

INSTITUTO NACIONAL DE ESTADÍSTICA (BOLIVIA)

Diseño de la muestra para la Encuesta de Hogares 2013

2014

UNIDAD DE MUESTREO Y LOGISTICA

Contenido

INTRODUCCIÓN	4
1. Aspectos Metodológicos de la Encuesta de Hogares 2013	4
1.1. Objetivos	4
1.2. Ámbito Geográfico	5
2. Universo de Estudio	5
2.1. Periodo de levantamiento de los datos	5
3. Características generales del diseño de muestra.....	5
4. Marco Muestral.....	6
5. Dominios de Estudio.....	6
6. Unidades de Muestreo.....	6
6.1. Selección de la muestra.	7
7. Cálculo del Tamaño de Muestra.....	7
7.1. Distribución de la Muestra por Departamento y Área Geográfica	8
7.2. Criterios de Conglomeración y Estratificación del Marco de Muestreo	9
Variable de Conglomeración	9
Variable de Estratificación.....	9
8. Factores de Expansión.....	10
8.1. Ajustes a los Factores Expansión	11
8.2. Errores de Muestreo	12
8.3. Incidencia de campo	13

Contenido de tablas

TABLA 1: Tamaño de Muestra Planificada y sus Coeficientes de Variación Esperados (CV).....	7
TABLA 2: Tamaño de Muestra Ejecutada y sus Coeficientes de Variación Observados (CV).....	8
TABLA 3: Distribución de la Muestra de Viviendas y UPMs por Departamento y Área Geográfica.....	8
TABLA 4: Distribución de los Conglomerados por Departamento	9
TABLA 5: Conformación de Estratos	10
TABLA 6: Incidencia de Campo por Departamento.....	13

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 MECOVI para el mejoramiento de las encuestas y de Condiciones de Vida. Con la culminación del programa, se inicio 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 y total rural.

A partir del 2011 en el que se realiza la desagregación a nivel nacional, total urbano y rural y departamental, para el 2012 y 2013 se han incluido las cuatro ciudades importantes de Bolivia (La Paz, Cochabamba, Santa Cruz y El Alto).

El documento abarca seis secciones, la primera contiene los aspectos metodológicos de la encuesta entre ellos los objetivos del diseño, segundo el universo de estudio, tercero características de la muestra, cuarto marco muestral, quinto dominios de estudio, sexto las unidades de la muestra, séptimo cálculo del tamaño de la muestra y la última sección los factores de expansión.

1. Aspectos Metodológicos de la Encuesta de Hogares 2013

1.1. Objetivos

Los objetivos del presente diseño de muestra son:

- El objetivo principal es el generar indicadores relacionados con el mercado laboral y otras características económicas, que permitirán 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 total urbano y total rural y a nivel departamental, en el caso de Beni y Pando se tendrán estimaciones conjuntas para ambos departamentos, para ser comparables con los datos obtenidos en la gestión 2012.
- Generar indicadores de ocupación económica para las ciudades capitales de los departamentos del eje central (La Paz, Cochabamba, Santa Cruz y El Alto).

1.2. Ámbito Geográfico

La encuesta de hogares 2013 cubre el área urbana y rural para todos los departamentos del país: Chuquisaca, La Paz, Cochabamba, Oruro, Potosí, Tarija, Santa Cruz, Beni y Pando.

2. Universo de Estudio

La investigación está dirigida a hogares establecidos en viviendas particulares con habitantes presentes, para las ciudades capitales, resto urbano y área rural de Bolivia, en el periodo definido para la ejecución de la encuesta en la gestión 2013.

Se excluye de esta investigación a la población que reside en viviendas colectivas, tales como hospitales, hoteles, cárceles, recintos militares, hogares de ancianos, unidades educativas, etc. Se incluyen a hogares que residen dentro de estos establecimientos, como ser porteros, cuidadores, etc.

2.1. Periodo de levantamiento de los datos

El levantamiento de información primaria corresponde al periodo de noviembre (21/11/13) a diciembre (22/12/13) de 2013.

