



Diseño de muestra Encuesta de Hogares 2015

UNIDAD DE MUESTREO Y LOGISTICA

Contenido

Introducción	1
Aspectos Metodológicos de la Encuesta de Hogares 2015	1
1. Objetivos	1
2. Necesidad del rediseño de la muestra.....	1
3. Ámbito Geográfico	2
4. Universo de Estudio	2
5. Periodo de levantamiento de los datos.....	2
6. Características del Marco Muestral.....	2
7. Marco Muestral.	2
8. Criterios de Conglomeración y Estratificación del Marco Muestral	3
Variable de Conglomeración.....	3
Variable de Estratificación	4
9. Características generales del diseño de muestra.	5
10. Dominios de Estudio.	5
11. Nivel de desagregación de la información.....	6
12. Unidades de Muestreo.	6
13. Selección de la muestra.	6
14. Cálculo del Tamaño de Muestra.....	6
15. Distribución de la Muestra por Departamento y área geográfica.....	7
16. Factores de Expansión.....	8
17. Ajustes a los Factores Expansión.....	9
18. Estimadores y Errores de Muestreo.....	10
19. Incidencia de campo	10

Contenido de tablas

Tabla 1: Estructura del Marco Muestral	2
Tabla 2: Marco Muestral – 2012 - Número de UPM por área geográfica según número de manzanas ..	3
Tabla 3: Marco Muestral 2012	4
Tabla 4: Marco Muestral 2012 -Estrato geográfico	4
Tabla 5: Reagrupación geográfica	4
Tabla 6: Coeficiente de variación (CV) del tamaño de muestra planificada	7
Tabla 7: Distribución de la muestra de viviendas y UPM por departamento y área geográfica	8
Tabla 8: Coeficiente de variación (CV) del tamaño de muestra ejecutada	10
Tabla 9: Incidencia de campo por departamento	11

Introducción

Las encuestas de hogares se desarrollaron a partir de 1980 con las encuestas permanentes de hogares, las cuales se remplazaron por las encuestas integradas de hogares EIH, a partir de 1999 a 2002 se desarrollaron encuestas puntuales del programa para el mejoramiento de las encuestas y la Medición de Condiciones de Vida MECOVI. Con la culminación del programa MECOVI, se inició la encuesta continua de hogares durante 2003 y 2004, incluyendo temas generales del programa MECOVI así como el de presupuestos familiares y un módulo sobre gestión de gobierno y democracia, propuesto en las reuniones de trabajo de la comunidad Andina de Naciones. Desde el 2005 a 2009 se retoma las encuestas puntuales de corte transversal, manteniendo como temática de exploración "las condiciones de vida de la población" donde los resultados se muestran a nivel nacional, total urbano, total rural.

A partir del 2011 se realiza la desagregación a nivel Nacional, Nacional Urbano, Nacional Rural y departamental, donde los tamaños de muestra son calculados de manera independiente por departamento excepto para el departamento de Beni y Pando, donde las estimaciones deben de considerarse de manera conjunta.

Aspectos Metodológicos de la Encuesta de Hogares 2015

1. Objetivos

Objetivo General

- Suministrar información sobre las condiciones de vida de los hogares, a partir de la recopilación de información de variables socioeconómicas y demográficas de la población boliviana, necesaria para la formulación, evaluación, seguimiento de políticas y diseño de programas de acción en el área social.

Objetivos Específicos

- Generar indicadores relacionados con el empleo, desempleo y otras características económicas, que permitan conocer la evolución de la pobreza, el bienestar y la tendencia de las condiciones de vida de los hogares del país, con resultados a nivel nacional, con desagregación nacional urbano y nacional rural y a nivel departamental, excepto Beni y Pando que será de manera conjunta.
- Generar indicadores de ocupación para las ciudades capitales de los departamentos del eje central (La Paz, Cochabamba, Santa Cruz).
- Servir de fuente de información a instituciones públicas y privadas que permita realizar investigaciones futuras.

