



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Estudios de Postgrado
Maestría en Artes en Estadística Aplicada

**USO DE METODOLOGÍAS DE ESTIMACIÓN EN ÁREAS PEQUEÑAS PARA
EL CÁLCULO DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL DE
GUATEMALA IPM-GT A NIVEL DE DEPARTAMENTO**

Lic. Luis Fernando Castellanos Bonilla
Asesorado por el Msc. Adonai Navas García

Guatemala, enero de 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**USO DE METODOLOGÍAS DE ESTIMACIÓN EN ÁREAS PEQUEÑAS PARA
EL CÁLCULO DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL DE
GUATEMALA IPM-GT A NIVEL DE DEPARTAMENTO**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

LIC. LUIS FERNANDO CASTELLANOS BONILLA
ASESORADO POR EL MSC. ADONAI NAVAS GARCÍA

AL CONFERÍRSELE EL TÍTULO DE

MAESTRO EN ESTADÍSTICA APLICADA

GUATEMALA, ENERO DE 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Inga. Aurelia Anabela Cordova Estrada
EXAMINADOR	Mtro. Edgar Darío Álvarez Cotí
EXAMINADOR	Mtro. Edwin Adalberto Bracamonte Orozco
EXAMINADOR	Mtro. Luis Carlos Bolaños Méndez
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

USO DE METODOLOGÍAS DE ESTIMACIÓN EN ÁREAS PEQUEÑAS PARA EL CÁLCULO DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL DE GUATEMALA IPM-GT A NIVEL DE DEPARTAMENTO

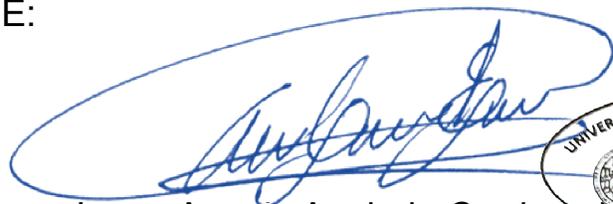
Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado, con fecha 5 de julio de 2021.

Luis Fernando Castellanos Bonilla

LNG.DECANATO.OI.077.2022

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Estudios de Posgrado, al Trabajo de Graduación titulado: **USO DE METODOLOGÍAS DE ESTIMACIÓN EN ÁREAS PEQUEÑAS PARA EL CÁLCULO DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL DE GUATEMALA IPM-GT A NIVEL DE DEPARTAMENTO**, presentado por: **Luis Fernando Castellanos Bonilla** , que pertenece al programa de Maestría en artes en Estadística aplicada después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:



Inga. Aurelia Anabela Cordova Estrada

Decana

Guatemala, enero de 2022

AACE/gaoc



Guatemala, enero de 2022

LNG.EEP.OI.077.2022

En mi calidad de Director de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor, verificar la aprobación del Coordinador de Maestría y la aprobación del Área de Lingüística al trabajo de graduación titulado:

“USO DE METODOLOGÍAS DE ESTIMACIÓN EN ÁREAS PEQUEÑAS PARA EL CÁLCULO DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL DE GUATEMALA IPM-GT A NIVEL DE DEPARTAMENTO”

presentado por **Luis Fernando Castellanos Bonilla** correspondiente al programa de **Maestría en artes en Estadística aplicada** ; apruebo y autorizo el mismo.

Atentamente,

“Id y Enseñad a Todos”

Mtro. Ing. Edgar Darío Álvarez Coñ
Director



Escuela de Estudios de Postgrado
Facultad de Ingeniería



Guatemala 5 de octubre 2021.

M.A. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Presente

M.A. Ingeniero Álvarez Cotí:

Por este medio informo que he revisado y aprobado el Informe Final del trabajo de graduación titulado **“USO DE METODOLOGÍAS DE ESTIMACIÓN EN ÁREAS PEQUEÑAS PARA EL CÁLCULO DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL DE GUATEMALA IPM-GT A NIVEL DE DEPARTAMENTO”** del estudiante **Luis Fernando Castellanos Bonilla** quien se identifica con número de carné **200022827** del programa de Maestría en Estadística Aplicada.

Con base en la evaluación realizada hago constar que he evaluado la calidad, validez, pertinencia y coherencia de los resultados obtenidos en el trabajo presentado y según lo establecido en el *Normativo de Tesis y Trabajos de Graduación aprobado por Junta Directiva de la Facultad de Ingeniería Punto Sexto inciso 6.10 del Acta 04-2014 de sesión celebrada el 04 de febrero de 2014*. Por lo cual el trabajo evaluado cuenta con mi aprobación.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.

Atentamente,


MSc. Ing. Edwin Adalberto Bracamonte Orozco
Coordinador
Maestría en Estadística Aplicada
Escuela de Estudios de Postgrado

Guatemala, agosto de 2021.

M.A. Ing. Edgar Darío Álvarez Cotí

Director

Escuela de Estudios de Postgrado

Presente

Estimado M.A. Ing. Álvarez Cotí

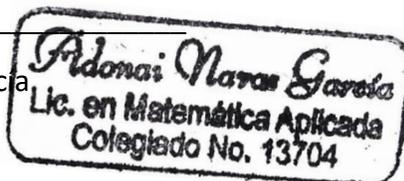
Por este medio informo a usted, que he revisado y aprobado el Trabajo de Graduación y el Artículo Científico: **“USO DE METODOLOGÍAS DE ESTIMACIÓN EN ÁREAS PEQUEÑAS PARA EL CÁLCULO DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL DE GUATEMALA IPM-GT A NIVEL DE DEPARTAMENTO”** del estudiante **Luis Fernando Castellanos Bonilla** del programa de Maestría en **Estadística Aplicada**, identificado con número de carné: **200022827**.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.

M.A. Lic. Adonai Navas García

Colegiado No. 13704

Asesor de Tesis



ACTO QUE DEDICO A:

יהוה (Yud-Hey-Vav-Hey)	Por regalarme el don de la vida un 20 de octubre, fecha en la que se celebra el Día Mundial de la Estadística, y por permitirme caracterizar la pobreza multidimensional de mi querida Guatemala.
Mi madre	Elvia Judith Bonilla Berger. Por enseñarme que el verdadero camino hacia la eternidad lo recorren las personas que aman a su prójimo como a sí mismas. Gracias por todo tu amor.
Mi padre	Jorge Luis Castellanos Estrada. Por enseñarme que la vida nunca sería fácil, pero que con valentía y esfuerzo se aprende a disfrutarla a pesar de los problemas. Te admiro por ser un papá ejemplar.
Mi hijo	Javier Alexander Castellanos Oliva. Porque el Eterno me regaló la oportunidad de ser tu mejor amigo. Te amo, nunca lo olvides.
Mi hermana	Paula Andree Castellanos Bonilla. Por todo el amor y apoyo incondicional. Gracias por cada momento vivido en familia.

Abuela

Marta Luz Estrada de Castellanos. Por todo el amor manifestado hacía el primogénito de su primogénito. Con todo mi amor y admiración.

Abuela y abuelos

Laura Berger Jaureguí (q. e. p. d.), Luis Alberto Castellanos Pellecer (q. e. p. d.) y Gustavo Antonio Bonilla Cambara (q. e. p. d.). Por ser una importante influencia en mi vida y su amor incondicional. Viven por siempre en mi presente.

