

Encuesta de comportamiento lector 2014



www.plandelectura.cl



Gobierno
de Chile

www.gob.cl



Plan Nacional
de la Lectura

MANUAL DE ANÁLISIS DE DATOS

ENCUESTA DE COMPORTAMIENTO LECTOR 2014

CONSEJO NACIONAL DE LA CULTURA Y LAS ARTES

**Consejo Nacional del Libro y la Lectura
Plan Nacional de la Lectura**

Santiago, abril 2015

INDICE

Presentación.....	01
I. ANTECEDENTES METODOLÓGICOS.....	02
Recomendaciones.....	02
Formato Base de datos.....	02
1. Metodología de la Encuesta.....	03
2. Cuestionarios.....	05
II. ELEMENTOS PARA EL ANÁLISIS DE LOS DATOS.....	07
1. Factor de expansión.....	07
2. Muestras complejas y errores muestrales.....	07
3. Base de datos.....	08
3.1. Edición de Base de datos.....	09
3.2. Variables construidas.....	09
3.3. Codificación de Preguntas Abiertas.....	11
3.4. Codificación ocupación y actividad de la empresa.....	12
3.5. Variable Grupo Socioeconómico.....	13
3.6. Tipología de Lectores.....	14
Referencias.....	16

PRESENTACIÓN

El siguiente manual tiene por objetivo facilitar la lectura y uso de la **Base de datos** de la Encuesta de Comportamiento Lector 2014. Está destinado, especialmente, a estudiantes, académicos, especialistas, agentes culturales y otros que requieran información sobre el estado de la lectura en el país.

Este manual entrega información sobre elementos metodológicos y sobre el análisis de los datos de la Encuesta. En la primera parte, se presentan: los antecedentes generales, para comprender los objetivos de esta medición y los responsables de la ECL 2014 (créditos), para que sean siempre citados en los documentos y publicaciones que resulten de su utilización. En su segunda parte, se entregan los elementos para el análisis de los resultados, que incluyen: factor de expansión, muestras complejas y errores muestrales, como también la estructura y edición de la base de datos, variables construidas y codificación de preguntas abiertas.

En síntesis, este manual es una guía para utilizar la base de datos de la Encuesta de Comportamiento Lector 2014.

I. ANTECEDENTES METODOLÓGICOS

En la Encuesta de Comportamiento Lector 2014 se mantiene el carácter nacional y se aborda una muestra representativa de la población entre 9 y 65 años de edad, residentes habituales de viviendas particulares, esta vez solo para zonas urbanas y con representatividad regional. Esto es a través de dos cuestionarios, uno aplicado a la población de entre 9 y 14 años y otro a la población de entre 15 a 65 años.

Recomendaciones

Para citar la Encuesta en documentos y publicaciones que resulten de su utilización, siempre usar el siguiente texto en la Introducción:

- "Esta investigación utilizó información de la Encuesta de Comportamiento Lector 2014. Se agradece al Consejo Nacional de la Cultura y las Artes, propietario intelectual de la Encuesta, haber permitido disponer de la base de datos. Todos los resultados de la presente investigación son de responsabilidad del autor y en nada compromete al Consejo Nacional de la Cultura".

Y el siguiente texto en las referencias bibliográficas:

- CNCA, Consejo Nacional de la Cultura y las Artes (2014). *Encuesta de Comportamiento Lector 2014*. Santiago: Consejo Nacional de la Cultura y las Artes.

Formato de la Base de datos

La base de datos se encuentra en formato .sav, el cual corresponde al paquete estadístico SPSS versión 19.0.

1. METODOLOGÍA DE LA ENCUESTA

La **población objetivo** considera a todas las personas de 9 a 65 años de edad que residan habitualmente en viviendas particulares ocupadas localizadas en zonas urbanas de las 15 regiones de Chile.

El **marco muestral** deriva del Instituto Nacional de Estadísticas (INE), actualizado al segundo semestre del año 2008.