3. Características generales del diseño de muestra.

El diseño muestral para esta encuesta es estratificado, por conglomerados y polietápico. A continuación se hace la descripción de cada una de estas características:

- Es probabilístico, porque la probabilidad de selección es conocida y diferente de cero para cada elemento objeto de estudio, lo cual permite elaborar estimadores de la población objetivo.
- Es estratificado¹, porque se tiene muestras independientes para cada grupo de estudio de la población boliviana, estas agrupaciones permiten obtener un mejor control en la distribución de la muestra.
- Es por conglomerados, tanto para las Unidades Primarias de muestreo (conglomerados), como para las Unidades Secundarias de Muestreo (Viviendas y/o hogares).

¹ El efecto de la conformación de estratos es *reducir la varianza* de los estimadores, y mejora la precisión en la estimación de indicadores.

- La selección se hizo en tres etapas:
 - i. Primera etapa: Selección de conglomerados (UPM'S).
 - ii. Segunda etapa: Selección aleatoria de segmentos censales (USM'S).
 - iii. Tercera etapa: Selección de las viviendas/hogares particulares.

4. Marco Muestral.

Se emplea el Marco Muestral construido en base a la información del Censo Nacional de Población y Vivienda del 2001, siendo utilizado a partir del año 2002 en las encuestas a hogares.

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.

5. Dominios de Estudio.

De acuerdo con las estimaciones del diseño muestral, la muestra podría proporcionar información desagregada a los siguientes niveles:

Nivel de desagregación de la información

- a) Nacional.
- b) Urbano y Rural.
- c) Departamental (en el caso de Beni y Pando se tiene una estimación conjunta).

La desagregación de la información para el indicador de ocupación es para las siguientes ciudades capitales:

- a) Ciudades Capitales (La Paz, Cochabamba, Santa Cruz y la ciudad de El Alto).

6. Unidades de Muestreo.

Tanto para el Área Amanzanada como para el Área Dispersa:

- UPM - La unidad primaria de muestreo es el sector censal seleccionado.
- USM - La unidad secundaria de muestreo es el segmento censal seleccionado dentro del sector censal.
- UTM - La unidad terciaria de muestreo conforman las viviendas/hogares particulares dentro del segmento censal seleccionado.

6.1. Selección de la Muestra.

Tanto para el Área Amanzanada y el Área Dispersa:

- En la primera etapa, se utiliza la Selección Sistemática con probabilidad proporcional al tamaño (PPT), en función al número de viviendas.
- En la segunda etapa, la selección es de tipo aleatoria irrestricta.
- En la última etapa, la selección es sistemática con arranque aleatorio

7. Cálculo del Tamaño de Muestra.

Para el cálculo del tamaño de la muestra, se consideran los indicadores más determinantes y apropiados para cada estrato o dominio de estudio, los mismos son: **pobreza por ingreso, pobreza extrema por ingreso y tasa de desempleo abierta.**

Los datos de referencia para el cálculo del tamaño de muestra, fueron en base a indicadores de pobreza y mercado laboral, obtenidos con la información de la Encuesta de Hogares 2011 y 2012.

Se procedió de manera independiente en cada uno de los dominios para el cálculo del tamaño de muestra, que permita dar estimaciones con una confiabilidad adecuada. En el caso de las ciudades capitales (La Paz, Cochabamba y Santa Cruz) y El Alto se empleó el coeficiente de variación de la tasa de desempleo (ver tabla 1); y para el resto se consideró el indicador de pobreza extrema por ingreso. Se admite un error de muestreo máximo del 10% en estos estimadores.