Mi tío

Daniel Arturo Bonilla Berger (q. e. p. d.). Por todos los recorridos en su automotor, cuando en mis primeros años de vida se hizo costumbre dar una vuelta por las calles de Guatemala, condición *sine qua non* hubiese conciliado el sueño. Siempre te llevaré en mi corazón.

Mi primo

Daniel Edilberto Bonilla Abril (q. e. p. d.). Por tu cariño sincero y amistad incondicional. Siempre estarás presente en mis pensamientos.

Mi hermano

Gustavo Daniel Montenegro Bonilla. Por todos los momentos musicales como Poesía Callejera.

Mi familia

Con todo mi amor, respeto y admiración.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por ser la casa de estudios en donde me formé como profesional responsable y comprometido con el desarrollo y futuro de Guatemala.
Facultad de Ingeniería	Por abrirme las puertas para la profesionalización en el campo de la Estadística Aplicada, entre otras cosas.
Mis amigos de la Facultad	La sociedad de estadísticos muertos de Guatemala, por todos los momentos compartidos en esta transformadora experiencia estadística.
Lic. Adonai Navas García	Por el apoyo brindado en la elaboración del presente trabajo de tesis.
Juan Francisco López Jiménez	Por el apoyo brindado en la diagramación de mapas de pobreza multidimensional.
Marvin Isaac Reyes López	Por el apoyo brindado en el manejo de la información censal a través del programa REDATAM.

Mis compañeros

Del Instituto Nacional de Estadística INE (presentes y ausentes). Por la amistad y el apoyo brindado durante los 18 años que he laborado en la Institución.

Ing. Edwin Bracamonte

Por todo el apoyo brindado durante el desarrollo de la maestría. Gracias por compartir sus conocimientos y experiencia.

**Lic. Roberto Antonio
Molina Cruz**

Por el apoyo manifestado en la etapa de formulación del protocolo de tesis.

**Dra. Aura Marina
Rodríguez de Peña**

Por el apoyo brindado durante el desarrollo de seminario 3.

2.1.2.5.	Diseño de encuestas y diseño de muestras.....	18
2.1.2.6.	Distribución de muestreo.....	18
2.1.2.7.	Estimador	19
2.1.2.8.	Varianzas de los estimadores	20
2.1.2.9.	Estimación puntual y por intervalos de confianza	21
2.1.2.10.	Propiedades y precisión de los estimadores.....	24
2.1.2.11.	Formas básicas de selección de muestras probabilísticas.....	26
2.1.2.12.	Marco de muestreo	27
2.1.2.13.	Tamaño de la muestra	28
2.1.2.14.	Muestras complejas	31
2.1.2.15.	Estimación en dominios.....	32
2.1.3.	Estimación en áreas pequeñas	32
2.1.3.1.	Demanda de estimaciones en áreas pequeñas.....	33
2.1.3.2.	Límites para la desagregación de datos estadísticos.....	33
2.1.3.3.	Metodologías para superar los límites de la desagregación	35
2.1.4.	Métodos directos para la desagregación de datos.....	36
2.1.4.1.	Estimador de Horvitz-Thompson (HT).	36
2.1.4.2.	Estimador de Hájek (HA).....	37
2.1.4.3.	Estimadores GREG y de calibración	38
2.1.5.	Metodologías de estimación en áreas pequeñas	41

2.1.5.1.	Estimador post-estratificado sintético ..	41
2.1.5.2.	Estimador sintético de regresión a nivel de área	43
2.1.5.3.	Estimador sintético de regresión a nivel de individuo	45
2.1.5.4.	Estimadores compuestos	47
2.1.5.5.	EBLUP basado en el modelo <i>Fay-Herriot</i>	51
2.1.5.6.	EBLUP basado en el modelo con errores anidados.....	57
2.1.5.7.	Modelo ELL.....	61
2.1.5.8.	Mejor predictor empírico bajo el modelo con errores anidados	64
2.1.5.9.	Método jerárquico Bayes bajo el modelo con errores anidados	65
2.1.5.10.	Métodos basados en modelos lineales generalizados mixtos	66
2.2.	Teoría general para la medición de la pobreza	69
2.2.1.	Privaciones y teorías normativas del bienestar.....	69
2.2.1.1.	Utilitarismo	69
2.2.1.2.	La justicia como imparcialidad de Rawls.....	70
2.2.1.3.	El enfoque de funcionamientos y capacidades de SEN	70
2.2.2.	Metodologías utilizadas para medir la pobreza.....	71
2.2.2.1.	Método de necesidades básicas insatisfechas NBI	72
2.2.2.2.	Método de línea de pobreza	75
2.2.2.3.	Método de medición integrada.....	76

2.2.2.4.	Método multidimensional.....	76
2.2.3.	Índice de Pobreza Multidimensional de Guatemala IPM-Gt	77
2.2.3.1.	Metodología del IPM-Gt	77
2.2.3.2.	Diseño de la medida de pobreza multidimensional IPM-Gt	79
2.2.3.3.	Tasas de privaciones no censuradas de los indicadores del IPM-Gt	83
3.	PRESENTACIÓN DE RESULTADOS.....	87
3.1.	Objetivo 1: Tasas de privación no censuradas de los indicadores del IPM-Gt a nivel de departamento	87
3.1.1.	Acceso a servicios de salud	87
3.1.2.	Seguridad alimentaria y nutricional	89
3.1.3.	Embarazo en adolescentes.....	91
3.1.4.	Cuidado prenatal	93
3.1.5.	Asistencia escolar	95
3.1.6.	Años de escolaridad	97
3.1.7.	Rezago educativo.....	99
3.1.8.	Cuidado infantil.....	101
3.1.9.	Empleo informal	103
3.1.10.	Trabajo infantil.....	105
3.1.11.	Materiales de la vivienda	107
3.1.12.	Hacinamiento	109
3.1.13.	Combustible para cocinar.....	111
3.1.14.	Acceso al agua.....	113
3.1.15.	Energía eléctrica	115
3.1.16.	Recolección de basura.....	117
3.1.17.	Sanearamiento	119

3.2.	Objetivo 2: Niveles de incidencia e intensidad de la pobreza de la población guatemalteca a nivel de departamento.....	121
3.3.	Objetivo 3: Índice de Pobreza Multidimensional IPM-Gt a nivel de departamento	125
4.	DISCUSIÓN DE RESULTADOS	127
4.1.	Tamaños de muestra departamentales	127
4.2.	Uso de variables auxiliares, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	133
4.3.	Calidad estadística de las estimaciones.....	136
	CONCLUSIONES	141
	RECOMENDACIONES.....	143
	REFERENCIAS	145
	APÉNDICES	151

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	CV de la proporción muestral p en función del tamaño muestral n , para cada valor de la verdadera proporción p	35
2.	Tasas de privación no censuradas por indicador al 2014	85
3.	Acceso a servicios de salud, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	88
4.	Seguridad alimentaria y nutricional, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	90
5.	Embarazo en adolescentes, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	92
6.	Cuidado prenatal, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	94
7.	Asistencia escolar, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	96
8.	Años de escolaridad, estimador EBLUP bajo el modelo <i>Fay Herriot</i> ...	98
9.	Rezago educativo, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	100
10.	Cuidado infantil, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	102
11.	Empleo informal, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	104
12.	Trabajo infantil, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	106
13.	Materiales de la vivienda, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	108
14.	Hacinamiento, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	110
15.	Combustible para cocinar, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	112
16.	Acceso al agua, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	114
17.	Energía eléctrica, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	116

