

SEMINARIO DE POBLACIÓN Y SALUD I

24 de mayo 2011
Sesión 3 Parte 2

Cecilia Gayet
Claudio Dávila

Encuestas DHS: revisión del cuestionario e introducción a
una base de país

DHS Colombia 2010

[http://www.measuredhs.com/help/D
atasets/index.htm](http://www.measuredhs.com/help/Datasets/index.htm)

Pasos para análisis de encuestas.

- Pasos principales para analizar una encuesta son:
 - Revisar el documento metodológico de la encuesta
 - Objetivos de la encuesta
 - Representatividad
 - Tipo de muestreo (aleatorio simple, por conglomerados, estratificado, una etapa o varias, etc.)
 - Tamaño de la muestra ([pdf](#))
 - Número de entrevistas y su cobertura o tasa de respuesta
 - Necesita ponderadores
 - Factores de expansión
 - Datos faltantes (ver cómo los trataron)

Pasos para análisis de encuestas.

- Revisar el cuestionario o los cuestionarios
 - Ver qué módulos tiene
- Buscar las preguntas del cuestionario en la base de datos o en las bases de datos
- Cómo fueron operacionalizadas las respuestas
- Análisis exploratorio de datos (frecuencias iniciales de ciertas variables como sexo, edad, estado civil o marital, etc.)

Objetivo.

- El objetivo general de la ENDS 2010 es obtener información actualizada sobre aspectos demográficos y de salud de la población, relacionados con los objetivos de la política nacional de salud sexual y reproductiva.

Población de estudio.

- Cobertura nacional con representatividad urbana y rural, seis regiones (Caribe, Oriental, Central, Bogotá, Pacífica, y Amazonía-Orinoquía), 16 subregiones y para cada uno de los departamentos de Colombia en forma independiente.

Cuadro de las regiones y subregiones de la ENDS 2010, Colombia

Cuadro 1.1 Regiones y subregiones en 1990, 1995, 2000, 2005 y 2010

Departamentos y ciudades en la clasificación de regiones y subregiones en 1990, 1995, 2000, 2005 y 2010.

Región	Subregiones en 2005 y 2010	Subregiones en 1995 y 2000	Subregiones en 1990
Caribe	Guajira, Cesar, Magdalena Barranquilla (Área Metropolitana), Atlántico sin Barranquilla, San Andrés, Bolívar Norte Bolívar Resto, Sucre, Córdoba	Idem Atlántico, Bolívar Norte Idem	Idem Atlántico Bolívar, Sucre, Córdoba
Oriental	Santanderes Boyacá, Cundinamarca, Meta	Idem Idem	Idem Idem
Bogotá	Bogotá (sin Soacha)	Bogotá (con Soacha)	Idem
Central	Medellín (Área Metropolitana) Antioquia sin Medellín Caldas, Risaralda, Quindío Tolima, Huila, Caquetá	Idem Idem Idem Idem	Idem Idem Idem Idem
Pacífica	Cali (Área Metropolitana) Valle del Cauca (sin Cali y sin Zona Costera) Cauca y Nariño (sin la Costa Pacífica) Litoral Pacífico: Chocó; Zona Costera del Valle, Cauca y Nariño	Idem Idem Idem Idem	Idem Valle del Cauca (sin Cali) Chocó, Cauca, Nariño na
Orinoquía y Amazonía	Cabeceras de Arauca, Casanare, Guainía, Vichada, Amazonas, Putumayo, Guaviare, Vaupés	na	Idem

Nota: En la presentación de resultados en 1990, Cali y Medellín no se mostraron por separado.

na: No aplica

Población de estudio.

- Primero: se recopiló información sobre viviendas y hogares en lo referente a su infraestructura, composición y características específicas de cada uno de sus miembros.
- Segundo: información completa de las mujeres en edad fértil (13 a 49 años de edad), su esposo o compañero y c/u de los hijos menores de cinco años.

Población de estudio.

- Tercero: se obtuvieron datos sobre citología cérvico-uterina y mamografía en mujeres de 18 a 69 años de edad, que hubieran tenido actividad sexual.
- Cuarto: información sobre la salud de las personas de 60 años y más.
- Quinto: medidas antropométricas (peso y talla) de todos los menores de 5 años y sus madres.

Muestra.

- Universo de estudio constituido por el 99% de la población civil residente en hogares particulares de las zonas urbana y rural del país.
- La muestra para la ENDS 2010 es probabilística, de conglomerados, estratificada y polietápica. La selección en cada una de las etapas fue estrictamente aleatoria.

Muestra.

- Se entrevistaron:
 - 51,447 hogares de 258 municipios con una cobertura de 92%;
 - 53,521 mujeres en edad fértil (13-49 años) cobertura de 94%;
 - 13,887 mujeres entre 50 y 69 años con una cobertura de 94%;
 - 162,331 personas de 0 a 64 años con una cobertura del 85%, a los cuales se les tomaron medidas antropométricas;
 - 17,574 adultos de 60+ con una cobertura del 92%.

Hogares y personas en la muestra y la tasa de respuesta ENDS 2010.

