

# ENCUESTA DE HOGARES 2011

## METODOLOGÍA DE DISEÑO DE MUESTRA

### 1. Introducción.

El Instituto Nacional de Estadística, dando continuidad a la serie de encuestas a hogares que viene llevando a cabo desde 1978, planificó la ejecución de la Encuesta de Hogares 2011 con objetivos y características similares de los diseños de muestra de las encuestas de hogares puntuales que se han realizado en el país bajo el programa MECOVI desde el año 1.999.

Adicionalmente, considerando la coyuntura actual del Estado Boliviano, en esta versión de la Encuesta de Hogares, se ha previsto responder a requerimientos del proceso de consolidación de las autonomías departamentales que se desarrolla en el país, poniendo a disposición resultados a nivel de departamento.

### 2. Objetivos del Diseño de Muestra.

Los objetivos del presente diseño de muestra son:

- Generar indicadores 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 urbano y rural y a nivel departamental, en el caso de Beni y Pando se tendrán estimaciones conjuntas para ambos departamentos.
- Generar indicadores de ocupación para las ciudades capitales de los departamentos del eje central (La Paz, Cochabamba y Santa Cruz).

### 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 2011.

La presente investigación excluye a los hogares constituidos en viviendas colectivas tales como hospitales, hoteles, cárceles, recintos militares, hogares de ancianos, etc.

#### 4. Características generales del diseño de muestra.

EL diseño muestral aplicado combina la estratificación estadística y la conglomeración de áreas geográficas, con el muestreo multietápico, por tanto, puede ser caracterizado como:

- 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;
- 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 2001, denominadas conglomerados;
- estratificado, porque previo a la selección de conglomerados, estos son clasificados en estratos con el propósito de mejorar la precisión de los estimadores y permitir un mejor control de la distribución de la muestra;
- multietápico, porque los elementos pertenecientes a la muestra se seleccionan en dos etapas en el área amanzanada; Etapa 1: Selección de conglomerados (UPM's); Etapa 2: selección de viviendas dentro la UPM elegida.

En área dispersa la selección se realiza en tres etapas; Etapa 1: Selección de conglomerados UPM's; Etapa 2: selección de segmento(s) censales dentro la UPM elegida; Etapa 3: selección de viviendas dentro del segmento(s) elegido.

#### 5. Marco Muestral.

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

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

### Estructura del Marco Muestral

Tipo de Variable	Variable
Variables de División Política Administrativa	Departamento: descripción y código
	Provincia: descripción y código
	Sección municipal: descripción y código
Variables de organización municipal administrativa rural	Cantón: descripción y código
	Comunidad: descripción y código
Variables de Organización Municipal administrativa urbana	Zona
	Manzano
Variables de organización Censal	Zona Censal
	Sector Censal
	Segmento Censal
Variables para el diseño de muestra	Conglomerados - Unidad Primaria de Muestreo - UPM
	Estratos - Estrato Geográfico - Estrato estadístico (NBI)
	Número de viviendas ocupadas
	Número de viviendas desocupadas
	Número de Hombres
	Número de Mujeres

Los planos y mapas cartográficos constituyen parte del marco muestral para la encuesta, pues 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.

#### 6. 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 y 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).

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

## 7. Unidades de Muestreo.

### Área Amanzanada

- UPM - La unidad primaria de muestreo es el conglomerado (UPM+Cantón+Sector Censal)
- 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 (UPM+Cantón+Sector Censal)
- USM - La unidad secundaria de muestreo conforman los segmentos censales dentro de las UPM seleccionadas.
- UTM - La unidad terciaria de muestreo conforman las viviendas particulares dentro del segmento censal seleccionado.

## 8. 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 y segunda 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.

## 9. Cálculo del Tamaño de Muestra.

El tamaño de muestra ha sido calculado en base a indicadores de pobreza y empleo obtenidos con la información de la Encuesta de Hogares 2007.

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.