18.	Figura 18.Recolección de basura, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	118
19.	Saneamiento, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	120
20.	Incidencia de pobreza multidimensional departamental, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	122
21.	Intensidad de la pobreza multidimensional departamental, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	124
22.	Índice de Pobreza Multidimensional IPM-Gt a nivel de departamento, estimador EBLUP bajo el modelo <i>Fay Herriot</i>	126
23.	Apéndice 1. Sintaxis para replicar el IPM-Gt	151

TABLAS

I.	Variables del estudio.....	XXXII
II.	Sintaxis estimaciones directas.....	XXXVIII
III.	Correlación entre recolección de basura y variables auxiliares ...	XXXVIII
IV.	Sintaxis EBLUP bajo el modelo <i>Fay Herriot</i>	XL
V.	Fórmulas para el tamaño de muestra sin reposición	30
VI.	Fórmulas para el tamaño de muestra con reposición	30
VII.	IPM-Gt: Matriz de dimensiones, indicadores, umbrales y pesos	80
VIII.	Acceso a servicios de salud, según ENCOVI 2014	87
IX.	Seguridad alimentaria y nutricional, según ENCOVI 2014.....	89
X.	Embarazo en adolescentes, según ENCOVI 2014	91
XI.	Cuidado prenatal, según ENCOVI 2014	93
XII.	Asistencia escolar, según ENCOVI 2014.....	95
XIII.	Años de escolaridad, según ENCOVI 2014	97
XIV.	Rezago educativo, según ENCOVI 2014	99
XV.	Cuidado infantil, según ENCOVI 2014.....	101
XVI.	Empleo informal, según ENCOVI 2014.....	103

XVII.	Trabajo infantil, según ENCOVI 2014	105
XVIII.	Materiales de la vivienda, según ENCOVI 2014	107
XIX.	Hacinamiento, según ENCOVI 2014	109
XX.	Combustible para cocinar, según ENCOVI 2014	111
XXI.	Acceso al agua, según ENCOVI 2014	113
XXII.	Energía eléctrica, según ENCOVI 2014	115
XXIII.	Recolección de basura, según ENCOVI 2014	117
XXIV.	Saneamiento, según ENCOVI 2014	119
XXV.	Incidencia de pobreza multidimensional departamental, según ENCOVI 2014	121
XXVI.	Intensidad de la pobreza multidimensional departamental, según ENCOVI 2014	123
XXVII.	Índice de Pobreza Multidimensional IPM-Gt a nivel de departamento, según ENCOVI 2014	125
XXVIII.	Tamaños de muestra parte 1, según ENCOVI 2014	127
XXIX.	Tamaños de muestra parte 2, según ENCOVI 2014	128
XXX.	Tamaños de muestra parte 3, según ENCOVI 2014	129
XXXI.	Tamaños de muestra parte 4, según ENCOVI 2014	130
XXXII.	Tamaños de muestra parte 5, según ENCOVI 2014	131
XXXIII.	Tamaños de muestra parte 6, según ENCOVI 2014	132
XXXIV.	Variables auxiliares utilizadas, estimador EBLUP bajo el modelo <i>Fay Herriot</i> 134	
XXXV.	Evaluación del número de estimaciones aceptables	137

LISTA DE SÍMBOLOS

Símbolo	Significado
\tilde{u}_d	<i>Best Linear Unbiased Predictor BLUP</i> de u_d .
Q	Complemento de la proporción poblacional.
c_{di}	Constante de heteroscedasticidad para el individuo i en el área o dominio d .
k_{di}	Constantes conocidas que representan la posible heteroscedasticidad de las áreas o dominios d .
S^2	Cuasivarianza poblacional.
$\sigma(\hat{\theta})$	Desviación estándar de la distribución de muestreo.
S	Desviación estándar de la muestra.
A_d	Diagonal de constantes conocidas que representan la posible heteroscedasticidad de las áreas o dominios d .
\bar{y}_c	Distribución teórica de todos los valores posibles de un estimador.
v_d	Efecto aleatorio del área o dominio d .
u_d^*	Efectos aleatorios para cada área o dominio d , mediante procedimiento <i>Bootstrap</i> .
s_d	El conjunto de elementos de la muestra del área o dominio d .
U_d	El conjunto de unidades en la población estadística que están en el área o dominio d .
N	Elementos de una población completa.

e_{di}^*	Error a nivel de individuo i en el área o dominio d , mediante procedimiento <i>Bootstrap</i> .
e	Error absoluto de muestreo.
e_d	Error de muestreo del área o dominio d .
$\sigma_{\bar{y}}$	Error de muestreo.
$g_{d3}(\hat{\sigma}_u^2)$	Error debido a la estimación de la varianza σ_u^2 .
$g_{d2}(\hat{\sigma}_u^2)$	Error debido a la estimación del vector de coeficientes de regresión β para área o dominio d .
$g_{d1}(\hat{\sigma}_u^2)$	Error debido a la predicción del efecto aleatorio del área o dominio d .
e_r	Error relativo de muestreo.
$C_{1,x}$	Error relativo de muestreo.
$E(\bar{y})$	Esperanza matemática de la media.
\hat{Y}	Estimación del total.
$\hat{\theta}(s)$	Estimación particular de θ dada la muestra s .
\tilde{Y}_{di}^{BLUP}	Estimador BLUP basado en modelo de errores anidados.
$\tilde{\theta}_d^{FH}$	Estimador BLUP <i>Fay Herriot</i> del área o dominio d .
\hat{Y}_d^C	Estimador compuesto de la media del área o dominio d .
HA	Estimador de Hájek
\hat{Y}_d^{HA}	Estimador de Hájek para la media del área o dominio d .
$\hat{\theta}_{HH}$	Estimador de Hansen y Hurwitz.
$\hat{\theta}_{HT}$	Estimador de Horvitz y Thompson.
HT	Estimador de Horvitz y Thompson.

$\widehat{var}_{\pi}(\widehat{Y}_d)$	Estimador de la varianza que no depende de las probabilidades de inclusión de segundo orden.
$\widehat{MSE}_{\pi}(\widehat{Y}_d^{SYN})$	Estimador del error cuadrático medio del modelo post-estratificado sintético del área o dominio d .
\widehat{Y}_d^{DIR}	Estimador directo de la media del área o dominio d .
$\widehat{\theta}_d$	Estimador directo en el área o dominio d .
$\widehat{\theta}_d^{FH}$	Estimador EBLUP basado en el modelo <i>Fay Herriot</i> para el área o dominio d .
$\widetilde{\theta}_d^B(\theta)$	Estimador general del mejor predictor empírico bajo el modelo de errores anidados.
\widehat{Y}_d^{GREG}	Estimador <i>generalized regression</i> de la media del área o dominio d .
$\widehat{Y}_d^{j,HA}$	Estimador Hájek de la media del grupo j .
\widehat{Y}_d^{PS-SYN}	Estimador post-estratificado sintético de la media del área o dominio d .
\widehat{Y}_d^{SYN}	Estimador sintético de la media del área o dominio d .
$\widehat{\theta}_d^{REG1-SYN}$	Estimador sintético de regresión a nivel de área o dominio d .
$\widehat{\theta}_d^{REG2-SYN}$	Estimador sintético de regresión a nivel de individuo para el área o dominio d .
w_i	Factor de expansión de la unidad i -ésima.
w_{di}	Factor de expansión del individuo i en el área o dominio d .
f	Fracción de muestreo.
$\widehat{\theta}$	Función que asocia a cada muestra s un valor numérico.
g	Función vínculo (biyectiva, con derivada continua).
M_o	Índice de Pobreza Multidimensional.

