y más.

ración, en el segundo semestre de 1985, y entró definitivamente en vigencia a partir de 1986.²

Un aspecto digno de destacar es que aun cuando el origen y aplicación del PIDEH hasta la fecha se ha ligado principalmente a la obtención de estadísticas laborales, no es menos cierto que en su desarrollo se consideró explícitamente el objetivo de construir una infraestructura muestral que tuviera propósitos más amplios. Esto se ve corroborado, por ejemplo, en el hecho de que el diseño muestral de la IV Encuesta de Presupuestos Familiares, cuyo levantamiento se inició en noviembre de 1987, se basa en dicha infraestructura.

El objetivo de este trabajo es entregar una visión panorámica de los principales aspectos metodológicos del PIDEH. Particularmente, y por tratarse de temas menos conocidos, se intenta profundizar los fundamentos estadísticos asociados al diseño de la muestra y la estructura de los estimadores y, además, a la organización del trabajo en terreno y otros aspectos operativos.

2. PIDEH: DESCRIPCION DEL DISEÑO MUESTRAL

2.1. Variables laborales investigadas

Como es sabido, la ENE se orienta, principalmente, a obtener información sobre la evolución de la fuerza de trabajo, los ocupados y los desocupados, a nivel nacional y para los distintos niveles de estimación definidos.

La fuerza de trabajo es toda la población de 15 o más años que se encuentra ocupada o desocupada. Por otra parte, se considera como ocupados a todas las personas de 15 años o más que durante la semana anterior a la de la toma de la encuesta habían trabajado una hora o más como empleados u obreros por remuneración o ganancia³ o como trabajadores por cuenta propia;⁴ a aquellas personas que habían trabajado 15 horas o más como familiares no remunerados; y aquellos que, teniendo empleo, no habían trabajado en la referida semana, por encontrarse temporalmente ausentes de su trabajo.⁵ Por último, los desocupados resultan de considerar a los cesantes, es

²La primera estimación y publicación oficial, a nivel de todo el país, corresponde al período noviembre de 1985-enero de 1986.

³Empleadores, empleados sector público, empleados sector privado, obreros sector público, obreros sector privado, personal de servicios puertas adentro, personal de servicios puertas afuera, PEM, POJH y otros.

⁴Independientes.

⁵Problemas de salud, problemas de trabajo, vacaciones, problemas personales, mal tiempo, conflictos laborales u otros.

decir, a todas aquellas personas que, habiendo trabajado antes en un empleo regular y teniendo deseos de trabajar, habían hecho esfuerzos definidos para conseguir empleo durante los dos meses precedentes a la fecha de la entrevista, sin tener éxito; y a los que buscan trabajo por primera vez, es decir, a las personas que habían hecho los mismos esfuerzos del grupo anterior, pero que nunca habían trabajado antes.

Además de la conceptualización básica, antes reseñada, la encuesta contiene preguntas que permiten clasificar a los ocupados y desocupados de acuerdo a la rama de actividad económica a que pertenecen,⁶ el grupo ocupacional⁷ y la categoría de la ocupación.⁸

En el caso de los desocupados, se incluyen también preguntas que permiten identificar la actividad realizada para conseguir empleo, el número de semanas que han buscado trabajo, el tipo de empleo buscado y, específicamente, para el caso de los cesantes, el período en que trabajaron por última vez y el motivo del abandono del empleo anterior.

En cuanto a los ocupados, se investiga el empleo en más de una ocupación, el tiempo efectivo y la jornada normal de trabajo, tanto en la ocupación principal como en la otra ocupación, y si el tiempo total trabajado en una semana es inferior a 35 horas, se consulta por la razón de lo anterior y por el deseo de trabajar más horas.

Por último, la encuesta identifica para cada encuestado las variables de ser jefe o no jefe de hogar, relación de parentesco con éste, sexo, edad, nivel educacional y estado civil. Lo anterior permite construir relaciones y cruces entre variables personales y variables de tipo laboral.

2.2. Niveles de estimación

En cuanto a los niveles de estimación cubiertos por la muestra, puede indicarse que en el PIDEH se adopta un esquema relativamente similar al definido en la M.H.R., aunque obviamente adaptado, como se señaló en la introducción, a los cambios en la distribución espacial de la población y a la nueva División Político-Administrativa del país.

⁶Agricultura y pesca; minas y canteras; industrias manufactureras; electricidad, gas y agua; construcción; comercio; transporte, almacenaje y comunicaciones; servicios financieros; servicios comunales, sociales y personales; actividades no bien especificadas.

⁷Profesionales, técnicos y afines; gerentes administrativos y directores; empleados de oficina y afines; vendedores y afines; agricultores, ganaderos, pescadores, cazadores y afines; conductores y afines; artesanos y operarios; otros artesanos y operarios; obreros y jornaleros; trabajadores en servicios personales; otros trabajadores.

⁸Empleadores, cuenta propia, empleados, obreros, personal de servicio, familiares no remunerados, otros.

Específicamente, y como regla general, las estimaciones nacionales se realizan componiendo cálculos independientes, para los niveles regional y provincial, distinguiendo al interior de estos últimos las denominadas ciudades autorrepresentadas o de inclusión forzosa —categoría IF— constituida por ciudades o grupos de ciudades contiguas con 40 mil o más habitantes al Censo de 1982;⁹ el resto del área urbana —nivel RAU— formado a partir de todos los centros urbanos que al Censo de 1982 contaban con menos de 40 mil habitantes; y el área rural —categoría R— constituida por el conjunto de entidades de población clasificadas como tales en el mismo Censo. Cabe anotar que en aquellas provincias en las cuales no existen niveles IF, el área urbana completa se agrupa en la categoría U (urbana).

La clasificación anterior contiene algunos casos especiales, como las ciudades de Chuquicamata y Coihaique, que se clasifican como IF, a pesar de no cumplir con el mínimo de habitantes especificado para esta categoría; los niveles RAU de las Regiones I, XI y XII, que corresponden al total regional; la agrupación de las provincias de El Loa y Antofagasta en un solo nivel RAU, en la Región II; y los niveles de R en las regiones I, II, III, XI y XII, que también corresponden al total regional.

En síntesis, la muestra PIDEH contempla un total de 103 niveles de estimación, en comparación con los 79 niveles de la M.H.R. El detalle comparado entre ambas muestras, así como un intento de homologación de sus respectivos niveles de estimación, se ilustra en la tabla 1.

2.3. Población objetivo

La población objetivo está constituida por toda la población del país residente en viviendas particulares. 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.

Por su parte, la población objetivo en cuanto a ámbito geográfico es todo el territorio continental de Chile, con exclusión de las áreas de difícil acceso,¹⁰ determinadas en una fase previa a la elaboración del marco muestral.

⁹XV Censo de Población y IV de Vivienda de abril de 1982.

¹⁰Las áreas de difícil acceso son aquellas que por razones de orden climático, topográfico o por ausencia de vías y medios de comunicación expeditos permanecen aisladas parcial o totalmente en el año.