Fórmula empleada:

$$n \geq \frac{Z_{1-\alpha/2}^2 CV_{(x)}^2}{\varepsilon_{REL}^2 + Z_{1-\alpha/2}^2 \frac{CV_{(x)}^2}{N}} * Deff * Tnr$$

Con los siguientes valores:

$Z_{1-\alpha/2}^2$  : Nivel de confianza 95%

$CV_{(x)}^2$  : Coeficiente de variación de la(s) variable(s)

$\varepsilon_{REL}^2$  : Error de muestreo admisible 10%

### Distribución de la Muestra 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
Chuquisaca	612	276	48	288	51	23	4	24
La Paz*	2,196	1,380	216	600	183	115	18	50
Cochabamba*	1,656	696	432	528	138	58	36	44
Oruro	588	288	84	216	49	24	7	18
Potosí	816	168	204	444	68	14	17	37
Tarija	636	240	240	156	53	20	20	13
Santa Cruz*	2,088	1,080	612	396	174	90	51	33
Beni	480	96	276	108	40	8	23	9
Pando	408	204	0	204	34	17	0	17
<b>NACIONAL</b>	<b>9,480</b>	<b>4,428</b>	<b>2,112</b>	<b>2,940</b>	<b>790</b>	<b>369</b>	<b>176</b>	<b>245</b>

\* En las ciudades capitales de los departamentos de La Paz, Cochabamba y Santa Cruz se empleó el coeficiente de variación de la tasa de desempleo.

Se adjunta en el CD el plan de selección de la muestra para cada uno de los departamentos, con estos se puede replicar la distribución y selección de la muestra.

## 10. Criterios de Conglomeración y Estratificación del Marco de Muestreo

### 10.1. Variable de Conglomeración

La variable de conglomeración que el INE aplica en las Encuestas a Hogares, ha sido conformada con información y cartografía del Censo Nacional de Población y Vivienda del 2001.

#### Nro. Sectores por UPM

Número de Sectores	Número de UPMs	Porcentaje
1	11.108	66,2
2	5.520	32,9
3	142	0,8
4	16	0,1
5	4	0,0
Total	16.790	100,0

Para la construcción de los conglomerados se han agrupado en áreas geográficas de uno hasta cinco sectores censales, de tal forma de obtener un número similar de viviendas en cada grupo. A estos conglomerados se denominan Unidades Primarias de Muestreo (UPM).

El Marco Muestral esta constituido por 16.790 conglomerados. Aproximadamente las UPMs contienen 80 a 150 viviendas en el área amanzanada y de 150 a 350 viviendas en el área dispersa.

Tal como muestra el cuadro anterior, el 33.8% (5.682) de los conglomerados incluyen a más de un Sector Censal, lo cual implica que se incrementen las etapas de selección o que los segmentos seleccionados dentro una UPM se encuentren a bastante distancia ocasionando problemas en el operativo de campo.

Adicionalmente, se debe considerar que estas UPMs han cambiado considerablemente desde el censo 2001 a la fecha, la mayoría de los sectores del área amanzanada han crecido notablemente, mientras que algunas UPMs del área rural dispersa han disminuido de población, que para diferenciar a los sectores se toma en cuenta la variable de división organización municipal administrativa (cantón) la cual esta inmersa en cada UPM, y previo a la variable de organización censal (sector censal).

Para evitar estas dificultades, se determinó tomar como conglomerado de primera etapa de selección de la muestra a los Sectores Censales.

Con esta modificación, el Marco Muestral para la Encuesta de Hogares 2011 está conformado 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.

### Distribución de los Conglomerados por Departamento

Departamento	Nro. de Conglomerados (Sectores)	Porcentaje
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
Total	22.658	100

#### 10.2. Variable de Estratificación

Con el propósito de obtener mayor precisión en los estimadores es necesario estratificar los conglomerados de selección en la primera etapa. Para este fin, dentro de cada departamento, se han utilizado los estratos que se han empleado en las encuestas a hogares desde el año 2002 al 2007 y el año 2009, el estrato de conglomeración de población y el de niveles de pobreza.

