



Diseño de muestra Encuesta de Hogares 2021

Diciembre 2021

UNIDAD DE MUESTREO Y LOGISTICA

Contenido

1.	Introducción	4
2.	Aspectos Metodológicos de la Encuesta de Hogares 2021	4
3.	Diseño muestral.....	4
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.....	5
4.	Nivel de desagregación de la información	6
5.	Marco Muestral	6
5.1	Criterios de Conglomeración y Estratificación del Marco Muestral	7
5.1.1	Variable de Conglomeración	7
5.1.2	Variable de Estratificación	7
5.1.3	Variable de Estratificación Socio-Económico	8
6.	Tamaño de la muestra	8
6.1	Distribución de la Muestra por Departamento y Área Geográfica.....	10
7.	Selección de la muestra.....	10
8.	Factores de expansión.	10
8.1.1	Por Incidencias Finales.....	11
8.1.2	Por Proyecciones de Población al 2021.....	11
9.	Estimaciones y Errores de Muestreo.....	12

Índice de tablas

Tabla 1: Estructura del Marco Muestral.....	6
Tabla 2: Distribución de los Conglomerados o UPM según departamento	7
Tabla 3: Estrato geográfico – Marco Muestral.....	7
Tabla 4: Agrupación del estrato geográfico	8
Tabla 5: Estrato socioeconómico – Marco Muestral.....	8
Tabla 6: Tamaño de muestra EH-21	9
Tabla 7: Coeficiente de variación (CV) del tamaño de muestra planificada.....	9
Tabla 8: Distribución de la muestra de viviendas y UPM según departamento y área geográfica	10

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.

La Encuesta de Hogares 2021, tiene como objetivo proporcionar información: demográfica y estadísticas e indicadores socioeconómicos de la población boliviana, necesarias para la formulación, seguimiento, evaluación de políticas y programas dentro del PDES 2021-2025.

2. Aspectos Metodológicos de la Encuesta de Hogares 2021

2.1 Objetivo General del diseño muestral

El objetivo principal del diseño muestral de la EH-2021 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 2021.

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-2021 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**, el marco muestral construido en base al CNPV 2012 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 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)*. 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 alfanuméricamente a nivel nacional (urbana y rural).

La estructura del Marco Muestral 2012 está definida por las siguientes variables:

Tabla 1: Estructura del Marco Muestral

Tipo de Variable	Nombre de la variable	Variable
División Política Administrativa	I02_DEPTO	Departamento: código
	I03_PROV	Provincia: código y descripción
	NOMBRE_PROVINCIA	
	I04_MUN	Municipio: código y descripción
	NOMBRE_MUNICIPIO	
Organización municipal administrativa	I04_MUN	Municipio: código y descripción
	NOMBRE_MUNICIPIO	
	ID_UPM	UPM

Para el diseño de muestra	Categoría_URBRUR	Estrato geográfico
	URBRUR	Área Urbana o Rural
	Vivienda_ocu	Número de viviendas Ocupadas
	totperso	Total de personas
	totalHombre	Total de personas hombres
	totalMujer	Total de personas mujeres

Fuente: Instituto Nacional de Estadística

Los planos y mapas cartográficos forman parte del Marco Muestral 2012, estos identifican las áreas correspondientes a las Unidades Primarias de Muestreo – UPM. Los mapas se utilizan como guía para la ubicación de las brigadas de encuestadores y supervisores en el terreno.

5.1 Criterios de Conglomeración y Estratificación del Marco Muestral

5.1.1 Variable de Conglomeración

El Marco Muestral 2012, está constituido por 102.593 conglomerados (UPM) a nivel nacional. La distribución de los conglomerados por departamento, se detallan en el cuadro de abajo:

Tabla 2: Distribución de los Conglomerados o UPM según departamento

Departamento	Frecuencia	Porcentaje
Total	102.593	100
Chuquisaca	5.362	5,2
La Paz	29.610	28,9
Cochabamba	18.041	17,6
Oruro	7.384	7,2
Potosí	8.519	8,3
Tarija	4.578	4,5
Santa Cruz	23.690	23,1
Beni	4.350	4,2
Pando	1.059	1

Fuente: Instituto Nacional de Estadística

5.1.2 Variable de Estratificación

Con el propósito de obtener precisión en los estimadores, en la EH 2021 se utiliza los siguientes estratos geográficos y estratos estadísticos socio-económicos.

Tabla 3: Estrato geográfico – Marco Muestral

Estrato Geográfico	Descripción
--------------------	-------------

1	Ciudades Capitales
2	Ciudades Intermedias (10.000 y más habitantes)
3	Centros Poblados (2000 a 9.999 habitantes)
4	Localidades con características amanzanadas (250 a 1.999 habitantes)
5	Áreas Dispersas (menos de 250 habitantes)
6	Ciudad de El Alto
7	Áreas Dispersas (mayor a 2.000 habitantes)

Fuente: Instituto Nacional de Estadística

Los estratos mencionados son clasificados en:

Tabla 4: Agrupación del estrato geográfico

Estrato Geográfico	Descripción	Reagrupación geográfica
1	Ciudades Capitales	1. Urbano
2	Ciudades Intermedias (10.000 y más habitantes)	
3	Centros Poblados (2.000 a 9.999 habitantes)	
6	Ciudad de El Alto	
4	Localidades con características amanzanadas (250 a 1.999 habitantes)	2. Rural
5	Áreas Dispersas (menos de 250 habitantes)	
7	Áreas Dispersas (mayor a 2.000 habitantes)	

Fuente: Instituto Nacional de Estadística

5.1.3 Variable de Estratificación Socio-Económico

Para la construcción de la variable de estratificación socio-económica se consideró las siguientes 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.

