

1. DISEÑO MUESTRAL

1.1 Objetivos del Diseño de la Muestra

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 total urbano y total rural y a nivel departamental.
- Generar indicadores de ocupación para las ciudades capitales de los departamentos del eje central (La Paz, Cochabamba, Santa Cruz) y la ciudad de El Alto.
- Servir de fuente de información a instituciones públicas y privadas que permitan realizar investigaciones futuras.

1.2 Componentes Básicos

Población objetivo: De acuerdo a los objetivos de la investigación, la unidad de observación es el **Hogar**.

Por otra parte, el cuestionario de la encuesta explora al objeto persona, delimitándolo por grupos de edad que se ubican en sus diferentes secciones.

Cobertura geográfica: El ámbito geográfico de la encuesta de hogares 2014 es el área urbana y rural de todos los departamentos del país: Chuquisaca, La Paz, Cochabamba, Oruro, Potosí, Tarija, Santa Cruz, Beni y Pando.

1.3 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 2014.

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.; se incluyen a hogares que, forman un grupo independiente y además residen dentro de estos establecimientos, ejemplo: porteros, cuidadores, etc.

1.4 Unidades de Observación, Análisis y Muestreo

De acuerdo a los objetivos de la investigación, la unidad de observación es el Hogar.

Las unidades de análisis son: el Hogar como unidad de consumo; los Miembros de hogar en cuanto a características sociodemográficas, educación, salud y ocupación y la vivienda en cuanto a características de tamaño, servicios y condiciones físicas.

La unidad de muestreo en su última etapa, es la vivienda particular ocupada, ésta tiene permanencia fija en el tiempo y espacio, característica que la habilita para ser utilizada como unidad de selección en el diseño muestral. La unidad de muestreo, de acuerdo a las áreas de estudio, se conforma de la siguiente manera:

Área Amanzanada

- 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.

Área dispersa

- UPM.- La unidad primaria de muestreo es el conglomerado.
- USM.- Es la comunidad dentro de la UPM seleccionada.
- UTM.- La unidad terciaria de muestreo conforman las viviendas particulares dentro de la comunidad seleccionada.

1.5 Marco Muestral

Se emplea para la encuesta de hogares 2014 el marco muestral construido en base a la información del CNPV-2012.

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

Tabla : 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 • PROV	Provincia: código y descripción
	• I04_SECC • MUN	Sección municipal: código y descripción
	• I04_SECC • MUN	Cantón: código y descripción
Variables de organización municipal administrativa rural	• I06_CIUCOM • NOMBRE_COMUNIDAD	Comunidad: código y descripción
Variables de Organización Municipal administrativa urbana	• I04_SECC • MUN	Sección municipal: código y descripción
	• ID_UNICO	Manzana
Variables para el diseño de muestra	• ID_UNICO_UPM • upm • NUM_UPM • ADAREA	UPM dado por la unidad de cartografía UPM normalizado UPM dado por la unidad de Muestreo Manzana o comunidad en el Área
	• AREA_CATEG6 • URBRUR	Estrato geográfico estadístico Área Urbana o Rural
	• nvivienda • vivOcup • vivDesoc	Número de viviendas Totales Número de viviendas Ocupadas Número de viviendas Desocupadas
	• TOTPERS_VIV • TOT_HOMBRES_VIV • TOT_MUJERES_VIV	Total de persona Total de personas hombres Total de personas mujeres

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.

Características generales del diseño de muestra.

El diseño muestral aplicado combina la estratificación geográfica en la primera etapa (UPM), utilizando para ello muestreo multietápico 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) Multietápico, porque los elementos pertenecientes a la muestra se seleccionan en dos etapas en el área amanzanada; Etapa 1: Selección de conglomerados (UPMS); Etapa 2: selección de viviendas dentro la UPM elegida. En el área dispersa la selección de la muestra se realizó en tres etapas ya que existen dos o más comunidades agrupadas en una UPM; Etapa 1: Selección de conglomerados UPMS; Etapa 2: selección de la comunidad dentro la UPM elegida; Etapa 3: selección de viviendas dentro de la comunidad elegida.

Selección de la muestra.

Área Amanzanada

- 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.

Área Dispersa

- En la primera etapa se utiliza una selección sistemática con probabilidad proporcional al tamaño (PPT) de viviendas.
- La segunda etapa se utiliza una selección aleatoria de la comunidad dentro la UPM.
- En la última etapa la selección es sistemática con arranque aleatorio.

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) y El Alto.