TABLA 1

HOMOLOGACION DE NIVELES DE ESTIMACION M.H.R. Y PIDEH

PIDEH				M.H.R.					
Región	Región		Estrato		Región		Estrato		
	Urb.	Rur.	Prov.	IF	RAU	U	R	U	R
I. TARAPACA	X				X.	X.	X.	X.	X.
Arica				X.					
Iquique				X.				X.	X.
II. ANTOFAGASTA	X				X.	X.	X.	X.	X.
Tocopilla			X			X.			
Calama				X.					
Chuquicamata				X.					X.
Antofagasta				X.					
El Loa/Antofagasta					X.				
III. ATACAMA	X				X.	X.	X.	X.	X.
Chañaral			X						
Copiapó (ciudad)				X.					
Copiapó			X		X.				
Huasco			X		X.				
IV. COQUIMBO	X	X			X.	X.	X.	X.	X.
Prov. Elqui	X		X		X.				
- Coquimbo/La Serena				X.					X.
Prov. Limarí	X		X		X.				
- Ovalle				X.					
Prov. Choapa						X.	X.	X.	X.

Continuación tabla 1

		PIDEH					M.H.R.									
Región	Región	Región			Estrato		Región	Región			Estrato					
		Urb.	Rur.	Prov.	IF	RAU		U	R	Urb.	Rur.	Prov.	IF	RAU	U	R
V. VALPARAISO	X	X					V. VALPARAISO	X								
Prov. Petorca					X.		Prov. Petorca						X.			X.
Prov. Los Andes					X.		Prov. Los Andes						X.			X.
Prov. San Felipe					X.		Prov. San Felipe						X.			X.
Prov. Quillota	X					X.	Prov. Quillota						X.			X.
- Quillota/La Calera/ La Cruz																
Prov. Valparaíso	X					X.	Prov. Valparaíso	X					X.			X.
- Gran Valparaíso					X.		- Gran Valparaíso									
Prov. San Antonio	X					X.	Prov. San Antonio	X					X.			X.
- San Antonio (ciudad)						X.	- San Antonio (ciudad)						X.			X.
VI. DEL LIBERTADOR GRAL.							VI. DEL LIBERTADOR GRAL.									
B. O'HIGGINS	X						B. O'HIGGINS	X								
Prov. Cachapoal	X					X.	Prov. O'Higgins	X					X.			X.
- Rancagua						X.	- Rancagua									
Prov. Colchagua							Prov. Colchagua									
Prov. Cardenal Caro																
VII. DEL MAULE	X						VII. DEL MAULE	X								
Prov. Curicó	X					X.	Prov. Curicó	X					X.			X.
- Curicó (ciudad)						X.	- Curicó (ciudad)									
Prov. Talca	X					X.	Prov. Talca	X					X.			X.
- Talca (ciudad)						X.	- Talca (ciudad)									
Prov. Linares	X					X.	Prov. Linares						X.			X.
- Linares (ciudad)						X.	Prov. Linares									
Prov. Cauquenes							Prov. Cauquenes									X.

PIDEH					M.H.R.									
Región	Región		Estrato		Región	Región		Estrato						
	Urb.	Rur.	Prov.	IF		RAU	U	R	Urb.	Rur.	Prov.	IF	RAU	U
VIII. DEL BIOBIO	X	X			VIII. DEL BIOBIO	X	X							
Prov. Ñuble	X			X.	Prov. Ñuble	X			X.			X.		X.
- Chillán				X.	- Chillán				X.					X.
Prov. Concepción	X			X.	Prov. Concepción	X			X.			X.		X.
- Gran Concepción				X.	- Gran Concepción				X.					X.
- Lota/Coronel				X.	- Lota/Coronel				X.					X.
Prov. Arauco				X.	Prov. Arauco				X.					X.
Prov. Biobío	X			X.	Prov. Biobío	X			X.			X.		X.
- Los Angeles				X.	- Los Angeles				X.					X.
IX. DE LA ARAUCANIA	X	X			XI. DE LA ARAUCANIA	X	X							
Prov. Malleco				X.	Prov. Malleco				X.					X.
Prov. Cautín	X			X.	Prov. Cautín	X			X.			X.		X.
- Temuco				X.	- Temuco				X.					X.
X. DE LOS LAGOS	X	X			X. DE LOS LAGOS	X	X							
Prov. Valdivia	X			X.	Prov. Valdivia	X			X.			X.		X.
- Valdivia				X.	- Valdivia				X.					X.
Prov. Osorno	X			X.	Prov. Osorno	X			X.			X.		X.
- Osorno				X.	- Osorno				X.					X.
Prov. Llanquihue	X			X.	Prov. Llanquihue	X			X.			X.		X.
- Puerto Montt				X.	- Puerto Montt				X.					X.
Prov. Chiloé, Palena				X.	Prov. Chiloé				X.					X.
XI. AISEN DEL GENERAL					XI. AISEN DEL GENERAL									
C. IBÁÑEZ DEL C.	X			X.	C. IBÁÑEZ DEL C.	X			X.			X.		X.
Coihaique (ciudad)				X.	Coihaique (ciudad)				X.					X.

Continuación tabla 1

PIDEH		M.H.R.					
Región	Región		Estrato				
	Urb.	Rur.	Prov.	IF	RAU	U	R
XII. DE MAGALLANES Y LA ANTARTICA CHILENA							
Punta Arenas (ciudad)					X.		
REGION METROPOLITANA	X	X					
Prov. Santiago			X.				
Prov. Chacabuco							X.
Prov. Cordillera	X						
- Puente Alto					X.		
Prov. Maipo	X						
- San Bernardo					X.		
Prov. Melipilla							X.
Prov. Talagante							X.
Gran Santiago							X.
XII. DE MAGALLANES Y LA ANTARTICA CHILENA							
Punta Arenas (ciudad)							X.
REGION METROPOLITANA	X						
- Gran Santiago							X.

2.4. Marco muestral

Una característica distintiva del diseño es que contempla un marco muestral de áreas que se basa en el conjunto de mapas regionales, provinciales, comunales y distritales del INE que abarca a todo el territorio nacional. Dichos mapas contienen una delimitación de las áreas urbanas y rurales, y una demarcación de las áreas de difícil acceso.

Adicionalmente, se utilizó información sobre población y viviendas del Censo de 1982. Cabe señalar que las localidades urbanas fueron objeto de un proceso de actualización postcensal, el que se realizó en los primeros meses de 1985.

Sobre la base de todos estos antecedentes, se procedió a la partición del territorio de los distintos niveles de estimación en áreas geográficas denominadas secciones. Tales secciones se conformaron a partir de una determinada magnitud de población y un determinado número de viviendas, y son de un tamaño variable entre los distintos niveles de estimación. Este tema se trata con más detalle en un acápite posterior.

2.5. Estratificación

El diseño muestral contempla un proceso de estratificación, el que se define de acuerdo a criterios fundamentalmente geográficos. Estos criterios, a su vez, se relacionan estrechamente con los niveles de estimación considerados.

En efecto, las regiones y/o provincias se consideran como áreas independientes de estimación a partir de las cuales se generan los valores de las diferentes variables a nivel nacional.

Por su parte, a nivel de cada región y/o provincia, se definen los estratos IF, RAU, U y R, ya descritos antes, los cuales permiten componer los respectivos totales ya sea provinciales o regionales, o urbanos y rurales.

Por último, cabe mencionar que, en el caso de la Provincia de Santiago, los estratos IF corresponden a las comunas de dicha provincia.¹¹

¹¹ Santiago; Independencia; Recoleta; Estación Central; Conchalí; Huechuraba; Quilicura; Renca; Quinta Normal; Pudahuel; Cerro Navia; Lo Prado; Maipú; Cerrillos; La Cisterna; Lo Espejo; El Bosque; La Granja; La Pintana; San Ramón; San Miguel; Pedro Aguirre Cerda; San Joaquín; La Florida; Ñuñoa; Macul; Peñalolén; La Reina; Providencia; Las Condes; Vitacura y Lo Barnechea.