2. Necesidad del rediseño de la muestra

Para la Encuesta de hogares 2015 se utilizó el marco conformado en base al último Censo de Población y Vivienda 2012 (CNPV-2012), que contiene una variable de estratificación geográfica, por lo que el nuevo marco muestral tiene:

Un alto grado de homogeneidad dentro cada estrato geográfico y heterogeneidad entre los estratos.

3. **Ámbito Geográfico**

El ámbito geográfico de la encuesta de hogares 2015 es el área urbana y rural de todos los departamentos del país: Chuquisaca, La Paz, Cochabamba, Oruro, Potosí, Tarija, Santa Cruz, Beni y Pando.

4. **Universo de Estudio**

La investigación está dirigida al conjunto de hogares establecidos en viviendas particulares ocupadas de las ciudades capitales, resto urbano y área rural de Bolivia, en el periodo definido para la ejecución de la encuesta en la gestión 2015.

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.

5. **Periodo de levantamiento de los datos**

El levantamiento de información primaria corresponde al periodo del 16 de noviembre al 31 de diciembre de 2015.

6. **Características del Marco Muestral.**

La información del Marco Muestral (MM) se basa en el Censo de Población y Vivienda de 2012 (CNPV-2012), la Actualización Cartográfica Multipropósito (ACM-2010-2012) y el Censo Nacional Agropecuario de 2013 (CNA-2013).

La conformación de las Unidades Primarias de Muestreo (UPM) en el área urbana y rural se realizó de acuerdo con los límites geográficos.

7. **Marco Muestral.**

La estructura del Marco Muestral está definida por las siguientes variables:

Tabla 1: Estructura del Marco Muestral

Tipo de Variable	Nombre de la variable	Variable
Variables de División Política Administrativa	I02_DEPTO	Departamento: código y descripción
	I03_PROV	Provincia: código y descripción
	NOMBRE_PROVINCIA	
	I04_MUN	Municipio: código y descripción
	NOMBRE_MUNICIPIO	
Variables de organización municipal administrativa rural	I04_MUN	Cantón: código y descripción
	NOMBRE_MUNICIPIO	Comunidad: código y descripción
	I06_CIUCOM	
	NOMBRE_COMUNIDAD	
Variables de Organización Municipal administrativa urbana	I04_MUN	Municipio: código y descripción
	NOMBRE_MUNICIPIO	
	ID_UNICO	Manzana

Variables para el diseño de muestra	ID_UNICO_UPM	UPM dado por la unidad de cartografía
	upm	UPM normalizado
	NUM_UPM	UPM dado por la unidad de Muestreo
	Categoría_URBRUR	Estrato geográfico
	URBRUR	Área Urbana o Rural
	Vivienda_ocu	Número de viviendas Ocupadas
	totperso	Total de persona
	totalHombre	Total de personas hombres
totalMujer	Total de personas mujeres	

Fuente: Instituto Nacional de Estadística

Los planos y mapas cartográficos constituyen parte del marco muestral para la encuesta, mediante el cual se identifican las áreas correspondientes a las unidades primarias de muestreo seleccionadas y sirven de guía para ubicación en el terreno y el desarrollo del plan de recorridos de brigadas de encuestadores y supervisores.

8. Criterios de Conglomeración y Estratificación del Marco Muestral

Variable de Conglomeración

A partir del 2014 las encuestas de hogares, el Marco Muestral 2012 está basado en la información del censo de población y vivienda 2012, CNPV-2012 y está constituido por 102593 conglomerados (UPM).

El 67% (68.854) de los conglomerados incluyen a una sola manzana y el 33% (33.739) de los conglomerados incluyen a dos o más manzanas (tabla 5), lo cual implica que se incrementen las etapas de selección o que las manzanas o comunidades seleccionadas dentro una UPM se encuentren a bastante distancia ocasionando problemas en el operativo de campo.