El **diseño muestral** se puede caracterizar como un muestreo de áreas probabilístico, estratificado geográficamente y multietápico.

A continuación se explicita la ficha metodológica resumen:

Tabla 1. Ficha Metodológica ECL 2014

Diseño del estudio	Estudio cuantitativo
Grupo objetivo	Hombres y mujeres de 9 a 65 años, residentes en zonas urbanas de las 15 regiones del país.
Instrumento	Cuestionario estructurado, compuesto por preguntas cerradas de carácter simple y múltiple.
Tipo de recolección de información	Encuesta presencial en hogares, aplicada por encuestadores (PAPI). ¹
Marco muestral	Marco de muestreo de manzanas actualizado al segundo semestre de 2008, entregado por INE (MM2008-B) y complementario para zona semiurbana (MS2002). Marco muestral de 310 comunas.
Diseño muestral	Nacional (15 regiones en zonas urbanas). Muestra de 105 comunas.
Tipo de muestreo	Muestreo de áreas probabilístico, estratificado geográficamente y multietápico, donde se seleccionaron: <ul style="list-style-type: none"> ➤ Unidad de Primera Etapa (UPE): 56 Comunas con inclusión forzosa (grandes centros urbanos, capital regional o parte del Gran Santiago) y 49 al azar por estratificación. ➤ Unidad de Segunda Etapa (USE): Manzanas censales seleccionadas con probabilidad proporcional al tamaño (total de viviendas). ➤ Unidad de Tercera Etapa (UTE): Viviendas ocupadas seleccionadas con igual probabilidad dentro de la manzana. ➤ Unidad de Cuarta Etapa (UCE): Personas entre 9 y 65 años de edad seleccionadas con igual probabilidad dentro de la vivienda.
Tamaño muestral	6.990 encuestas completas.
Error muestral (95% nivel de confianza) ²	<ul style="list-style-type: none"> ➤ Error absoluto de $\pm 1,2\%$ a nivel nacional, bajo supuesto de muestreo aleatorio simple (MAS), para una proporción de 50% a nivel de confianza de 95%. ➤ Error absoluto de $\pm 1,7$ a nivel nacional, bajo muestreo aleatorio simple, para la variable "Lee 15 minutos al menos una vez al día" a un nivel de significancia de 95%. Efecto de diseño estimado de 1,97 para la variable "Lee 15 minutos al menos una vez al día", considerando el diseño complejo de la encuesta.
Factores de expansión y ponderación de la muestra	El ponderador fue calculado considerando las probabilidades de selección, el ajuste de no respuesta y la posestratificación por sexo, edad, región y nivel educacional. Para los ajustes de no respuesta y posestratificación se utilizaron las técnicas de <i>propensity score</i> y raking, respectivamente.
Tasa de respuesta y rechazo	Tasa de respuesta (RR2): 84,1% Tasa de rechazo (REF2): 6,2%
Factores de expansión y ponderación de la muestra	El ponderador fue calculado considerando las probabilidades de selección, el ajuste de no respuesta y la posestratificación por sexo, edad, región y nivel educacional. Para los ajustes de no respuesta y posestratificación se utilizaron las técnicas de <i>propensity score</i> y raking, respectivamente.
Tasa de respuesta y rechazo	Tasa de respuesta (RR2): 84,1% Tasa de rechazo (REF2): 6,2%

¹ PAPI: Paper and Pencil Interview.

2. CUESTIONARIOS

La Encuesta de Comportamiento Lector 2014 consta de dos cuestionarios: **Cuestionario Forma A** (15 a 65 años de edad) y **Cuestionario Forma B** (9 a 14 años de edad). Los módulos que contienen ambos cuestionarios son:

- 0. Módulo 0
 - A. Módulo Tiempo libre, lectura en general y Tenencia de libros en el hogar.
 - B. Módulo Libros
 - C. Módulo Diarios impresos
 - D. Módulo Diarios digitales
 - E. Módulo Revistas
 - F. Módulo Historietas y/o comics
 - G. Módulo Formación del hábito lector
 - H. Módulo Formación lectora en el hogar actual
 - I. Módulo Autopercepción lectora
 - J. Módulo Acceso a lugares de lectura
 - K. Módulo Uso de internet
 - L. Módulo Composición del hogar
 - M. Módulo Entrevistado (ocupación e ingresos)
 - N. Módulo Jefe de hogar (ocupación e ingresos)
 - O. Módulo ingresos del hogar

De estos módulos, diez son obligatorios (A, G, H, I, J, K, L, M, N, O) esto es, se aplican a todos los encuestados, y cinco dependen de la respuesta al "Perfil del Encuestado", una sección inicial que contiene preguntas de los módulos B, C, D, E y F y que permiten decidir la aplicación de esos módulos.

Para el caso de los niños, niñas y/o adolescentes encuestados de 9 a 17 años de edad, los Módulos L, M, N y O deben ser contestados por uno de los padres del seleccionado o su responsable principal (legal).

Si bien, ambos cuestionarios contienen los mismos módulos, poseen las siguientes diferencias: el **Cuestionario Forma A** posee más preguntas que el **Cuestionario Forma B** en algunos módulos y en algunas preguntas varían los ítems de respuesta. Por su parte, la Forma B fue elaborada con un fraseo más informal (en vez de "usted" se utilizó "tú").

II. ELEMENTOS PARA EL ANÁLISIS DE LOS DATOS

1. FACTOR DE EXPANSIÓN

Para analizar los datos de la Encuesta y realizar inferencias poblacionales, se debe utilizar el factor de expansión, que debe estar siempre con la ponderación activa. El factor de expansión es identificado en la base de datos como "FEXP", el cual expande y pondera la muestra a la población de 9 a 65 años, residentes habituales en viviendas particulares ocupadas localizadas en zonas urbanas de las 15 regiones de Chile.

2. MUESTRAS COMPLEJAS Y ERRORES MUESTRALES

Debido a que esta Encuesta posee un diseño muestral complejo, se recomienda la utilización de las herramientas que entregan los software de análisis de datos estadísticos para las encuestas complejas, considerando la utilización de las variables "VARSTRAT" (pseudo estrato) y "VARUNIT" (pseudo conglomerados).

Estas le permitirán obtener márgenes de error para cada variable, obteniendo así el estimador y su intervalo de confianza.

La muestra 2014 fue diseñada para producir estimaciones estadísticamente válidas a nivel nacional y en las 15 regiones del país. Si se desea producir estimaciones a menores niveles de desagregación, se recomienda evaluar su confiabilidad estadística utilizando criterios como el tamaño de la muestra, coeficiente de variación, intervalos de confianza y error muestral.

3. BASE DE DATOS

La base de datos de la ECL 2014 es única para ambos cuestionarios. Su estructura se define por: primero, se presentan variables de identificación; luego, están las variables ordenadas según la presentación del cuestionario (entre las cuales se incluyen variables codificadas posteriormente a partir de preguntas abiertas) y se finaliza con las variables construidas en la fase de análisis³. Las etiquetas de las variables corresponden a las preguntas de los cuestionarios.

Debido a las diferencias entre preguntas de ambos cuestionarios, se tomaron las siguientes decisiones:

- Se agrega una letra "a" minúscula, seguida de un punto seguido, en las preguntas que solo son del Cuestionario Forma A. Ejemplo: "a.A6".
- Se agrega una letra "b" minúscula, seguida de un punto seguido, en las preguntas que sólo aparecen en el Cuestionario Forma B. Ejemplo: "b.B9B".
- Se agrega una letra "c" minúscula, seguida de un punto seguido, en las preguntas que son consultadas en ambas formas de cuestionario, pero que en la Forma B no aparecen todas las categorías de respuesta como posibilidades de elección. Ejemplo: "c.A1A".
- Las variables que no se anteceden de ninguna letra minúscula aparecen en ambos cuestionarios exactamente igual.