TABLA 1: Tamaño de Muestra Planificada y sus Coeficientes de Variación Esperados (CV)

Departamento/Ciudad	Número de viviendas (muestra planificada 2013)	Coeficiente de Variación esperado (CV) en base a la Gestión 2012		
		Pobreza por Ingreso	Pobreza extrema o indigencia por ingreso	Población Desocupada
NACIONAL	9.876	0,6	1,0	3,3
Chuquisaca	432	1,7	2,7	
La Paz (ciudad)	1.080	2,3	4,5	8,0
La Paz (resto)	504	1,9	2,9	
El Alto	1.140	1,9	3,5	9,0
Cochabamba (ciudad)	1.392	1,7	3,7	6,8
Cochabamba (resto)	468	2,0	3,4	
Oruro	492	3,3	5,1	
Potosí	456	1,8	2,4	
Tarija	888	3,3	11	
Santa Cruz (ciudad)	1.380	2,3	4,6	7,8
Santa Cruz (resto)	732	1,6	3,0	
Beni	432	2,6	4,7	
Pando	480	2,6	4,4	

FUENTE: Encuesta a Hogares-2012

Posteriormente de concluir el operativo de campo, se procede a obtener la incidencia en campo y el ajuste de los coeficientes de variación de cada una de los departamentos.

TABLA 2: Tamaño de Muestra Ejecutada y sus Coeficientes de Variación Observados (CV)

Departamento/Ciudad	Número de viviendas (muestra ejecutada 2013)	Coeficiente de Variación (CV) en base a la gestión 2013		
		Pobreza por Ingreso	Pobreza extrema o indigencia por ingreso	Población Desocupada
NACIONAL	9.553	0,7	1,3	3,0
Chuquisaca	432	2,3	3,3	12,7
La Paz (ciudad)	979	2,8	5,3	8,9
La Paz (resto)	501	1,9	3,1	48,6
El Alto	1.097	2,2	4,0	9,6
Cochabamba (ciudad)	1.353	2,0	4,6	7,1
Cochabamba (resto)	464	2,5	4,6	15,0
Oruro	490	3,7	6,6	12,7
Potosí	456	2,3	3,3	12,2
Tarija	886	4,0	9,6	13,5
Santa Cruz (ciudad)	1.326	2,8	6,6	7,5
Santa Cruz (resto)	692	2,4	4,6	10,1
Beni	410	2,8	6,7	14,0
Pando	467	3,1	5,2	14,0

FUENTE: Encuesta a Hogares-2013

7.1. Distribución de la Muestra por Departamento y Área Geográfica

La tabla siguiente presenta la distribución de la muestra ejecutada por departamento y área geográfica, de la Encuesta de Hogares 2013:

TABLA 3: Distribución de la Muestra de Viviendas y UPMs por Departamento y Área Geográfica

DEPARTAMENTO/CIUDAD	Numero de Viviendas en la muestra				Número de UPMs de la muestra			
	Total	Ciudad	Resto Urbano	Rural	Total	Ciudad	Resto Urbano	Rural
NACIONAL	9.553	6.003	1.233	2.317	823	523	105	195
Chuquisaca	432	216	24	192	36	18	2	16
La Paz	2.577	979	58	443	227	90	5	37
El Alto*		1.097			95	95		
Cochabamba	1.817	1.353	132	332	155	116	11	28
Oruro	490	263	48	179	41	22	4	15
Potosí	456	108	108	240	38	9	9	20
Tarija	886	372	310	204	74	31	26	17
Santa Cruz	2.018	1.303	340	375	176	115	29	32
Beni	410	92	213	105	36	8	19	9
Pando	467	220		247	40	19	0	21

FUENTE: Encuesta a Hogares-2013

7.2. Criterios de Conglomeración y Estratificación del Marco de Muestreo

Variable de Conglomeración

La variable de conglomeración que el INE aplica a partir de la gestión 2011 en el Marco Muestral, está conformada por 22.658 conglomerados (UPM + Cantón + Sector Censal). Estos conglomerados contienen entre 40 y 120 viviendas en el área amanzanada y de 60 a 280 viviendas en el área dispersa. A continuación se presenta la distribución de los conglomerados en el país.