Cuadro 1.2 Hogares y personas en la muestra de la ENDS 2010

Número de hogares, número de entrevistas y tasas de respuesta, según zona y región de residencia, Colombia 2010

Característica	Zona		Región						Total
	Urbana	Rural	Caribe	Oriental	Bogotá	Central	Pacífica	Orinoquía y Amazonía	
Entrevistas de hogar									
Hogares seleccionados	45,444	18,745	12,312	10,472	5,037	16,497	9,365	10,506	64,189
Hogares ocupados	40,384	15,754	11,013	9,404	4,482	14,332	8,187	8,720	56,138
Hogares entrevistados	36,412	15,035	10,276	8,480	3,604	13,145	7,528	8,414	51,447
Tasa de respuesta de hogares	90.2	95.4	93.3	90.2	80.4	91.7	92.0	96.5	91.6
Entrevistas de mujeres 13-49 años									
Mujeres elegibles	41,488	15,398	12,239	9,101	4,054	13,917	8,172	9,403	56,886
Mujeres entrevistadas	38,885	14,636	11,474	8,407	3,690	13,096	7,737	9,117	53,521
Tasa de respuesta	93.7	95.1	93.7	92.4	91.0	94.1	94.7	97.0	94.1
Entrevistas de mujeres 50-69 años									
Mujeres elegibles para citología	11,010	3,835	3,123	2,400	1,076	4,266	2,322	1,658	14,845
Mujeres entrevistadas para citología	10,281	3,606	2,893	2,235	977	4,033	2,156	1,593	13,887
Tasa de respuesta	93.4	94.0	92.6	93.1	90.8	94.5	92.9	96.1	93.5
Antropometría (0-64 años)									
Número de personas elegibles	131,215	59,692	42,300	30,461	12,170	45,537	26,738	33,701	190,907
Número de personas medidas	109,798	52,533	36,835	23,676	9,459	38,831	23,483	30,047	162,331
Tasa de respuesta antropometría	83.7	88.0	87.1	77.7	77.7	85.3	87.8	89.2	85.0
Adulto Mayor (60 años y más)									
Personas elegibles	13,275	5,851	4,140	3,297	1,235	5,403	3,031	2,020	19,126
Personas entrevistadas	12,105	5,469	3,640	3,054	1,085	5,050	2,810	1,935	17,574
Tasa de respuesta	91.2	93.5	87.9	92.6	87.9	93.5	92.7	95.8	91.9 ₁₂

Ponderadores.

- Son factores de ajuste aplicados para cada caso, con la finalidad de ajustar por diferencias en la probabilidad de selección y entrevista entre casos en una muestra.
- La muestra se elige con una probabilidad desigual para expandir el número de casos disponibles (y por tanto reducir la variabilidad) para ciertas áreas o subgrupos para los cuales las estadísticas son necesarias.

Ponderadores.

- Existen dos tipos principales de ponderadores en las DHS: ponderadores de hogares (HV005) y de individuos (V005).
- También pueden haber ponderadores muestrales para subconjuntos de la muestra, tales como en el módulo de hombres o en el de violencia doméstica (D005).
- Son calculados utilizando las probabilidades de selección muestral para cada hogar y las tasas de respuesta para los hogares y para individuos.

Variables encontradas en las distintas bases de datos por módulo.

Variables in the recode file begin with one or two letters followed by one, two, or three digits and in some cases followed by a letter. Following is a list describing the general variable name conventions.

- HVxxx - Household standard variables
 - HAxX - Anthropometry and anemia for women
 - HCxx - Anthropometry and anemia for children
 - SHxxx - Household, country-specific.
- Vxxx - Women standard variables
 - Bxx - Birth history
 - Mxx - Pregnancy, postnatal care, and breastfeeding
 - Hxx - Immunization and health
 - HWxx - Anthropometry for children of eligible women
 - MMxx - Maternal mortality (optional)
 - DVxx - Domestic violence (optional)
 - Sxxx - Women, country-specific
- MVxxx - Men standard variables
 - SMxx - Men, country-specific

Problemas con los datos.

- Las mayores fuentes de problemas: reporte parcial o incompleto de información y respuestas inconsistentes a diferentes preguntas.
- Para datos faltantes (missing) hay tres soluciones:
 - Dejar la pregunta en blanco
 - Asignar un código especial a la pregunta indicando que no hubo respuesta en el cuestionario
 - Deducir una respuesta.

Problemas con los datos.

- Para respuestas inconsistentes hay tres posibilidades:
 - Dejar los datos como están
 - Asignar un código especial a la pregunta indicando que la respuesta fue inconsistente con otra información reportada
 - Deducir una respuesta.

Problemas con los datos.

- Imputar información es el proceso de atribuir una característica a un caso basándose en características conocidas de la población en general y en otras características particulares del caso en cuestión.
- Hay dos pasos:
 - se usa un conjunto de reglas para restringir las posibles respuestas que se pueden atribuir a cada caso
 - el método de elección de respuesta se aplica cuando hay mayor posibilidad de reducir el conjunto de respuestas aceptables.

