

Diseño de muestra Encuesta de Hogares 2020

Diciembre 2020

UNIDAD DE MUESTREO Y LOGÍSTICA

Contenido

1.	Introducción	4
2.	Aspectos Metodológicos de la Encuesta de Hogares 2020	4
3.	Diseño muestral.....	5
3.1	Universo	5
3.2	Tipo de Muestreo	5
3.3	Unidades básicas de investigación	6
3.4	Unidad de análisis	6
3.5	Unidad de muestreo.....	6
4.	Nivel de desagregación de la información	6
5.	Marco Muestral	7
5.1	Criterios de Conglomeración y Estratificación del Marco Muestral.....	8
5.1.1	Variable de Conglomeración	8
5.1.2	Variable de Estratificación	8
5.1.3	Variable de Estratificación Estadística Socio-Económico	9
6.	Tamaño de la muestra	9
6.1	Distribución de la Muestra por Departamento y Área Geográfica.....	10
7.	Selección de la muestra	11
8.	Factores de expansión.	11
8.1.1	Por Incidencias Finales.....	12
8.1.2	Por Proyecciones de Población al 2020.....	12
9.	Estimaciones y Errores de Muestreo.....	13
9.1	Tasa de respuesta.....	13

Índice de tablas

Tabla 1: Estructura del Marco Muestral	7
Tabla 2: Distribución de los Conglomerados o UPM según departamento	8
Tabla 3: Estrato geográfico – Marco Muestral	8
Tabla 4: Agrupación del estrato geográfico	9
Tabla 5: Estrato socioeconómico – Marco Muestral	9
Tabla 6: Tamaño de muestra EH-20	10
Tabla 7: Coeficiente de variación (CV) del tamaño de muestra planificada.....	10
Tabla 8: Distribución de la muestra de viviendas y UPM según departamento y área geográfica	11
Tabla 9: Tasa de Respuesta según departamento y área.....	13

1. 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 para el mejoramiento de las encuestas y la Medición de Condiciones de Vida MECOVI. Con la culminación del programa MECOVI, se inició 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, 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.

Es en la gestión 2016 y 2019, en cumplimiento de la Ley Nro. 548 y la política de fortalecimiento y actualización de las estadísticas e indicadores sobre trabajo de Niñas, Niños y Adolescentes, el Instituto Nacional de Estadística, realiza la implementación de la Encuesta de Niñas, Niños y Adolescentes que realizan una Actividad Laboral o Trabajan, en coordinación con el Ministerio de Trabajo, Empleo y Prevención Social y el Ministerio de Planificación y Desarrollo.

La EH-2020, al ser un instrumento del INE, tiene como objetivo proporcionar estadísticas e indicadores socioeconómicos y demográficos de la población boliviana, necesarias para la formulación, evaluación, seguimiento de políticas y diseño de programas de acción contenidas en el PDES 2016 – 2021.

2. Aspectos Metodológicos de la Encuesta de Hogares 2020

2.1 Objetivo General del diseño muestral

El objetivo principal del diseño muestral de la EH-2020, es determinar el tamaño de muestra, para recopilar información de los hogares, que permitirá obtener resultados confiables.

2.2 Objetivo Especifico

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 cálculo de los factores de expansión.

3. Diseño muestral

El objetivo principal del diseño muestral de la EH-2020, es determinar el tamaño de muestra, para recopilar información de los hogares, que permitan obtener resultados a nivel departamental excepto para los departamentos de Beni y Pando.

3.1 Universo

La investigación está dirigida al conjunto de hogares establecidos en viviendas particulares ocupadas dentro del territorio de Bolivia, en el periodo definido para la ejecución de la encuesta en la gestión 2020.

3.2 Tipo de Muestreo

El diseño de la muestra para la EH-2020, posee las características de un muestreo probabilístico, estratificado y bi-etápico.

- a) **Probabilístico**, 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, estos, son clasificados con el propósito de mejorar la precisión de los estimadores y permite un mejor control en la distribución de la muestra.
- d) **Bi-etápico**, porque los elementos pertenecientes a la muestra se seleccionan en dos etapas:
 - Etapa 1: Selección de conglomerados (UPM), la selección se la realiza proporcional al tamaño de las viviendas.

- Etapa 2: selección de viviendas dentro la UPM elegida, la selección de esta segunda etapa es de manera sistemática con arranque aleatorio.