TABLA 4: Distribución de los Conglomerados por Departamento

Departamento	Nro. de Conglomerados	Porcentaje
NACIONAL	22.658	100,0
Chuquisaca	1.121	4,9
La Paz	7.565	33,4
Cochabamba	3.914	17,3
Oruro	1.205	5,3
Potosí	1.743	7,7
Tarija	1.006	4,4
Santa Cruz	5.196	22,9
Beni	787	3,5
Pando	121	0,5

Variable de Estratificación

Con el propósito de obtener mayor precisión en los estimadores en la Encuesta de Hogares 2013, se estructura y combina los conglomerados de selección en la primera etapa. Dentro de cada departamento, utilizando la combinación de variables del estrato geográfico (1: Ciudades capitales incluyendo El Alto; 2: Ciudades intermedias con 10,000 y más habitantes; 3: Centros poblados de 2,000 hasta menos de 10,000 habitantes; 4: Localidades con características amanzanadas de 250 hasta menos de 2,000 habitantes; 5: Áreas dispersas con menos de 250 habitantes) y el estrato de pobreza NBI (1: Alto; 2: Medio Alto; 3: Medio Bajo; 4: Bajo).

TABLA 5: Conformación de Estratos

Estrato (s)	Descripción	Reagrupación
Geográfico		
1	Ciudades Capitales y ciudad de El Alto.	1. Urbana
2	Ciudades Intermedias con 10,000 y más habitantes.	
3	Centros Poblados de 2,000 a menos de 10,000 habitantes.	
4	Localidades con características amezanadas de 250 a menos de 2.000 habitantes.	2. Rural
5	Áreas Dispersas de menos de 250 habitantes.	
Pobreza NBI		
1	Alto, son aquellas unidades muestrales que se encuentran entre el umbral y las necesidades básicas satisfechas.	1. Alto (no pobre)
2	Medio Alto, son unidades muestrales que están en el umbral de pobreza.	
3	Medio Bajo, son unidades muestrales que están en la clasificación moderada de pobreza.	2. Bajo (pobre)
4	Bajo, son unidades muestrales que están entre la indigencia y marginalidad de pobreza.	

Para la Encuesta de Hogares 2013 se considera la combinación de estas variables, que permite corregir las variaciones de las estimaciones de los estratos.

Mediante la combinación de los estratos originales de carácter geográfico y de pobreza, se han conformado los nuevos estratos. Conformando $2 \times 2 = 4$ estratos por departamento, y haciendo un total de 36 estratos para los nueve departamentos.

8. Factores de Expansión

En términos generales el factor de expansión es el inverso de la probabilidad de selección de la vivienda y el factor de expansión final, viene incluido con el ajuste de la tasa de no-respuesta.

Considerando los resultados de las incidencias de campo en la actualización cartográfica y luego tomando el número de viviendas encuestadas por UPM, se detallan a continuación las fórmulas matemáticas que describe el factor de expansión final de la siguiente forma:

Factores de expansión para el Área Amanzanada o el Área Dispersa

$$P(viv_{ikjh}) = \left(\frac{A_h N_{jh}}{N_h} \right) \cdot \left(\frac{b_{jh}}{B_{jh}} \right) \cdot \left(\frac{VE_{jh}}{VOE_{jh}} \right) = \left(\frac{A_h N_{jh}}{N_h} \right) \cdot \left(\frac{b_{jh}}{B_{jh}} \right) \cdot \left(\frac{c}{VOE_{jh}} \right)$$

Donde:

- $P(viv_{ikjh})$: Probabilidad de seleccionar la i-ésima vivienda, de la k-ésima USM, de la j-ésima UPM, en el estrato h.
- 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.
- B_{jh} : Número total de USM's (grupo de segmentos censales) en la j-ésima UPM del estrato h.
- b_{jh} : Número de USM's seleccionados en la j-ésima UPM, del estrato h.
- VE_{jh} : Total de viviendas elegidas (viviendas elegidas ocupadas) en la(s) b_{jh} USM'S seleccionadas de la j-ésima UPM, del estrato h, donde, $VE_{kjh} = c$.
- VOE_{jh} : Total de viviendas objeto de estudio (viviendas ocupadas) listadas actualizadas en la en la(s) b_{jh} USM'S seleccionadas, de la j-ésima UPM, del estrato h.
- c : Número fijo de viviendas seleccionadas en la última etapa, $c = 12$.

