

**UNIVERSIDAD DE LOS ANDES
FACULTAD DE ECONOMIA
CENTRO DE ESTUDIOS SOBRE DESARROLLO ECONOMICO – CEDE**



**ENCUESTA LONGITUDINAL COLOMBIANA
DE LA UNIVERSIDAD DE LOS ANDES – ELCA**

**GUÍA DEL USUARIO
BASES DE DATOS ELCA 2013**

ENERO DE 2016

Tabla de contenido

1.	Estructura de archivos	3
1.1.	Bases de datos de la encuesta 2013	5
2.	Descripción de la documentación.....	5
2.1.	Cuestionarios	5
2.2.	Diccionario de datos	6
2.3.	Manual integrado de conceptos, codificación y captura de datos formularios urbano y rural primer seguimiento 2013.....	6
3.	Identificadores	6
3.1.	Nivel de observación hogar	6
3.2.	Nivel de observación persona.....	7
3.3.	Nivel de observación comunidad	7
3.4.	Identificación de los hogares panel	7
4.	Otros datos.....	10
4.1.	Factores de expansión.....	10
4.2.	Codificación de valores faltantes.....	10
5.	Migración	10
6.	Confidencialidad.....	11

1. Estructura de archivos

Las bases de datos de la Encuesta Longitudinal Colombiana de la Universidad de los Andes (ELCA) se entrega a los usuarios dividida según la unidad de análisis (Ver tablas 1 y 2). Se recomienda a los usuarios revisar conjuntamente los cuestionarios, las bases de datos, los diccionarios de datos y los manuales de conceptos y codificación al analizar las variables y su construcción.

Tabla 1. Bases de Datos Zona Urbana

Zona	Unidad de Análisis	Bases	Capítulos por módulo incluidos en cada base de datos	
Urbano	Hogares	Hogar	Capítulos I, III, V y IV (Preguntas 406-407) del Módulo 1. Hogar.	
		Gastos del Hogar	Capítulo IV (Excepto preguntas 406-407) del Módulo 1. Hogar.	
		Choques	Capítulo VI del Módulo 1. Hogar.	
	Personas	Personas		Capítulo II del Módulo 1. Hogar
				Capítulos I-IX del Módulo 2. Jefe del Hogar y Cónyuge.
				Capítulos V (Preguntas 501-509), VI, VIII y IX del Módulo 3. Niños(as) de 0 a 13 años de seguimiento.
				Capítulos I-III del Módulo 4. Personas de Contexto (Personas de 0 a 64 años, que no sean de seguimiento).
				Capítulos I-IV del Módulo 5. Jóvenes (Personas de seguimiento de 10 a 13 años).
				Capítulos I-IV del Módulo 6. Módulo de Política (Jefe de hogar o Cónyuge según selección).
		Niños (Restringida)	Capítulos I, II, III, IV, V (Preguntas 510-514) y VII del Módulo 3. Niños(as) de 0 a 13 años de seguimiento. Resultados de las prueba antropométrica, cognitiva y socioemocional.	

	Comunidad	Comunidad	Capítulos I-IV del Formulario de Comunidad Urbana.
--	-----------	-----------	--

Tabla 2. Bases de Datos Zona Rural

Zona	Unidad de Análisis	Bases	Módulos	
Rural	Hogares	Hogar	Capítulos I, III, V y IV (Preguntas 406-407) del Módulo 1. Hogar.	
		Gastos del Hogar	Capítulo IV (Excepto preguntas 406-407) del Módulo 1. Hogar.	
		Choques	Capítulo VI del Módulo 1. Hogar.	
		Activos Productivos	Capítulo VI del Módulo 7. Unidad de Producción Agropecuaria – UPA.	
		Producción (Restringida)	Capítulos IV y V del Módulo 7. Unidad de Producción Agropecuaria – UPA.	
	Personas	Personas		Capítulo II del Módulo 1. Hogar
				Capítulos I-IX del Módulo 2. Jefe del Hogar y Cónyuge.
				Capítulos V (Preguntas 501-509), VI, VIII y IX del Módulo 3. Niños(as) de 0 a 13 años de seguimiento.
				Capítulos I-III del Módulo 4. Personas de Contexto (Personas de 0 a 64 años, que no sean de seguimiento).
				Capítulos I-IV del Módulo 5. Jóvenes (Personas de seguimiento de 10 a 13 años).
				Capítulos I-IV del Módulo 6. Módulo de Política (Jefe de hogar o Cónyuge según selección).
		Niños (Restringida)	Capítulos I, II, III, IV, V (Preguntas 510-514) y VII del Módulo 3. Niños(as) de 0 a 13 años de seguimiento. Resultados de la prueba antropométrica, cognitiva y socioemocional.	

