

2 CENSOS, 10 AÑOS DE ENCUESTAS A HOGARES EN BOLIVIA TIEMPO DE REPONDERAR Y DEFINIR UN DISEÑO MUESTRAL COMPARABLE PARA LOS INDICADORES DE BIENESTAR¹

Lic. Chirino Gutiérrez, Álvaro

✉ achirino@aru.org.bo

RESUMEN

Este documento propone un diseño muestral para el periodo 2002 - 2012 de la serie de encuestas a hogares que llevo a cabo el Instituto Nacional de Estadística de Bolivia, el diseño se orienta en lograr un rendimiento muestral optimo en los indicadores de pobreza moderada y pobreza extrema a nivel departamental (primera unidad sub-nacional) y nacional. Existen dos principales motivaciones para realizar este nuevo diseño; el primero, los diseños muestrales existentes en las encuestas de la serie difieren en varias características, como ser: la estratificación, la conglomeración, el marco muestral, la asignación de la muestra, el tamaño de muestra, et.al. Estas variaciones no permiten una lectura consistente de la evolución de los indicadores de bienestar a lo largo de estos años. La segunda motivación, es la pertinencia, el 2012 se realizó el último Censo de Población y Vivienda, los resultados de este censo muestran una estructura diferente a las proyecciones, que fueron la base de las estimaciones de las Encuestas a hogares. Y siguiendo las recomendaciones de la Naciones Unidas (2009) sobre las oportunidades de un nuevo censo, este es un momento idóneo para sugerir un nuevo diseño que traerá consigo la reponderación de la muestra.

PALABRAS CLAVE

Reponderación; Bolivia; Encuestas a hogares; Indicadores de Bienestar.

1. INTRODUCCIÓN

El Instituto Nacional de Estadística de Bolivia (INE), al igual que la mayoría de los países de la región, desarrolló los últimos años, encuestas a hogares (EH) orientadas a medir bienestar, esta práctica tuvo el incentivo inicial del Programa MECOVI del Banco Mundial. El 2001 Bolivia llevó a cabo el Censo Nacional de Población y Vivienda (CNPV) y como uno de los resultados naturales del censo, la información obtenida por éste, se convirtió en la base para el desarrollo de las encuestas, en noviembre de 2012 se realizó nuevamente el CNPV. En el periodo intercensal (periodo entre censos) se realizaron un total de 10 rondas de encuestas, desde el 2002 al 2012, algunas

particularidades en este periodo son; el 2003 y 2004 se realizó una encuesta de tipo continua destinada a construir una nueva canasta familiar para el IPC², de 2002 a 2005 se empleó una nueva muestra como marco muestral, el 2010 no se realizó la encuesta, a partir del 2011 la encuesta casi duplicó el tamaño de muestra de viviendas.

Los cuadros 1 y 2 presentan un resumen acerca del diseño muestral del periodo intercensal, se verifica que en la serie no existe una relación respecto al diseño de un año a otro, es más, casi ningún año es similar en diseño al otro, esto motiva a pensar que las estimaciones que provienen de estas encuestas no son comparables completamente. Así también, los resultados del CNPV-2012 muestran que

1 Este documento fue presentado en el *60th World Statistics Congress* desarrollado en Rio de Janeiro, Brasil en Julio 2015.

2 Índice de Precios al Consumidor

la estructura del país no es la misma a la CNPV-2012 registro un total de 10,059,856 habitantes, más de 800 mil habitantes de diferencia. 2012 expande a una población de alrededor 10,874,551 habitantes, sin embargo, el

Cuadro 1
Diseños muestrales de las Encuestas a Hogares en Bolivia 2002 a 2006

Año		2002	2003-2004	2005	2006
Marco		Muestra Maestra, CNPV-2001	Muestra Maestra, CNPV-2001	Muestra Maestra, CNPV-2001	CNPV-2001
Etapas	Urbano	2	2	2	2
	Rural	3	3	3	3
Cluster	Etapas 1	Grupo de sectores censales Segmentos			
	Etapas 2	Grupo de sectores censales Segmentos			
Estratificación, Ira. Etapa		Población, 5 Niveles	Población, 5 Niveles	Población, 5 Niveles	Area y NBI, 8 levels
Mecanismo de selección	Etapas 1	πps , Viviendas			
	Etapas 2	πps , Viviendas			
	Ultima etapa	Sistemática	Sistemática	Sistemática	Sistemática
Muestra	UPM	679	547	166	314
	Hogares	5746	8767	3865	4006
	Personas	24933	38500	16895	16511