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

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

La información del *Marco Muestral (MM)* 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.

El *MM-2012* cubre todo el territorio nacional incluyendo las áreas de difícil acceso. Las unidades primarias de muestreo (UPM) son aéreas geográficas que se encuentran codificadas alfanuméricamente a nivel nacional (urbana y rural). Las unidades secundarias de muestreo (USM) están constituidas por las viviendas, cubriendo la totalidad de las viviendas particulares¹.

La conformación de las Unidades Primarias de Muestreo (UPM), en el área urbana y rural, está realizada de acuerdo con los límites geográficos y estadísticos bajo los siguientes criterios de Agrupación, División y Unicidad.

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

Tabla 1: 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	Provincia: código y descripción
	NOMBRE_PROVINCIA	
	I04_MUN	Municipio: código y descripción
	NOMBRE_MUNICIPIO	
Variables de organización municipal administrativa rural	I04_MUN	Municipio: código y descripción
	NOMBRE_MUNICIPIO	Comunidad: código y descripción
	I06_CIUCOM	
	NOMBRE_COMUNIDAD	
Variables de Organización Municipal administrativa urbana	I04_MUN	Municipio: código y descripción
	NOMBRE_MUNICIPIO	
	ID_UNICO	Manzana
Variables para el diseño de muestra	ID_UNICO_UPM	UPM dado por la unidad de cartografía
	UPM	UPM normalizado
	NUM_UPM	UPM dado por la unidad de Muestreo
	Categoría_URBRUR	Estrato geográfico
	URBRUR	Área Urbana o Rural
	Vivienda_ocu	Número de viviendas Ocupadas
	totperso	Total de personas

¹ Vivienda particular es aquella que está habitada hasta por tres hogares, una vivienda con más de tres hogares es reconocida como vivienda colectiva (por ejemplo: hospitales, cuarteles, etc.)

	totalHombre	Total de personas hombres
	totalMujer	Total de personas mujeres

Fuente: Instituto Nacional de Estadística

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 - UPM seleccionadas y se utiliza como guía para ubicación en el terreno de las brigadas de encuestadores y supervisores.

5.1 Criterios de Conglomeración y Estratificación del Marco Muestral

5.1.1 Variable de Conglomeración

A partir del 2014, en las encuestas de hogares se utiliza el Marco Muestral 2012, la cual se encuentra basada en la información del Censo de Población y Vivienda 2012 (CNPV-2012), y 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 2020 se utilizaron los siguientes estratos geográficos y estratos estadísticos socio-económicos.

Tabla 3: Estrato geográfico – Marco Muestral

Estrato Geográfico	Descripción
1	Ciudades Capitales
2	Ciudades Intermedias
3	Centros Poblados
4	Localidades con características ameznadas < 2.000 habitantes

5	Áreas Dispersas
6	Ciudad de El Alto
7	Áreas Dispersas > 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	
3	Centros Poblados	
6	Ciudad de El Alto	
4	Localidades con características amanzanadas con menores a 2.000 habitantes	2. Rural
5	Áreas Dispersas	
7	Áreas Dispersas, mayores a 2.000 habitantes	

Fuente: Instituto Nacional de Estadística

5.1.3 Variable de Estratificación Estadística Socio-Económico

Para la construcción de la variable de estratificación estadística socio-económica se consideró variables estructurales del Censo de Población y Vivienda 2012, entre ellas: 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-2020 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 Estadístico	Descripción	Reagrupación Estadístico
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, determinantes para el cálculo del tamaño de muestra fueron: la tasa de desempleo y pobreza, obtenidos con la información de la Encuesta de Hogares 2019.

El tamaño de muestra determinada para la EH 2020 fue de 11292 viviendas, considerando muestras independientes para cada uno de los departamentos, ver tabla 6.

Tabla 6: Tamaño de muestra EH-20

Muestra 2020	
Departamento	EH 2020
NACIONAL	11.292
Chuquisaca	852
La Paz	2.568
Cochabamba	2.136
Oruro	828
Potosí	708
Tarija	936
Santa Cruz	1.968
Beni	708
Pando	588

Fuente: Instituto Nacional de Estadística

Para garantizar la precisión en los estimadores para la encuesta a hogares, se calculó el tamaño de muestra en cada uno de los dominios de estudio, considerando el indicador de pobreza. 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.