$e_{r\alpha}$	Margen de error relativo ($e_{r\alpha} = \lambda_{\alpha}CV$).
X_d	Matriz del área o dominio d para un vector x_{di} de p variables auxiliares relacionadas con el individuo i .
\bar{y}	Media de la distribución de muestreo.
\bar{y}_{U_d}	Media de las unidades del área o dominio d .
\bar{X}_d	Media del área o dominio d para un vector x_{di} de p variables auxiliares relacionadas con el individuo i .
\bar{Y}^j	Media del grupo j .
\bar{Y}_d^j	Media del grupo j en el área o dominio d .
\bar{Y}_d^j	Media del post-estrato j del área o dominio d .
\widehat{Y}_d	Media estimada de la población del área o dominio d .
\bar{Y}_d	Media poblacional del área o dominio d .
\bar{y}_{da}	Media ponderada de la variable de respuesta Y para el área o dominio d .
y_i	Medida de la unidad i -ésima.
x_i	Medida de la unidad poblacional i -ésima.
ξ	Modelo de superpoblación.
A^*	Número de censos <i>Bootstrap</i> .
N_d	Número de unidades de la población en U_d .
N^j	Número de unidades del grupo j .
N_d^j	Número de unidades del grupo j en el área o dominio d .
θ_d	Parámetro poblacional del área o dominio d .
Φ_d^*	Peso óptimo otorgado al estimador directo del área o dominio d .
ϕ_d	Peso otorgado al estimador directo del área o dominio d .

\mathbf{a}_{di}	Peso ponderado de la variable de respuesta para la observación i -ésima del área o dominio d .
\mathbf{h}_{di}	Pesos calibrados para el individuo i en el área o dominio d .
γ_d	Ponderador del efecto aleatorio EBLUP <i>Fay Herriot</i> para el área o dominio d .
π_i	Probabilidad de inclusión de la unidad i -ésima.
$\pi_{d,ij}$	Probabilidad de inclusión de segundo orden.
P_c	Probabilidad de materialización de los valores posibles de un estimador.
P_i	Probabilidad de selección de la unidad i -ésima.
\widehat{X}_d	Promedio estimado del área o dominio d para un vector \mathbf{x}_{di} de p variables auxiliares relacionadas con el individuo i .
p	Proporción muestral.
P	Proporción poblacional.
$G_{di}(\mathbf{h}_{di}, \mathbf{w}_{di})$	Pseudo distancia ji-cuadrado.
$B(\widehat{\theta})$	Sesgo del estimador.
n_d	Tamaño de la muestra del área o dominio d .
n	Tamaño de la muestra.
\widehat{N}	Tamaño estimado de la población.
H	Tasa de incidencia de pobreza multidimensional.
A	Tasa de intensidad de pobreza multidimensional.
ϵ_{di}	Término de error independiente del modelo para el individuo i en el área o dominio d .
ϵ_d	Término de error independiente en el área o dominio d .
\mathbf{u}_i	Unidad i -ésima de la población.

t_{α}	Valor crítico de la distribución t de Student para un nivel de confianza asumido.
λ_{α}	Valor crítico de la normal unitaria a nivel α .
\tilde{e}_{di}	Valor estimado para el individuo i en el área o dominio d mediante el estimador <i>generalized regression</i> .
Y_{di}	Valor observado de la característica Y en el individuo i del área o dominio d .
Y_{di}^*	Valores <i>Bootstrap</i> de la variable de respuesta para el individuo i en el área o dominio d .
$y_d^{*(a)}$	Valores de la variable de respuesta en el área o dominio d en el censo <i>Bootstrap</i> a .
X_i	Variable aleatoria.
E_{di}	Variable que mide el poder adquisitivo del individuo i en el área o dominio d .
ψ_d	Varianza poblacional del área o dominio d .
σ^2	Varianza poblacional.
α	Vector de coeficientes de regresión basado en modelos lineales generalizados mixtos.
β_d	Vector de coeficientes del modelo de regresión para el área o dominio d .
\bar{x}_{da}	Vector de medias ponderadas de la variable de respuesta X .
x_{di}	Vector de p variables auxiliares del área o dominio d .
y_d	Vector de variables de respuesta del área o dominio d .
\hat{B}_d	Vector del estimador de mínimos cuadrados ponderados para el área o dominio d .

GLOSARIO

Área pequeña	Es un área o dominio de estudio para el cual el tamaño de muestra específico no es suficientemente grande para obtener estimaciones confiables.
<i>Benchmarking</i>	Propiedad que satisface que al sumar los totales estimados a nivel de área o dominio d , se obtiene el total estimado para la región mediante el estimador directo.
BID	Banco Interamericano de Desarrollo.
BM	Banco Mundial.
BLUP	<i>Best Linear Unbiased Predictor.</i>
Campo de estimación	Subgrupos de análisis para los que se desean estimaciones con igual fiabilidad.
CEPAL	Comisión Económica para América Latina y el Caribe.
CV	Coefficiente de variación o error relativo de un estimador.
DESA	Departamento de Asuntos Económicos y Sociales de la Secretaría de las Naciones Unidas.

EBLUP	<i>Empirical Best Linear Unbiased Predictor.</i>
ECM	Error cuadrático medio.
ECMR	Error cuadrático medio relativo o estándar.
EE	Error estándar de la distribución de muestreo.
ELL	Método de Elbers, Lanjouw y Lanjouw.
ENCOVI	Encuesta Nacional de Condiciones de Vida.
ENEI	Encuesta Nacional de Empleo e Ingresos.
ENIGFAM	Encuesta Nacional de Ingresos y Gastos Familiares.
FGT	Foster, Greer y Thorbecke.
GLMM	<i>Generalized linear mixed models.</i>
GREG	<i>Generalized regression.</i>
INE	Instituto Nacional de Estadística.
INEGI	Instituto Nacional de Estadística y Geografía.
IPM-Gt	Índice de Pobreza Multidimensional de Guatemala.

ISI	International Statistical Institute.
Jamovi	Programa de uso libre para cálculos estadísticos bajo el lenguaje estadístico de R.
LFS	Labor Force Survey, del Reino Unido.
MED	Metas Estratégicas de Desarrollo.
MECOVI	Programa para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida en América Latina y el Caribe.
MIDES	Ministerio de Desarrollo Social.
MINFIN	Ministerio de Finanzas Públicas.
NBI	Necesidades básicas insatisfechas.
ODS	Objetivos de Desarrollo Sostenible.
ONS	Office for National Statistics, del Reino Unido.
OPHI	Oxford Poverty and Human Development Initiative.
PND	Plan Nacional de Desarrollo: K´atun Nuestra Guatemala 2032.

REDATAM	Recuperador de datos para áreas pequeñas por microcomputador.
RStudio	Es un entorno de desarrollo integrado (IDE) para el lenguaje de programación R.
SAE	<i>Small Area Estimation.</i>
SEGEPLAN	Secretaría de Planificación y Programación de la Presidencia de Guatemala.
SEN	Sistema Estadístico Nacional.
SIEH	Sistema integrado de Encuestas de Hogares.
SPSS	Statistical Package for Social Sciences.
U.S. Census Bureau	Oficina de Censo de Estados Unidos de América.