Fuente: Elaboración propia

Cuadro 2
Diseños muestrales de las Encuestas a Hogares en Bolivia 2007 a 2012

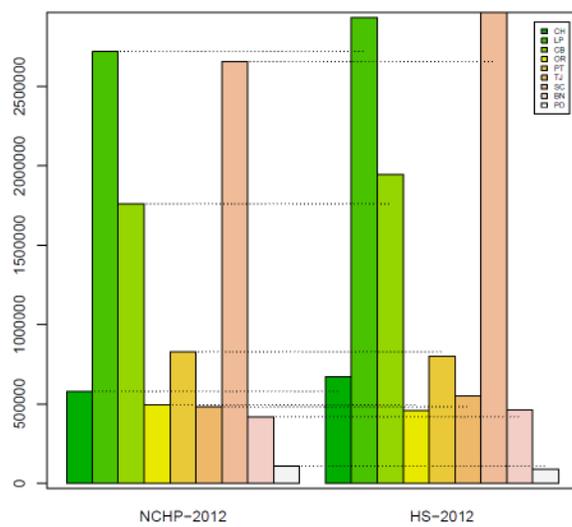
2007	2008	2009	2011	2012
CNPV-2001	CNPV-2001	CNPV-2001	CNPV-2001	CNPV-2001
2	2	2	2	2
3	3	3	2	2
Grupo de sectores censales	Grupo de sectores censales	Grupo de sectores censales	sector censal	sector censal
Grupo de segmentos	Grupo de segmentos	Grupo de segmentos	Grupo de segmentos	Grupo de segmentos
Área y NBI, 8 niveles	población y consumo	Área y NBI, 8 niveles	Población y NBI, 20 niveles	Población y NBI, 20 niveles
Sistemática πps , viviendas	Sistemática πps , hogares			
Desconocido	Desconocido	Desconocido		
Sistemática	Sistemática	Sistemática	Sistemática	Sistemática
309	304	301	790	729
4069	3856	3932	8851	8360
16804	15030	15665	33821	31935

Fuente: Elaboración propia

Un resultado aún más importante que el nacional son los referidos a los departamentales, estas son las primeras unidades sub nacionales de Bolivia y son los primeros actores locales de la política pública, es por ello que es necesaria una información

adecuada al nivel del departamento, la estructura reflejada por la EH-2012 a nivel departamental difiere de lo obtenido en el CNPV-2012, esta diferencia se la presenta en la Figura 1, donde se observa claramente que existen departamentos con sub y sobre cobertura.

Figura 1
Población estimada de la EH-2012 vs Resultados del CNPV - 2012



Fuente: Elaboración propia

2 Censos, 10 años de encuestas a hogares en Bolivia tiempo de reponderar y definir un diseño muestral comparable para los indicadores de bienestar

Este documento plantea un diseño muestral expost, con el fin de poder generar estimaciones comparables y coherentes a los resultados del CNPV-2012 de toda la serie, el diseño muestral final es el resultado de un contraste en el rendimiento de muestreo de cinco propuestas alternativas de diseño y un ajuste en los factores de expansión para que se adecuen al ritmo de crecimiento del periodo intercensal 2001 a 2012, el análisis de rendimiento muestral se lo hace en base a los indicadores de pobreza moderada y extrema a nivel de los departamentos.

El documento presenta en la sección 2 la metodología empleada, en esta se describen las fuentes de datos empleadas, la definición de pobreza en las EH, el tratamiento para los componentes del muestreo como ser la estratificación y conglomeración, la construcción de las probabilidades teóricas y el factor de expansión y la proyección empleada para el periodo intercensal, en la sección 3 se muestran los resultados referentes al rendimiento de los diseños planteados, la evolución de la pobreza y el diseño muestral final. Finalmente, en la sección 4 se dan las conclusiones del trabajo.

1.1 OBJETIVO

El objetivo general del trabajo es:

“Desarrollar un diseño muestral comparable para los indicadores de bienestar provenientes de la serie de encuestas a hogares 2002 a 2012, ajustado a los patrones de crecimiento poblacional de los Censos de Población y Vivienda 2001 y 2012” .

