

**Metodología de diseño muestral:  
Encuesta de Demografía y Salud 2016**

**INSTITUTO NACIONAL DE ESTADISTICA  
UNIDAD DE MUESTREO Y LOGÍSTICA**

**LA PAZ – BOLIVIA**

**2016**

## ÍNDICE

- 1. Introducción, 1**
- 2. Antecedentes, 1**
- 3. Objetivos del diseño muestral, 2**
  - 3.1. Objetivo general, 2**
  - 3.2. Objetivos específicos, 2**
- 4. Diseño muestral, 2**
  - 4.01. Universo de estudio, 3**
  - 4.02. Ámbito geográfico, 3**
  - 4.03. Periodo de levantamiento de los datos, 3**
  - 4.04. Unidades de la encuesta, 3**
  - 4.05. Características del marco muestral, 3**
  - 4.06. Tipo de muestreo, 6**
  - 4.07. Dominios de estudio, 6**
  - 4.08. Nivel de desagregación de la información, 7**
  - 4.09. Selección de la muestra, 7**
  - 4.10. Constitución y distribución del tamaño muestral, 8**
  - 4.11. Cálculo del tamaño muestral, 8**
  - 4.12. Distribución de la muestra dentro de los dominios, 12**
- 5. Probabilidades y ponderaciones de muestreo, 15**
  - 5.1. Ponderaciones de diseño y ponderaciones de muestreo, 15**
  - 5.2. Ponderaciones de diseño, 15**
  - 5.3. Corrección por falta de respuesta por unidad y cálculo de las ponderaciones de muestreo, 17**
    - 5.3.1. Ponderación de muestreo para viviendas, 18**
    - 5.3.2. Ponderación de muestreo para individuos, 19**
  - 5.4. Corrección por proyección de población, 19**
- 6. Bibliografía, 20**

## 1. Introducción

La *Encuesta de Demografía y Salud (EDSA) 2016*, que es llevada a cabo por el *INE* en conjunto con el *Ministerio de Salud*, tiene como principal objetivo la recopilación y análisis de información estadística oportuna, actualizada del país y diversas áreas geográficas para los diversos indicadores de salud y nutrición. Un segundo objetivo es el permitir un análisis de tendencia, esto es, los niveles de los diferentes indicadores obtenidos con la *EDSA-2016* puedan compararse con los niveles de los mismos indicadores estimados a partir de las anteriores *ENDSA*. Un tercer objetivo es proporcionar información comparable a nivel internacional en los temas de salud, nutrición, mortalidad, fecundidad, regulación de la fecundidad, ITS y VIH/SIDA, entre otros temas. De esta manera disponer de información de base que permita medir el estado de situación de la población en las variables investigadas, para el diseño y evaluación de políticas en salud a cargo del Ministerio de Salud y a nivel inter-institucional.

En Bolivia se realizaron las Encuestas Nacionales de Demografía y Salud los años 1989, 1994, 1998, 2003 y 2008, proporcionando información acerca de la salud Materno – Infantil, planificación familiar, fecundidad, mortalidad materna (exceptuando 1998 y 2008), mortalidad infantil y en la niñez, lactancia, nutrición, enfermedades de transmisión sexual y violencia contra las mujeres y niños. Es así que la *EDSA-2016* es una de las principales fuentes de información para reportar sobre el progreso y cumplimiento de los *Objetivos de Desarrollo del Milenio (ODM)*; consecuentemente existe la necesidad de realizar nuevamente la *EDSA* en el año 2016.

El presente documento describe el diseño muestral para la *EDSA-2016*, como tal su propósito es precisar la metodología que se ha ido desarrollando en las predecesoras *ENDSA*. Es así que presenta aproximaciones y/o estimaciones para las tasas de respuesta, errores muestrales usados para la determinación de los tamaños de muestra, selección de la muestra y ponderaciones de diseño como muestrales.

## 2. Antecedentes

Las *Encuestas Nacionales de Demografía y Salud (ENDSA)* se han ido realizando en cinco ocasiones, años 1989, 1994, 1998, 2003 y 2008, las cuales se han enmarcado metodológicamente en el *Programa Measure DHS*, dentro de este marco son llamadas rondas. El diseño muestral para estas fue probabilística, estratificada, por conglomerados y multi-etápica, donde los tamaños de muestra en las distintas versiones, llegaron a fijarse alrededor de 20'000 viviendas con 1'000 UPM y 20 viviendas.

**Tabla 1.**

Bolivia: Tamaños de muestra calculados en cada ronda de la ENDSA, 1989 - 2008.			
Año	Tamaño de muestra calculado	Numero de UPM	Número de hogares por UPM
2008	20'000	1'000	20
2003	20'000	1'000	20
1998	13'000	821	16
1994	10'000	625	alrededor de 16
1989	10'000	690	

Fuente: Documentos de resultados de cada ENDSA 1989, 1994, 1998, 2003 y 2008.

El marco muestral utilizado para el diseño y selección de la muestra de las últimas dos *ENDSA*, 2003 y 2008, fue preparado con información y material del *Censo Nacional de Población y Vivienda* del año 2001.

Para la muestra de la última *ENDSA-2008* se llegó a emplear una metodología estandarizada en función a las experiencias de las *ENDSA* precedentes y fue diseñada para obtener estimaciones aceptables de los diversos indicadores a niveles de desagregación: total del país, total áreas de residencia (urbana y rural), total para regiones geográficas (altiplano, valle y llano) y para cada uno de los nueve departamentos (Chuquisaca, La Paz, Cochabamba, Oruro, Potosí, Tarija, Santa Cruz, Beni y Pando).

### **3. Objetivos del diseño muestral**

#### **3.1. Objetivo general**

Implementar un diseño muestral eficiente mediante una metodología adecuada y acorde con los lineamientos para el cumplimiento de los propósitos y objetivo de la *EDSA-2016* que son “*Obtener y proveer información sobre los niveles actuales de fecundidad, mortalidad, morbilidad, conocimiento uso y fuentes de obtención de métodos de planificación familiar, salud materno - infantil, nutrición, educación y vivienda*”<sup>1</sup>.

#### **3.2. Objetivos específicos**

- Establecer el tamaño de la muestra, así como su distribución;
- Definición del marco muestral;
- Implementación del método de selección de la muestra;
- Cálculo de probabilidades de selección y cálculo de los factores de expansión (ponderadores muestrales).

### **4. Diseño muestral**

Con el fin de lograr la comparabilidad, la conservación de la continuidad de los análisis, la coherencia, la mejor calidad en los resultados de la encuesta y la eficiencia, el diseño muestral para la *EDSA-2016* mantiene similitud a los diseños empleados en las

<sup>1</sup> Perfil del proyecto “*Encuesta de Demografía y Salud 2016*”.

precedentes *ENDSA*. Ergo, el cálculo del tamaño de muestra y su selección deberán ser adecuados, de modo que se enmarquen e los presupuestos planificados y lograr una buena precisión en los diferentes dominios.