RESUMEN

El propósito de la presente investigación fue generar estimaciones desagregadas de los 20 indicadores priorizados por el Índice de Pobreza Multidimensional de Guatemala, IPM-Gt (2018) a nivel de departamento.

El objetivo general consistió en estimar los principales indicadores de pobreza multidimensional de la población guatemalteca a nivel de departamento, mediante una evaluación comparativa entre el estimador directo del diseño complejo de la ENCOVI 2014 y la metodología de estimación en área pequeña EBLUP basado en el modelo *Fay-Herriot*, para obtener estimaciones precisas y confiables.

El enfoque del estudio realizado es mixto, porque se apoya en la revisión de la literatura al alcance con relación a las metodologías de estimación en áreas pequeñas, medición de la pobreza y, además, toma como base para el cálculo de los indicadores del IPM-Gt a nivel de departamento los datos provenientes de la Encuesta Nacional de Condiciones de Vida ENCOVI 2014, la cual fue estructurada bajo un esquema de muestreo complejo. Su alcance es descriptivo y su diseño es no experimental, mientras que el tipo de estudio aplicado es descriptivo correlacional.

Se obtuvo estimaciones precisas y estadísticamente aceptables de los 20 indicadores priorizados por el Índice de Pobreza Multidimensional de Guatemala IPM-Gt a nivel de departamento, los cuales se presentan en cuadros y mapas temáticos.

En los cuadros se incluye el estimador puntual basado en el estimador directo del diseño complejo de la ENCOVI 2014, el error de muestreo (EE), el error relativo de muestreo (CV), la estimación puntual basada en el EBLUP bajo el modelo *Fay Herriot*, el error cuadrático medio estimado y, por último, el error cuadrático medio relativo o estándar estimado.

Los mapas temáticos presentan la intensidad de los indicadores estimados a través de gradaciones de color, según el departamento de pertenencia del indicador. El número de cifras significativas en los cuadros y mapas varía según la naturaleza del estimador (hasta un máximo de 8 cifras).

En conclusión, se estimó los principales indicadores de pobreza multidimensional de la población guatemalteca a nivel de departamento, obteniendo un 56 % de estimaciones aceptables bajo el estimador directo basado en el diseño muestral complejo de la ENCOVI 2014 y el 91 % mediante la metodología de estimación en área pequeña EBLUP bajo el modelo *Fay Herriot*, dados los criterios de calidad adoptados para el estudio.

De acuerdo con los resultados, se recomienda utilizar las estimaciones obtenidas mediante la metodología de estimación en áreas pequeñas, estimador EBLUP bajo el modelo *Fay Herriot*, como estadísticas oficiales de los niveles de pobreza multidimensional de la población guatemalteca a nivel de departamento.

PLANTEAMIENTO DEL PROBLEMA

- Contexto general

En el marco del Sistema Integrado de Encuestas de Hogares (SIEH) del Instituto Nacional de Estadística (INE), a partir del año 2000 se implementó un conjunto de encuestas que tenían como principal objetivo, conocer y evaluar las condiciones de vida de la población guatemalteca, así como determinar los niveles de pobreza existentes y los factores que los determinan.

Esto como parte del Programa para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida en América Latina y el Caribe (MECOVI). El MECOVI fue un programa regional bajo el auspicio del Banco Mundial (BM), el Banco Interamericano de Desarrollo (BID) y la Comisión Económica para América Latina y el Caribe (CEPAL), cuyo fin era fortalecer los sistemas de encuestas de hogares de los institutos nacionales de estadística de América Latina y el Caribe.

Los altos índices de pobreza y desigualdad que caracterizan a la sociedad guatemalteca plantean la necesidad de encarar el diseño y puesta en práctica de una serie de encuestas de hogares. Por esta razón, el INE a través de la Encuesta Nacional de Condiciones de Vida (ENCOVI) adoptó la medición indirecta de la pobreza por el método de líneas de pobreza por consumo. Esta fue replicada durante los años 2000, 2006, 2011 y 2014.

Según estas mediciones, la incidencia de pobreza ha oscilado en el país manteniéndose entre el 50 y el 60 por ciento. Sin embargo, la metodología empleada por la ENCOVI no permite conocer en qué aspectos son privados los pobres o qué tan intensa es esa pobreza en dimensiones específicas.

En este contexto, el desarrollo de indicadores de carencias en dimensiones distintas al ingreso, que complementaran las mediciones tradicionales de la pobreza, tomó impulso en países de América Latina y el Caribe, con el desarrollo de la metodología Alkire-Foster (2009) de *Oxford Poverty and Human Development Initiative* (OPHI). Esta situación, llevó a concebir metas más exigentes y a considerar el abordaje de la pobreza desde la óptica multidimensional.

En el año 2014, Guatemala adoptó el Plan Nacional de Desarrollo Nuestra Guatemala K'atun 2032 (PND). Subsiguientemente, en el año 2015 la Asamblea General de las Naciones Unidas aprobó la Agenda de Desarrollo Sostenible. El documento “Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible” fue aprobado en la Asamblea General celebrada en septiembre de 2015, y definió una serie de objetivos globales, conocidos como los Objetivos de Desarrollo Sostenible (ODS).

Derivado de lo anterior, la Secretaría de Planificación y Programación de la Presidencia de Guatemala (SEGEPLAN) con el acompañamiento de la Unidad de Desarrollo Social de la Sede Subregional de la Comisión Económica para América Latina en México implementaron una metodología que permitió articular los objetivos y metas de la Agenda de Desarrollo Sostenible 2030 con los propios de los instrumentos de planificación del país. De dicho ejercicio surge una agenda de país que resume 10 prioridades nacionales y 16 Metas Estratégicas de Desarrollo (MED).

En este marco, el Gobierno de Guatemala, entre junio del 2017 y septiembre del 2018, a través de un comité técnico interinstitucional, conformado por personal técnico del Ministerio de Desarrollo Social (MIDES), el Instituto Nacional de Estadística (INE), el Ministerio de Finanzas Públicas (MINFIN) y la Secretaría de Programación y Planificación de la Presidencia (SEGEPLAN), definen el Índice de Pobreza Multidimensional para Guatemala IPM-Gt, el cual se compone de 5 dimensiones y 17 indicadores. Para su cálculo se utilizó la Encuesta de Condiciones de Vida (ENCOVI) para el año 2014.

Finalmente, el informe del IPM-Gt fue publicado en diciembre de 2018 con estimaciones a nivel nacional, regional y para sus respectivas áreas geográficas urbano y rural. En esa ocasión, se observó que a pesar de que la ENCOVI 2014 consideró dominios de estudio departamentales, este nivel de estimación no fue considerado en el cálculo del IPM-Gt, ya que, al generar estimaciones a este nivel, se obtienen estimaciones imprecisas en algunos indicadores con errores relativos de muestreo superiores al 15 %.

Esta situación pone de manifiesto la importancia que tiene para la toma de decisiones informada, obtener estimaciones de las dimensiones e indicadores del IPM-Gt a nivel de departamento.

- Descripción del problema

Las encuestas oficiales que realiza el Instituto Nacional de Estadística (INE), en el marco del Sistema Integrado de Encuestas de Hogares (SIEH), están diseñadas para producir datos estadísticos a un nivel de agregación determinado (nacional, urbano metropolitano, rural nacional, etc.), sin embargo, la precisión de las estimaciones decrece sustancialmente cuando se trata de realizar

inferencias para grupos poblacionales que representan desagregaciones para las cuales la encuesta no fue diseñada (ejemplos: resultados para pueblos indígenas, estimaciones a nivel municipal, etc.).