2. METODOLOGÍA

Con la finalidad de obtener un diseño muestral comparable para la serie de encuestas del

2002 a 2012 se optó por elaborar 5 distintos diseños muestrales y comparar el rendimiento de cada uno de ellos para los indicadores de pobreza moderada y extrema, para ello, se eligen los datos, se establece la definición de pobreza que seguirá el documento, se definen las características de conglomeración y estratificación, se realizan proyecciones para el periodo intercensal, se construyen las probabilidades teóricas y finalmente se elaboran los ponderadores muestrales ajustados a las proyecciones.

2.1 DATOS

Las bases de datos y la documentación metodológica de las encuestas a hogares fueron obtenidas del banco de datos sociales del portal del INE³, las bases de datos del Censo de Población y Vivienda 2001 que se emplea como la base del marco muestral de la serie, corresponde a la que distribuye el departamento de comunicación del INE . Finalmente, la información del CNPV-2012 fue obtenida del portal en versión Redatam del Censo⁴ .

2.2 POBREZA

Todas las encuestas a hogares del periodo de interés definen a la pobreza en base al ingreso y a su ubicación por encima o por debajo de la línea de pobreza, siguiendo la metodología de Foster, James, J. Greer, and Eric Thorbecke. 1984.

Donde la pobreza moderada para un individuo se define como:

$$p_i = I(y_i < z_k)$$

3 <http://www.ine.gob.bo:8081/Webine10/enchogares1.aspx>

4 <http://datos.ine.gob.bo/binbol/RpWebEngine.exe/Portal?LANG=ESP>

Donde $I(\cdot)$ es una función indicatriz respecto al ingreso (y_i) del individuo i y la línea de pobreza moderada (z_k) de la región k del país.

De similar forma se define a la pobreza extrema como:

$$pe_i = I(y_i < ze_k)$$

Donde $I(\cdot)$ es una función indicatriz respecto al ingreso (y_i) del individuo i y la línea de pobreza extrema (ze_k) de la región k del país.

El estudio emplea como un indicador agregado, a la incidencia de pobreza extrema y moderada, este se define dentro de un área k , como:

$$P_0 = \frac{\sum_k p_i}{N_k}$$
$$P_{0e} = \frac{\sum_k pe_i}{N_k}$$

Donde P_0 es la incidencia (headcount) de pobreza moderada y P_{0e} la incidencia de pobreza extrema.

En el estudio se emplea a partir del 2006 las variables de pobreza reportadas en las bases de datos, para los años inferiores se emplea la información de Hernani and Villarroel (2012), esto a razón de la ausencia de las variables de pobreza en las bases de datos antes del 2006.

2.3 CLUSTER

Durante el periodo de la serie de encuestas existieron 2 unidades primarias de muestreo (PSU), la primera estuvo desde el 2002 al 2009 y fue una agrupación de sectores censales que en promedio tenía 120 viviendas, en las encuestas a hogares de 2011 y 2012 las unidades primarias de muestreo fueron

los sectores censales que en promedio tenían 90 viviendas, todas estas unidades en base al CNPV-2001, tal como se mostró en el cuadro 1, algunos años de encuesta existieron dos etapas en el área rural, sin embargo, no existe la documentación necesaria para conocer el criterio de formación y selección de estas áreas, es por ello que para fines de este trabajo se define la existencia de 2 etapas; la selección del cluster y la selección del hogar.

La PSU será la denominada UPM, esto debido a su presencia en la mayoría de las encuestas y que en las EH 2011 y 2012 se las puede identificar sin dificultad.