#### **4.01. Universo de estudio**

La investigación está dirigida al conjunto de hogares establecidos en viviendas particulares ocupadas de los nueve departamentos del país del área urbana y rural donde residen **mujeres en edad fértil de 14 a 49 años de edad, menores de cinco años de edad y varones de 15 a 64 años de edad**. Se excluye de la investigación a la población residente en viviendas colectivas tales como hospitales, hoteles, cárceles, recintos militares, hogares de ancianos, unidades educativas, etc.; pero se incluyen a hogares que forman un grupo independiente y además residen dentro de estos establecimientos, ejemplo: porteros, cuidadores, etc.

#### **4.02. Ámbito geográfico**

La encuesta cubrirá el área urbana y rural de todos los nueve departamentos del país: Chuquisaca, La Paz, Cochabamba, Oruro, Potosí, Tarija, Santa Cruz, Beni y Pando.

#### **4.03. Periodo de levantamiento de los datos**

El levantamiento de información primaria corresponde al periodo del *3 de mayo* al *27 de septiembre de 2016*.

#### **4.04. Unidades de la encuesta**

- **Unidades de observación.-** El hogar establecido en viviendas particulares; mujeres en edad fértil de 14 a 49 años de edad; menores de cinco años de edad; y varones de 15 a 64 años de edad.
- **Unidades de información.-** El jefe de hogar<sup>2</sup> o informante calificado, quien debe tener por lo menos 12 años de edad.
- **Unidades de muestreo.-** Son las UPM constituidas en el *Marco Muestral Maestro 2012*; y las viviendas particulares ocupadas que pertenecen a estas UPM.
- **Unidades de la muestra.-** Son las UPM y viviendas seleccionadas para la muestra.
- **Unidades de análisis.-** Estos se constituirán en los hogares, y miembros del hogar para cuales se relevará información.

#### **4.05. Características del marco muestral**

La información del *Marco Muestral Maestro (MMM)* está basada en el *Censo de Población y Vivienda de 2012 (CNPV-2012)*, la *Actualización Cartográfica Multipropósito (ACM-*

---

<sup>2</sup> Persona reconocida como jefe o jefa por los miembros del hogar, que independientemente de su edad, sexo o estado civil, generalmente es la que toma las decisiones en el hogar.

2010-2012) y el *Censo Nacional Agropecuario de 2013 (CNA-2013)*. Por tanto el *MMM-2012* se describe como un marco de áreas y listas.

El *MMM-2012* cubre todo el territorio nacional no existiendo áreas de difícil acceso. Codifica las aéreas geográficas en unidades primarias de muestreo (UPM), constituyéndose bien delimitadas en el área rural dispersa; área rural amanzanada; área urbana y periferias (centros poblados que se encuentran en el área urbana y rural a la vez). Las unidades secundarias de muestreo (USM) están designas a las viviendas, cubriendo la totalidad de las viviendas particulares<sup>3</sup>.

La conformación de las Unidades Primarias de Muestreo (UPM), en el área urbana y rural, esta realizada de acuerdo con los límites geográficos y estadísticos con los siguientes criterios:

- **Agrupación.-** Consiste en la unión de dos o más manzanas o comunidades dentro del área urbana o rural para la conformación de una UPM.
- **División.-** Consiste en la división estricta de manzanas en el área urbana para la conformación de UPM, exclusivamente en las ciudades capitales de cada departamento.
- **Unicidad.-** Consiste en una relación de uno-a-uno estricta entre un manzano o comunidad con la UPM conformada.

Para realizar la agrupación, unicidad y división de manzanas (o comunidades) se utilizó el criterio del número de viviendas en la manzana (o comunidad) de acuerdo al siguiente detalle:

- **Manzanas con 0 a 11 viviendas.-** Son manzanas que generalmente están localizadas en las áreas de crecimiento urbano y áreas alejadas como periferias. Estas se unieron entre si y en algunos casos con otras manzanas que se encuentran en el rango de 12 a 120 viviendas para la generación de una sola UPM. Respecto a las comunidades el procedimiento de agrupación de UPM fue análogo.
- **Manzanas con 12 a 120 viviendas.-** Estas manzanas, según experiencia en operativo de campo, tienen una carga de trabajo ideal para una brigada de encuestadores. Respecto a las comunidades con estas cantidades de viviendas se aplicó el mismo criterio.
- **Manzanas con 121 a 240 viviendas.-** Estas manzanas fueron consideradas, por los técnicos regionales, para su verificación, actualización y respectiva división en trabajo de campo y gabinete, solo si era necesario. Respecto a las comunidades con estas cantidades de viviendas no se aplicó ni considero división alguna.
- **Manzanas con 241 viviendas o más.-** Estas manzanas fueron sujetas a división con colaboración de los técnicos de trabajo de campo. Para el caso de las comunidades con estas cantidades de viviendas también se aplicó una división.

---

<sup>3</sup> Vivienda particular es aquella que está habitada hasta por tres hogares, una vivienda con más de tres hogares es reconocida como vivienda colectiva (por ejemplo: hospitales, cuarteles, etc.)

De esta manera, además de la estratificación estadística, el *MMM-2012* mantiene similitud al anterior *MMM-2001* en la forma que implementa Unidades de Muestreo de Primera etapa (UPM) constituidas por: una manzana (mayormente) o conjunto de manzanas o en parte de una manzana (en algunos casos), esto en el área urbana; y una comunidad (mayormente) o parte de una comunidad (en algunos casos), esto en el área rural.

**Tabla 2.**

<b>Bolivia, MMM-2012: Distribución de las UPM según departamento y área</b>			
	<b>Área</b>	<b>Viviendas</b>	<b>UPMs</b>
	<b>Total</b>	<b>2'778'697</b>	<b>102'593</b>
<b>Bolivia</b>	<b>Urbana</b>	1'812'825	71'310
	<b>Rural</b>	965'872	31'283
Chuquisaca	Urbana	73'769	3'005
	Rural	74'648	2'357
La Paz	Urbana	540'218	20'611
	Rural	306'289	8'999
Cochabamba	Urbana	327'519	12'495
	Rural	183'405	5'546
Oruro	Urbana	90'490	4'370
	Rural	59'298	3'014
Potosí	Urbana	92'354	3'806
	Rural	147'138	4'713
Tarija	Urbana	81'569	3'346
	Rural	43'846	1'232
Santa Cruz	Urbana	525'582	19'797
	Rural	114'440	3'893
Beni	Urbana	67'965	3'258
	Rural	25'062	1'092
Pando	Urbana	13'359	622
	Rural	11'746	437

Fuente: INE/CNPV-2012

Así, las 102'593 UPM conformadas fueron estratificadas con el empleo de indicadores construidos sobre las siguientes dimensiones: *Características de la vivienda; Acceso a servicios básicos; Hacinamiento; Capital Humano: Educación; Patrimonio de los hogares; y Capacidad para la generación de ingresos.*

El método de estratificación utilizado tiene por objeto formar estratos con mínima varianza buscando agrupar a los elementos que más se parezcan entre sí o que estén más cercanos, siguiendo el criterio de similitud establecido y que a su vez permita diferenciar un estrato de otro. La técnica utilizada es una combinación entre el *Análisis Factorial* como primera instancia y el *Algoritmo de las k-medias* para el número y conformación de los estratos y el proceso de estratificación se realizó mediante el algoritmo numérico llamado *CCC (Cubic Cluster Criterion, criterio cúbico de agrupamiento)*, implementado en los paquetes estadísticos *R* y *SAS*, el cual permite que las agrupaciones generadas, de acuerdo con las dimensiones de indicadores, estén distribuidas de manera uniforme. Así, el número de estratos óptimo sugerido por el algoritmo es cuatro.