3.1. Edición de Base de Datos

³ El nombre de las variables, creadas posteriormente para el análisis, se encuentran en el Manual de la Base de Datos. En este se presentan todas las variables de la base en orden, y dan cuenta del punto desde el cual se incluyen variables construidas para el análisis de datos.

Las respuestas entregadas por los encuestados en la ECL 2014 fueron revisadas y validadas con el fin de verificar la consistencia de los datos. Es importante destacar que los casos válidos que derivan de una pregunta dicotómica excluyen el código "no sabe", pero aquellos que provienen de una pregunta nominal con más de dos categorías incluyen ese código. Por otro lado, los casos válidos que derivan de una pregunta filtro excluyen la no respuesta.

Sin embargo, pese al análisis existen algunas inconsistencias que habitualmente se encuentran en la mayoría de las encuestas. Cuando se observó una inconsistencia relevante se generaron variables para el análisis que corrigieran este elemento, manteniendo siempre las variables originales.

Por otro lado, se detectaron variables introducidas en el cuestionario que no entregaron resultados adecuados, por lo cual se recomienda no utilizar por parte de investigadores. Este es el caso de las preguntas K9 y K10⁴ que consultan las horas diarias de lectura, pero dentro de las categorías de respuestas no se incluye el valor cero, por lo cual la respuesta de los encuestados podría estar sobre estimada.

3.2. Variables Construidas

Debido a que el Módulo L del cuestionario contiene la información del encuestado y de todos los integrantes de su hogar respecto a educación y actividad principal, se crearon variables adicionales que permitieran identificar la posición en la que fue consignado el:

- Encuestado "L00".

⁴ Existen personas que responden leer 24 horas. Si bien, en la práctica es imposible que alguien esté esa cantidad de tiempo leyendo activamente, puede haberse dado porque aparece como categoría de respuesta en el cuestionario y/o podría haber respondido bajo la premisa de utilizar pasivamente internet las 24 horas ya que mediante celulares se encuentran conectados continuamente a la red. Esta pregunta no se acompañó de instrucciones al encuestador respecto a la definición de "utilización de internet" que permitieran excluir esta posibilidad.

- Jefe de hogar "L0".

Se realizó, en el caso del encuestado, a partir de la coincidencia de los datos iniciales de este establecidos en la Hoja de Ruta 3 (utilizada para el empadronamiento hogar) y los datos del Módulo L; y para el jefe de hogar a partir de la pregunta "L3".

Además, se crearon todas las variables del módulo L, a partir de las variables "L7" y "L8"⁵, para el entrevistado y para el jefe de hogar, con el fin de facilitar el manejo de esos datos. Luego de lo anterior, se construyeron las variables de educación para, a partir:

- Entrevistado "edu_entrevistado" y "edu_entrevistado_3"
- Jefe de hogar "edu_jh" y "edu_jh_3"

Se incluyen las variables categóricas de edad utilizadas para el análisis de datos, así como las tres variables que constituyeron indicadores claves en la presente encuesta, a saber:

- "Frec_Lectura_seg" referente a la lectura diaria por 15 a 20 minutos continuos de cualquier tipo de material y en cualquier soporte.
- "B4AR_tot" lectura de al menos 1 libro por motivos de estudio y/o trabajo en los últimos 12 meses.
- "B4BR_tot" lectura de al menos 1 libro por motivos de entretenimiento y/u ocio en los últimos 12 meses.

⁵ Estas variables se derivan de dos preguntas incluidas en los cuestionarios y usualmente consideradas para cualquier cuestionario sociodemográfico que consulte por educación, a saber, "¿Cuál es el nivel educacional más alto alcanzado por usted?" (L7 en la presente encuesta) y "¿Completó el nivel educacional anteriormente declarado?" (L8 en la presente encuesta) (INE, 2011). De esta manera, se genera una variable final que combina ambas preguntas y permite dar cuenta de las personas que han completado o no cada nivel educacional (de acuerdo a los parámetros chilenos).