2.4 LA ESTRATIFICACIÓN

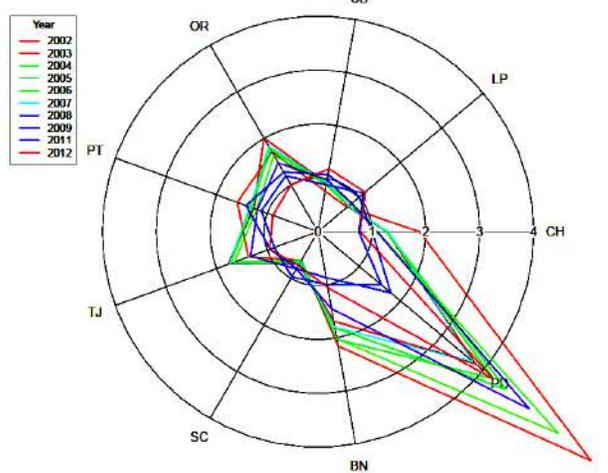
La estratificación es la forma central de la cual depende la estructura de un diseño muestral complejo, a lo largo de la serie de encuestas se han empleado diversos estratos, todos los documentos metodológicos previos al 2011 definen los estratos a nivel nacional y no incluyen un tratamiento a nivel departamental, sin embargo, el analizar las bases de datos se encuentra que existió un criterio de selección basada en los departamentos, esto debido al patrón de sobre y sub representación de algunos departamentos respecto a una estructura proporcional del marco muestral, la figura 2 muestra el patrón, donde los anillos representan el cociente entre la muestra asignada y la necesaria, se observa claramente el caso del departamento de Pando, que casi siempre recibió más de 3 veces la muestra necesaria, esto pone en evidencia que existió una selección basada en los departamentos.

Para el trabajo se definen 5 tipos de estratos a nivel de las PSU, cada uno combinado con los 9 departamentos, esto a razón de la Figura 2, los estratos son:

2 Censos, 10 años de encuestas a hogares en Bolivia tiempo de reponderar y definir un diseño muestral comparable para los indicadores de bienestar

- **Población:** 5 niveles al interior de los departamentos; ciudades capitales, ciudades intermedias, centros poblados mayores, centros poblados menores y área dispersa.
- **Geográfico:** 3 niveles; Altiplano, Valles y Llano.
- **Consumo:** 12 niveles al interior de los departamentos, fue construido en

Figura 2
Sub y sobre muestreo en la serie EH 2002 - 2012



Fuente: Elaboración propia

base a un modelo de regresión basado en las HS-1999 A 2002

básicas insatisfechas.

- **Componentes principales:** 5 niveles al interior de los departamentos, fue construido en basa a la técnica de los componentes principales empleando información del CNPV-2001.
- **NBI:** 4 niveles al interior de los departamentos, fue construido en base a la técnica del índice de necesidades

2.5 PROYECCIONES

Para las proyecciones de población en el periodo intercensal se empleó tasas de crecimiento exponencial, esta se define como:

$$r = \frac{\ln(P_{2012}) - \ln(P_{2001})}{Y_{2012-2001}}$$

Donde P_t corresponde a la población censada el año t y al tiempo $Y_{2012-2001}$ transcurrido

Cuadro 3

Tasa de crecimiento anual y población del CNPV 2001 y 2012 por departamento y área

Departamento	NCPH-2001			NCPH-2012			Tasa de crecimiento anual (%)		
	Urbano	Rural	Total	Urbano	Rural	Total	Urbano	Rural	Departamento
Chuquisaca	218126	313396	531522	283123	298224	581347	2.325	-0.442	0.799
La Paz	1552146	797739	2349885	1814148	905196	2719344	1.390	1.126	1.302
Cochabamba	856409	599302	1455711	1200912	561849	1762761	3.013	-0.575	1.706
Oruro	236110	156341	392451	316757	177830	494587	2.619	1.148	2.062
Potosí	239083	469930	709013	336412	491681	828093	3.044	0.403	1.384
Tarija	247736	143490	391226	314510	169008	483518	2.127	1.459	1.888
Santa Cruz	1545648	483823	2029471	2160579	497183	2657762	2.985	0.243	2.404
Beni	249152	113369	362521	308690	113318	422008	1.910	-0.004	1.354
Pando	20820	31705	52525	53831	56605	110436	8.467	5.166	6.624
Total	5165230	3109095	8274325	6788962	3270894	10059856	2.436	0.452	1.742

Fuente: Elaboración propia

en años desde el 2001 al 2012, la tasa corresponde al 21 de Noviembre de 2012 (día del CNPV-2012). La fórmula empleada para la proyección es:

$$Pr_i = P_{2001} * e^{r*a}$$

Donde P_i corresponde a la proyección para el año i y es el tiempo transcurrido en años desde la base 2001 al año de proyección, el cuadro 3 muestra las tasas de crecimiento anual por departamento y área.