2.6. Configuración de las secciones

Un aspecto clave del diseño, por basarse en un marco de áreas, es el proceso de definición, conformación y delimitación física de las secciones, en los distintos estratos y niveles de estimación.

La división de los estratos en secciones se efectuó en planos actualizados correspondientes a los mismos. En el caso urbano, tales planos contenían la siguiente información: crecimiento postcensal, decrecimiento postcensal, áreas de probable crecimiento futuro, áreas de probable decrecimiento futuro, límites urbano-rural, áreas de exclusión, numeración de las manzanas —incluyendo a aquellas que corresponden a crecimiento postcensal— número de viviendas por manzana, población por manzana y categoría de la manzana. Este último aspecto se profundiza más adelante. Por su parte, en las áreas rurales, la información de los planos daba cuenta de los límites urbano-rural, áreas de difícil acceso, áreas de exclusión, demarcación de las localidades y entidades rurales, y número de viviendas y población por entidad.

El tamaño de las secciones, variable que, como se sabe influye, tanto en los resultados de las estimaciones como en sus correspondientes varianzas, se definió en términos de población y viviendas. La superficie territorial no constituyó un criterio considerado relevante para tal definición.

Como regla general, tanto para los estratos urbanos como rurales, se trató de obtener secciones con un determinado tamaño medio, en términos de número de viviendas. Dicho tamaño fue directamente proporcional al tamaño de los estratos y al tamaño de las localidades que pertenecen a los mismos, e inversamente proporcional a la dispersión geográfica de tales localidades dentro del estrato.

Adicionalmente, y teniendo presente que la densidad de población por vivienda puede ser diferente en los distintos estratos, provincias y regiones, el criterio anterior se complementó incorporando como variable decisional, en cuanto a tamaño de las secciones, el número de personas por viviendas ocupadas. En general, el tamaño de la sección también se define como directamente proporcional a esta variable.

En la tabla 2 se presenta una relación del tamaño de las secciones —en términos de viviendas y población— en los diferentes estratos, por región y provincia. Se aprecia que el tamaño de los estratos IF se acerca, en general, a las 300 viviendas; resultando este tamaño superior a la magnitud promedio de los restantes estratos RAU, U, R. Este hecho, así como la variabilidad que se observa en el tamaño de los estratos entre las distintas provincias y re-

giones, no hacen más que corroborar los criterios sobre la base de los cuales se definió tal variable.

Otros aspectos fundamentales que se tuvieron presente en la construcción de las secciones fueron su estabilidad a través del tiempo, la facilidad para su identificación en terreno y su homogeneidad y compacidad interna. Esta última consideración se justifica por la necesidad de mantener una relativa estabilidad en las estimaciones al efectuar el proceso de rotación de la muestra.

En el caso de las secciones urbanas, sus límites fueron generalmente calles y respetaron los límites de las comunas. En las secciones rurales, la delimitación se realizó a partir de elementos naturales: ríos, esteros, quebradas, etc., elementos indirectos: caminos, canales, líneas de alta tensión, líneas férreas, etc., o, en ausencia de los anteriores, sobre la base de líneas imaginarias. Al igual que en el caso urbano, dichos límites respetaron los límites comunales.

En cuanto a la homogeneidad y compacidad de las secciones, puede señalarse que, en los niveles urbanos, las secciones debieron incluir manzanas de sólo una de tres categorías: categoría A, de nivel socioeconómico alto; categoría B, de nivel socioeconómico medio o de poblaciones nuevas para empleados y obreros; categoría C, de nivel socioeconómico bajo o de poblaciones marginales.¹² Estas categorías se definieron de acuerdo a las características de la construcción de las viviendas, información que, a su vez, se obtuvo en el proceso de empadronamiento del Censo de 1982.

En forma análoga al caso urbano, las secciones rurales debieron abarcar entidades rurales compactas.

2.7. Unidades de muestreo y unidades de información

El diseño muestral del PIDEH es bietápico. Las unidades de la primera etapa son secciones que corresponden, como ya se señaló, a áreas geográficas bien delimitadas y cuyo tamaño se expresa en términos de una determinada magnitud de población y un determinado número de viviendas.

Las unidades de la segunda etapa son las viviendas particulares que existen dentro de las secciones. Las viviendas particulares que finalmente resultan seleccionadas en la muestra se consideran como conglomerados, proce-

¹²Excepcionalmente una sección podía incluir manzanas de dos categorías, siempre que la categoría dominante comprendiera al menos el 70 por ciento del total de viviendas de la sección.

TABLA 2

TAMAÑOS MEDIOS ESTIMADOS DE SECCIONES, POR ESTRATOS
(número de viviendas y población por sección)

	IF		RAU		U		R	
	Tamaño medio viviendas	Tamaño medio población						
I. DE TARAPACA								
Arica	274	1.434	79	272			57	143
Iquique	294	1.570						
II. DE ANTOFAGASTA								
Tocopilla					87	364		
Calama	320	1.426						
Chuquibambilla	260	1.294						
Antofagasta	290	1.457						
El Loa/Antofagasta			97	336				
III. DE ATACAMA								
Chañaral					70	293		
Copiapó (ciudad)	304	1.420						
Copiapó			80	303				
Hualde					77	335		
IV. DE COQUIMBO								
Prov. Elqui			84	634			94	498
- Coquimbo/La Serena	279	1.230						
Prov. Limari			89	371			84	443
- Ovalle	277	1.192						
Prov. Choapa					88	411	99	497
V. DE VALPARAISO								
Prov. Petorca					209	805	135	744
Prov. Los Andes					206	779	146	1.025
Prov. San Felipe					221	753	139	1.038
Prov. Quillota							137	816
- Quillota/La Calera/La Cruz	289	1.322	213	897				
Prov. Valparaíso:								
- Gran Valparaíso:	292	1.339	195	1.339			119	571
Prov. San Antonio			59	334				
- San Antonio (ciudad)	275	1.226					126	717

Continuación tabla 2

	IF		RAU		U		R	
	Tamaño medio viviendas	Tamaño medio población						
VI. DEL LIBERTADOR GRAL. B. O'HIGGINS								
Prov. Cachapoal	282	1.405	96	312			113	702
- Rancagua								
Prov. Colchagua					83	332	105	713
Prov. Cardenal Caro					90	256	88	495
VII. DEL MAULE								
Prov. Cunicó	295	1.401	193	741			110	598
- Curicó (ciudad)								
Prov. Talca	283	1.303	217	853			83	553
- Talca (ciudad)								
Prov. Linares	288	1.372	184	793			112	597
- Linares (ciudad)								
Prov. Cauquenes					150	703	62	537
VIII. DEL BIOBIO								
Prov. Ñuble	303	1.537	94	432			93	549
- Chillán								
Prov. Concepción	302	1.461	93	478			100	696
- Gran Concepción								
- Lota/Coronel	292	1.435						
Prov. Arauco								
Prov. Bío-bío	271	1.466	83	440	101	496	99	688
- Los Angeles							95	539
IX. DE LA ARAUCANIA								
Prov. Malleco					87	389	90	522
Prov. Cautín	284	1.374	83	427			92	470
- Temuco								