	Comunidad	Comunidad	Capítulos I-VII del Formulario de Comunidad Rural.
	Predio	Tierras (Restringida)	Capítulos A-D, II y III del Módulo 7. Unidad de Producción Agropecuaria – UPA.

1.1. Bases de datos de la encuesta 2013

Las bases de datos de la encuesta están disponibles en formato STATA versión 9. Para acceder a las bases de datos se debe registrar en la sección Bases de Datos de la página de la encuesta (encuestalongitudinal.uniandes.edu.co), luego podrán descargar las bases de datos.

2. Descripción de la documentación

En adición a las bases disponibles, los usuarios también tienen acceso a los cuestionarios, los diccionarios de datos y el manual integrado de conceptos, codificación y captura de datos. Dichos documentos los pueden encontrar en la sección Documentación de la página de la encuesta. Se recomienda que al analizar los datos se usen todos los documentos para entender mucho mejor la encuesta y la construcción de cada una de las variables.

2.1. Cuestionarios

Los cuestionarios contienen las preguntas que se realizaron en cada uno de los módulos de la encuesta. En la página pueden encontrar los cuestionarios del hogar, jefe y cónyuge, niños de seguimiento, personas de contexto, jóvenes de seguimiento, módulo de política y comunidad para ambas zonas. Además, para la zona rural, se encuentra el cuestionario de la Unidad de Producción Agropecuaria¹. Es importante resaltar que los cuestionarios no son iguales para ambas zonas de la encuesta, puede haber diferencias en algunas preguntas o en algunos módulos, dependiendo de la información que se busca recolectar en cada zona y sus particularidades.

¹ Ver Módulo 7. Unidad de Producción Agropecuaria – UPA.

2.2. Diccionario de datos

Los diccionarios muestran el nombre de las variables que hay en las bases de datos y los valores que toma cada una de ellas. Además, permiten saber la forma en la que están codificados los valores faltantes para cada una de las variables. Por otro lado, permiten identificar a que pregunta de la encuesta corresponde cada una de las variables y el módulo en el que se encuentra dicha pregunta. Es necesario que al analizar las bases de datos tengan en cuenta la descripción del diccionario junto a los cuestionarios. Esto es importante para entender la construcción de la encuesta y de los datos. Por ejemplo, al analizar los datos cuando hay un salto en las preguntas.

2.3. Manual integrado de conceptos, codificación y captura de datos formularios urbano y rural primer seguimiento 2013

El manual permite entender mejor la forma en la que se realizó cada una de las preguntas y contiene una explicación de las opciones o conceptos de algunas preguntas. También sirve para conocer la codificación de los valores de no sabe y no responde en las preguntas que tienen dicha opción.

3. Identificadores

Para poder desarrollar el pegado de los distintos módulos de la ELCA, con el objetivo de consolidar una base que agrupe la información que requiera el investigador, es necesario tener claridad sobre cuál es el identificador correcto de unidad de observación de cada módulo.

3.1. Nivel de observación hogar

Para las bases de hogar, gastos del hogar, choques y activos, como se mencionó, la unidad de observación es el hogar, por lo tanto, para poder desarrollar una base consolidada que agrupe estos tres módulos es necesario pegarlas según el identificador de hogar de 2013, que se denomina *llave*. En algunos casos la información está organizada en forma rectangular (*long*) por la forma en que se recolecta la información. Por ejemplo choques, gastos y activos de hogares rurales.

3.2. Nivel de observación persona

En el módulo de personas se tiene el identificador del individuo (*llaveper*) y el del hogar al que pertenece (*llave*): como varios individuos pueden pertenecer al mismo hogar (por ejemplo el jefe del hogar y cónyuge), se pueden tener varios identificadores de individuo distintos para un mismo hogar. Ahora, si el investigador está interesado en consolidar la información de personas con las de los módulos de hogar, la variable fundamental para desarrollar el pegado de las bases es el identificador del hogar *llave*, y como se mencionó, dado que varios individuos pueden pertenecer a un mismo hogar, se pegará la misma información del hogar para aquellas personas que tengan una misma *llave*.

3.3. Nivel de observación comunidad

Adicionalmente, si quiere pegarse la información del módulo de comunidades al de hogares, dado que varios hogares viven en una misma comunidad o barrio, la variable por la que se desarrolla el pegado de las bases debe ser el identificador de la comunidad: *consecutivo_c*. En este caso se pegará la misma información para los hogares que vivan en la misma comunidad. Es importante tener en cuenta que es posible que algunos hogares no tengan información de comunidad en 2013 porque para el desarrollo de la encuesta de comunidades debía haber mínimo 5 hogares en el barrio o vereda para que se realizara dicho módulo. En las zonas en las que había menos de 5 hogares, no se realizó el cuestionario de comunidades.