La base incluye también las variables más relevantes creadas para el análisis de datos de la presente encuesta que constituyen:

- Recodificaciones considerando el total de casos encuestados: Identificadas con el nombre de la variable de origen más el sufijo “_rec”, “_cat” o “_tot”.
- Rectificaciones de casos inconsistentes⁶: Específicamente las variables asociadas a “B4A”, “B4B”.
- Para las variables continuas la eliminación de casos fuera de rango o outliers (generalmente truncando en el percentil 99): Se identifican con el nombre de la variable de origen más “_trunc”.

Se informa a investigadores, estudiantes y usuarios utilizar para el análisis de lectura de libros las variables “B4AR_tot” y “B4BR_tot”⁷ y aquellas que están recodificadas⁸, con el fin de trabajar sobre el total de los encuestados y no solo sobre el total de alfabetos, esto es, porque la diferenciación de alfabetos/analfabetos no suele aparecer en otras encuestas, por lo cual se decidió mantener una parsimonia con otras investigaciones del tema.

3.3. Codificación de Preguntas Abiertas

Todas las preguntas abiertas fueron sometidas a codificación. En los casos de preguntas abiertas que obedecían a categorías de “Otro, ¿cuál?”, se revisó, primero, que la mención no estuviera contenida en categorías de respuesta ya existente; de ser así, se rectificó el código en una nueva variable la cual es señalizada con un guión bajo seguida del número dos (_2) al finalizar el

⁶ En esta serie de preguntas existieron personas que mencionaban haber leído menos libros en los últimos 12 meses que en el último mes. Las preguntas “B4A” en su conjunto terminaron con un 9,3% y las preguntas “B4B” con un 12,6%. Si bien, es una proporción que no invalida las preguntas, se optó por crear una variable solo para “B4A1” y “B4B1” donde se imputó la lectura para aquellos casos donde se mencionaba leer menos los últimos 12 meses que los últimos seis, tres o último mes.

⁷ “B4AR_tot” *lectura de al menos 1 libro por motivos de estudio y/o trabajo en los últimos 12 meses* y “B4BR_tot” *lectura de al menos 1 libro por motivos de entretenimiento y/u ocio en los últimos 12 meses*.

⁸ Considera las preguntas B1A, B1B y B4A1 para los motivos de estudio y trabajo; y las preguntas B1A, B1B y B4B1 para los motivos de entretenimiento y/u ocio.

nombre de la variable (por ejemplo, "A1A_2"). De esta manera, en la base de datos aparece tanto la variable original y la variable rectificada (por ejemplo: "A1A" y "A1A_2").

Luego de ello se codificaron las demás preguntas abiertas que no estaban contenidas en una categoría de respuesta pre existente, en una variable nueva del mismo nombre de la original, pero que termina con el número "2" (por ejemplo: "A1_OTRA2"). Esto a excepción de la codificación realizada para las variables de ocupación ("M6" y "N6") y actividad de la empresa ("M9" y "N9"), la cual es explicada en el siguiente apartado.

3.4. Codificación Preguntas Abiertas sobre ocupación y actividad de la empresa

Las preguntas "M6" y "N6" relacionadas a la ocupación o actividad principal del entrevistado (o responsable principal de los niños, niñas y/o adolescentes de 9 a 17 años encuestados) y del jefe de hogar respectivamente, fueron sometidas a una codificación según lo establecido en la Clasificación Internacional Uniforme de Ocupaciones CIUO (OIT, 2008). Por otro lado, las preguntas "M9" y "N9" relacionadas a la actividad de la empresa, industria o servicio donde el encuestado y jefe de hogar (respectivamente) mencionan trabajar, fueron codificadas mediante la Clasificación Industrial Internacional Uniforme de todas las actividades económicas CIIU (Naciones Unidas, 2009)⁹.