Continuación tabla 2

	IF		RAU		U		R	
	Tamaño medio viviendas	Tamaño medio población						
X. DE LOS LAGOS								
Prov. Valdivia			93	394			88	560
- Valdivia	286	1.595						
Prov. Osorno			157	902			84	489
- Osorno	270	1.436						
Prov. Llanquihue			170	694			108	537
- Puerto Montt	297	1.446						
Prov. Chiloé/Palena					91	465	98	552
XI. AISEN DEL GENERAL CARLOS IBÁÑES DEL C.								
- Coihaique (ciudad)	141	719	83	404			71	291
XII. DE MAGALLANES Y LA ANTARTICA CHILENA								
- Punta Arenas (ciudad)	151	742	84	386			43	275
REGION METROPOLITANA								
Prov. Santiago								
Prov. Chacabuco					208	809	133	715
Prov. Cordillera			168	822			111	868
- Puente Alto	282	1.254						
Prov. Maipo			204	856			156	1.066
- San Bernardo	258	1.312						
Prov. Melipilla					194	980	118	768
Prov. Talagante					180	831	118	970
Gran Santiago	304	1.419						

diéndose a investigar a todas las personas que tienen su residencia habitual en ellas y a encuestar a todas aquellas que tienen 15 años o más.

Dos aspectos que vale la pena anotar son, por una parte, el hecho de que las secciones seleccionadas son objeto de procesos de actualización —renovación de registros de viviendas, hogares y personas— a través del tiempo. Por otra parte, se contempla que las secciones seleccionadas permanezcan como unidades de muestreo de primera etapa a lo largo de toda la vida útil de la muestra,¹³ pero que las viviendas que se seleccionen dentro de dichas secciones sean sometidas a procesos de rotación a través del tiempo.

2.8. Determinación del tamaño de la muestra

La determinación del tamaño de la muestra en un diseño bietápico es, en general, un proceso relativamente complejo. Tal complejidad se explica, en parte, porque, a diferencia de, por ejemplo, un diseño de muestreo aleatorio simple es necesario incorporar en el análisis la variabilidad de la(s) variable(s) en estudio, tanto en las unidades de muestreo de primera como de segunda etapa. En el caso del PIDEH, se suma la complejidad asociada al hecho de que la infraestructura muestral es de propósitos amplios, y de que el diseño contempla un mayor número de niveles de estimación que la muestra anterior, para los cuales la información histórica de varianzas es limitada.

En este contexto, la determinación del tamaño de la muestra del PIDEH trata, en primer lugar, de ser compatible con los niveles de estimación y la importancia de los estratos definidos, y con las principales variables que son el objeto de estimación.

En relación con el primero, cabe notar que el diseño privilegia los niveles de estimación nacional y regional, sin descuidar los restantes estratos. En este sentido, se trató de obtener un tamaño de muestra, nacional y regional, que arrojara errores de muestreo de baja magnitud a dichos niveles.

Con respecto a las variables en estudio, el tamaño de la muestra se determinó considerando a la tasa de desocupación —definida tanto en relación con la fuerza de trabajo total como con la población total— como una variable principal. Sin embargo, considerando la variabilidad que presenta la tasa de desocupación, desde el punto de vista de su comportamiento económico, para la determinación del tamaño de muestra también se consideraron otras

¹³Salvo que se produzcan cambios significativos en la estructura de la población de las secciones o estratos. Cabe anotar que el proceso de seccionar de la muestra, considera áreas de probable crecimiento o decrecimiento.

variables que observan una mayor estabilidad —como, por ejemplo, la tasa de participación.

Operativamente, y en una primera aproximación, se supuso un esquema de muestreo aleatorio simple, utilizando para tal efecto la información histórica disponible sobre varianzas. En una segunda aproximación, se incorporó la mayor variabilidad que provoca el hecho de que el diseño muestral es bietápico. La apreciación de esa mayor variabilidad se basó en algunos estudios empíricos realizados por el propio INE.

Para fijar el número de unidades de muestreo de segunda etapa a investigar dentro de cada unidad de primera etapa, se adoptaron varios enfoques, los cuales se basaron, principalmente, en hipótesis acerca de la variabilidad entre las dos etapas, y en criterios operativos de realización del trabajo en terreno y recolección de datos. Dichos criterios llevaron, finalmente, a considerar un tamaño de muestra, a nivel de unidades de segunda etapa, constante e igual, en promedio, a 15 viviendas dentro de cada sección.

La distribución de los tamaños de la muestra nacionales y regionales entre las distintas provincias y estratos —urbanos y rurales— se realizó adoptando, en parte, criterios de proporcionalidad —número de viviendas y tamaño poblacional de los mismos— y de acuerdo a la frecuencia de las variables laborales investigadas, en particular, de la tasa de desocupación.

También se consideró la presencia e importancia de estratos IF, a nivel provincial y regional, por cuanto mientras mayor es esa presencia y relevancia, menor es la exigencia de tamaño muestral para los restantes estratos, suponiendo una determinada magnitud de error deseada a nivel regional y/o provincial.

El enfoque de afijación, desde niveles de estimación altos a niveles bajos, se complementó y compatibilizó con un criterio inverso de agregación de tamaños de muestra independientes, calculados para los diferentes estratos más bajos.

Otro criterio relevante para la determinación del tamaño de muestra fueron los recursos disponibles. A este respecto, puede mencionarse que si bien la idea fue desarrollar una muestra que tuviese una mayor dimensión que la M.H.R., las limitaciones de recursos impidieron definir un tamaño demasiado amplio, como, por ejemplo, es el caso en algunos países desarrollados.

Por último, cabe consignar que todas las consideraciones anteriores se complementaron con algunos criterios *ad hoc*, mediante los cuales se trató de dar una mayor representatividad a todos los estratos definidos. Uno de

estos criterios consistió, por ejemplo, en fijar un mínimo de secciones a seleccionar por estrato, para evitar situaciones en que los criterios de afijación llevaran a que determinados estratos no quedaran suficientemente representados.

En la tabla 3, se presenta un resumen de los tamaños muestrales finalmente adoptados a nivel nacional y en las distintas regiones, distinguiendo al interior de las mismas sus áreas urbana y rural.

El tamaño total de la muestra, a nivel del país, alcanza a 35.910 viviendas, las que se distribuyen en 2.394 secciones.

Un aspecto que llama la atención en la distribución de la muestra total es que la necesidad de entregar estimaciones de buena calidad a nivel regional, que constituye, como ya se ha señalado, uno de los objetivos relevantes del Programa, obliga a utilizar fracciones de muestreo relativamente más altas en el caso de regiones de baja densidad poblacional y en que la frecuencia de las variables en estudio —por ejemplo, de la tasa de desocupación— es menor. Ello se aprecia claramente en las Regiones I, II, III y XI y XII. Por otra parte, las fracciones relativas de muestreo en las áreas rurales son sistemáticamente superiores a las de las áreas urbanas.

