



Diseño Muestral Encuesta de Hogares 2022

Septiembre 2022

UNIDAD DE MUESTREO Y LOGISTICA

Contenido

1.	Introducción	4
2.	Aspectos Metodológicos de la Encuesta de Hogares 2022	4
3.	Diseño muestral	5
3.1	Universo.....	5
3.2	Tipo de Muestreo.....	5
3.3	Unidades básicas de investigación	5
3.4	Unidad de análisis.....	5
3.5	Unidad de muestreo	6
4.	Nivel de desagregación de la información	6
5.	Marco Muestral.....	6
6.	Tamaño de la muestra	7
6.1	Distribución de la Muestra por Departamento y Área Geográfica	8
7.	Selección de la muestra.....	9
8.	Factores de expansión.....	9
8.1.1	Por Incidencias Finales.....	9
8.1.2	Por Proyecciones de Población al 2022.....	10
9.	Ponderadores Normalizados (Sección Discriminación)	10
10.	Estimaciones y Errores de Muestreo.....	11
11.	Anominización de Códigos de UPM.....	11

Índice de tablas

Tabla 1: Marco Muestral 2012- Estrato Socioeconómico.....	7
Tabla 2: Tamaño de muestra de viviendas EH 2022.....	7
Tabla 3: Coeficiente de variación (CV) del tamaño de muestra planificada.....	8
Tabla 4: Distribución de la muestra de viviendas y UPM.....	8

1. Introducción

Las encuestas de hogares se desarrollaron bajo la siguiente cronología:

- En 1980, con las encuestas permanentes de hogares, que posteriormente fueron reemplazadas 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).
- En las gestiones 2003 y 2004 se inició la encuesta continua de hogares, con la culminación del programa MECOVI, 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 a la fecha se realiza las Encuestas de Hogares con desagregación a nivel nacional, nacional urbano, nacional rural y departamental, donde los tamaños de muestra son calculados de manera independiente por departamento excepto para los departamentos de Beni y Pando, donde las estimaciones deben ser considerados de manera conjunta.
- En la gestión 2020, debido a la pandemia la recolección de datos fue desarrollada considerando las normas generales de bioseguridad con el fin de precautelar la salud de los entrevistadores y del personal operativo, así mismo, el levantamiento de información se desarrolló en 2 fases: Fase 1, el operativo de campo consistió en listar a las Viviendas Urbanas; Fase 2, se aplicaron posteriormente los cuestionarios vía teléfono a las viviendas urbanas enlistadas y para el área rural y resto urbano se realizaron los Listados de Viviendas y entrevistas de manera consecutiva.
- A partir de la gestión 2021, se retoma el levantamiento del listado de viviendas y consecutivamente la entrevista (urbano y rural) en las viviendas seleccionadas de la UPM.

En lo general la encuesta tiene como objetivo generar información socioeconómica y demográfica de la población, medición de la pobreza en el país y cálculo de indicadores de las condiciones de vida de los hogares en Bolivia, para la formulación, evaluación, seguimiento de políticas y diseño de programas de acción contenidas en el Plan de Desarrollo Económico y Social y los Objetivos de Desarrollo Sostenible (ODS).

2. Aspectos Metodológicos de la Encuesta de Hogares 2022

2.1 Objetivo General del diseño muestral

El objetivo principal del diseño muestral de la EH-2022 es determinar el tamaño de muestra para recopilar información de los hogares, que permitirá obtener resultados confiables.

2.2 Objetivo Específico

Los objetivos específicos del diseño muestral son:

- Considerar los indicadores relacionados con el empleo, desempleo y la evolución de la pobreza, para el cálculo del tamaño de muestra, con resultados a nivel nacional, con desagregación nacional urbano y nacional rural y a nivel departamental, excepto para los departamentos de Beni y Pando que será de manera conjunta.
- Cálculo de probabilidades de selección y el cálculo de los factores de expansión.

3. Diseño muestral

3.1 Universo

Es el conjunto de hogares establecidos en viviendas particulares ocupadas dentro de los nueve (9) departamentos del país, en el periodo definido para la ejecución de la encuesta en la gestión 2022.

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.

3.2 Tipo de Muestreo

El diseño de la muestra para la EH-2022 posee las características de un muestreo probabilístico, estratificado y bi-etápico.

- **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**, las unidades objeto de estudio (viviendas/hogares) se encuentran agrupadas en áreas conformadas con fines de muestreo, estas agrupaciones se denominan conglomerados.
- **Estratificado**, porque previo a la selección de conglomerados, estos son clasificados con el propósito de mejorar la precisión de los estimadores y permitir un mejor control en la distribución de la muestra.
- **Bi-etápico**, 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

3.3 Unidades básicas de investigación

Las unidades básicas de la investigación son las viviendas particulares ocupadas.

3.4 Unidad de análisis

Las unidades de análisis son: el hogar como unidad de consumo, los miembros del hogar en cuanto a sus características sociodemográficas, ocupacionales y de ingreso.