En este contexto, y bajo el establecimiento del paradigma del desarrollo sostenible y el abordaje multidimensional de la pobreza, existe desinformación sobre la situación de los 22 departamentos de la República de Guatemala, con respecto a las 5 dimensiones y los 17 indicadores necesarios para la estimación del Índice de Pobreza Multidimensional de Guatemala (IPM-Gt), ya que la desagregación de los datos estadísticos a este nivel genera estimaciones imprecisas de los parámetros de interés.

Esta situación limita sustentar el proceso de decisión en la gestión técnico-política que permita la adecuada planificación, programación y asignación de recursos del Estado, así como el seguimiento y evaluación en función de las prioridades nacionales de desarrollo (priorizadas en el plan en el Plan Nacional de Desarrollo Nuestra Guatemala K'atun 2032 y la Agenda de Desarrollo Sostenible 2030) para el bienestar e igualdad de oportunidades con énfasis en la población en pobreza multidimensional.

- Formulación del problema
 - Pregunta central

¿Cuál es la metodología estadística que mejor se ajusta para realizar estimaciones precisas y confiables de los principales indicadores de pobreza multidimensional de la población guatemalteca a nivel de departamento?

- Preguntas auxiliares
 - ¿Qué metodología de estimación es adecuada para la desagregación de estimaciones precisas y confiables de las tasas de privación no censuradas de los indicadores del IPM-Gt a nivel de departamento?
 - ¿Cuál metodología de estimación es apropiada para la desagregación de estimaciones precisas y confiables de los niveles de incidencia e intensidad de la pobreza de la población guatemalteca a nivel de departamento?
 - ¿Cuál es la metodología que mejor se ajusta para la desagregación de estimaciones precisas y confiables del Índice de Pobreza Multidimensional IPM-Gt a nivel de departamento?
- Delimitación del problema

La población objeto de estudio para la estimación de los principales indicadores de pobreza multidimensional, lo conforman las personas que residen permanentemente en las viviendas particulares ubicadas dentro de los 22 departamentos de la República de Guatemala, con base en datos provenientes de la Encuesta Nacional de Condiciones de Vida, ENCOVI 2014.

OBJETIVOS

General

Estimar los principales indicadores de pobreza multidimensional de la población guatemalteca a nivel de departamento, mediante una evaluación comparativa entre el estimador directo del diseño complejo de la ENCOVI 2014 y la metodología de estimación en área pequeña EBLUP basado en el modelo *Fay-Herriot*, para obtener estimaciones precisas y confiables.

Específicos

1. Estimar las tasas de privación no censuradas de los indicadores del IPM-Gt a nivel de departamento, mediante la evaluación comparativa entre el estimador directo del diseño complejo de la ENCOVI 2014 y la metodología de estimación en área pequeña EBLUP basado en el modelo *Fay-Herriot*, para obtener estimaciones aceptables de los niveles de pobreza multidimensional de la población guatemalteca.
2. Estimar los niveles de incidencia e intensidad de la pobreza de la población guatemalteca a nivel de departamento, mediante la evaluación comparativa entre el estimador directo del diseño complejo de la ENCOVI 2014 y la metodología de estimación en áreas pequeñas EBLUP basado en el modelo *Fay-Herriot*, para obtener estimaciones aceptables de los niveles de pobreza multidimensional de la población guatemalteca.

3. Estimar el Índice de Pobreza Multidimensional IPM-Gt a nivel de departamento, mediante la evaluación comparativa entre el estimador directo del diseño complejo de la ENCOVI 2014 y la metodología de estimación en áreas pequeñas EBLUP basado en el modelo *Fay-Herriot*, para obtener estimaciones aceptables de los niveles de pobreza multidimensional de la población guatemalteca.

RESUMEN DEL MARCO METODOLÓGICO

A continuación, se exponen los criterios adoptados para proceso investigativo. Se describen las características principales del estudio, las variables de interés y las fases que se implementaron para el logro de los objetivos propuestos.

- Características del estudio

El enfoque del estudio realizado es mixto, porque está fundamentado en la revisión de la literatura al alcance relacionada con las metodologías de estimación en áreas pequeñas, medición de la pobreza y, además, se tomó como base para el cálculo de los indicadores del IPM-Gt a nivel de departamento los datos provenientes de la Encuesta Nacional de Condiciones de Vida ENCOVI 2014, la cual fue estructurada bajo un diseño de muestreo probabilístico y complejo.

El alcance es descriptivo, dado que se estimaron los principales indicadores de pobreza multidimensional para la población guatemalteca a nivel de departamento. El tipo de estudio aplicado es descriptivo correlacional.

El diseño adoptado fue no experimental, pues la información de los indicadores necesarios para el Índice de Pobreza Multidimensional IPM-Gt se analizó en su estado original sin ninguna manipulación; además, fue transversal pues se estudió con base en un solo período particular de medición.

- Unidades de análisis

La población en estudio está conforma por residentes permanentes en los 22 departamentos de la República de Guatemala, con base en la información recopilada por la Encuesta Nacional de Condiciones de Vida (ENCOVI 2014).

Tabla I. **Variabes del estudio**

Variable	Definición teórica	Definición operativa
Acceso a servicios de salud (X_i^{ass})	El hogar está privado en acceso a servicios de salud (uso efectivo de servicios de salud) si una persona mayor de 5 años se enfermó y no tuvo acceso a servicios de salud por las siguientes causas: 1. El lugar de atención está lejos. 2. Falta de dinero. 3. No hay medio de transporte. 4. No hay médicos/enfermeras. 5. No hablan en el idioma de la persona enferma. 6. La espera es muy larga. 7. No alcanzó número o cupo. 8. Otro. Indicador de privación de tipo cualitativo.	Filtrando la población mayor de 5 años se verifican las preguntas de opción múltiple consignadas en la boleta de la ENCOVI 2014 según las condiciones del IPM-Gt. Indicador de privación de escala nominal. $X_i^{ass} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$
Seguridad alimentaria y nutricional (X_i^{san})	El hogar se considera privado si se encuentra en los niveles de inseguridad alimentaria y nutricional moderada o severa definidos por la <i>Food Insecurity Experience Scale</i> (Escala Latinoamericana y Caribeña de Seguridad Alimentaria -ELCSA), en la cual: 1. Inseguridad alimentaria moderada: el hogar reportó entre 6 y 10 experiencias de inseguridad alimentaria, en hogares con menores de 18 años, y entre 4 y 6 en hogares sin menores de 18 años. 2. Inseguridad alimentaria severa: el hogar reportó entre 11 y 15 experiencias de inseguridad alimentaria, en hogares con menores de 18 años, y entre 7 y 8 en hogares sin menores de 18 años. Indicador de privación de tipo cualitativo.	Con base en las 16 preguntas de opción múltiple correspondientes a la escala ELCSA, se construye el indicador de seguridad alimentaria moderada y severa según las condiciones del IPM-Gt. Indicador de privación de escala nominal. $X_i^{san} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$

Continúa tabla I.