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

Departamento/Ciudad	Número de viviendas (muestra planificada 2020)	Coeficiente de Variación esperado (CV) en base a la Gestión 2019		
		CV de desempleo	CV de pobreza	CV de pobreza extrema
Chuquisaca	852	---	8,60%	17,50%
La Paz (Ciudad Capital)	984	8,10%	4,60%	13,40%
La Paz (No Ciudad Capital)	600	---		
La Paz - El Alto	984	---		
Cochabamba (Ciudad Capital)	1356	8,80%	4,90%	14,00%
Cochabamba (No Ciudad Capital)	780	---		
Oruro	828	---	8,60%	16,20%
Potosí	708	---	8,90%	16,00%
Tarija	936	---	6,20%	14,40%
Santa Cruz (Ciudad Capital)	1176	11,20%	6,20%	15,60%
Santa Cruz (No Ciudad Capital)	792	---		
Beni y Pando	1296		8,50%	26,90%
Total	11292			

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 por departamento y área geográfica, así como la distribución de las UPM.

Las UPM tienen una selección 12 viviendas.

Tabla 8: Distribución de la muestra de viviendas y UPM según departamento y área geográfica

Departamento	Distribución de Viviendas				Distribución de UPM			
	Total	Urbano	Resto Urbano	Rural	Total	Urbano	Resto Urbano	Rural
NACIONAL	11.292	6.900	1.716	2.676	941	575	143	223
Chuquisaca	852	384	96	372	71	32	8	31
La Paz	2.568	2.220	60	288	214	185	5	24
Cochabamba	2.136	1.680	120	336	178	140	10	28
Oruro	828	396	96	336	69	33	8	28
Potosí	708	144	180	384	59	12	15	32
Tarija	936	372	300	264	78	31	25	22
Santa Cruz	1.968	1.272	432	264	164	106	36	22
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

En las diferentes etapas la selección de las UPM y viviendas fueron de la siguiente manera:

- En la primera etapa (selección de UPM) fue considerando la Probabilidad Proporcional al Tamaño (PPT) de viviendas.
- En la última etapa la selección de viviendas fue sistemática con arranque aleatorio.

8. 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 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'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, $c = 12$.

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.

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 = 12

8.1.2 Por Proyecciones de Población al 2020.

Para estimar el total de población, se corrige el factor de expansión con las proyecciones de población que genere el INE para la gestión 2020, 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:

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

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 calcular los estimadores, es emplear 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.

9.1 Tasa de respuesta

La tasa de respuesta a nivel nacional es del 97,56%. En la Tabla 9. Se detalla la tasa de respuesta según departamento y área de la muestra ejecutada.

Tabla 9: Tasa de Respuesta según departamento y área

Departamento	Muestra planificada (viviendas)				Muestra Ejecutada (viviendas)				Tasa de respuesta (%)			
	Total	Urbano	Resto Urbano	Rural	Total	Urbano	Resto	Rural	Total	Urbano	Resto Urbano	Rural
							Urbano					
Nacional	11.292	6.900	1.716	2.676	11.016	6.651	1.700	2.665	97,56%	96,4%	99,1%	99,6%
Chuquisaca	852	384	96	372	845	378	96	371	99,18%	98,4%	100,0%	99,7%
La Paz	2.568	2.220	60	288	2.508	2.161	60	287	97,66%	97,3%	100,0%	99,7%
Cochabamba	2.136	1.680	120	336	2.056	1.601	120	335	96,25%	95,3%	100,0%	99,7%
Oruro	828	396	96	336	816	386	96	334	98,55%	97,5%	100,0%	99,4%
Potosí	708	144	180	384	700	140	177	383	98,87%	97,2%	98,3%	99,7%
Tarija	936	372	300	264	924	362	298	264	98,72%	97,3%	99,3%	100,0%
Santa Cruz	1.968	1.272	432	264	1.879	1.197	421	261	95,48%	94,1%	97,5%	98,9%
Beni	708	144	384	180	704	140	384	180	99,44%	97,2%	100,0%	100,0%
Pando	588	288	48	252	584	286	48	250	99,32%	99,3%	100,0%	99,2%

Fuente: Instituto Nacional de Estadística

Nota.- Se han excluido 40 boletas correspondiente a viviendas omitidas (20) y viviendas con más de un hogar (20), sin embargo se encuentran en la base publicada.