3.5 Unidad de muestreo

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:

- UPM - La unidad primaria de muestreo es el conglomerado de viviendas, en las ciudades amanzanadas puede estar compuesta por una o más manzanas y en área dispersa está compuesta por la comunidad o parte de ella.
- USM - La unidad secundaria de muestreo conforman las viviendas particulares dentro de las UPM.

4. Nivel de desagregación de la información

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

- Nacional
- Nacional Urbano
- Nacional Rural
- 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 a los planteados en el diseño, estará sujeto a la valoración de los errores muestrales.

5. Marco Muestral

Características del Marco Muestral MM-2012:

- Está basada con los datos del *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)*. Por tanto, el MM-2012 se describe como un marco de áreas y listas.
- Cubre todo el territorio nacional incluyendo las áreas de difícil acceso.
- Está constituido por unidades primarias de muestreo-UPM que son áreas geográficas codificadas alfa numéricas a nivel nacional (urbana y rural), con 102.593 conglomerados.
- El MM-2012 contiene la estratificación de áreas geográficas y socioeconómicas

Para la construcción de la variable de estratificación socioeconómica se consideró las variables estructurales del Censo de Población y Vivienda 2012: Características de la vivienda; Acceso a servicios básicos; Hacinamiento; Capital Humano; Educación; Patrimonio de los hogares; y Capacidad para la generación de ingresos.

El MM cuenta de cuatro (4) estratos socioeconómicos que se describen en la tabla siguiente:

Tabla 1: Marco Muestral 2012- Estrato Socioeconómico

Estrato socioeconómico	Descripción	Reagrupación socioeconómica
1	Bajo	1. Bajo
2	Medio Bajo	
3	Medio Alto	2. Alto
4	Alto	

Fuente: Instituto Nacional de Estadística

Los planos y mapas cartográficos digitales se constituyen parte del marco muestral (MM-2012) para la encuesta, mediante el cual se identifican las áreas correspondientes a las Unidades Primarias de Muestreo- UPM seleccionadas y se utilizan como guía para la ubicación en el terreno y el desarrollo del plan de recorridos de las brigadas de encuestadores y supervisores.

6. Tamaño de la muestra

Las variables que han sido, determinantes para el cálculo del tamaño de muestra para el presente estudio fueron: 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 de la Encuesta de Hogares 2021.

El tamaño de muestra para la EH 2022 es de 12.816 viviendas, considerando muestras independientes en cada uno de los departamentos, ver tabla 2.

Tabla 2: Tamaño de muestra de viviendas EH 2022
Según departamento

Muestra	
Departamento	Numero de Viviendas
Nacional	12.816
Chuquisaca	936
La Paz	3.048
Cochabamba	2.328
Oruro	948
Potosí	948
Tarija	996
Santa Cruz	2.316
Beni	708
Pando	588

Fuente: Instituto Nacional de Estadística

El cuadro de abajo muestra los coeficientes de variación obtenido para el tamaño de muestra planificado, considerando los dominios de estudio.

Tabla 3: Coeficiente de variación (CV) - tamaño de muestra planificado

Departamento/Ciudad	Número de viviendas (muestra planificada 2022)	Coeficiente de Variación (CV) obtenidos para:		
		desempleo	pobreza	pobreza extrema
Nacional	12.816	---	2,5%	6,2%
Chuquisaca	936	---	9,2%	10,0%
La Paz	3.048	7,7% *	10,8%	15,3%
Cochabamba	2.328	9,4% *	10,2%	15,2%
Oruro	948	---	13,3%	19,1%
Potosí	948	---	8,3%	10,0%
Tarija	996	---	12,6%	17,2%
Santa Cruz	2.316	10,2% *	13,5%	25,3%
Beni	708	---	14,4%	22,2%
Pando	588	---	6,3%	14,2%

Fuente: Instituto Nacional de Estadística

* Ciudades capitales de La Paz, Cochabamba y Santa Cruz

6.1 Distribución de la Muestra por Departamento y Área Geográfica

La Tabla 4. presenta la distribución de la muestra de viviendas y UPM por departamento y área geográfica.

Dentro de las UPM se selecciona 12 viviendas.

Tabla 4: Distribución de la muestra de viviendas y UPM según departamento y área geográfica

DEPARTAMENTO	Número de viviendas de la Muestra				Numero de UPM de la Muestra			
	Total	Urbano	Resto Urbano	Rural	Total	Urbano	Resto Urbano	Rural
Nacional	12.816	7.020	2.724	3.072	1.068	585	227	256
Chuquisaca	936	420	108	408	78	35	9	34
La Paz	3.048	2.184	456	408	254	182	38	34
Cochabamba	2.328	1.476	492	360	194	123	41	30
Oruro	948	444	108	396	79	37	9	33
Potosí	948	180	252	516	79	15	21	43
Tarija	996	396	324	276	83	33	27	23
Santa Cruz	2.316	1.488	552	276	193	124	46	23
Beni	708	144	384	180	59	12	32	15
Pando	588	288	48	252	49	24	4	21

Fuente: Instituto Nacional de Estadística

7. Selección de la muestra

Las etapas de selección son:

- En la primera etapa, se selecciona la UPM, de manera sistemática con Probabilidad Proporcional al Tamaño (PPT) de viviendas.
- En la segunda y última etapa, se seleccionan a las viviendas, de manera sistemática con arranque aleatorio.