Estrato Geográfico	Descripción	Estrato de Aglomeración de Población
1	Ciudades Capitales más la ciudad de El Alto	Ciudad capital
2	Ciudades Intermedias con 10,000 y más habitantes	Resto Urbano
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	Rural
5	Áreas Dispersas de menos de 250 habitantes	

El estrato de pobreza, denominado estrato estadístico, es resultado de un tratamiento del indicador de Necesidades Básicas Insatisfechas (NBI), con el método cluster, mediante el cual se caracterizó las UPMs por nivel de pobreza.

Estrato de Pobreza NBI	Descripción
------------------------	-------------

1	Alto, son aquellas unidades muestrales que se encuentran entre el umbral y las necesidades básicas satisfechas.
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.
4	Bajo, son unidades muestrales que están entre la indigencia y marginalidad de pobreza.

Mediante la combinación de las variables de niveles de conglomeración de población y Niveles de pobreza se han conformado estratos en cada departamento.

### 10.3. Post Estratificación

La post estratificación o ajuste de los estratos es un procedimiento similar a la estratificación tratada en el punto de arriba, con la diferencia que la post estratificación o ajuste de los estratos es realizada después de obtener los resultados. En el punto de arriba la estratificación constituía una herramienta para la selección de la muestra, Unidades primarias de muestreo, y fue realizada en la etapa del diseño. La post estratificación consiste en la obtención de nuevos grupos de elementos denominados post estratos o ajustes de los estratos, en donde estos nuevos post estratos poseen características homogéneas al interior y heterogéneas al exterior de los mismos, basadas en variables combinadas, geográficas y pobreza.

La post estratificación utiliza con más frecuencia para corregir las variaciones de las estimaciones de los estratos, con el propósito de afinar o ajustar factores de expansión.

Para este fin, se utilizó los estratos de estratificación del diseño previo al levantamiento de campo en las encuestas a hogares, el de conglomeración de población (geográfico) y el de niveles de pobreza (NBI).

Estrato Geográfico	Descripción	Estrato de Aglomeración de Población	Post Estratificación geográfica
1	Ciudades Capitales más la ciudad de El Alto	Ciudad capital	1. Urbano
2	Ciudades Intermedias con 10,000 y más habitantes	Resto Urbano	
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	Rural	2. Rural
5	Áreas Dispersas de menos de 250 habitantes		

Estrato de Pobreza NBI	Descripción	Post Estrato de pobreza
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.	

Mediante la combinación de las últimas columnas de los cuadros de arriba geográfica y pobreza se han conformado los post estratos para cada departamento.

## 11. Factores de Expansión

El factor de expansión preliminar 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 Amanzanada

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

donde:

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

**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.
- $B_{jh}$  : Número de USM's (grupo de segmentos censales) 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 segmentos censales), 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 (segmentos censales) 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 segmentos censales) de la j-ésima UPM, del estrato h.

Tomando en cuenta que en el área dispersa no se realizará la actualización de toda la UPM elegida en la primera etapa, sino del segmento seleccionado, entonces los valores  $N_{jh} = N'_{jh}$ , por tanto se compensan.

Además, considerando que  $B_{jh} = 1$ , la fórmula del área dispersa queda de la siguiente manera:

$$P(viv_{ikjh}) = \left( \frac{A_h N_{kjh}}{N_h} \right) \cdot \left( \frac{c}{VOE_{kjh}} \right)$$

En consecuencia, para el cálculo de las probabilidades se opera como si fuera en dos etapas.

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.

## 11.1. 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 2011**

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 2011, esta información se requiere 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**

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

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

## 12. Estimadores y Errores de Muestreo

Se ha señalado que el diseño empleado es complejo, por lo cual se requiere utilizar fórmulas especiales que consideren 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 y el STATA.

Las variables que se utilizarán en el procedimiento son:

- El Factor de expansión calculado de la manera que se indica en el acápite anterior.
- La variable Conglomerado que se encuentra en la base de datos de la muestra y el Marco.
- La variable Estrato combinado existente en la base de datos de la muestra y el Marco.