La clasificación se realizó a nivel de subgrupos principales (que en la práctica corresponde a un código de dos dígitos), a partir de la información contenida en las preguntas abiertas "M6" y "N6", complementadas con nivel educacional, tipo de empleado, relación contractual, actividad que realiza la empresa donde trabaja e ingreso (tanto del entrevistado para "M6"¹⁰ como del jefe de

⁹. La Clasificación CIUO corresponde a "un conjunto de tareas y cometidos desempeñados por una persona, o que se prevé que ésta desempeñe, incluido para un empleador o por cuenta propia" (OIT, 2008).

¹⁰. En el caso de los niños, niñas y adolescentes de 9 a 17 años encuestados, debido a que el módulo M es respondido por sus padres o responsables principales, cuando existía información que permitía identificarlo con

hogar para "N6"). De esta manera, en la base de datos, aparecerá tanto la variable derivada de una pregunta abierta ("M6" y "N6"), como la variable codificada posteriormente identificada como "M6_CODIGO_CIUO" y "N6_CODIGO_CIIU". Se incluyen además dos variables que resumen la codificación a un dígito correspondiente a "M6_CIUO" y "N6_CIIU" respectivamente.

Se clasificó la actividad productiva según códigos CIUU¹¹ a partir de la información contenida en las preguntas abiertas "M9" y "N9" complementadas igualmente con datos respecto a nivel educacional, tipo de empleado, relación contractual y ocupación principal (del entrevistado para "M9" y del jefe de hogar para "N9"). Las variables codificadas por este método son identificadas en la base de datos como

"M9_CODIGO_CIIU" y "N9_CODIGO_CIIU". Adicionalmente, se incluyen variables resumidas al código por sección denominadas "M9_CIIU" y "N9_CIIU" respectivamente.

3.5. Variables de Grupo Socioeconómico

Para la presente encuesta, DESUC construyó variables de nivel educacional para el entrevistado y para el jefe de hogar ("edu_entrevistado" y "edu_entrevistado_3"; "edu_jh" y "edu_jh_3"), a partir de las variables "L7" y "L8" del entrevistado y del jefe de hogar¹². De esta manera, se genera una variable final que combina ambas preguntas y permite dar cuenta de las

claridad se tomaron los datos de educación de ellos para la clasificación. En los casos que no fue posible identificarlos, la información sobre educación quedó en blanco, utilizando los demás campos para codificar.

¹¹ Por otro lado, la Clasificación CIIU corresponde a una clasificación internacional de referencia a las actividades productivas, cuyo propósito es ofrecer un conjunto de categorías de actividades que se puedan utilizar para la reunión y difusión de datos estadísticos de acuerdo con esas actividades (Naciones Unidas, 2009). Esta clasificación puede realizarse por sección (código de letra) división (código de dos dígitos), grupo (código de tres dígitos) y clase (código de cuatro dígitos).

¹² Las preguntas "L7" y "L8" derivan de dos preguntas incluidas en los cuestionarios y usualmente consideradas para cualquier cuestionario sociodemográfico que consulte por educación, a saber, "¿Cuál es el nivel educacional más alto alcanzado por usted?" (L7 en la presente encuesta) y "¿Completó el nivel educacional anteriormente declarado?" (L8 en la presente encuesta) (INE, 2011).

personas que han completado o no cada nivel educacional (de acuerdo a los parámetros chilenos).

Con el insumo de las variables de nivel educacional del jefe de hogar, se construyó el grupo socioeconómico, mediante un método establecido en la Dirección de Estudios Sociales DESUC basado en las recomendaciones de la Asociación de Investigación de Mercado (AIM, 2012), pero con algunas variaciones, utilizando la cantidad de bienes del hogar para quienes no se encuentran ocupados (estudiantes, dueñas de casa, jubilados, otros).

Sin embargo, estas variables no fueron utilizadas en el informe final debido a que se considera, por la naturaleza del estudio, que la variable de segmentación sea solamente nivel educacional del jefe de hogar. De esta manera, las variables de nivel socioeconómico no se incluyen en la base final.