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

1.6 Cálculo del Tamaño de la 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.

1.7 Distribución de la Muestra

El cuadro siguiente presenta la distribución de la muestra por departamento y área geográfica, de la Encuesta de Hogares 2014:

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

Departamento	Número de Viviendas de la muestra				Número de UPMS de la muestra			
	Total	Ciudad	Resto Urbano	Rural	Total	Ciudad	Resto Urbano	Rural
NACIONAL	10.404	6.636	1.392	2.376	867	553	116	198
Chuquisaca	588	228	48	312	49	19	4	26
La Paz	1.776	1.272	96	408	148	106	8	34
El Alto	1.416	1.416	0	0	118	118	0	0
Cochabamba	1.860	1.392	204	264	155	116	17	22
Oruro	492	240	48	204	41	20	4	17
Potosí	468	84	96	288	39	7	8	24
Tarija	804	300	216	288	67	25	18	24
Santa Cruz	2.112	1.380	444	288	176	115	37	24
Beni	420	108	204	108	35	9	17	9
Pando	468	216	36	216	39	18	3	18

4.6 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 2002 hasta el 2009 en las Encuestas a Hogares, ha sido conformada con información y cartografía del Censo Nacional de Población y Vivienda del 2001. El Marco Muestral utilizado hasta el 2009, está constituido por 16.790 conglomerados (UPMS) y en el Marco Muestral utilizado en base al CNPV-2012 está constituido por 95,999 UPMS.

El 71% (68.136) de los conglomerados incluyen a una sola manzana y el 29% (27.863) 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.

A continuación se observa el número de UPMS por Departamento.

Tabla : Distribución de los conglomerados o UPM por departamento

Departamento	Frecuencia	Porcentaje
Chuquisaca	5.051	5.3
La Paz	26.829	27.9
Cochabamba	17.501	18.2
Oruro	6.883	7.2
Potosí	8.153	8.5
Tarija	4.280	4.5
Santa Cruz	22.229	23.2
Beni	4.143	4.3
Pando	930	1.0
Total	95.999	100.0

Variable de Estratificación

Con el propósito de obtener precisión en los estimadores en la encuesta de Hogares 2014 se estructuro y combino los conglomerados de selección en la primera etapa. Dentro de cada departamento, se ha utilizado la combinación de las variables 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).

A continuación se observa la descripción del Estrato conformado por la cantidad de población.

Tabla : Estrato geográfico

Estrato Geográfico	Descripción
1	Ciudades Capitales
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 ameznadas de 250 a menos de 2.000 habitantes
5	Áreas Dispersas de menos de 250 habitantes
6	Ciudad de El Alto

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

Tabla : Reagrupación geográfica

Estrato Geográfico	Descripción	Reagrupación geográfica
1	Ciudades Capitales más la ciudad de El Alto	1. Urbano
2	Ciudades Intermedias con 10.000 y más habitantes	
3	Centros Poblados de 2.000 a menos de 10,000 habitantes	
6	Ciudad de El Alto	
4	Localidades con características ameznadas de 250 a menos de 2.000 habitantes	2. Rural
5	Áreas Dispersas de menos de 250 habitantes	

4.7 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:

Área Ameznada

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

Donde:

- $P(viv_{ijh})$: Probabilidad de seleccionar la i-ésima vivienda de la j-ésima UPM, del estrato h de una UPM del área ameznada.
- 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, $c = 12$.

Área Dispersa

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

Donde:

- $P(viv_{ikjh})$: Probabilidad de seleccionar la i-ésima vivienda de la k-ésima USM, de la j-ésima UPM, del estrato h en una UPM del área dispersa.
 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} : USM's (grupo de comunidades) seleccionados en la j-ésima UPM del estrato h, se tomará uno solo, entonces $B_{jh} = 1$.
 N_{kjh} : Número de viviendas en la k-ésima USM (grupo de comunidades), de la j-ésima UPM del estrato h.
 N'_{jh} : Número de viviendas en la j-ésima UPM del estrato h.
 c : Número fijo de viviendas seleccionadas en la última etapa, $c = 12$.
 VE_{kjh} : Total de viviendas elegidas (viviendas elegidas ocupadas) en la k-ésima USM (comunidad) de la j-ésima UPM, del estrato h, donde, $VE_{kjh} = c$.
 VOE_{kjh} : Total de viviendas objeto de estudio (viviendas ocupadas) listadas actualizadas en la k-ésima USM (grupo de comunidades) de la j-ésima UPM, del estrato h.

Tomando en cuenta que en el área dispersa se realiza la actualización en la segunda etapa, comunidad seleccionada, entonces los valores de se compensan. $N_{jh} = N'_{jh}$

El procedimiento habitual indica que a los factores de expansión se aplique la corrección de las proyecciones de población, para que los mismos expresen el total de la población boliviana excepto las personas que habitan en viviendas colectivas que no son objeto de estudio en la encuesta.

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 2014

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 2014, esta información por 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 estarán dados por:

Área Amanzanada

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

Área Dispersa

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

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

4.8 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, entre 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.