Tabla 2: Marco Muestral – 2012 - Número de UPM por área geográfica según número de manzanas

Número de manzanas/ comunidad	Número de UPM						Total	
	Área Amanzanada			Área Dispersa			Frecuencia	Porcentaje
	Frecuencia	Porcentaje	Porcentaje acumulado	Frecuencia	Porcentaje	Porcentaje acumulado		
Total	82357	100		20236	100		102593	100
1	49718	60,4	60,4	19136	94,6	94,6	68854	67,1
2	14625	17,8	78,1	960	4,7	99,3	15585	15,2
3	7315	8,9	87	128	0,6	99,9	7443	7,3
4	4188	5,1	92,1	9	--	100	4197	4,1
5	2224	2,7	94,8	3	--		2227	2,2
6	1590	1,9	96,7	--	--		1590	1,5
7	1007	1,2	97,9	--	--		1007	1
8	703	0,9	98,8	--	--		703	0,7
9	459	0,6	99,4	--	--		459	0,4
10	313	0,4	99,7	--	--		313	0,3
11 a 15	215	0,3	100	--	--		215	0,2

Fuente: Instituto Nacional de Estadística

A continuación se observa el número de la UPM por Departamento.

Tabla 3: Marco Muestral 2012
Distribución de los conglomerados o UPM por departamento

Departamento	Frecuencia	Porcentaje
Total	102593	100
Chuquisaca	5362	5,2
La Paz	29610	28,9
Cochabamba	18041	17,6
Oruro	7384	7,2
Potosí	8519	8,3
Tarija	4578	4,5
Santa Cruz	23690	23,1
Beni	4350	4,2
Pando	1059	1

Fuente: Instituto Nacional de Estadística

Variable de Estratificación

Con el propósito de obtener precisión en los estimadores en la encuesta de Hogares 2015 se utilizó los siguientes estratos geográficos:

Tabla 4: Marco Muestral 2012 -Estrato geográfico

Estrato Geográfico	Descripción
1	Ciudades Capitales
2	Ciudades Intermedias
3	Centros Poblados
4	Localidades con características amanzanadas < 2.000 habitantes
5	Áreas Dispersas
6	Ciudad de El Alto
7	Áreas Dispersas > 2.000 habitantes

Fuente: Instituto Nacional de Estadística

Los estratos mencionados son clasificados en el área urbana y rural.

Tabla 5: Reagrupación geográfica

Estrato Geográfico	Descripción	Reagrupación geográfica
1	Ciudades Capitales	1. Urbano
2	Ciudades Intermedias	
3	Centros Poblados	
6	Ciudad de El Alto	
4	Localidades con características amanzanadas < 2.000 habitantes	2. Rural
5	Áreas Dispersas	
7	Áreas Dispersas > 2.000 habitantes	

Fuente: Instituto Nacional de Estadística

9. Características generales del diseño de muestra.

El diseño muestral aplicado es estratificación, por conglomerados y bi etapico, con las siguientes características:

- a) Probabilístico, porque la probabilidad de selección es conocida y diferente de cero para cada elemento objeto de estudio, lo cual permite realizar inferencia estadística al total de la población objetivo.
- b) Por conglomerados, debido a que las unidades objeto de estudio (viviendas/hogares) se encuentran agrupadas en áreas conformadas, con fines de muestreo, en base a la información y cartografía del Censo de Población y Vivienda 2012, denominadas conglomerados.
- c) Estratificado, porque previo a la selección de conglomerados, los estratos son clasificados con el propósito de mejorar la precisión de los estimadores y permitir un mejor control de la distribución de la muestra.
- d) Bi-etapico, porque los elementos pertenecientes a la muestra se seleccionan en dos etapas:
 - Etapa 1: Selección de conglomerados (UPM);
 - Etapa 2: selección de viviendas dentro la UPM elegida.

10. Dominios de Estudio.

Se garantiza generar estimaciones de indicadores de pobreza, con errores muestrales admisibles, para los siguientes niveles de desagregación:

- Nacional
- Nacional Urbano
- Nacional Rural
- Departamental, excepto Beni y Pando, de los cuales se darán estimaciones de manera conjunta.

Para los indicadores de ocupación, los resultados podrán ser desagregados a nivel Nacional y ciudades capitales de los departamentos del eje central (La Paz, Cochabamba y Santa Cruz).