2.6 PROBABILIDADES TEÓRICAS

Con las consideraciones de los conglomerados y los estratos se elaboran las probabilidades teóricas asociadas a las dos etapas del diseño, para la primera etapa, se asume una probabilidad proporcional al tamaño sin reposición, empleando el total de vivienda de las PSU registradas en el marco muestral del CNPV-2001, este tiene la forma de:

$$\pi_{Ii} = n_{Is} * \frac{t_{is}}{\sum_S t_{is}}$$

Donde π_{Ii} es la probabilidad de primera etapa para la PSU i , n_{Is} el tamaño de muestra de PSU en el nivel s de la estratificación y t_{is} el total de viviendas en la PSU i del nivel de la estratificación.

En general la probabilidad del hogar k , se define como el producto de la probabilidad de primera etapa y la última etapa.

$$\pi_{ki} = \pi_{Ii} * \frac{n_{IIIi}}{t_{is}}$$

Donde π_{ki} es la probabilidad para el individuo k de la PSU i , n_{IIIi} el tamaño de muestra de la última etapa.

Para el análisis se eliminaron las PSU con $n_{IIIi}=1$ para evitar el incremento de varianzas.

2.7 FACTORES DE EXPANSIÓN

La construcción de los factores de expansión para los 5 distintos tipos de estratos se inició con el factor teórico, este es el inverso de la probabilidad de selección:

$$W_{ki} = \left(\pi_{ki} * \frac{n_{IIIi}}{t_{is}} \right)^{-1}$$

Sobre este factor se incluyó un ajuste debido a las proyecciones P_i , este ajuste se lo hizo a nivel departamental y por área.

$$aW_{ki} = W_{ki} * \frac{P_l}{\sum_l W_{ki}}$$

Se acotó los factores finales en base al valor del percentil 99 de su distribución, esto siguiendo la recomendación de Naciones Unidas. (2009), con la finalidad de reducir la varianza de los estimadores de pobreza.

3. RESULTADOS

Esta sección presenta los resultados de los 5 tipos de diseño, mostrando la evolución de la pobreza, el rendimiento de muestreo y la elección del diseño final.

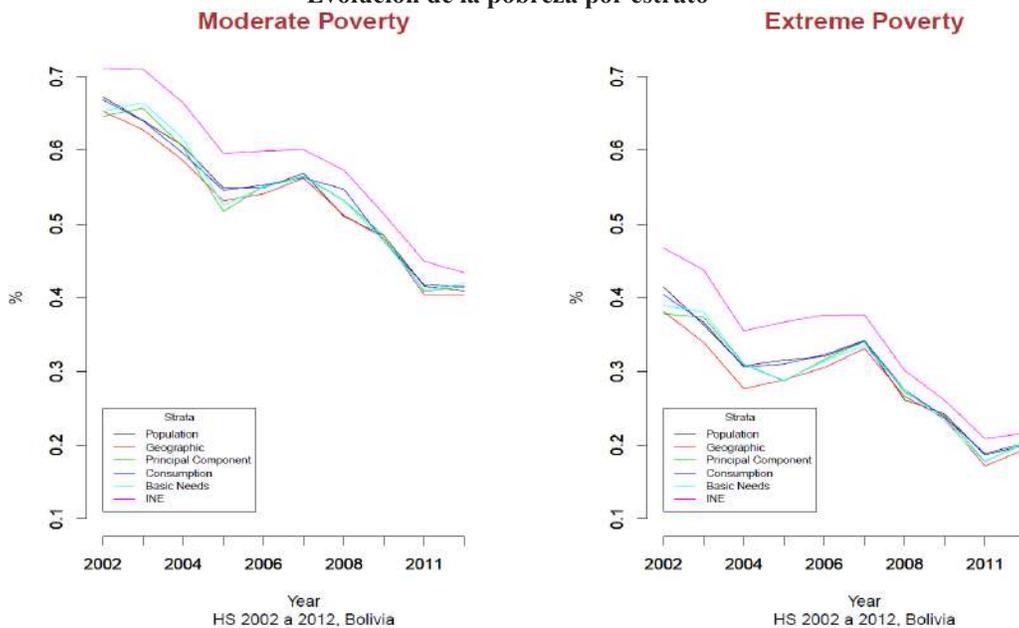
3.1 EVOLUCIÓN DE LA POBREZA

La figura 3, presenta la evolución de la pobreza para la serie de interés y considerando los 5 tipos de estratos definidos, se incluye la estimación oficial del INE. Se observa que todos los tipos de estratos estiman una incidencia menor a la oficial en ambos tipos de pobreza pero conservan la tendencia.