El marco muestral obtenido además se caracteriza como un marco muestral maestro el cuál se emplea para varias encuestas o varias rondas de una misma encuesta.

#### 4.06. Tipo de muestreo

Dado el marco muestral, la estratificación usada para la encuesta es incluida en cada uno de los departamentos, de esta manera se tienen dos tipos de estratificación para las UPM, por un lado una estratificación por área de residencia (urbana/rural) y por otro lado una estratificación estadística (cuatro niveles socio-económicos), conformándose esta última como una estratificación implícita.

Se determina el diseño de la muestra para la encuesta *EDSA-2016* en un muestreo probabilístico bi-etápico y estratificado por conglomerado en primera etapa. Las selecciones de primera etapa serán con Probabilidades Proporcionales al Tamaño (PPT) del número de viviendas; y en segunda etapa una muestra fija de viviendas con selección sistemática; esta última etapa es determinante para la ponderación de los conglomerados, y consiguientemente el cálculo de los errores. Adicionalmente, se emplea un sub-muestreo de viviendas para la recolección de información de todos los varones elegibles (y aplicación de la prueba de hemoglobina) en la vivienda seleccionada o sub-muestreada.

El diseño de la encuesta tiene las siguientes características:

**Probabilístico:** Porque las unidades de selección, UPM y viviendas, tienen una probabilidad conocida y distinta de cero de ser seleccionadas para formar parte de la muestra.

**Por conglomerados:** Debido a que las unidades objeto de estudio, viviendas, se encuentran agrupadas en áreas conformadas, con fines de muestreo y en base a la información y cartografía del CNA-2013, las cuales son las UPM (manzanas o comunidades).

**Estratificado:** Se agrupan las unidades de muestreo de primera etapa, UPM, en estratos lo cual permite mejorar la distribución de la muestra.

**Bi-etápico:** Porque se tiene dos etapas de selección, en la primera etapa se seleccionan las UPM (manzanas o comunidades) y en la segunda etapa se seleccionan las USM (viviendas) al interior de la UPM seleccionada.

**Sub-muestreo:** Consiste en seleccionar una fracción de entre las viviendas que son seleccionadas para la muestra por UPM para luego aplicar una encuesta a varones y realizar otras mediciones.

#### 4.07. Dominios de estudio

Un factor que determina los dominios de diseños es el tamaño de muestra total, ya que para cada dominio se deberá calcular un tamaño de muestra adecuado en función a indicadores básicos con un número mínimo de unidades de muestra para lograr una



precisión deseada, por tanto no es conveniente incrementar la cantidad de los dominios de estudio.

Los dominios de estudio se llegan a caracterizar como dominios de diseño<sup>4</sup>, los cuales son: los nueve (9) departamentos (Chuquisaca, La Paz, Cochabamba, Oruro, Potosí, Tarija, Santa Cruz, Beni y Pando); cada una de las tres (3) regiones naturales (altiplano, valle y llano); las dos (2) áreas de residencia (urbana y rural); y finalmente total del país.

#### **4.08. Nivel de desagregación de la información**

La precisión deseada para cada uno de los dominios, puede garantizarse para los indicadores de referencia con los que se calcula la muestra. Si se desea estudiar la información de manera desagregada para el resto de los indicadores o parámetros a estimar, se recomienda valorar la precisión sujeta a los errores muestrales, por tanto para el estudio de otras variables o indicadores es recomendable agregar la información mínimamente a nivel departamental.

#### **4.09. Selección de la muestra**

**Muestra de UPM:** Su selección se realiza de manera aleatoria e independiente con probabilidades proporcionales al tamaño (PPT), esto corresponde a la primera etapa bi-etápica.

**Muestra de viviendas (USM):** En la segunda etapa de la muestra bi-etápica, se seleccionan las viviendas enlistadas de manera sistemática y con arranque aleatorio, para el área urbana; y selección sistemática por corridas, para el área rural.

**Sub-muestra de viviendas:** Para la selección de estas viviendas se toma una fracción de las viviendas que fueron seleccionadas para la entrevista a mujeres, en la sub-muestra obtenida se aplican otras mediciones adicionales y especiales como la medición de hemoglobina a mujeres y niños/niñas y entrevistas a todos los varones elegibles de esa vivienda.

La unidad de selección de la muestra es la vivienda particular ocupada, al interior de la vivienda se identifica el hogar (o en algunos casos: hogares) y las personas miembros del hogar. La entrevista se realiza comenzando con un informante por hogar que corresponde generalmente al jefe de hogar, a través del informante se recolectan datos de la vivienda y de todos los miembros del hogar que son residentes habituales de la vivienda, incluyendo a aquellos que pasaron la noche anterior en la vivienda. En el caso de dos o tres hogares, dentro de la vivienda, se recolecta información de todos estos hogares.

---

<sup>4</sup> Un dominio de diseño consiste en una sub-población que puede ser identificado en el marco de la muestra y por lo tanto puede ser manejado de manera independiente en los procedimientos de muestreo y tamaño de la muestra, por lo general consiste de áreas geográficas o unidades administrativas.

#### **4.10. Constitución y distribución del tamaño muestral**

El tamaño de muestra para cada dominio de diseño (área de residencia), que coincide con el estrato geográfico urbano y rural, es calculado de manera separada con dos indicadores de aproximación. Luego la muestra de viviendas calculada para cada área de residencia es distribuida por departamento (los nueve departamentos) mediante una asignación potencial; posteriormente, con las asignaciones de viviendas obtenidas por departamento, se continua distribuyendo la muestra a nivel de estrato geográfico (el cual esta categorizado, también, en función a la cantidad de población), mediante una asignación proporcional; el siguiente paso es traducir estas cantidades de viviendas en términos de cantidades de UPM, las cuales son distribuidas a cada estrato socio-económico, esta asignación se distribuye de manera proporcional. La implementación de una estratificación socio-económica para cada UPM distingue al *MMM-2012*, ergo, el presente diseño muestral para la *EDSA-2016*, respecto a diseños de las anteriores *ENDSA*.

#### **4.11. Cálculo del tamaño muestral**

Como el propósito del diseño muestral es alcanzar mejores niveles de precisión para las estimaciones por dominios, para conseguir este logro, el tamaño muestral fue determinado en forma independiente para cada una de las áreas de residencia: urbana y rural.