11. Nivel de desagregación de la información

Generar indicadores que permitan conocer la evolución de la pobreza, el bienestar y la tenencia de las condiciones de vida de los hogares del país a nivel:

- a) Nacional
- b) Nacional Urbano
- c) Nacional Rural
- d) Departamental (En el caso de Beni y Pando se obtendrán estimaciones conjuntas)

Presentar indicadores de ocupación para las siguientes ciudades capitales: La Paz, Cochabamba y Santa Cruz.

Realizar desagregaciones mayores de los resultados estará sujeto a la valoración de los errores muestrales.

12. Unidades de Muestreo.

Las unidades de muestreo son:

- UPM - La unidad primaria de muestreo es el conglomerado de viviendas.
- USM - La unidad secundaria de muestreo conforman las viviendas particulares dentro de las UPM seleccionadas.

13. Selección de la muestra.

Etapas:

- En la primera etapa se utiliza la Selección Sistemática con probabilidad proporcional al tamaño (PPT) de viviendas.
- En la última etapa la selección es sistemática con arranque aleatorio.

14. Cálculo del Tamaño de Muestra.

Las variables que han sido, y son, determinantes del cálculo de tamaños muestrales apropiados para cada estrato y dominio de estudio son: la ***tasa de desempleo y pobreza***.

Los datos de referencia para el cálculo del tamaño de muestra fueron en base a indicadores de pobreza y empleo obtenidos con la información de la Encuesta de Hogares 2013.

Para garantizar la precisión en los estimadores, se procedió de manera independiente en cada uno de los dominios de estudio. En el caso de las ciudades capitales de los departamentos de La Paz, Cochabamba y Santa Cruz se empleó el coeficiente de variación de la tasa de desempleo; en el resto se consideró el indicador de pobreza. Se admite un error de muestreo máximo del 10% en estos estimadores.

Tabla 6: Coeficiente de variación (CV) del tamaño de muestra planificada

Departamento/Ciudad		Número de viviendas (muestra planificada 2015)	Coeficiente de Variación esperado (CV) en base a la Gestión 2014		
			CV de desempleo	CV de pobreza	CV de pobreza extrema
Nacional	NACIONAL	10656	3,0	1,4	2,0
Chuquisaca	Chuquisaca	516		4,5	5,5
La Paz	La Paz - Ciudad	984	8,8	5,8	9,2
	La Paz - Resto	816		3,2	4,0
	La Paz - El Alto	900	11,3	5,5	8,8
Cochabamba	Cochabamba - Ciudad	1344	7,1	4,5	8,7
	Cochabamba - Resto	792		4,6	6,6
Oruro	Oruro	576		6,9	10,2
Potosí	Potosí	480		4,7	5,6
Tarija	Tarija	972		7,0	13,9
Santa Cruz	Santa Cruz - Ciudad	1488	7,2	5,6	11,6
	Santa Cruz - Resto	840		5,1	8,4
Beni	Beni	492		6,6	12,5
Pando	Pando	456		7,1	10,3

Fuente: Instituto Nacional de Estadística

15. Distribución de la Muestra por Departamento y área geográfica

La Tabla 7, presenta la distribución de la muestra de viviendas por departamento y área geográfica, así como la distribución de la UPM.

En cada una de las UPM se selecciona 12 viviendas.

En las tres ciudades de: La Paz, Cochabamba y Santa Cruz, se considera las conurbaciones, en la distribución de la muestra, La conurbación de La Paz contempla las ciudades de: La Paz ciudad, El Alto, Achocalla y Viacha; Conurbación de Cochabamba: Cochabamba ciudad, Vinto, Tiquipaya, Sacaba, Colcapirua, Quillacollo y Sipe Sipe; y la Conurbación de Santa Cruz: Santa Cruz ciudad y la Guardia.