2 Censos, 10 años de encuestas a hogares en Bolivia tiempo de reponderar y definir un diseño muestral comparable para los indicadores de bienestar

Figura 3

Evolución de la pobreza por estrato



Fuente: Elaboración propia

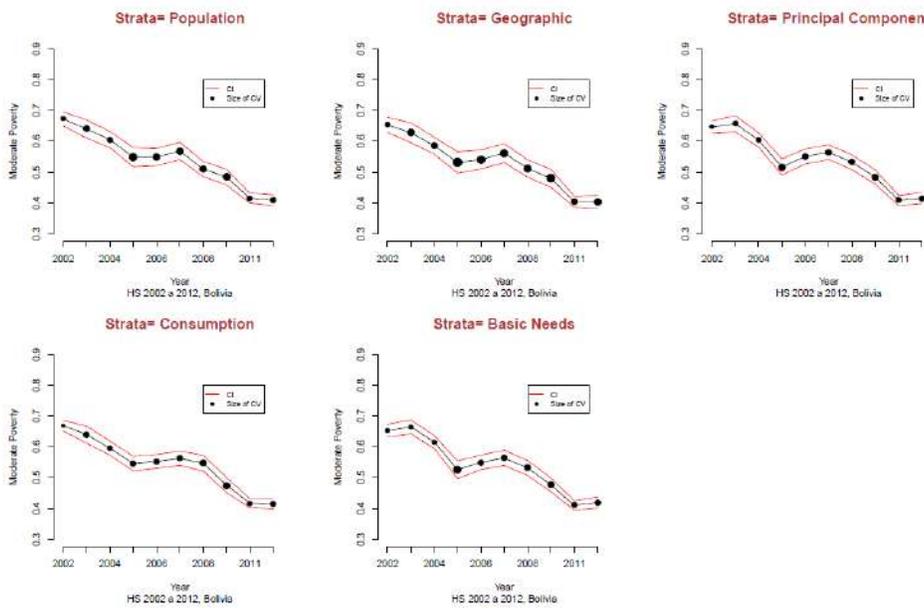
3.2 DESEMPEÑO DEL DISEÑO MUESTRAL

Para el rendimiento del muestreo se considera el coeficiente de variación muestral y el

efecto de diseño para los dos tipos de pobreza a nivel nacional y departamental, las figuras 4, 5 muestran la evolución de la pobreza moderada y extrema por el tipo de diseño e incluye el coeficiente de variación y los intervalos de confianza al 95 %.

Figura 4

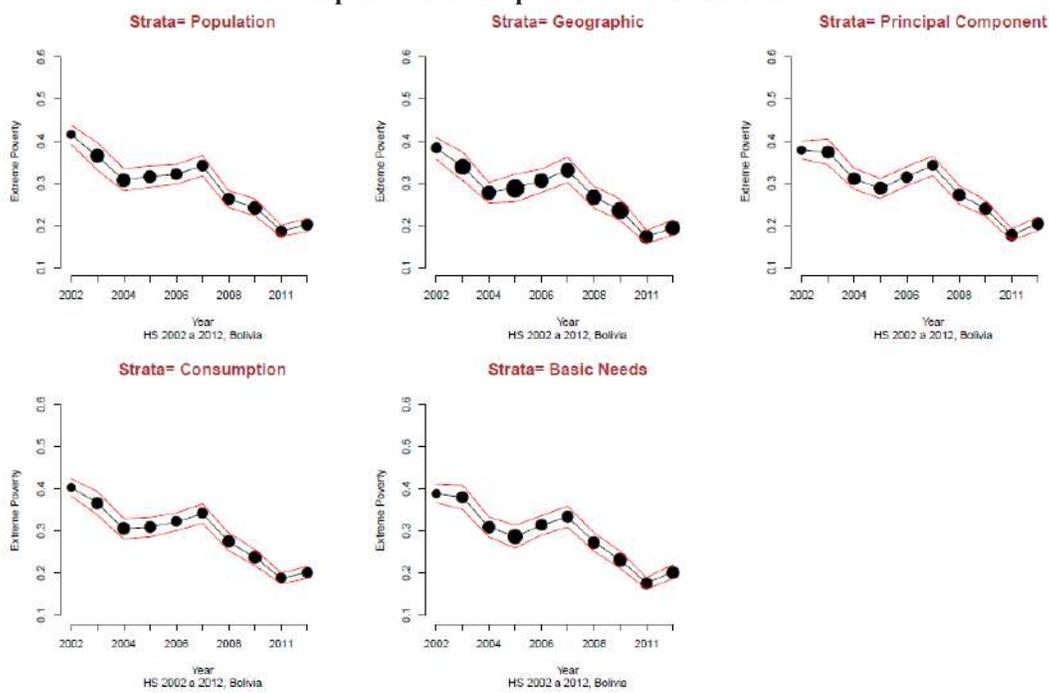
Desempeño del diseño para la Pobreza Moderada