8. Factores de expansión.

El factor de expansión es el inverso de la probabilidad de selección de la vivienda. Para la construcción del factor de expansión final se considera: el ajuste de la no-respuesta, los resultados de incidencias de campo y el total poblacional proyectado para la presente gestión, los factores de expansión son calculados una vez realizada la encuesta.

Las fórmulas del cálculo de las probabilidades, vienen dadas por:

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

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 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, $c = 12$.

8.1 Ajustes a los Factores de Expansión.

8.1.1 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: $\frac{c}{V_{jh}}$

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 son 12, dentro de la UPM seleccionada.

8.1.2 Por Proyecciones de Población al 2022.

Para estimar el total de población, el factor de expansión considera las proyecciones de población que genera el INE para la presente gestión, por departamento.

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.

9. Ponderadores Normalizados (Sección Discriminación)

La sección de discriminación del cuestionario se aplica a personas de 15 años o más, en la que se selecciona a una persona del hogar que habita en la vivienda seleccionada mediante el método *KISH*.

El ponderador normalizado para la muestra de la sección de discriminación considera que la suma de las ponderaciones sea del total de registros, mientras que la media de la variable de ponderación es igual a uno.

Para el cálculo del ponderador normalizado:

F_{jh} es igual w_k^*

$$w_k^* = nper_k * w_k$$

Dónde:

$nper_k$: Total de personas elegibles de 15 años o mas de la k – esima vivienda

w_k : Factor de expansión de la encuesta Hogares

Seguidamente se realiza un ajuste por proyección de población considerando 18 Grupos (nueve (9) departamentos y dos (2) áreas dentro del departamento).

$$w_k^{**} = w_k^* \frac{N_g}{\hat{N}_{g,\pi}} \quad g = 1, 2, \dots, 18$$

N_g : Proyecciones de población de 15 años y mas del g – esimo grupo

$\hat{N}_{g,\pi}$: Estimación poblacional de 15 años y mas del g – esimo grupo

Posteriormente, se obtiene el factor ajustado:

$$w_k^{***} = \frac{w_k^{**}}{nper_k}$$

Finalmente, la normalización se realiza por área, entonces la fórmula empleada está dada por:

$$w_{kj}^{Norm} = \frac{w_{kj}^{***}}{\text{Promedio}(w_{kj}^{***})} = \frac{w_{kj}^{***}}{\frac{\sum_{k=1}^{n_j} w_{kj}^{***}}{n_j}} = \frac{w_{kj}^{***} * n_j}{\sum_{k=1}^{n_j} w_{kj}^{***}}$$

Donde:

n_j : Tamaño de la j – esima área (1 = Urbano, 2 = Rural)

w_{kj}^{Norm} : Ponderador normalizado del k – esimo registro j – esima área

Debido a que los ponderadores normalizados carecen de escala, son válidos para la obtención de estimaciones relativas.

10. Estimaciones y Errores de Muestreo.

Como el diseño muestral empleado es complejo, se requiere utilizar fórmulas especiales que consideran los efectos de los conglomerados y la estratificación al momento de obtener las estimaciones de los indicadores de la encuesta.

Para el cálculo de los intervalos de confianza, coeficiente de variación, error estándar y efecto de diseño, entre otros, es necesario aplicar el plan del diseño de la muestra, considerando los factores de expansión.

Se recomienda valerse de algún software estadístico que permita realizar el cálculo de los errores muestrales aplicando el diseño de muestra empleado.

Se considera los siguientes estándares para evaluación de los errores muestrales (CV):

Optimo: CV <10%

Bueno: CV entre el 10% a 15%

Regular: CV entre el 16% a 20%

**No significativo: CV >20%

**Indicador sujeto a evaluación.

11. Anonimización de Códigos de UPM.

El término anonimizado implica, además de haber retirado los identificadores directos del informante, como ser los nombres, carnets de identidad, teléfono, direcciones, entre otros, se realiza la anonimización de los identificadores geográficos de la UPM, con el fin de resguardar la información proporcionada por los informantes y reducir al mínimo el riesgo de identificar a los mismos. Así mismo, se considera la **Ley N° 1405 vigente, sobre el Secreto Estadístico, donde la información proporcionada por los informantes, no podrá ser revelada en forma individual y solo podrá ser publicada de forma anonimizada.**