Tabla 7: Distribución de la muestra de viviendas y UPM por departamento y área geográfica

Departamento		Número de Viviendas de la muestra				Número de UPM de la muestra			
		Total	Ciudad	Resto Urbano	Rural	Total	Ciudad	Resto Urbano	Rural
Nacional	NACIONAL	10656	6024	2316	2316	888	502	193	193
Chuquisaca	Chuquisaca	516	228	60	228	43	19	5	19
La Paz	La Paz Ciudad/Conurbación	1284	984	300	--	107	82	25	--
	La Paz - Resto	516	--	96	420	43	--	8	35
	La Paz - El Alto	900	900	--	--	75	75	--	--
Cochabamba	Cochabamba Ciudad/Conurbación	1728	1344	384	--	144	112	32	--
	Cochabamba - Resto	408	--	108	300	34	--	9	25
Oruro	Oruro	576	264	72	240	48	22	6	20
Potosí	Potosí	480	96	120	264	40	8	10	22
Tarija	Tarija	972	384	312	276	81	32	26	23
Santa Cruz	Santa Cruz Ciudad/Conurbación	1584	1488	96	--	132	124	8	--
	Santa Cruz - Resto	744	--	468	276	62	--	39	23
Beni	Beni	492	108	264	120	41	9	22	10
Pando	Pando	456	228	36	192	38	19	3	16

Fuente: Instituto Nacional de Estadística

16. Factores de Expansión

El factor de expansión es el inverso de la probabilidad de selección de la vivienda. El factor de expansión final, tiene el ajuste de la no-respuesta, considerando los resultados de incidencias de campo y el total poblacional proyectado para ese año, por esta razón los factores de expansión deben ser calculados una vez realizada la encuesta.

Las fórmulas con las que se calculará las probabilidades, vienen dadas por:

$$P(viv_{ijh}) = \left(\frac{A_n N_{jh}}{N_h} \right) \cdot \left(\frac{VE_{jh}}{VOE_{jh}} \right) = \left(\frac{A_n N_{jh}}{N_h} \right) \cdot \left(\frac{c}{VOE_{jh}} \right)$$

Donde:

$P(viv_{ijh})$: Probabilidad de seleccionar la i-ésima vivienda de la j-ésima UPM, del estrato h de una UPM del área amanzanada.

- A_h : Número de UPM's seleccionadas del estrato h.
- N_h : Número de viviendas del estrato h.
- N_{jh} : Número de viviendas en la j-ésima UPM del estrato h.
- VE_{jh} : Total de viviendas elegidas (viviendas elegidas ocupadas) en la j-ésima UPM, del estrato h, es decir, $VE_{jh} = c$
- VOE_{jh} : Total de viviendas objeto de estudio (viviendas ocupadas) listadas en la j-ésima UPM, del estrato h.
- c : Número fijo de viviendas seleccionadas en la última etapa, $c = 12$.

17. Ajustes a los Factores Expansión

Por Incidencias Finales

Los factores de expansión se corrigen por incidencias de campo, es decir, que se utiliza los valores de la muestra efectiva por UPM.

El ajuste se realiza multiplicando el factor básico por la relación: $\frac{c}{V_{jh}}$

V_{jh} : es la muestra efectiva, número de viviendas con encuestas válidas en la j-ésima UPM del estrato h.

c : es el número de viviendas planificadas para ser encuestadas = 12

Por Proyecciones de Población al 2015

Para estimar el Total de hogares y el Total de población, se corrige el factor de expansión con las proyecciones de población que genere el INE para la gestión 2015, esta información por cada dominio de estudio.

La corrección se realiza multiplicando por: $\frac{P_h}{\hat{P}_h}$

Dónde:

P_h : Población en el h-ésimo dominio, según la proyección.

\hat{P}_h : Población en el h-ésimo dominio, a la que expande la encuesta.

Los factores de expansión finales estarán dados por:

$$F_{jh} = \frac{1}{P(Viv_{ijh})} \times \frac{c}{V_{jh}} \times \frac{P_h}{\hat{P}_h}$$

El resultado constituye el Factor de Expansión final que debe ser utilizado para la generación de indicadores.

18. Estimadores y Errores de Muestreo

Se ha señalado que el diseño empleado es complejo, por lo cual se requiere utilizar fórmulas especiales que consideran los efectos de la estratificación y la conglomeración al momento de obtener los resultados de la encuesta.

Para calcular los estimadores, es fundamental que la base de datos de la encuesta esté expandida, vale decir, que se empleen los factores de expansión. Para el cálculo de los intervalos de confianza, coeficiente de variación, error estándar, efecto de diseño, en otros, es necesario aplicar el diseño de la muestra antes de utilizar la base de datos en la generación de indicadores.