2.9. Selección de la muestra

La selección de la muestra se realiza en dos etapas, utilizando procedimientos similares en cada uno de los estratos definidos. Dentro de cada estrato, la selección de las unidades de primera etapa se efectuó en forma sistemática, a partir de un arranque aleatorio, y con una probabilidad de selección proporcional a una medida de tamaño, midiendo éste por medio del número de viviendas que tenían las mismas al Censo de 1982.¹⁴

En términos formales, la probabilidad de inclusión de la *i*-ésima unidad de primera etapa, $P_h(i)$, es igual:

$$P_h(i) = n_h \cdot \frac{M_{hi}}{M_h}, \quad (1)$$

donde *h* representa el índice del estrato F, RAU, U, R respectivo; n_h , el número de secciones en la muestra del estrato *h*; M_{hi} , el número total de viviendas en la sección *i* del estrato *h*, al Censo de 1982; y M_h es el total de viviendas en el estrato *h* al mismo Censo.

¹⁴Incluyendo un proceso de actualización postcensal de localidades urbanas.

TABLA 3

TAMAÑOS Y FRACCIONES DE MUESTREO DEL PIDEH
(total nacional, regiones y niveles urbano y rural)

Región	(1) Población (a junio 1986)	(2) Tasa de desempleo (julio-sept. 1986)	(3) Total de secciones en la población	(4) Total de secciones en la muestra	(5) Total estimado de viviendas en la población	(6) Total de viviendas en la muestra	(4)/(3)	(6)/(5)
TOTAL NACIONAL	12.068.595	10,6	13.670	2.394	2.498.820	35.910	17,5	1,4
Urbano	10.048.339	11,6	10.070	1.791	2.151.082	26.865	17,8	1,3
Rural	2.020.256	5,2	3.600	603	347.738	9.045	16,8	2,6
I. DE TARAPACA	294.282	11,3	273	87	58.225	1.305	31,9	2,2
Urbano	286.255	12,1	217	75	55.033	1.125	34,6	2,0
Rural	8.027	0,6	56	12	3.192	180	21,4	5,6
II. DE ANTOFAGASTA	362.850	10,3	389	114	77.027	1.710	29,3	2,2
Urbano	355.855	10,6	362	105	75.974	1.575	29,0	2,1
Rural	6.995	1,0	27	9	1.053	135	33,3	12,8
III. DE ATACAMA	196.852	4,4	465	90	45.457	1.350	19,4	3,0
Urbano	181.974	4,6	389	78	41.353	1.170	20,1	2,8
Rural	14.878	2,5	76	12	4.104	180	15,8	4,4
IV. DE COQUIMBO	444.254	9,5	680	150	91.621	2.250	22,1	2,5
Urbano	333.413	10,7	443	105	70.333	1.575	23,7	2,2
Rural	110.841	5,9	237	45	21.288	675	19,0	3,5
V. DE VALPARAISO	1.288.459	11,4	1.181	297	278.911	4.455	25,1	1,6
Urbano	1.150.437	11,8	1.017	240	256.787	3.600	23,6	1,4
Rural	138.022	8,5	164	57	22.124	855	34,8	3,9
VI. DEL LIBERTADOR	613.783	8,7	1.107	165	128.576	2.475	14,9	1,9
B. O'HIGGINS	352.382	10,2	722	102	87.361	1.530	14,1	1,8
Urbano	261.401	6,5	385	63	41.215	945	16,4	2,3
Rural								

Región	(1) Población (a junio 1986)	(2) Tasa de desempleo (julio-sept. 1986)	(3) Total de secciones en la población	(4) Total de secciones en la muestra	(5) Total estimado de viviendas en la población	(6) Total de viviendas en la muestra	(4)/(3)	(6)/(5)
VII. DEL MAULE	754.192	11,3	977	222	152.323	3.330	22,7	2,2
Urbano	431.118	13,3	419	135	97.230	2.025	32,2	2,1
Rural	323.074	8,4	558	87	55.093	1.305	15,6	2,4
VIII. DEL BIOBIO	1.606.054	8,9	2.046	327	312.315	4.905	16,0	1,6
Urbano	1.256.771	10,4	1.436	246	254.435	3.690	17,1	1,5
Rural	349.283	3,3	610	81	57.880	1.215	13,3	2,1
IX. DE LA ARAUCANIA	708.133	7,7	1.337	150	141.872	2.250	11,2	1,6
Urbano	428.073	10,9	757	99	88.794	1.485	13,1	1,7
Rural	280.060	2,9	580	51	53.078	765	8,8	1,4
X. DE LOS LAGOS	881.284	5,7	1.305	246	169.048	3.690	18,9	2,2
Urbano	534.616	8,1	661	153	108.440	2.295	23,1	2,1
Rural	346.668	2,1	644	93	60.608	1.395	14,4	2,3
XI. AISEN DEL GRAL. C. IBÁÑEZ DEL CAMPO	71.956	5,3	156	57	15.028	855	36,5	5,7
Urbano	57.677	6,5	107	45	11.549	675	42,1	5,8
Rural	14.279	1,4	49	12	3.479	180	24,5	5,2
XII. DE MAGALLANES Y ANTARTICA CHILENA	142.592	7,0	246	63	28.903	945	25,6	3,3
Urbano	132.708	7,3	210	51	27.355	765	24,3	2,8
Rural	9.884	4,9	36	12	1.548	180	33,3	11,6
R. METROPOLITANA	4.703.904	12,7	3.508	426	999.514	6.390	12,1	0,6
Urbano	4.547.060	12,8	3.330	357	976.438	5.355	19,7	0,6
Rural	156.844	9,7	178	69	23.076	1.035	38,8	4,5

Una vez seleccionadas estas unidades, se procedió a actualizar su tamaño —número de viviendas— obteniéndose M'_{hi} viviendas actualizadas por sección, quedando así definido el marco muestral para proceder a la selección de las unidades de segunda etapa.

Por su parte, la selección de las unidades de segunda etapa —viviendas dentro de las secciones seleccionadas— se efectuó también mediante un procedimiento sistemático, de una forma tal que cada vivienda tuviese la misma probabilidad de selección.

Formalmente:

$$P(j/i) = \frac{m_{hi}}{M'_{hi}}, \quad (2)$$

donde j hace referencia a una vivienda; i a la sección; m_{hi} es el número de viviendas a seleccionar dentro de la i -ésima unidad primaria de muestreo; y M'_{hi} , el total actualizado de viviendas dentro de la sección i .

2.10. Distribución de la muestra

La muestra total (2.394 secciones) se distribuye en un período trimestral, dividiéndose en tres muestras de un tamaño aproximadamente similar, cada una de las cuales se asigna a uno de los tres meses. Por ende, las consultas de cada una de estas muestras más pequeñas duran un mes.

De esta forma, resulta que todas las viviendas encuestadas en el mes t , vuelven a serlo en el mes $t+3$ y cada vivienda de la muestra es encuestada sólo una vez por trimestre.

Las estimaciones para el período (mes) t se realizan utilizando la información de los meses t , $t-1$ y $t-2$.

2.11. Rotación de la muestra

Con fines de actualización de la información obtenida y para evitar el cansancio de los informantes, se contempla un proceso de renovación periódica de las viviendas a encuestar de las secciones de la muestra. Para ello, se han definido 18 turnos rotantes de alrededor de 133 secciones cada uno.¹⁵ De esta forma, cada mes y de cada una de las secciones del turno correspondiente se vuelven a seleccionar 15 viviendas.

¹⁵La asignación de las secciones entre los turnos de rotación se realiza al azar, mediante un procedimiento sistemático.

Por medio de este sistema, la muestra completa se habrá rotado después del período de un año y medio. Se contempla que este proceso sea permanente a lo largo de toda la vida útil de la muestra.