Variable	Definición teórica	Definición operativa
Embarazo en adolescentes (X_i^{ea})	El hogar se considera privado si hay alguna niña o adolescente menor de 20 años embarazada o que ha estado embarazada. Indicador de privación de tipo cualitativo.	Se filtra la edad estipulada para las mujeres en el indicador y se verifica la condición descrita en el IPM-Gt. Indicador de privación de escala nominal. $X_i^{ea} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$
Cuidado prenatal (X_i^{cp})	El hogar se considera privado si hay al menos una mujer de 12 a 49 años que ha estado embarazada en los cinco años anteriores a la realización de la encuesta, que en su último embarazo no asistió a por lo menos 4 controles de embarazo. Indicador de privación de tipo cualitativo.	Filtrando en la base de datos a las mujeres de 12 a 49 años, se verifican las condiciones descritas en el IPM-Gt. Indicador de privación de escala nominal. $X_i^{cp} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$
Asistencia escolar (X_i^{ae})	El hogar es privado cuando hay al menos un niño o adolescente de 7 a 15 años que no asiste a la escuela, o hay al menos un adolescente de 16 a 17 años que no asiste a la escuela y no ha completado 9 años de estudios. Indicador de privación de tipo cualitativo.	Con base en las edades estipuladas en el indicador, se filtra la matriz de datos y se verifican las condiciones descritas en el IPM-Gt. Indicador de privación de escala nominal. $X_i^{ae} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$
Años de escolaridad (X_i^{aes})	El hogar se considera privado si hay al menos un adulto (18 años o más) con menos de los años de escolaridad según las edades siguientes: 1. 18 a 32 años: 9 años de escolaridad. 2. 33 a 64 años: 6 años de escolaridad. 3. Mayores de 65 años: no saber leer y escribir. Variable cualitativa.	Se filtra en la base de datos la población de 18 o más años y se verifican las condiciones descritas en el IPM-Gt. Indicador de privación de escala nominal. $X_i^{aes} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$
Rezago educativo (X_i^{re})	El hogar se considera privado cuando hay al menos un niño o adolescente de 8 a menos de 18 años que presenta más de dos años de rezago escolar, estimado según su edad y el grado en el que se inscribió. Indicador de privación de tipo cualitativo.	Filtrando la base de datos para el rango de edades considerado en el indicador y se verifican las condiciones descritas en el IPM-Gt. Indicador de privación de escala nominal. $X_i^{re} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$
Cuidado infantil (X_i^{ci})	El hogar se considera privado si hay al menos un niño menor de 7 años que permanece con el padre o la madre en su lugar de trabajo, o con vecinos o amigos, o permaneció la mayor parte del tiempo solo. Indicador de privación de tipo cualitativo.	Se filtra la población menor de 7 años y se verifican las condiciones descritas por el IPM-Gt. Indicador de privación de escala nominal. $X_i^{ci} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$

Continúa tabla I.

Variable	Definición teórica	Definición operativa
Empleo informal (X_i^{ei})	El hogar se considera privado cuando hay al menos una persona ocupada en alguna de las siguientes categorías: 1. Empleadores, empleados y obreros de empresas de menos de 6 personas. 2. Todos los trabajadores por cuenta propia o autónoma (excluyendo profesionales y técnicos, medido como 12 años de escolaridad o más). 3. Todos los familiares no remunerados. 4. Ocupados en servicio doméstico. Indicador de privación de tipo cualitativo.	Con base en el capítulo de empleo de la ENCOVI se verifican las condiciones descritas por el IPM-Gt. Indicador de privación de escala nominal. $X_i^{ei} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$
Trabajo infantil (X_i^{ti})	El hogar se considera privado si existe al menos un niño de 15 años o menos que trabaja, o cuando existe al menos un adolescente de 16 a 17 años que trabaja, no estudia y no tiene al menos 9 años de educación completos. Indicador de privación de tipo cualitativo.	Se filtra la base de datos conforme a las edades y se verifican las condiciones para esta variable por el IPM-Gt. Indicador de privación de escala nominal. $X_i^{ti} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$
Materiales de la vivienda (X_i^{mv})	El hogar se considera privado si la vivienda que habita está construida con materiales inapropiados, bajo los siguientes criterios: 1. Se consideran privadas las viviendas con piso de tierra. 2. Se consideran privadas las viviendas, sin importar el material del piso, cuando el material de pared está hecho de adobe, lámina metálica, bajareque, lepa, palo o caña, u otro. Indicador de privación de tipo cualitativo.	Con base en la información consignada en el capítulo de las características de la vivienda en la ENCOVI 2014, se verifica el cumplimiento de las condiciones descritas por el IPM-Gt. Indicador de privación de escala nominal. $X_i^{mv} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$
Hacinamiento (X_i^h)	El hogar se encuentra privado por hacinamiento cuando habitan más de tres personas en promedio por cuarto (sin considerar los cuartos utilizados para cocina, baños, pasillos, garajes y los dedicados a negocios). Indicador de privación de tipo cualitativo.	Se calcula el promedio de hacinamiento y se verifica el cumplimiento de las condiciones descritas por el IPM-Gt. Indicador de privación de escala nominal. $X_i^h = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$

Continúa tabla I.

Variable	Definición teórica	Definición operativa
Combustible para cocinar (X_i^{cc})	El hogar se considera privado si el combustible utilizado para cocinar es leña y no cuenta con una chimenea o una salida de escape para el humo. Indicador de privación de tipo cualitativo.	Con base en la información consignada en el capítulo de las características de la vivienda en la ENCOVI 2014, se verifica el cumplimiento de las condiciones descritas por el IPM-Gt. Indicador de privación de escala nominal. $X_i^{cc} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$
Acceso al agua (X_i^{aa})	El hogar se considera privado cuando no tiene acceso al agua según los siguientes criterios: 1. Área urbana: el agua es obtenida principalmente de chorro público, pozo perforado público o privado, río, lago, manantial, camión cisterna, agua de lluvia y otro. 2. Área rural: el agua es obtenida principalmente de río, lago, manantial, camión cisterna, agua de lluvia y otro. 3. Para ambas áreas: el hogar recibe agua por tubería, pero no tuvo servicio durante 15 días o más en el mes anterior a la encuesta. Indicador de privación de tipo cualitativo.	Con base en la información consignada en el capítulo de las características de la vivienda en la ENCOVI 2014, se verifica el cumplimiento de las condiciones descritas por el IPM-Gt. Indicador de privación de escala nominal. $X_i^{aa} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$
Energía eléctrica (X_i^{ee})	El hogar se considera privado cuando la vivienda que habita no está conectada a servicio de energía eléctrica, o cuando sí está conectada a red de distribución de energía eléctrica, pero no contó con el servicio durante más de un día en el mes anterior a la encuesta. Indicador de privación de tipo cualitativo.	Con base en la información consignada en el capítulo de las características de la vivienda en la ENCOVI 2014, se verifica el cumplimiento de las condiciones descritas por el IPM-Gt. Indicador de privación de escala nominal. $X_i^{ee} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$
Recolección de basura (X_i^{rb})	El hogar se considera privado si no tiene acceso a servicio de recolección de basura y la mayor parte de la basura la queman, la entierran, la tiran en cualquier lugar o respondieron "otra" forma de eliminación de la basura. Indicador de privación de tipo cualitativo.	Con base en la información consignada en el capítulo de las características de la vivienda en la ENCOVI 2014, se verifica el cumplimiento de las condiciones descritas por el IPM-Gt. Indicador de privación de escala nominal. $X_i^{rb} = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$

Continúa tabla I.