Para el cálculo del tamaño de la muestra, se consideran dos indicadores más determinantes y apropiados para cada dominio de estudio, los mismos son: proporción de mujeres que sufrieron muerte infantil en un periodo de cinco años en mujeres de 15 a 49 años de edad<sup>5</sup>, como indicador de aproximación a la mortalidad infantil; y proporción de mujeres de 15 a 49 años de edad embarazadas encontradas en un momento dado, como indicador de aproximación a la fecundidad. Buscando con estos indicadores de referencia la importancia demográfica, valor moderado y cobertura poblacional moderada (es decir, se aplican a una proporción considerable de la población). Los datos de referencia para la estimación de los indicadores fueron calculados en base a los datos de la *ENDSA-2008*; las tablas resumen de las proporciones se muestran a continuación.

---

<sup>5</sup> La *ENDSA-2008* tenía como población objeto de estudio entre las mujeres a aquellas que tienen edad entre 15 a 49 años de edad, por tanto se utilizan este grupo para obtener los datos de referencia.

**Tabla 3.**

Bolivia: Algún menor de 5 años fallecido en el historial de nacimientos de mujeres de 15 a 49 años de edad (calculado con ponderadores de-normalizados), ENDSA-2008										
Tipo de residencia		Estimación	Error típico	Intervalo de confianza al 95%		Coeficiente de variación	Efecto del diseño (Def)	Raíz cuadrada del efecto del diseño (Fdd=Def)	Recuento no-ponderado	
				Inferior	Superior					
Urbano	Población	Si	22'744	2'250	18'335	27'154	.099	1.521	1.233	141
		Total	1'662'548	14'790	1'633'558	1'691'537	.009	2.622	1.619	10'905
	% del total	Si	1.4%	0.1%	1.1%	1.7%	.099	1.516	<b>1.231</b>	141
		Total	100.0%	0.0%	100.0%	100.0%	.000	.	.	10'905
Rural	Población	Si	39'165	2'853	33'573	44'758	.073	1.430	1.196	241
		Total	855'835	11'356	833'577	878'094	.013	1.545	1.243	6'034
	% del total	Si	4.6%	0.3%	4.0%	5.3%	.071	1.420	<b>1.192</b>	241
		Total	100.0%	0.0%	100.0%	100.0%	.000	.	.	6'034

Fuente: INE/ENDSA-2008

**Tabla 4.**

Bolivia: Mujeres de 15 a 49 años de edad actualmente embarazadas (calculado con ponderadores de-normalizados), ENDSA-2008										
Tipo de residencia		Estimación	Error típico	Intervalo de confianza al 95%		Coeficiente de variación	Efecto del diseño (Def)	Raíz cuadrada del efecto del diseño (Fdd=Def)	Recuento no-ponderado	
				Inferior	Superior					
Urbano	Población	Si	81'213	4'485	72'421	90'004	.055	1.733	1.317	521
		Total	1'662'548	14'790	1'633'558	1'691'537	.009	2.622	1.619	10'905
	% del total	Si	4.9%	0.3%	4.4%	5.4%	.054	1.702	<b>1.305</b>	521
		Total	100.0%	0.0%	100.0%	100.0%	.000	.	.	10'905
Rural	Población	Si	58'218	3'399	51'555	64'881	.058	1.376	1.173	405
		Total	855'835	11'356	833'577	878'094	.013	1.545	1.243	6'034
	% del total	Si	6.8%	0.4%	6.1%	7.6%	.057	1.367	<b>1.169</b>	405
		Total	100.0%	0.0%	100.0%	100.0%	.000	.	.	6'034

Fuente: INE/ENDSA-2008

Se procedió a calcular el tamaño de muestra de manera independiente en cada uno de los dominios, con errores estándar relativos esperados para las estimaciones de las variables de referencia. Consecuentemente, para el caso del área de residencia rural en los nueve departamentos se empleó el error estándar relativo<sup>6</sup> (o coeficiente de variación muestral) deseado para la proporción de mujeres que sufrieron muerte infantil en un periodo de cinco años en mujeres de 15 a 49 años de edad; y para el área de residencia urbana se consideró el error estándar relativo (o el coeficiente de variación muestral) deseado para la proporción de mujeres embarazadas encontradas en un momento. El tamaño total de la muestra necesaria es la suma de los dos tamaños de muestra necesarios en los dominios exclusivos (dominios de primer nivel).

El cálculo del tamaño muestral se presenta en pasos, en la *Tabla 6* se muestra los resultados del proceso completo donde se incluyen los parámetros estimados a partir de los datos de la *ENDSA-2008*, los niveles de precisión requeridos (error estándar relativo

<sup>6</sup> El error estándar relativo describe la cantidad de error de la muestra en relación con el nivel del indicador y es independiente de la escala del indicador a estimar.

deseado y error estándar relativo esperado) y los tamaños muestrales (bajo muestreo aleatorio simple, bajo diseño complejo y sobre-dimensionado por no-respuesta y sobre-dimensionado por número de personas elegibles en la vivienda).

A continuación se describe cada uno de los pasos realizados para definir el total de viviendas a encuestar en cada dominio.

Sean:

$\hat{p}$ : la estimación de la proporción de mujeres que sufrieron muerte infantil en un periodo de cinco años en mujeres 15 a 49 años de edad;

$\hat{r}$ : la estimación de la proporción de mujeres de 15 a 49 años de edad embarazadas encontradas en un momento;

$Deff(\hat{p})$ : efecto de diseño asociado a la estimación de una proporción  $p$ , que se puede interpretar como el aumento o disminución en la varianza, debido a considerar un muestreo complejo en vez de un muestreo aleatorio simple de viviendas;

$Fdd(\hat{p})$ : factor de diseño utilizado como factor de ajuste para corregir el error estándar de un estimador<sup>7</sup>, se calcula como la raíz del  $Deff(p)$ ;

$ee(\hat{p})$ : error estándar de la estimación de la proporción  $p$ , igual a la raíz cuadrada de la varianza estimada de la estimación;

$eer(\hat{p})$ : error estándar relativo de la estimación de la proporción  $p$ , igual a la raíz cuadrada de la varianza estimada de la estimación dividido por el valor esperado del estimador.

**a)** Se obtienen estimaciones de la proporción de mujeres que sufrieron muerte infantil en un periodo de cinco años en mujeres de 15 a 49 años de edad; como también, de la proporción de mujeres embarazadas encontradas en un momento. Así como los errores estándar y los errores estándar relativos asociados a las estimaciones y el efecto del diseño asociado a estos, a nivel de área de residencia. Estas estimaciones se incluyen en *Tabla 3* y *Tabla 4*.