3.4. Identificación de los hogares panel

La fortaleza de una encuesta longitudinal como la ELCA consiste en poder observar cómo cambian a lo largo del tiempo las características del universo de observación elegido. En particular, para la ELCA es interesante observar qué tanto cambian las características de hogares y personas.

Como los hogares cambian su composición a lo largo del tiempo, debido a que los hijos en un hogar crecen y deciden formar su propio hogar, o a una separación del jefe del hogar y el cónyuge, el identificador del hogar necesariamente cambia a lo largo del tiempo. Queda a decisión del investigador si quiere pegar exclusivamente ciertos hogares (por ejemplo identificando aquellos que se separaron a lo largo del tiempo).

Para construir el panel de datos teniendo el hogar como unidad de observación se debe utilizar la variable *consecutivo*, dado que dicha variable es la que permite identificar el hogar en 2010 y 2013. Al momento de construir el panel se debe tener en cuenta que los hogares pueden dividirse y migrar de zona.

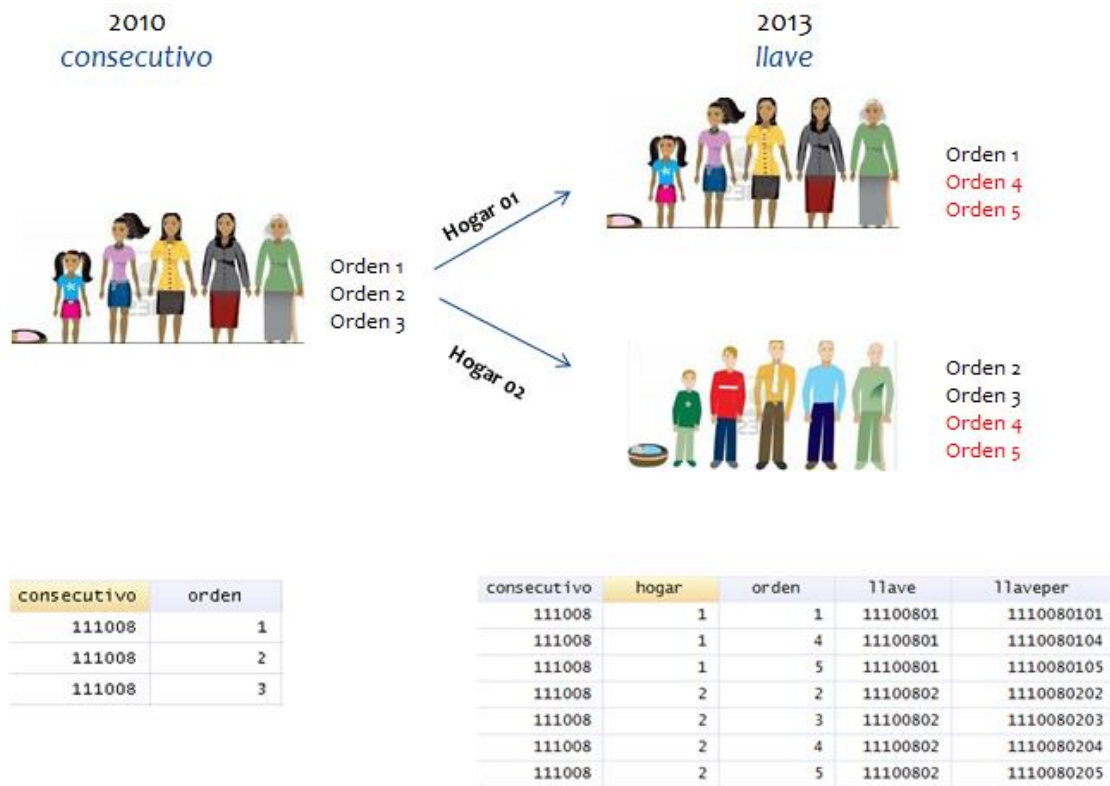
Hogares que no se dividen: En la figura 1, se puede observar un hogar que no se dividió entre 2010 y 2013. En 2013 se incluye la variable *hogar* que toma el valor 1 para todos los hogares que no se dividen puesto que el jefe del hogar no se separa, es decir, la variable *hogar* toma el *orden* del jefe de hogar en 2010 que corresponde a 1. Este hogar, aunque no se divide, tiene en 2013 nuevas personas de contexto (orden 4 y 5).

Figura 1. Hogar que no se divide



Hogares que se dividen: Para los hogares que se dividen se debe usar el identificador de *llave* que está compuesto por *consecutivo* y *hogar*. En este caso, se debe tener en cuenta que la variable *consecutivo* es igual a la de 2010. En la Figura 2 se puede observar un hogar que se dividió. En el primer hogar se mantuvo la persona de seguimiento con *orden* 1 en 2010, mientras que en el segundo hogar se mantuvieron las personas de seguimiento con *orden* 2 y 3 en 2010.