Esquemáticamente, la rotación de la muestra puede representarse como sigue:

1, 2, 3, . . . , 18 Marzo (1987)	2, 3, 4, . . . , 18, (1) Abril	3, 4, 5, . . . 18, 1,(2) Mayo
4, 5, 6, . . . 18, 1, 2,(3) Junio	5, 6, 7, . . . 18, 1, 2, 3,(4) Julio	6, 7, 8, . . . 18, 1, 2, 3, 4,(5) Agosto
.....		1, 2, . . . , (18) Septiembre (1988)

donde los números de ordenamiento 1, 2, . . . , 18 representan los turnos de rotación de 133 secciones cada uno, hasta completar el total de 2.394 secciones una vez transcurrido un año y medio desde el inicio del proceso. El número entre paréntesis indica la fracción rotada en cada período, es decir, el turno —de 133 secciones— en el cual se efectúa una nueva selección de viviendas.

Cabe mencionar que el primer turno de rotación comenzó en abril de 1987, por este motivo en el diagrama aparece en dicho mes el número uno entre paréntesis. A modo de ejemplo, se puede constatar que en agosto del mismo año se rotaron las secciones del quinto turno, habiéndose rotado hasta esa fecha los turnos 1, 2, 3 y 4.

3. ESTIMADORES Y ERRORES DE MUESTREO

3.1. Alcances previos

Antes de presentar los estimadores utilizados, conviene realizar algunos comentarios que contribuyan a clarificar dicha presentación.

En primer lugar, de acuerdo a los criterios definidos, la obtención de estimaciones para parámetros tales como la tasa de ocupación, desocupación, participación u otras, en los distintos estratos y niveles de estimación, se calculan como cocientes de estimaciones separadas para el total de ocupados, desocupados, integrantes de la fuerza de trabajo o la variable que corresponde usar.

Por otra parte, los estimadores de estos totales consideran las características del diseño: estratificado, bietápico, probabilidades de selección de unidades primarias proporcionales a una medida de tamaño, etc. Pero, adicionalmente, se incorpora el hecho de que se dispone de información exógena a la muestra sobre proyecciones intercensales de población, tanto para la población total como para esta misma variable dividida en subpoblaciones de

acuerdo a características como sexo, tramos de edad y otros criterios, en los distintos estratos y niveles de estimación.¹⁶

3.2. Ajuste por sexo y edad y justificación de los estimadores de razón

Particularmente relevante es la clasificación de la población de acuerdo a la edad menor o mayor o igual de 15 años, que, como se mencionó antes, es una edad límite definitoria de las distintas variables laborales investigadas; y la partición de la población según sexo, que, como se verá más adelante, es un ajuste que se incorpora explícitamente en el proceso de estimación de los totales —estratales, provinciales, regionales y nacionales— de las diferentes variables.

La incorporación de la información exógena mencionada otorga una mayor flexibilidad para la definición de los estimadores, por cuanto permite construir estimaciones a partir de estimadores de razón, que, como se sabe, pueden generar importantes ganancias en términos de una menor variabilidad en el proceso de estimación.

3.3. Estimación de totales por estratos, sólo con información muestral

Supóngase que se desea estimar el total de ocupados, desocupados, pertenecientes a la fuerza laboral u otro en un estrato determinado, que constituye el nivel de estimación más reducido. Matemáticamente se procede como sigue.

Si se denota por y_{hij} el total de personas con la característica laboral analizada en un estrato h ($h = 1, 2, \dots, L$), sección i ($i = 1, 2, \dots, n_h$), vivienda j ($j = 1, 2, \dots, m_{hi}$) de la muestra, se puede definir:

$$y_{hi} = \sum_{j=1}^{m_{hi}} y_{hij}, \quad (3)$$

que corresponde al total de personas con la característica laboral que entrega la muestra de viviendas de la sección i ;

$$\bar{y}_{hi} = \frac{\sum_{j=1}^{m_{hi}} y_{hij}}{m_{hi}} \quad (4)$$

¹⁶Estas proyecciones se obtienen en el marco de un Convenio INE—CELADE.

o promedio muestral por vivienda de la respectiva sección y estrato;

$$\hat{y}_{hi} = M'_{hi} \cdot \bar{y}_{hi} \quad (5)$$

que expande el total de la característica en estudio a la sección i -ésima del estrato analizado, donde el factor M'_{hi} es el tamaño poblacional actualizado de viviendas de la sección i , estrato h ; y,

$$\hat{Y}_h^{(i)} = \frac{M'_{hi} \cdot \bar{y}_{hi}}{Z_{hi}}, \quad (6)$$

donde $Z_{hi} = \frac{M_{hi}}{M_h}$ indica la probabilidad de selección de la sección i -ésima del estrato considerado.

La última expresión ($\hat{Y}_h^{(i)}$), que corresponde a un total estimado —a partir de la muestra de viviendas de una sección i cualquiera— de personas que poseen la característica en estudio, en el estrato h , permitiría generar un estimador para dicho total considerando todas las secciones de la muestra, como sigue:

$$\hat{Y}_h = \frac{1}{n_h} \cdot \sum_{i=1}^{n_h} \hat{Y}_h^{(i)}, \quad (7)$$

donde n_h es el número de secciones seleccionadas del estrato considerado.

Sin embargo, dada la disponibilidad de información exógena de población, y con el objetivo de ganar una precisión adicional en la estimación, se incorpora el concepto de estimador de razón y el ajuste según sexo.

3.4. Los estimadores finalmente considerados.

Conservando la nomenclatura del acápite anterior, se define el siguiente estimador de razón:

$$\hat{R}_{hs} = \frac{\hat{Y}_{hs}}{\hat{X}_{hs}} = \frac{\sum_{i=1}^{n_h} \frac{M'_{hi} \cdot \bar{y}_{his}}{Z_{hi}}}{\sum_{i=1}^{n_h} \frac{M'_{hi} \cdot \bar{x}_{his}}{Z_{hi}}}, \quad (8)$$

donde s hace referencia a una de las dos subpoblaciones consideradas: hombres (de 15 o más años) o mujeres (de 15 o más años); \hat{Y}_{hs} , al total estima-

do de personas de la correspondiente subpoblación antes definida, que tienen la característica laboral investigada; y \hat{X}_{hs} , al total de personas de la misma subpoblación, estimado a partir de la muestra.

De esta forma, la expresión:

$$\hat{Y}_{hs} = \hat{R}_{hs} \cdot X_{hs}, \quad (9)$$

donde X_{hs} es la proyección exógena del número total de personas de la subpoblación s en el estrato, proporcionaría la estimación del total de personas que poseen la característica laboral analizada, referida a la subpoblación s , en el estrato respectivo.

Por último, el estimador del total de ocupados, desocupados, pertenecientes a la fuerza laboral u otra variable, en el estrato, surge de considerar:

$$\hat{Y}_h = \sum_{s=1}^2 \hat{Y}_{hs}, \quad (10)$$

donde $s = 1$, indica, por ejemplo, "hombres de 15 años o más" y $s = 2$, indica "mujeres de 15 años o más".

3.5. Estimadores para totales provinciales o regionales, y para los totales a nivel nacional

Si se supone la existencia de L estratos a nivel de una provincia y/o región, para generar estimadores de totales a esos niveles de estimación se procede a agregar los valores de los estratos, utilizando un procedimiento análogo a la metodología especificada en el acápite anterior.