**b)** Con las estimaciones anteriormente obtenidas, se calcula un tamaño de muestra considerando un muestreo aleatorio simple mono-etápico, mediante la siguiente expresión (este proceso es para una proporción estimada):

$$n = \frac{\left(\frac{1}{p}\right) - 1}{eer(\hat{p})^2}$$

la cuál proviene de:

$$eer(\hat{p}) = \frac{\sqrt{\frac{p(1-p)}{n}}}{p}$$

donde el numerador proviene de (asumiendo que  $N$  es grande):

<sup>7</sup> *Los Errores de Muestreo en las Encuestas Complejas*, Medina; CEPAL

$$ee(\hat{p}) = \sqrt{(1-f) \left(\frac{1}{n}\right) \left(\frac{N}{N-1}\right) p(1-p)}$$

**c)** El tamaño de muestra de diseño complejo para la *EDSA-2016* se puede calcular inflando el tamaño de muestra calculado anteriormente, mediante el uso de un efecto de diseño ( $Deff^8$ ), donde se define como:

$$Deff = \left( \frac{ee(\hat{p}) \text{ bajo un diseno muestral complejo}}{ee(\hat{p}) \text{ bajo un diseno muestral aleatorio simple}} \right)^2$$

Por consiguiente, el tamaño de la muestra neto necesario para un muestreo complejo de conglomerados con el mismo error estándar relativo es igual a:

$$n = Deff \times \frac{\left(\frac{1}{\hat{p}}\right) - 1}{eer(\hat{p})^2}$$

**d)** Entonces la expresión para calcular el tamaño final de la muestra en términos de la cantidad de viviendas teniendo en cuenta la falta de respuesta está dada por:

$$n = Deff \times \frac{\left(\frac{1}{\hat{p}}\right) - 1}{eer(\hat{p})^2} \times \frac{1}{(R_i \times R_v \times b)}$$

o

$$n = Fdd^2 \times \frac{\left(\frac{1}{\hat{p}}\right) - 1}{eer(\hat{p})^2} \times \frac{1}{(R_i \times R_v \times b)}$$

donde:

$n$ : es el tamaño de muestra de viviendas;

$Deff$ : es el efecto de diseño;

$Fdd$ : es el factor de diseño;

$\hat{p}$ : es la proporción estimada;

$eer(\hat{p})$ : es el error estándar relativo deseado;

$R_i$ : es la tasa de respuesta individual;

$R_v$ : es la tasa de respuesta bruta de las viviendas; y

$b$ : es el número de personas elegibles por hogar particular ocupado registrado en el *CNPV-2012* (mujeres de 14 a 49 años de edad).

El número promedio mujeres de 14 a 49 años de edad por vivienda particular ocupada es obtenido a partir de los datos del *CNPV-2012*, el cual da un promedio de 1.08 en el área urbana; 0.75 en el área rural; y 0.96 a nivel nacional. La “tasa de respuesta bruta de las

<sup>8</sup> Un valor  $Deff$  de 1.0 indica que el diseño de la muestra es tan eficiente como una muestra aleatoria simple, mientras que un valor superior a 1.0 indica el incremento en el error de muestreo debido a la utilización de un diseño más complejo y estadísticamente menos eficiente.

viviendas” es el número de viviendas entrevistadas sobre el número seleccionado de viviendas. El resumen de estas tasas, para la *ENDSA-2008*, se detalla en *Tabla 5*.

**Tabla 5.**

Rendimiento de la muestra, ENDSA-2008						
	Hogares			Mujeres		
	Urbano	Rural	Total	Urbano	Rural	Total
Seleccionados	11'865	8'138	20'003	11'391	6'264	17'655
Encontrados	11'698	8'097	19'795	--	--	--
Entrevistados	11'485	8'079	19'564	10'905	6'034	16'939
Tasa de respuesta	96.8%	99.3%	97.8%	95.7%	96.3%	95.9%

Fuente: INE/ENDSA-2008

**Tabla 6.**

Bolivia: Tamaños de muestra calculados para dominio (Área de residencia), EDSA-2016															
Área de residencia	Proporción estimada ENDSA-2008		Error estándar relativo (eer) deseado	Error estándar (ee) esperado		Factor de diseño estimado(1) (Fdd)		Tamaño de muestra individual neto (m.a.s.)		Tamaño de muestra individual neto (muestreo complejo)	Tasa de respuesta individual (R <sub>i</sub> )	Tasa de respuesta bruta de viviendas (R <sub>v</sub> )	Núm de personas elegibles por hogar (mujeres de 15 a 49 años de edad) (b)	Tamaño de muestra de hogares, EDSA-2016	Tamaño de muestra calculado para ENDSA-2008
	Proporción de mujeres con muerte infantil(2) (p)	Proporción de mujeres actualmente embarazadas (r)		Proporción de mujeres con muerte infantil	Proporción de mujeres actualmente embarazadas	Proporción de mujeres con muerte infantil	Proporción de mujeres actualmente embarazadas	Proporción de mujeres con muerte infantil	Proporción de mujeres actualmente embarazadas						
<b>Urbano(*)</b>	0.014	0.049	0.063	0.001	0.003	1.23	1.30	18'398	4'969	8'456	0.96	0.97	1.08	8'481	11'865
<b>Rural(**)</b>	0.046	0.068	0.081	0.004	0.005	1.19	1.17	3'218	2'114	4'570	0.96	0.99	0.75	6'337	8'138
<b>Total</b>	<b>0.025</b>	<b>0.055</b>	--	--	--	--	--	--	--	<b>13'026</b>	--	--	--	<b>14'818</b>	<b>20'003</b>

(1) Calculado con ponderaciones de-normalizadas en la base de datos de ENDSA-2008.

(2) Se refiere a la proporción de mujeres que reportaron haber sufrido al menos un caso de mortalidad infantil en su historia de nacimientos en los últimos 5 años antes de la entrevista de la ENDSA-2008.

(3) Se refiere a la proporción de mujeres que se encuentran embarazadas en el momento de la entrevista de la ENDSA-2008.

(b) Obtenido del CNPV-2012

(\*) y (\*\*) Para el tamaño de muestra del área urbana se utiliza como indicador de referencia la proporción de mujeres actualmente embarazadas y para el área rural la proporción de mujeres con muerte infantil.

#### 4.12. Distribución de la muestra dentro de los dominios

Esta asignación tiene por objeto reforzar la eficiencia de muestreo a nivel de dominio y la reducción de errores de muestreo. Después de haber calculado el tamaño de muestra por dominio de diseño, se distribuye la muestra a los dominios de estudio que son los nueve departamentos (asignación potencial), y estrato geográfico (asignación proporcional); posteriormente una vez establecido el total de viviendas para cada uno de estos, en función a la cantidad de “toma de muestra” de viviendas por conglomerado (UPM), se convierte el tamaño de muestra de viviendas a número de conglomerados y luego se distribuye estas UPM a cada estrato estadístico socio-económico con una asignación proporcional. Esta conversión de la muestra de viviendas a muestra de UPM es debida a que la estratificación socio-económica, del *MMM-2012*, está realizada a nivel de UPM.

Generalmente es ideal diseñar las encuestas de manera que cada unidad final de selección en la población tenga la misma probabilidad de ser seleccionada, el resultado de eso es una muestra auto-ponderada. Los resultados de una muestra con este diseño auto-ponderado pueden ser tratados como directamente representativos a la totalidad de la población que se estudia y para su análisis no es necesario aplicar ponderación.