Para la encuesta EH-2021 se considera la combinación de estas variables (estrato geográfico y estadístico socio-económico), que permite corregir las variaciones de las estimaciones de los estratos.

Tabla 5: Estrato socioeconómico – Marco Muestral

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

Fuente: Instituto Nacional de Estadística

6. Tamaño de la muestra

Las variables que han sido, usadas para el cálculo del tamaño de muestra fueron: la tasa de desempleo y pobreza, obtenidos con la información de la Encuesta de Hogares 2020.

El tamaño de muestra para la EH 2021 es de 12.948 viviendas, considerando muestras independientes para cada uno de los departamentos, ver tabla 6.

Tabla 6: Tamaño de muestra EH-21

Muestra 2021	
Departamento	EH 2021
Total, Nacional	12.948
Chuquisaca	1.008
La Paz	3.084
Cochabamba	2.304
Oruro	1.044
Potosí	936
Tarija	1.020
Santa Cruz	2.256
Beni	708
Pando	588

Fuente: Instituto Nacional de Estadística

Para garantizar la precisión en los estimadores para la Encuesta a Hogares 2021, se calculó el tamaño de muestra en cada uno de los dominios de estudio, considerando el indicador de pobreza y la tasa de desempleo (esta última tasa se emplea para el cálculo del tamaño de muestra en las ciudades capitales de La Paz, Cochabamba y Santa Cruz), obtenidas en la gestión 2020. El cuadro de abajo muestra los coeficientes de variación esperados de acuerdo al tamaño de muestra de cada departamento.

Tabla 7: Coeficiente de variación (CV) del tamaño de muestra planificada

Departamento/Ciudad	Número de viviendas (muestra planificada 2021)	Coeficiente de Variación esperado (CV)		
		CV de desempleo	CV de pobreza	CV de pobreza extrema
Total	12.948	---	2,75%	5,53%
Chuquisaca	1.008	---	6,83%	12,80%
La Paz (Ciudad Capital)	1.188	12,61%	5,58%	12,92%
La Paz (Resto)	816	---		
La Paz - El Alto	1.080	13,87%		
Cochabamba (Ciudad Capital)	1.452	14,68%	5,58%	12,26%
Cochabamba (Resto)	852	---		
Oruro	1.044	---	9,54%	13,74%
Potosí	936	---	6,03%	10,04%
Tarija	1.020	---	9,20%	13,78%
Santa Cruz (Ciudad Capital)	1.380	11,11%	9,72%	18,09%
Santa Cruz (Resto)	876	---		
Beni	708	---	9,15%	18,46%
Pando	588	---	10,01%	20,27%

Fuente: Instituto Nacional de Estadística

6.1 Distribución de la Muestra por Departamento y Área Geográfica

La Tabla 8. Presenta la distribución de la muestra de viviendas y UPM por departamento y área geográfica.

Dentro de las UPM se selecciona 12 viviendas.

Tabla 8: 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 (Capital/Conurbación)	Resto Urbano	Rural	Total	Urbano (Capital/Conurbación)	Resto Urbano	Rural
Total Nacional	12.948	7.080	2.760	3.108	1.079	590	230	259
Chuquisaca	1.008	456	120	432	84	38	10	36
La Paz	3.084	2.268	432	384	257	189	36	32
Cochabamba	2.304	1.452	492	360	192	121	41	30
Oruro	1.044	492	120	432	87	41	10	36
Potosí	936	192	240	504	78	16	20	42
Tarija	1.020	408	324	288	85	34	27	24
Santa Cruz	2.256	1.380	600	276	188	115	50	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, se realiza de manera sistemática con Probabilidad Proporcional al Tamaño (PPT) de viviendas.
- En la segunda y última etapa, se selecciona 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 ese año, 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: $\frac{c}{V_{jh}}$

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 2021.

Para estimar el total de población, se corrige el factor de expansión con las proyecciones de población que genere el INE para la gestión, 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:

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

9. Estimaciones y Errores de Muestreo.

Como el diseño muestral empleado es complejo, 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 el cálculo de los intervalos de confianza, coeficiente de variación, error estándar, 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; entre los programas estadísticos más utilizados para realizar este procedimiento están el SPSS.

10. 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 (UPM), con el fin de resguardar la información y reducir al mínimo el riesgo de identificar a los informantes mediante los códigos de la UPM relacionados con la cartografía estadística, y por la ley vigente, garantizar la confidencialidad y el secreto estadístico.