Se recomienda valerse de algún software estadístico que permita realizar el cálculo de los errores muestrales aplicando el diseño de muestra empleado; entre los programas estadísticos más utilizados para realizar este procedimiento están el SPSS.

19. Incidencia de campo

Una vez finalizado el operativo de campo se procede a realizar el ajuste de los coeficientes de variación de cada una de los departamentos, así como la incidencia de campo.

Tabla 8: Coeficiente de variación (CV) del tamaño de muestra ejecutada

Departamento/Ciudad		Número de viviendas (muestra ejecutada 2015)	Coeficiente de Variación esperada (CV) en base a la muestra ejecutada por departamento		
			CV de desempleo	CV de pobreza	CV de pobreza extrema
Nacional	NACIONAL	10171	3,0	1,4	2,1
Chuquisaca	Chuquisaca	496		4,6	5,6
La Paz	La Paz - Ciudad	919	9,1	6,0	9,5
	La Paz - Resto	797		3,2	4,0
	La Paz - El Alto	879	11,5	5,6	8,9
Cochabamba	Cochabamba - Ciudad	1270	7,3	4,7	9,0
	Cochabamba - Resto	742		4,7	6,8
Oruro	Oruro	558		7,0	10,3
Potosí	Potosí	464		4,7	5,7
Tarija	Tarija	951		7,1	14,1
Santa Cruz	Santa Cruz - Ciudad	1393	7,5	5,8	11,9
	Santa Cruz - Resto	802		5,2	8,6
Beni	Beni	487		6,6	12,5
Pando	Pando	413		7,4	10,8

Fuente: Instituto Nacional de Estadística

Según la definición de la Real Academia de la Lengua Española define incidencia como "acontecimiento que sobreviene en el curso de un asunto o negocio y tiene con él alguna conexión". Esta definición genérica, en el ámbito de la investigación a través de encuestas, se refiere fundamentalmente a los acontecimientos que tienen lugar entre el número de muestra de viviendas planificadas para la recogida de información y lo que realmente se obtiene. Es decir, sería la diferencia entre lo planificado y lo ejecutado; entre la muestra teórica y la realmente conseguida.

A continuación se presenta la tabla 9, que refleja las incidencias de campo de la encuesta por Departamento, la ciudad de El Alto, y área urbana y rural:

Tabla 9: Incidencia de campo por departamento

Departamento	Muestra Planificada				Muestra Planificada				Incidencia de Campo en %			
	Total	Ciudad Capital	Resto Urbano	Rural	Total	Ciudad Capital	Resto Urbano	Rural	Total	Ciudad Capital	Resto Urbano	Rural
NACIONAL	10656	6024	2316	2316	10171	5707	2228	2236	95%	95%	96%	97%
Chuquisaca	516	228	60	228	496	216	60	220	96%	95%	100%	96%
La Paz Ciudad	984	984	--	--	919	919	--	--	93%	93%	--	--
La Paz - Resto	816	--	396	420	797	--	389	408	98%	--	98%	97%
La Paz - El Alto	900	900	--	--	879	879	--	--	98%	98%	--	--
Cochabamba Ciudad	1344	1344	--	--	1270	1270	--	--	94%	94%	--	--
Cochabamba - Resto	792	--	492	300	742	--	458	284	94%	--	93%	95%
Oruro	576	264	72	240	558	261	70	227	97%	99%	97%	95%
Potosí	480	96	120	264	464	90	118	256	97%	94%	98%	97%
Tarija	972	384	312	276	951	371	302	278	98%	97%	97%	101%
Santa Cruz Ciudad	1488	1488	--	--	1393	1393	--	--	94%	94%	--	--
Santa Cruz - Resto	840	--	564	276	802	--	534	268	95%	--	95%	97%
Beni	492	108	264	120	487	108	262	117	99%	100%	99%	98%
Pando	456	228	36	192	413	200	35	178	91%	88%	97%	93%

Fuente: Instituto Nacional de Estadística