Sin embargo en la práctica, no es viable emplear un diseño auto-ponderado complejo. Una consecuencia desventajosa, en el trabajo, de este tipo de diseño es el costo de la encuesta, el cual puede ser muy diferente en los distintos dominios/estratos de la encuesta. Entonces, la afijación de la “toma de muestra” (Número de viviendas a entrevistar por UPM) puede tener una gran influencia sobre el presupuesto total de la encuesta. Con un tamaño fijo de “toma de muestra”, una pequeña toma de muestra por conglomerado es favorable para la precisión de la encuesta debido a la reducción del efecto de diseño. Sin embargo, tampoco debe ser tan pequeño pues puede llegar a ser costoso, debido a que se necesitarían más conglomerados<sup>9</sup>.

Una vez determinado el tamaño total de la muestra y asignado a los diferentes dominios de la encuesta, se decide cuántas “tomadas de muestra” deben ser entrevistadas por conglomerado y luego convertir el tamaño de muestra del dominio a número de conglomerados.

El área de residencia es un factor importante para la determinación de la “toma de muestra” por conglomerado. Al momento de definir esto se consideran dos recomendaciones para encuestas de este tipo<sup>10</sup>; por un lado se recomienda una “toma de muestra” de alrededor de 25 viviendas por grupo rural, mientras que en las áreas urbanas, la ventaja de costos de una "toma" grande es generalmente menor, por eso se recomienda una "toma" de unas 20 viviendas<sup>11</sup>; por otro lado se recomienda mantener un mínimo de 35 UPM por dominio de estudio para el análisis adecuado de la información.

La experiencia de anteriores *ENDSA* llegó a conformar una toma de muestra, única en todos los dominios por conglomerado, de hasta 20 viviendas. Precisamente esta cantidad es usada en las dos últimas *ENDSA* (2003 y 2008). Considerando lo anterior y la logística para el trabajo de campo, se usará una toma de muestra única para ambas áreas de residencia (urbana y rural), la cual será 16 viviendas por conglomerado (UPM). El resumen del cálculo de la asignación del número de viviendas por departamento, como el cálculo de tamaño de “toma de muestra” de viviendas por conglomerado es mostrado en las siguientes tablas.

---

<sup>9</sup> El número de conglomerados afecta el presupuesto de la encuesta más que el tamaño total de la muestra debido a los viajes entre los conglomerados durante la recolección de datos, que representa una parte importante de los costos de campo en las zonas rurales.

<sup>10</sup> *Sampling and Household Listing Manual*, Measure DHS, 2012.

<sup>11</sup> Puesto que en general y en la realidad las UPM del área urbana son más homogéneas-dentro pero más heterogéneas-entre.

**Tabla 7.**

Bolivia: Distribución de la muestra de viviendas* a los departamentos, mediante el método de afiliación potencial, EDSA-2016											
Área de residencia		Asignación potencial $\alpha = 0.5$									
		Departamento									
		Chuquisaca	La Paz	Cochabamba	Oruro	Potosí	Tarija	Santa Cruz	Beni	Pando	Total
<b>Urbano</b>	Viviendas en marco	74'401	543'729	330'051	91'617	93'169	81'982	529'389	68'598	13'544	1'826'480
	Asignación potencial	645	1'745	1'359	716	722	677	1'721	620	275	8'480
<b>Rural</b>	Viviendas en marco	75'801	308'844	186'557	60'444	149'012	44'266	115'465	25'292	11'821	977'502
	Asignación potencial	643	1'298	1'009	574	902	491	794	371	254	6'336
<b>Total</b>	Viviendas en marco	150'202	852'573	516'608	152'061	242'181	126'248	644'854	93'890	25'365	2'803'982
	Asignación potencial	1'288	3'043	2'368	1'290	1'624	1'168	2'515	991	529	14'816

(\*) Viviendas particulares ocupadas

**Tabla 8.**

Bolivia: Distribución de la muestra de UPM* según área y departamento, EDSA-2016									
Departamento	Área urbana				Área rural			Total	
	Ciudades capitales	Ciudades intermedias (10'000 y más hab)	Centros poblados (2'000 a menos de 10'000 hab)	Ciudad de El Alto	Total urbano	Localidades con características amanzanadas (250 a menos de 2'000 hab)	Áreas dispersas (menos de 250 hab)		Total rural
Chuquisaca	35		7		42	22	19	41	83
La Paz	44	10	7	50	111	42	40	82	193
Cochabamba	44	33	10		87	28	36	64	151
Oruro	38	5	4		47	20	17	37	84
Potosí	23	15	8		46	25	32	57	103
Tarija	25	16	3		44	20	12	32	76
Santa Cruz	70	28	11		109	30	21	51	160
Beni	14	21	5		40	9	15	24	64
Pando	15		3		18	6	11	17	35
<b>Total</b>	<b>308</b>	<b>128</b>	<b>58</b>	<b>50</b>	<b>544</b>	<b>202</b>	<b>203</b>	<b>405</b>	<b>949</b>

(\*) 1 UPM = 16 viviendas

**Tabla 9.**

Bolivia: Distribución de la muestra de viviendas según área y departamento, EDSA-2016									
Departamento	Área urbana				Área rural			Total	
	Ciudades capitales	Ciudades intermedias (10'000 y más hab)	Centros poblados (2'000 a menos de 10'000 hab)	Ciudad de El Alto	Total urbano	Localidades con características amanzanadas (250 a menos de 2'000 hab)	Áreas dispersas (menos de 250 hab)		Total rural
Chuquisaca	560		112		672	352	304	656	1'328
La Paz	704	160	112	800	1'776	672	640	1'312	3'088
Cochabamba	704	528	160		1'392	448	576	1'024	2'416
Oruro	608	80	64		752	320	272	592	1'344
Potosí	368	240	128		736	400	512	912	1'648
Tarija	400	256	48		704	320	192	512	1'216
Santa Cruz	1'120	448	176		1'744	480	336	816	2'560
Beni	224	336	80		640	144	240	384	1'024
Pando	240		48		288	96	176	272	560
<b>Total</b>	<b>4'928</b>	<b>2'048</b>	<b>928</b>	<b>800</b>	<b>8'704</b>	<b>3'232</b>	<b>3'248</b>	<b>6'480</b>	<b>15'184</b>



## 5. Probabilidades y ponderaciones de muestreo

Con el fin de que toda inferencia estadística extraída de los datos de la encuesta sea válida, la representatividad de la muestra debe ser tomada en cuenta. En términos generales, los pesos de muestreo se utilizan para hacer la muestra más como la población objetivo. Los análisis deben utilizar las ponderaciones muestrales calculadas para cada hogar/vivienda<sup>12</sup> entrevistada y para cada persona (mujer, varón o niño/a) entrevistada. Precisamente, un ponderador de muestreo sirve como un factor de inflación que extrapolará la muestra a la población objetivo.