Figura 2. Hogar que se divide



La variable llave, como se puede observar, es diferente para cada uno de los hogares pues cambia el valor que toma la variable *hogar*. En nuestro ejemplo, se observa que en el primer hogar se mantuvo la persona de seguimiento con *orden 1* en 2010, mientras que en el segundo hogar se mantuvieron las personas de seguimiento con *orden 2* y *3* en 2010. Es importante tener en cuenta que la variable *hogar* toma el *orden* de las personas de seguimiento. En el primer hogar, por consiguiente, la variable *hogar* toma el valor de 1. En el segundo hogar, en donde hay más de una persona de seguimiento, dicha variable toma el **menor** valor del *orden* de las personas de seguimiento que se mantuvieron en el hogar, en este caso *orden 2*.

Para una construcción más sencilla del panel de datos se debe utilizar como unidad de observación al individuo, pegando las bases de personas de 2010 con las de 2013, por medio del identificador *llave_id_lb*. Es importante resaltar que solo aquellas personas que estaban en el 2010 tienen *llave_id_lb*, es decir, si la persona es nueva en el 2013 no tendrá dicho identificador y no se puede armar el panel para dichas personas.

4. Otros datos

Por otro lado, es importante tener en cuenta algunos temas sobre el funcionamiento de la encuesta: factores de expansión, migración de los hogares y codificación de los valores faltantes.

4.1. Factores de expansión

Cada una de las observaciones de la muestra de la ELCA tiene un peso diferente dentro de la misma. Por lo tanto, al realizar análisis de estadísticas descriptivas como la media, varianza, o las frecuencias, se recomienda tener en cuenta el peso que tiene cada una de las observaciones de la muestra. Para ello, se deben usar los factores de expansión que están dados por *fexpers* para las personas y *fexhog* para los hogares. Dichas variables identifican el peso que tiene cada persona o cada hogar dentro de la muestra.

4.2. Codificación de valores faltantes

En algunos casos, no es posible obtener la información de los hogares por falta de conocimiento de la misma o el hogar decide no responder una pregunta. En estos casos, las respuestas se encuentran codificadas como “No sabe” o “No responde/No informa”. En los datos dichas respuestas se encuentran codificadas con varios valores: 88, 99, 888 o 999, para conocer el valor que toman dichas respuestas en las preguntas es necesario revisar los diccionarios de cada uno de los módulos.

5. Migración

Por el carácter longitudinal de la ELCA, se debe tener en cuenta que es posible que los hogares migren por diversas razones: empleo, educación o violencia, entre otras. Esta migración se puede dar dentro del mismo municipio (ej. migrar de la zona rural a la cabecera municipal) o entre municipios (pueden ser del mismo departamento o no).

Es importante tener en cuenta esto cuando los hogares cambian de zona urbana a rural y viceversa, pues la información que se recoge en las dos rondas puede ser diferente dado que los cuestionarios de la zona urbana y la zona rural tienen algunas diferencias. Por lo tanto, si se desea utilizar un panel de hogares en donde la información que se recoge en las dos rondas sea la misma, se deben eliminar los hogares que cambian de zona.

6. Confidencialidad

Debido a que la principal prioridad de una encuesta longitudinal consiste en el seguimiento de los hogares y personas a lo largo del tiempo, se debe procurar un cuidado excesivo por la confidencialidad de los datos para que los individuos y familias, que muy amablemente sacrifican varias horas de su tiempo para la realización de la encuesta, no sientan violada la privacidad de su información ni sientan que los datos que ofrecen puedan ponerlos en peligro.

Es por esto que en la información georreferenciada de la localización, los municipios y los departamentos de residencia de los hogares y la información personal y de contacto de los individuos es confidencial. No obstante, se incluyen unos identificadores falsos del departamento (*id_dpto*) y el municipio (*id_mpio*) para que se puedan realizar análisis con dichas variables sin tener que acceder a los datos restringidos.

Por otra parte, los módulos de Tierras, Producción Agrícola y Niños no son de libre circulación y por tanto, no se encuentran en la página web de la ELCA. Para acceder a estas bases de datos es necesario ser profesor de planta o estudiante de la Facultad de Economía de la Universidad de los Andes. En el segundo caso, el estudiante debe tener la aprobación de un profesor de planta de la Facultad de Economía. Los estudiantes y el profesor deberán firmar un acuerdo de confidencialidad para acceder a los datos, y se deben comprometer a usarlos únicamente para el desarrollo de la investigación para la cual solicitan los datos y no de otras investigaciones.