Fuente: Elaboración propia

Figura 5

Desempeño del diseño para la Pobreza Extrema



Fuente: Elaboración propia

Se observa que uno de los estratos con menor rendimiento es el geográfico y que el de consumo y componentes son los de mejor rendimiento. El cuadro 4 presenta un resumen

del rendimiento a nivel departamental y nacional para la pobreza moderada y extrema.

Se observa que estrato de consumo es el que

Cuadro 4

Rendimiento promedio del muestreo por indicador y estrato EH 2002 - 2012

Pobreza Moderada				
Estrato	Nacional		Departamento	
	CV	DEFF	CV	DEFF
Población	2.364	13.180	8.567	10.295
Geográfico	2.679	16.077	9.427	12.735
Componentes principales	2.155	10.749	8.007	8.521
NBI	2.120	10.460	7.880	8.667
Consumo	2.070	10.028	7.719	7.918
Pobreza extrema				
Estrato	Nacional		Departamento	
	CV	DEFF	CV	DEFF
Población	3.883	12.512	15.032	10.505
Geográfico	4.767	17.117	18.081	13.918
Componentes principales	3.837	11.824	15.268	9.675
NBI	3.937	12.353	15.356	9.995
Consumo	3.595	10.593	14.336	8.626

Fuente: Elaboración propia

tienen un mejor rendimiento, seguido por el de componentes principales, luego el de necesidades básicas, el de población y el de más bajo rendimiento es el geográfico. El efecto de diseño es un indicador interesante de analizar, es superior a 7 en todos los casos,

esto se debe a la alta homogenización en las PSU.

3.3 DISEÑO FINAL

Por lo desarrollado se recomienda adoptar el diseño de muestra basado en el estrato de

consumo, puesto que es el que logra un mejor rendimiento en los indicadores de muestreo.

4. CONCLUSIÓN

Las conclusiones son:

- El diseño muestral que mejora los indicadores de rendimiento de la muestra es el diseño basado en el estrato de consumo.
- Existen en general efectos de diseños altos, es necesario reducir la muestra al interior de las PSU e incluir más PSU en la muestra.
- Los patrones de pobreza moderada y extrema están sobre estimados con el diseño oficial que no se ajusta al ritmo de crecimiento que existió en Bolivia en el periodo intercensal.
- Para el siguiente periodo intercensal es necesario tener una planificación previa del diseño muestral para toda la serie, logrando establecer efectivamente un sistema integrado de encuestas nacionales.



BIBLIOGRAFÍA

- Cochran, W. G. (1953). *Técnicas de Muestreo*. Mexico: John Wiley and Sons
- Foster, James, J. Greer, and Eric Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica* 52 (3): 761–65
- Hernani W., Villarroel P. (2012) *Evolution of poverty in Bolivia: a multidimensional approach*
- INE, Bolivia. (2002-2012). *Documento Metodológico, Encuesta a Hogares 2002 - 2012*. La Paz, Bolivia.
- Kish, L. (1955). *Survey Sampling*. New York: John Wiley and Sons. Lohr, S. L. (2000). *Muestreo: Diseño y Análisis*. Naciones Unidas. (1986). *Sampling frames and sample design for integrated household survey programmers*.
- Naciones Unidas. (2009). *Diseño de muestras para encuestas de hogares Directrices prácticas* (p. 252). Nueva York.
- Rosén, B. (1997). *On sampling with probability proportional to size*. J. Statistics, Plann Inference. Särndal, Swensson, y Wretman. (1992). *Model Assisted Survey Sampling*. Canada: Springer, Verlang