### 5.1. Ponderaciones de diseño y ponderaciones de muestreo

Es necesario calcular tanto las ponderaciones del diseño y ponderaciones de muestreo (o ponderaciones de la encuesta), tanto para las viviendas e individuos (mujeres, varones, niños/as). La ponderación del diseño de una unidad de muestreo es la inversa de la probabilidad global con la que fue seleccionada la unidad en la muestra. La ponderación de muestreo de una unidad de muestreo es la ponderación del diseño, pero, corregido por la falta de respuesta o de otras calibraciones.

Dado que no es posible disponer un marco de muestreo a nivel individual para la *EDSA-2016* (así como en anteriores *ENDSA* y otras encuestas en general), se utiliza un procedimiento de muestreo complejo<sup>13</sup> de múltiples etapas, conglomeración y estratificación, con una combinación de muestreo PPT en la primera etapa y un muestreo sistemático de igualdad probabilidad en la segunda etapa. Esto implica que todas las personas elegibles de la misma vivienda comparten la misma ponderación de diseño, que es la misma ponderación del diseño de la vivienda. Por lo tanto, la ponderación del diseño de la vivienda es la **ponderación básica** para la encuesta. El resto de los ponderadores se calculan en base a este ponderador de diseño. El cálculo de la ponderación de muestreo para la *EDSA-2016*, es corregido por falta de respuesta por unidad<sup>14</sup> (esto es, una unidad de muestreo no es entrevistada en absoluto) a nivel de estrato. Si se empleara una corrección de falta de respuesta por unidad a nivel de UPM aumentaría la variabilidad de las ponderaciones de muestreo y por lo tanto aumentarían los errores de muestreo.

### 5.2. Ponderaciones de diseño

Se calculan las ponderaciones de diseño basándose en las probabilidades de muestreo separadas para cada etapa de muestreo y para cada conglomerado.

Sean los subíndices:

$h$ : correspondiente al estrato;

$i$ : correspondiente a las UPM;

---

<sup>12</sup> Dado que en el marco muestral es de viviendas particulares y ocupadas, en adelante se usará solamente como unidad de muestreo la vivienda.

<sup>13</sup> La conglomeración utilizada para la implementación de la eficiencia; y la estratificación para la reducción de errores de muestreo.

<sup>14</sup> El otro tipo de corrección, que es de falta de respuesta por ítem, es dejado a los usuarios de los datos, debido a que esta corrección es especificada según la variable de estudio.

$j$ : correspondiente a las USM.

Sean:

$P_{1hi}$ : probabilidad de muestreo de primera etapa de la UPM  $i$  en el estrato  $h$ ;

$P_{2hi}$ : probabilidad de muestreo de segunda etapa dentro el  $i$ -ésimo conglomerado (selección de viviendas);

$n_h$ : número (o cantidad) de UPM seleccionadas en el estrato  $h$ ;

$N_h$ : número (o cantidad) total de UPM en el estrato  $h$ ;

$M_{hi}$ : medida de tamaño de la UPM  $i$  utilizada en la selección de la primera etapa (la medida de tamaño es el número de viviendas en la UPM de acuerdo con el marco de muestreo);

$\sum_{i=1}^{N_h} M_{hi}$ : medida total de tamaño en el estrato  $h$ .

La probabilidad de seleccionar la  $i$ -ésima UPM es:

$$P_{1hi} = n_h \times \frac{M_{hi}}{\sum_{i=1}^{N_h} M_{hi}}$$

Sean:

$L_{hi}$ : número de viviendas que figuran en el listado de viviendas en la UPM  $i$  del estrato  $h$ ;

$t_{hi}$ : número de viviendas seleccionadas<sup>15</sup> en la UPM  $i$  del estrato  $h$  (como en la encuesta se toma una muestra fija de viviendas por UPM, entonces  $t_{hi} = m$ ).

La probabilidad de selección de segunda etapa para cada vivienda en la UPM  $i$  es:

$$P_{2hi} = \frac{m'}{L_{hi}} = \frac{16}{L_{hi}}$$

Por tanto, la probabilidad general o global de selección de cada una de las viviendas en la UPM  $i$  del estrato  $h$  es el producto de las probabilidades de selección de las dos etapas:

$$P_{hi} = P_{1hi} \times P_{2hi}$$

La ponderación de diseño para cada vivienda en la UPM  $i$  del estrato  $h$  es la inversa de su probabilidad de selección en general:

$$d_{hi} = \frac{1}{P_{hi}} = 1 / \left( n_h \times \frac{M_{hi}}{\sum_{i=1}^{N_h} M_{hi}} \times \frac{16}{L_{hi}} \right)$$

Para la sub-muestra de viviendas donde se aplica la encuesta a varones y las mediciones de hemoglobina a mujeres de 14 a 49 años de edad y niños/as de hasta cinco años de edad, se tiene:

$$d_{hi}^v = \frac{1}{P_{hi}^v} = 1 / \left( n_h \times \frac{M_{hi}}{\sum_{i=1}^{N_h} M_{hi}} \times \frac{6}{L_{hi}} \right)$$

<sup>15</sup> La cantidad de viviendas seleccionadas para la sub-muestra de entrevistas a varones, de 15 a 64 años, es 6 viviendas; a diferencia de la selección para entrevista a mujeres, de 14 a 49 años, que es 16 viviendas.

donde el superíndice  $v$  hace referencia a los cálculos para las viviendas de la sub-muestra a varones y medición de hemoglobina.

### 5.3. Corrección por falta de respuesta por unidad y cálculo de las ponderaciones de muestreo

La falta de respuesta es inevitable en cualquier encuesta, y diferentes unidades tienen diferentes comportamientos de respuesta. Para la corrección de no-respuesta por unidad (viviendas; o personas: mujeres, varones, niños/as y prueba de hemoglobina), se calcula una tasa de respuesta para cada grupo de respuesta homogéneo, luego se infla la ponderación de diseño dividiéndola por la tasa de respuesta para cada grupo de respuesta. Para la *EDSA-2016* se usa el estrato de muestreo como el grupo de respuesta debido a que en la estratificación se reagrupan unidades de muestreo relativamente homogéneas. Aunque sería posible utilizar la UPM como un grupo de respuesta, pero la desventaja es que las tasas de respuesta pueden variar mucho a nivel de UPM, lo que aumentaría la variabilidad de la ponderación de muestreo; que a su vez aumenta la varianza del muestreo.

Asumiendo que los grupos de respuesta coinciden con los estratos de muestreo, para calcular las ponderaciones de muestreo primero se calculan las distintas tasas de respuesta por la falta de respuesta por unidad. En los siguientes cálculos se usan razones ponderadas porque el objetivo es corregir la ponderación del diseño para compensar la falta de respuesta, por tanto, la ponderación de diseño debe estar involucrada.

#### a) Tasa de respuesta a nivel de UPM

Sean:

$n_h$ : número (o cantidad) de UPM seleccionadas en el estrato  $h$ ;

$n_h^*$ : número (o cantidad) de UPM entrevistadas.