Matemáticamente:

$$\hat{Y}_{ps} = \sum_{h=1}^L \hat{R}_{hs} \cdot X_{hs}, \quad (11)$$

correspondería al total poblacional estimado con la característica en estudio, referido a una de las dos subpoblaciones ya definidas, en la provincia o región analizada; y,

$$\hat{Y}_p = \sum_{s=1}^2 \hat{Y}_{ps}, \quad (12)$$

sería el total de personas ocupadas, desocupadas, pertenecientes a la fuerza laboral u otra variable, en una determinada provincia o región.

El mismo procedimiento indicado se extendería para generar estimaciones de totales a nivel nacional.

3.6. Cálculo periódico de las estimaciones

Las estimaciones se realizan considerando la muestra completa y sobre la base de períodos —móviles— trimestrales, en los que se incluye el mes en proceso y los dos meses precedentes. La información exógena sobre proyecciones de población, que se utiliza en los cálculos, se refiere al mes central.

3.7. Una acotación acerca de la calidad de las estimaciones

Una idea de la precisión del proceso de estimación antes presentado, puede obtenerse a partir del análisis de las varianzas, coeficientes de variación y errores absolutos y relativos de muestreo asociados a las correspondientes estimaciones.

TABLA 4

**PRECISION RELATIVA DE LA ESTIMACION DE LA TASA DE PARTICIPACION*
A PARTIR DEL PIDEH Y DE LA M.H.R.**

*(medida como el cociente entre los coeficientes de variación
del PIDEH y la M.H.R.)*

Regiones	Precisión relativa (C.V. PIDEH/C.V.M.H.R.)
I	0,702
II	0,808
III	1,577
IV	0,723
V	1,056
VI	0,772
VII	0,771
VIII	0,646
IX	0,654
X	0,698
XI	1,610
XII	0,687
Región Metropolitana	0,734

*Período julio-septiembre 1985, para las Regiones V y Metropolitana; y trimestre octubre-diciembre 1985, para las restantes regiones. En el caso del PIDEH, el primer período indicado corresponde a agosto-octubre 1985.

Aprovechando la existencia de estimaciones para la tasa de participación obtenidas en forma independiente, pero para el mismo período, a partir de la M.H.R. y el PIDEH, y de información acerca de las varianzas de dichas estimaciones, calculadas con fórmulas correspondientes a sus respectivos diseños, en la tabla 4 se ilustra una relación de precisión relativa entre ambas muestras.

La referida tabla permite apreciar que, exceptuando las regiones III, V y XI, las estimaciones del PIDEH resultan más precisas que las correspondientes a la M.H.R.

4. ORGANIZACION DEL TRABAJO EN TERRENO Y OTROS ASPECTOS OPERATIVOS

Dentro del proceso operativo del PIDEH se pueden distinguir cinco etapas principales: las tareas de ajuste, el levantamiento, la revisión, la codificación y el procesamiento electrónico de los datos.

4.1. Ajuste

Las denominadas labores de ajuste son las que dan inicio y cierran cada período de levantamiento. Ellas incluyen, por una parte, la preparación y el despacho del material a las oficinas regionales y provinciales del INE. Dicho material está constituido por las Tarjetas de Registro de Hogares (T.R.H.),¹⁷ cuestionarios de Encuesta, croquis y material cartográfico, cartas de presentación para los jefes de hogar y otros elementos necesarios para que el entrevistador se ubique en terreno y realice exitosamente su trabajo.

Por otra parte, la etapa de ajuste también incluye las labores de recepción, verificación y control del material, una vez realizadas las encuestas, y de preparación y organización del material para la etapa de codificación, lo cual implica analizar la secuencia y los contenidos de las anotaciones realizadas por encuestadores y supervisores.

Con respecto al control de encuestas, cabe destacar que se distinguen dos situaciones: las encuestas logradas y las encuestas no logradas. Esta última categoría se divide, a su vez, en tres clases: no logradas tipo A; no logradas tipo B y no logradas tipo C. Las encuestas no logradas tipo A se consideran como susceptibles de ser logradas a través de una nueva visita a la vivienda y se originan en factores como: moradores ausentes, rechazo, sector

¹⁷Formulario que acompaña a la Encuesta y del cual se obtiene información referida al hogar, su ubicación, las características socioeconómicas de las personas que lo constituyen y la caracterización de la vivienda.

inaccesible y otros análogos. Por su parte, las no logradas tipo B, son consideradas como encuestas no realizables en el presente, y se originan en motivos como: viviendas desocupadas temporalmente, en reparación y otras (viviendas colectivas, de veraneo, etc.). Por último, las encuestas no logradas tipo C, corresponden a viviendas que ya no existen como tales en el momento de la entrevista y se subdividen en viviendas demolidas, destinadas a otros usos y otras.

4.2. Levantamiento de la encuesta

La etapa de levantamiento es una de las más importantes dentro del proceso operativo del PIDEH. Para efectuarla, el INE cuenta con un equipo de encuestadores y supervisores que se distribuyen a lo largo de todo el país, donde los centros de operación son las oficinas regionales y provinciales del Instituto. El número de encuestadores, a nivel nacional, alcanza a poco menos de un centenar, de los cuales aproximadamente un 15 por ciento trabajan en la Región Metropolitana y el 85 por ciento restante se distribuye en las otras regiones del país.

Los programas de levantamiento de la encuesta, a nivel de toda la muestra, son trimestrales, con períodos de levantamiento mensuales para cada una de las partes en que se divide dicha muestra total. Para tal efecto, se usa un sistema de cargas de trabajo semanales —4 ó 5 semanas por mes, según programación— que incluyen trabajo de levantamiento en los cuatro niveles de estimación IF, RAU/U, R. Dentro de cada semana, la distribución promedio aproximada por estratos es de 65 por ciento de secciones IF, 20 por ciento de secciones RAU/U y 15 por ciento de secciones R.

Particularmente relevante es el proceso de supervisión de la encuesta, en la fase de levantamiento. A este respecto, se distinguen dos niveles: la supervisión nacional y la supervisión regional.

La supervisión nacional es llevada a cabo por un equipo a cargo de un Coordinador de Supervisión Nacional, y abarca a todas las regiones y estratos IF, RAU/U y R dentro de ellas.

El énfasis principal de esta supervisión está puesto en la capacitación y el perfeccionamiento del personal que lleva a cabo el trabajo de terreno, y la superación de los errores no muestrales. Entre éstos cabe mencionar: errores de cobertura (encuestas no logradas tipo A, B y C), errores de falta de respuestas (encuestas incompletas u omisiones), errores de respuesta (mala secuencia, insinuación de respuestas u otras), errores de enumeración de secciones (direcciones incorrectas, viviendas fuera de la sección, errores de con-

cepto —vivienda versus hogar, etc.), falta de antecedentes cartográficos en las secciones enumeradas y otros errores.

En cuanto a la supervisión regional, ella es realizada por personal de las oficinas regionales y provinciales del INE. El principal objetivo de esta supervisión es la capacitación y apoyo del trabajo del encuestador, la repetición aleatoria de encuestas con fines de control, la verificación y repetición de encuestas no logradas por el encuestador, y la coordinación de todo el trabajo realizado.

4.3. Revisión y codificación

La primera de éstas consiste en la revisión del contenido general y las relaciones de secuencia, dependencia y continuidad de las preguntas de la encuesta. Como se sabe, la secuencia de las preguntas está determinada por indicaciones como: salte a, pase a, termine y otras combinaciones, las cuales apuntan a que el encuestador realice el trabajo con seguridad, agilidad y evitando cansar al informante.