Valores missing y otros códigos.

Missing Values and Other Special Codes

In DHS, a missing value is defined as a variable that should have a response, but because of interview errors the question was not asked. For example, questions 111 and 112 must be answered by women with no education or only primary school education. If the interviewer incorrectly marked question 110 as secondary school or higher when in fact the respondent only completed primary school, questions 111 and 112 do not get asked. The general rule for DHS data processing is that “UNDER NO CIRCUMSTANCES AN ANSWER SHOULD BE MADE UP;” instead a missing value will be assigned. The data will be kept as missing in the data file and no imputation for this kind of questions is assigned. As will be discussed below, only the seven dates related to women’s chronological events are imputed.

Missing values in general are codes 9, 99, 999, 9999, etc. depending on the variable number of digits. There are however some important background variables where the missing code is not accepted:

- Geographical variables such as Urban/rural (HV025, V025, MV025, V102, MV102), regions (HV024, V024, V101, MV101), and in general any other variable whose value can be established by the sample design
- Level of education for women and men in the individual questionnaire (V106, MV106)
- Current use of contraception for women (V312)
- Current marital status of women (V501)
- Variables related to the woman’s birth history (V201 to V210, B0, B4, B5, B9).

It is possible that earlier phases of DHS may have missing values in some of these variables.

Another special code used for data cleaning purposes is code “Inconsistent.” This code is generally used by people in the secondary editing group, when a value or code is not plausible. For example, dates for vaccinations recorded as having occurred before the birth of the child. The value is not missing but is not possible. The secondary editing group is instructed to find out what piece of information is wrong (day, month, or year) and assign code “Inconsistent” to it. This of course is done after checking for clues that could lead to correct the problem.

Inconsistent codes are 7, 97, 997, etc., depending on the variable number of digits.

Codes 8, 98, 998 are assigned to “don’t know” responses. These codes are normally pre-coded in the questionnaires, but they are consistently used throughout the recode file.

Bases de datos.

- Hay 5 bases de datos de la ENDS Colombia 2010:
 - Hogares (cohr60): tiene una observación por hogar. Incluye información de quién pertenece al hogar, pero no tiene información de los cuestionarios individuales. La unidad de análisis es el hogar.
 - Nacimientos (cobr60): cada observación por cada hijo alguna vez nacido por mujer elegible. En esencia es la historia completa de nacimientos de las mujeres entrevistadas. Puede usarse para calcular indicadores de salud, así como tasas de mortalidad y fecundidad. La unidad de análisis son los hijos nacidos de las mujeres elegibles.

Bases de datos.

- Individuos (coir60): observación por mujer elegible. Contiene la información del cuestionario individual. Los programas de mortalidad y fecundidad de la DHS utilizan estos datos. La unidad de análisis es la mujer.
- Miembros del hogar (copr60): observación por cada miembro del hogar. Incluye variables como sexo, edad, escolaridad, talla y peso, etc. También incluye las características del hogar en el que vive el individuo. La unidad de análisis es el miembro del hogar.

Bases de datos.

- Niños (cokr60): observación por cada niño nacido en los últimos 5 años de las mujeres elegibles. Información relacionada con el embarazo del niño, cuidados postnatales, vacunación y salud. También incluye información de la madre. Se utiliza para calcular indicadores de salud como cobertura de vacunación, reciente ocurrencia de diarrea, fiebre o tos en los niños pequeños y tratamiento de enfermedades de la niñez. La unidad de análisis son los niños de mujeres elegibles nacidos en los últimos 5 años.

Cuestionarios.

- Se utilizaron 5 cuestionarios en la ENDS 2010:
 - Cuestionario de hogar para ser aplicado en cada hogar visitado (página 708 [pdf](#) resumen resultados).
 - Cuestionario individual para mujeres en edad fértil (13-49 años).
 - Cuestionario sobre prevención de cáncer de cuello uterino y mama (mujeres 18-69 años).
 - Cuestionario de adulto mayor, para hombres y mujeres de 60 años y más.
 - Cuestionario de peso y talla, para niños menores de cinco años y sus madres.

Módulos.

- En la base de hogares se tienen:
 - Variables de identificación
 - Datos de control
 - Personas del hogar (características básicas)
 - Tipo de familia
 - Condiciones habitacionales del hogar
 - Problemas de salud
 - Hospitalizados

Módulos.

- En la base de individuos se tienen:
 - Antecedentes de la entrevistada
 - Reproducción
 - Anticoncepción
 - Embarazo, parto y puerperio
 - Lactancia y alimentación complementaria
 - Inmunización y salud
 - Nupcialidad y comportamiento sexual
 - Preferencias de fecundidad
 - Antecedentes del marido y trabajo de la mujer
 - Prevención del cáncer
 - SIDA y otras I.T.S.
 - Violencia doméstica (página 772)

Módulos.

- Para adultos mayores se tienen:
 - Ingresos y dependencia económica
 - Vínculo familiar
 - Discapacidad
 - Salud Mental - Depresión
 - Estado de salud
 - Accesibilidad a los servicios de salud