Por tanto, la tasa de respuesta a nivel de UPM en el estrato  $h$  es:

$$R_{UPM|h} = \frac{n_h^*}{n_h}$$

#### b) Tasa de respuesta a nivel de vivienda

Sean:

$m_{hi}$ : número (o cantidad) de viviendas encontradas en la UPM  $i$  del estrato  $h$ ;

$m_{hi}^*$ : número (o cantidad) de viviendas entrevistadas en el conglomerado  $i$  del estrato  $h$ .

La tasa de respuesta de las viviendas en el estrato  $h$  se calcula por:

$$R_{viv|h} = \frac{\sum_{i \in S_h} d_{hi} \times m_{hi}^*}{\sum_{i \in S_h} d_{hi} \times m_{hi}}$$

donde  $d_{hi}$  es la ponderación de diseño de cada vivienda en la UPM  $i$  del estrato  $h$ .

Para la sub-muestra de viviendas donde se aplica la encuesta a varones y las mediciones de hemoglobina a mujeres de 14 a 49 años de edad y niños/as de hasta cinco años de edad, se tiene:

$$R_{Viv|h}^v = \frac{\sum_{i \in \mathcal{S}_h} d_{hi}^v \times m_{hi}^{v*}}{\sum_{i \in \mathcal{S}_h} d_{hi}^v \times m_{hi}^v}$$

donde el superíndice  $v$  hace referencia a los cálculos para las viviendas de la sub-muestra a varones y medición de hemoglobina.

### c) Tasa de respuesta por individuos

Individuos a quienes se les entrevista, los cuales pueden ser mujeres de 14 a 49 años de edad o varones de 15 a 64 años de edad o la sub-muestra de individuos a quienes se aplica la medición de hemoglobina.

Sean:

$k_{hi}$ : número (o cantidad) de personas elegibles que se encuentran en la UPM  $i$  del estrato  $h$ ;

$k_{hi}^*$ : número (o cantidad) de personas entrevistadas en la UPM  $i$  del estrato  $h$ .

La tasa de respuesta individual en el estrato  $h$  se calcula por:

$$R_{Pers|h} = \frac{\sum_{i \in \mathcal{S}_h} d_{hi} \times k_{hi}^*}{\sum_{i \in \mathcal{S}_h} d_{hi} \times k_{hi}}$$

donde  $d_{hi}$  es la ponderación de diseño de cada vivienda en la UPM  $i$  del estrato  $h$ .

Para la sub-muestra de viviendas donde se aplica la encuesta a varones y las mediciones de hemoglobina a mujeres de 14 a 49 años de edad y niños/as de hasta cinco años de edad, se tiene:

$$R_{Pers|h}^v = \frac{\sum_{i \in \mathcal{S}_h} d_{hi}^v \times k_{hi}^{v*}}{\sum_{i \in \mathcal{S}_h} d_{hi}^v \times k_{hi}^v}$$

donde el superíndice  $v$  puede variar según se obtengan cálculos para los varones; o mujeres a quienes se aplica la medición de hemoglobina; o niños/as a quienes se aplica la medición de hemoglobina dentro las viviendas de la sub-muestra a varones.

#### 5.3.1. Ponderación de muestreo para viviendas

La ponderación de muestreo de viviendas de la UPM  $i$  en el estrato  $h$  se calcula dividiendo la ponderación del diseño de la vivienda entre el producto de la tasa de respuesta de la UPM y la tasa de respuesta de las viviendas, para cada uno de los estratos de muestreo:

$$D_{hi} = \frac{d_{hi}}{R_{UPM|h} \times R_{Viv|h}}$$

Para la sub-muestra de viviendas donde se aplica la encuesta a varones y las mediciones de hemoglobina a mujeres de 14 a 49 años de edad y niños/as de hasta cinco años de edad, se tiene:

$$D_{hi}^v = \frac{d_{hi}^v}{R_{UPM|h} \times R_{Viv|h}^v}$$

donde el superíndice  $v$  hace referencia a los cálculos para las viviendas de la sub-muestra a varones y medición de hemoglobina.

### 5.3.2. Ponderación de muestreo para individuos

La ponderación de muestreo individual de la UPM  $i$  en el estrato  $h$  se calcula dividiendo la ponderación de muestreo de viviendas por tasa de respuesta individual, o de forma equivalente, dividiendo la ponderación de diseño de la vivienda por el producto de la tasa de respuesta de la UPM, la tasa de respuesta de las viviendas y la tasa de respuesta individual, para cada uno de los estratos de muestreo:

$$W_{hi} = \frac{D_{hi}}{R_{Pers|h}} = \frac{d_{hi}}{R_{UPM|h} \times R_{Viv|h} \times R_{Pers|h}}$$

Los anteriores cálculos se aplican directamente para el cálculo de las ponderaciones de muestreo para la encuesta a mujeres.

En cuanto a las ponderaciones de muestreo para varones así como para la medición de hemoglobina a mujeres y niños/as se pueden calcular de manera similar, a partir de la ponderación de diseño de la sub-muestra de viviendas para la encuesta a varones.

$$W_{hi}^v = \frac{D_{hi}^v}{R_{Pers|h}^v} = \frac{d_{hi}^v}{R_{UPM|h} \times R_{Viv|h}^v \times R_{Pers|h}^v}$$

### 5.4. Corrección por proyección de población

Además si se quiere estimaciones de totales de individuos o personas, ajustadas a las proyecciones de población para un periodo dado, se puede aplicar un factor de corrección por crecimiento de población según la población objeto de análisis: varones de 15 a 49 años de edad; mujeres de 14 a 49 años de edad o de 15 a 49 años de edad; niños y niñas de hasta cinco años de edad.

Sean:

$\tilde{K}_{Dom,t,mujer(14-49)}$ : Una proyección de población para mujeres de 14 a 49 años de edad para un momento dado (por ejemplo: octubre/2016) y para un dominio dado (por ejemplo: La Paz-urbano).

$\hat{K}_{Dom,t,mujer(14-49)}$ : La estimación del total de la población de mujeres de 14 a 49 años de edad, obtenida con datos de la encuesta con el anterior ponderador.

$W_{f:Dom,t}$ : La ponderación final ajustada por proyección de población para mujeres de 14 a 49 años de edad, para un momento y dominio dados.

Entonces el ponderador final se obtiene mediante la siguiente expresión:

$$W_{f:Dom,t} = W_{hi} \times \frac{\tilde{K}_{Dom,t,mujer(14-49)}}{\bar{K}_{Dom,t,mujer(14-49)}}$$

El ponderador se aplica a cada estrato dentro del dominio que es proyectada la población de estudio.

## 6. Bibliografía

INE. (2009). Encuesta Nacional de Demografía y Salud 2008.

Lohr, S. L. (2010). Sampling: Design and Analysis.

Measure DHS, (2012). Demographic and Health Survey: Sampling and Household Listing Manual.

NN.UU. (2009). Diseño de muestras para encuestas de hogares: directrices prácticas, Estudios de métodos Serie F, No. 98.

Pérez Lopez Cesar. (2000). Técnicas de Muestreo Estadístico: Teoría, Práctica y Aplicaciones.