Para el cálculo de probabilidades de selección de las viviendas/hogares particulares, se opera como si fuera en dos etapas. El estimador de razón que sirve para el cálculo de los factores de expansión, se elabora para cada UPM, de la muestra.

8.1. Ajustes a los Factores Expansión

Por Incidencias Finales

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

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

Donde:

- 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 2013

Para estimar el total de hogares y el total de población, es necesario corregir el factor de expansión con las proyecciones de población que genere el INE para el año 2013, esta información se requiere para cada dominio de estudio.

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

Donde:

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 tanto para el Área Amanzanada y Área Dispersa están dados por:

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

Este resultado que es el factor de expansión final que sirve para calcular estimaciones a nivel población total de cada dominio de estudio, como también para estimar valores cuantitativos como, ingresos, gastos.

En relación a datos cualitativos, el factor de expansión que se utiliza es el porcentaje y/o tasa que en términos generales su estructura es similar al dado por los resultados de la muestra misma.

8.2. Errores de Muestreo

La precisión de las estimaciones obtenidas a partir de una muestra dada, se mide mediante los errores estándar de las mismas. De manera análoga al error de medición, el error total de una estimación obtenida es igual a la diferencia entre el estimado y el valor real. Dicha diferencia se encuentra compuesta por dos partes: el error sistemático y el error aleatorio.

El primer tipo de error se relaciona con el marco muestral, el encuestador y la tasa de no respuesta. Por otra parte el error aleatorio o error muestral, se encuentra estrechamente relacionado con el hecho de que no se está encuestado a toda la población, sino a un subconjunto de la misma. En líneas habituales, el diseño muestral empleado en un estudio como la Encuesta a Hogares 2013 influye en la precisión de las estimaciones que se vayan a obtener, ya sean éstos: promedios, proporciones o totales.

Se tienen procedimientos manuales y/o software especializado para el cálculo de los errores muestrales. Por la complejidad del cálculo es preferible utilizar software como son SPSS y/o STATA, los que contienen módulos exclusivos para el cálculo de los mismos, los siguientes parámetros son relacionados con el cálculo de errores muestrales:

- i. El estimador de la variable a ser calculada (Promedio, porcentaje o total).
- ii. La desviación estándar y/o varianza del estimador de la variable calculada.
- iii. El error estándar estimado.
- iv. El efecto de diseño, que es el cociente de la varianza del diseño utilizado entre la varianza del muestreo aleatorio simple para un mismo tamaño de muestra.
- v. Intervalo confidencial al 95 % de confianza del estimador de la variable utilizada.
- vi. Coeficiente de variación del estimador de la variable utilizada.
- vii. Número de casos muestrales de la variable utilizada.

Todos estos parámetros pueden ser calculados en forma inmediata una vez que se defina la estructura básica del diseño muestral utilizado en la encuesta.

8.3. Incidencia de campo

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

TABLA 6: Incidencia de Campo por Departamento

DEPARTAMENTO / CIUDAD	Muestra planificada			Muestra Ejecutada			Incidencia de campo (%)		
	Total	Urbano	Rural	Total	Urbano	Rural	Total	Urbano	Rural
NACIONAL	9.876	7.536	2.340	9.553	7.236	2.317	96,7	96,0	99,0
Chuquisaca	432	240	192	432	240	192	100,0	100,0	100,0
La Paz	1.584	1.140	444	1.480	1.037	443	93,4	91,0	99,8
El Alto*	1.140	1.140		1.097	1.097		96,2	96,2	
Cochabamba	1.860	1.524	336	1.817	1.485	332	97,7	97,4	98,8
Oruro	492	312	180	490	311	179	99,6	99,7	99,4
Potosí	456	216	240	456	216	240	100,0	100,0	100,0
Tarija	888	684	204	886	682	204	99,8	99,7	100,0
Santa Cruz	2.112	1.728	384	2.018	1.643	375	95,5	95,1	97,7
Beni	432	324	108	410	305	105	94,9	94,1	97,2
Pando	480	228	252	467	220	247	97,3	96,5	98,0