De este proceso de revisión, así como de todas las demás etapas operativas, surgen elementos que alimentan y perfeccionan el trabajo de las restantes fases.

En cuanto a la codificación de la encuesta, se distinguen dos tipos de preguntas: las preguntas cerradas o precodificadas, en que cada respuesta considerada tiene un código impreso, y las preguntas abiertas que no tienen código y es necesario asignárselo según corresponda. También se considera la posibilidad de preguntas parcialmente cerradas, como cuando se incluye "otros (especificar)". Los conceptos incluidos en "otros" deben codificarse en forma separada.

4.4. Procesamiento electrónico de la encuesta

La etapa de procesamiento puede ser subdividida en cuatro actividades o procesos principales: ingreso de datos, validación, cálculo de ponderaciones y obtención de tabulados finales.

El ingreso de la información contenida en los cuestionarios es un proceso continuo a lo largo del mes, que considera un control manual estricto de la cobertura geográfica del material recepcionado.

Por otra parte, la validación de la información —detección y corrección de errores de la encuesta— considera dos fases. En la primera, se procede a depurar la identificación geográfica de los datos —regiones, estra-

tos y secciones. Para ello, se confecciona un listado exhaustivo de todas las viviendas, hogares y personas dentro de los mismos, obteniéndose, además, información agregada sobre el número de viviendas encuestadas por sección y estrato. Si aparece algún código de sección, estrato o región inexistente o que no corresponda, ello queda indicado con una señal de error.

Un segundo trabajo de validación corresponde a la depuración de los códigos correspondientes a variables definidas a partir de características personales y de características económicas, el que se realiza en forma automática. A este respecto, se considera una validación de rangos —chequeo de códigos erróneos en lo que respecta a los valores permitidos para las diferentes variables— y una validación de consistencia lógica de la información recogida en la encuesta.

Por su parte, el proceso de cálculo de ponderaciones requiere disponer de la muestra completa para el trimestre en estudio. Para ello, se trabaja con el mes en proceso y los dos meses precedentes, además de las proyecciones de población correspondientes al mes central.

El cálculo de ponderaciones supone, en primer lugar, la obtención de factores generales para cada estrato y, en una segunda etapa, el cálculo del ponderador final a nivel de personas, por sexo y edad para cada sección, el cual es asignado a los registros —personas— de la muestra.

Para visualizar más claramente la estructura de estos ponderadores, notemos que de las expresiones (11) y (12) de la sección 3, se puede deducir:

$$\hat{Y} = \sum_{s=1}^2 \sum_{h=1}^L \hat{R}_{hs} X_{hs}, \quad (13)$$

expresión que es equivalente a:

$$\hat{Y} = \sum_{s=1}^2 \sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \sum_k \frac{M'_{hi}}{Z_{hi} m_{hi} \cdot n_h} \frac{X_{hs}}{\hat{X}_{hs}} y_{hsijk} \quad (14)$$

Por lo tanto, la ponderación para y_{hsijk} , que corresponde a la característica del individuo k de la vivienda j , seleccionada en la sección i , y que pertenece a la subpoblación s (1 ó 2) del estrato h -ésimo, viene dada por:

$$\frac{M'_{hi} X_{hs}}{Z_{hi} \cdot m_{hi} \cdot n_h \cdot \bar{X}_{hs}}$$

(15)

Por último, debe mencionarse la actividad de obtención de tabulados finales. Las tablas 5 y 6 ilustran el conjunto de cuadros simples y de doble entrada obtenidos a partir de la información que entrega la encuesta.

TABLA 5
ENCUESTA NACIONAL DEL EMPLEO:
VARIABLES Y TABULADOS PUBLICADOS

CUADRO	AREAS DE ESTIMACION						VARIABLES																		
	País	Región	Provincia	Urbano	Rural	Area de estimación	Población total	Menores de 15 años	15 años y más	Fuerza de trabajo	Ocupada	Desocupada	Cesante	Busca trabajo por primera vez	Sexo	Fuera de la fuerza de trabajo	Alfabetismo	Años de estudio cursados	Rama de actividad económica	Grupo de ocupación	Categoría en la ocupación	Método utilizado para buscar trabajo	Tasas de desocupación por rama de actividad	Tasas de desocupación por grupo de ocupación	
1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X									
2	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X									
3	X	X		X	X			X	X	X							X	X							
4	X	X	X	X	X	X				X									X						
5	X	X	X	X	X	X				X										X					
6	X	X	X	X	X	X				X										X					
7	X	X	X	X	X	X					X			X					X						
8	X	X	X	X	X	X					X			X						X					
9	X	X	X	X	X	X					X			X						X					
10	X	X	X	X	X	X					X											X			
11	X	X	X	X	X	X																	X		
12	X	X	X	X	X	X																		X	

TABLA 6

ENCUESTA NACIONAL DEL EMPLEO: VARIABLES
Y TABULADOS PUBLICADOS
(doble entrada)

AREAS DE ESTIMACION		VARIABLES																																		
Cuadro		País	Región	Provincia	Urbano	Rural	Area de estimación	Población total	Menores de 15 años	15 años y más	Fuerza de trabajo	Ocupada	Desocupada	Cesante	Busca trabajo por primera vez	Sexo	Fuera de la fuerza de trabajo	Alfabetismo	Años de estudio currado	Rama de actividad económica	Grupo de ocupación	Categoría de la ocupación	Método utilizado para buscar trabajo	Tasas de desocupación por rama de actividad	Tasas de desocupación por grupo de ocupación	Tramos de edad	Tipo de educación	Horas semanales trabajadas	Actividad agrícola y no agrícola	Razon para trabajar menos de 35 horas semanales	Tiempo que busca trabajo en semanas	Tiempo desde último empleo	Motivo para dejar de trabajar			
1	X	X					X	X	X	X	X	X	X	X	X	X	X																			
2	X	X			X	X			X	X	X	X	X	X	X	X	X	X	X							X										
3	X	X		X	X							X					X			X	X															
4	X	X		X		X						X					X			X	X															
5	X	X										X	X	X		X			X	X		X														
6	X	X										X	X	X		X		X	X		X	X														
7	X	X										X	X	X		X		X	X		X	X														
8	X	X										X	X	X		X		X	X		X	X														
9	X	X										X	X	X		X		X	X		X	X														
10	X	X		X								X				X		X	X		X	X			X		X									
11	X	X		X	X							X				X		X	X		X	X														
12	X	X		X								X				X		X	X		X	X														
13	X	X			X							X				X		X	X		X	X														
14	X	X										X				X		X	X		X	X			X		X									
15	X	X										X				X		X	X		X	X			X		X									
16	X	X										X	X	X		X		X	X		X	X			X		X									
17	X	X										X	X	X		X		X	X		X	X			X		X									
18	X	X										X	X	X		X		X	X		X	X			X		X									
19	X	X										X	X	X		X		X	X		X	X			X		X									
20	X	X										X	X	X		X		X	X		X	X			X		X									
21	X	X										X	X	X		X		X	X		X	X			X		X						X	X		
22	X	X										X	X	X		X		X	X		X	X			X		X								X	X
23	X	X										X	X	X		X		X	X		X	X			X		X								X	X
24	X	X										X	X	X		X		X	X		X	X			X		X								X	X
For Reg.											X				X					X	X															X