3.6. Tipología de Lectores

Para observar de manera sintética la multiplicidad de perfiles lectores, se elaboró una tipología que permitiera sintetizar la información considerando las siguientes dimensiones de medición del comportamiento lector:

- **Diversidad:** tipos de textos y materiales leídos por los entrevistados. Para ello se trabaja con los cuatro materiales consultados, a saber, libros, diarios, revistas e historietas o cómics.
- **Frecuencia:** intensidad lectora para cada tipo de texto o material.
- **Soporte:** lectura en formato digital e impreso de cada texto y material.

De esta manera, se trabajó con las siguientes variables¹³:

¹³ Todas las variables mencionadas fueron elaboradas a partir de la recodificación de variables de prevalencia, frecuencia y alfabetismo. Por ejemplo, para el caso de "Frecuencia de lectura de libros impresos", se utilizaron las variables P0 (respecto a si sabe leer o escribir), B1A (sobre lectura de libros impresos en los últimos 12 meses) y B5A (frecuencia de lectura de libros en formato impreso en los últimos 12 meses). A partir de ellas,

- Frecuencia de lectura de libros impresos.
- Frecuencia de lectura de libros digitales.
- Frecuencia de lectura de diarios impresos.
- Frecuencia de lectura de diarios digitales.
- Frecuencia de lectura de revistas impresas.
- Frecuencia de lectura de revistas digitales.
- Frecuencia de lectura de historietas y/o cómics impresos.
- Frecuencia de lectura de historietas y/o cómics digitales.

Las tipologías fueron elaboradas mediante análisis de *cluster* o conglomerados del tipo *two-step cluster* (análisis de dos pasos). Este se utiliza en caso de trabajar con bases de datos que posean un alto número de encuestas y funciona con una medida de distancia entre los casos. Para este procedimiento se prefiere que las variables categóricas tengan una distribución multinomial, sin embargo, el algoritmo es capaz de entregar resultados plausibles aun cuando no se cumplan estos supuestos¹⁴.

Bajo el procedimiento de *two-step cluster* se realizaron una serie de pruebas para la construcción de conglomerados, sin embargo, todos presentaron medidas de ajuste menos idóneas¹⁵.

se elaboraron variables que identificaban las siguientes frecuencias: "al menos una vez al día", "al menos una vez a la semana", "al menos una vez al mes", "menos de una vez al mes", "no lee ese material o texto / es analfabeto". La no respuesta (no sabe/no responde) es clasificada como *missing*, para que no constituya un valor válido en las tipologías.

¹⁴ Las variables categóricas refieren a aquellas que no poseen un orden de magnitud. La distribución multinomial es una generalización de la distribución binomial y puede tener más de dos resultados posibles. Un algoritmo corresponde a un proceso de operaciones ordenado que permite llegar a una solución.

¹⁵ La medida de silueta de cohesión y separación (que varía de -1 a +1) no superó el valor 0,2; esto significa que los *clusters* no son categorías totalmente cohesionadas. Solo cuando se excluyen casos atípicos se logra aumentar la medida de ajuste; sin embargo, esto provoca una pérdida de casi el 50 % de los casos de la muestra, lo que no resulta recomendable.

Referencias

INE. (2011). *Propuesta de variables educacionales a incluir en cuestionarios sociodemográficos del INE*. Santiago: Instituto Nacional de Estadísticas.

Naciones Unidas. (2009). *Clasificación Industrial Internacional Uniforme de todas las actividades económicas (CIIU)*. Revisión 4. Nueva York: Naciones Unidas. Departamento de Asuntos Económicos y Sociales.

OIT. (2008). *Resolución sobre la actualización de la Clasificación Internacional Uniforme de Ocupaciones*. Recuperado el 1 de Septiembre de 2014, de <http://www.ilo.org/public/spanish/bureau/stat/isco/docs/resol08.pdf>.

OIT. (2013). *19° Conferencia Internacional de Estadísticos del Trabajo*. Ginebra: Oficina Internacional del Trabajo.