Variable	Definición teórica	Definición operativa
Saneamiento (X_i^s)	El hogar se considera privado si no cuenta con saneamiento mejorado, según los siguientes criterios: 1. Área urbana: la vivienda no está conectada a una red de drenajes o sí está conectado a una red de drenaje, pero su servicio sanitario es un excusado lavable, letrina o pozo ciego, o no tiene. 2. Área rural: la vivienda cuenta con servicio sanitario tipo letrina o pozo ciego, o no tiene. Indicador de privación de tipo cualitativo.	Con base en la información consignada en el capítulo de las características de la vivienda en la ENCOVI 2014, se verifica el cumplimiento de las condiciones descritas por el IPM-Gt. Indicador de privación de escala nominal. $X_i^s = \begin{cases} 1 & A_i \in C \\ 0 & A_i \notin C \end{cases}$

Fuente: elaboración propia con datos obtenidos del IPM-Gt.

- Fases del estudio

Básicamente el proceso de estimación de las variables involucradas en el cálculo del IPM-Gt a nivel de departamento se desarrolló en cinco fases principales, las cuales se describen a continuación:

- Fase 1: Revisión de literatura

La primera fase del proyecto estuvo orientada a la revisión de las fuentes bibliográficas y otros materiales que fueron útiles para los propósitos del estudio, entre estos: documentación relacionada con el Índice de Pobreza Multidimensional IPM-Gt, la Encuesta Nacional de Condiciones de Vida (ENCOVI 2014), los estimadores directos aplicados a muestreos complejos y las metodologías de estimación en áreas pequeñas, entre otros.

- Fase 2: Gestión o recolección de la información

En esta fase se descargó de la página web del Instituto Nacional de Estadística (INE), la base de datos correspondiente a la Encuesta Nacional de Condiciones de Vida (ENCOVI 2014). Por otro lado, en la misma página del INE se utilizó la plataforma REDATAM para el manejo de la base de datos del XII Censo Nacional de Población y VII de Vivienda 2018. Esta información se utilizó en la implementación del *Empirical Best Linear Unbiased Predictor* EBLUP bajo el modelo *Fay Herriot*, ya que permitió el manejo de información auxiliar en relación con las condiciones de vida de los guatemaltecos, la cual fue fundamental para la utilización de la librería *Small Area Estimation* (SAE) del paquete estadístico RStudio.

- Fase 3: Análisis de información

La primera parte del análisis consistió en replicar la metodología del Índice de Pobreza Multidimensional IPM-Gt a través del software estadístico SPSS versión 25. Posteriormente, haciendo uso del módulo de muestras complejas del software estadístico SPSS versión 25, se estimaron los 17 indicadores correspondientes a las tasas de privación no censuradas del IPM-Gt a nivel de departamento, los niveles de incidencia e intensidad de la pobreza de la población guatemalteca por departamento y, finalmente, el Índice de Pobreza Multidimensional IPM-Gt a nivel de departamento. La creación del plan de muestreo en SPSS utilizó el área geográfica como campo de tabulación dentro de un dominio de estudio departamental, las unidades primarias de muestreo y los factores de expansión. La sintaxis utilizada se describe a continuación:

Tabla II. **Sintaxis estimaciones directas**

Código en SPSS
<pre> DATASET ACTIVATE ConjuntoDatos1. * Descriptivos de muestras complejas. CSD DESCRIPTIVES /PLAN FILE='C:\Users\Owner\Desktop\Informe final tesis\Tareas\www.IPM.gt\plann luis fernando '+ 'con area.csaplan' /SUMMARY VARIABLES=hh_d_cens_salud hh_d_cens_segal hh_d_cens_emba hh_d_cens_prenatal hh_d_cens_asis hh_d_cens_escol hh_d_cens_reza hh_d_cens_cinfantil hh_d_cens_inf hh_d_cens_trabainf hh_d_cens_mat hh_d_cens_hacina hh_d_cens_coci hh_d_cens_agua hh_d_cens_elec hh_d_cens_basu hh_d_cens_san M0_c_privaciones multi_poor A_intensidad /SUBPOP TABLE=DEPTO DISPLAY=LAYERED /MEAN /STATISTICS SE CV /MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE. </pre>

Fuente: elaboración propia con datos de la ENCOVI 2014, programa SPSS versión 25.

Finalmente, se estimó los indicadores de pobreza multidimensional de la población guatemalteca utilizando la librería *Small Area Estimation* (SAE), de Isabel Molina, a través del paquete estadístico RStudio. Para poder desarrollar este procedimiento fue necesario evaluar el grado de correlación existente entre cada indicador del IPM-Gt y la información contenida en las bases de datos del XII Censo Nacional de Población y VII de Vivienda 2018 a nivel de departamento (variables auxiliares). A continuación, se presenta un ejemplo con la tasa no censurada “recolección de basura” obtenida a través del software estadístico de uso libre Jamovi en su versión 1.8.2.

Tabla III. **Correlación entre recolección de basura y variables auxiliares**

Variable auxiliar	Parámetro	Recolección de basura
Escolaridad	Pearson's r	-0.817
	p-value	< .001
No servicio de agua	Pearson's r	0.651
	p-value	0.001
No servicio sanitario	Pearson's r	0.546
	p-value	0.009
Población Económicamente Activa	Pearson's r	-0.687
	p-value	< .001
Población ocupada	Pearson's r	-0.662
	p-value	< .001
Población desocupada	Pearson's r	-0.394
	p-value	0.07

Continúa tabla III.

Variable auxiliar	Parámetro	Recolección de basura
NBI Precariedad ocupacional	Pearson's r	0.734
	p-value	< .001
Personas de 15 y más alfabetas	Pearson's r	-0.712
	p-value	< .001
Paredes adecuadas en la vivienda	Pearson's r	-0.723
	p-value	< .001
Techo adecuado de la vivienda	Pearson's r	-0.325
	p-value	0.14
Piso adecuado en la vivienda	Pearson's r	-0.807
	p-value	< .001
Conexión a energía eléctrica	Pearson's r	-0.682
	p-value	< .001
Radio	Pearson's r	-0.722
	p-value	< .001
Televisor	Pearson's r	-0.819
	p-value	< .001
Refrigeradora	Pearson's r	-0.686
	p-value	< .001
Moto	Pearson's r	-0.138
	p-value	0.539
Computadora	Pearson's r	-0.859
	p-value	< .001
Internet	Pearson's r	-0.809
	p-value	< .001

Fuente: elaboración propia con datos de la ENCOVI 2014, programa Jamovi 1.8.2.

De acuerdo con el análisis de correlación aplicado a nivel de indicador, se procedió con la estimación de los principales parámetros utilizando el EBLUP bajo el modelo *Fay Herriot*. En este sentido, para determinar la estimación más aceptable se realizaron algunos ensayos experimentales, variando en cada caso la variable auxiliar utilizada (atendiendo al grado de correlación existente), lo que modificaba en cada ensayo el error cuadrático medio relativo obtenido, seleccionando finalmente la variable auxiliar que presentaba los valores más bajos de ECMR. Para el caso del estimador en áreas pequeñas EBLUP bajo el modelo *Fay Herriot* se fijó el 20 % como valor máximo para considerar una

estimación como aceptable. La siguiente tabla muestra un ejemplo de la sintaxis utilizada para indicador “recolección de basura”.

Tabla IV. **Sintaxis EBLUP bajo el modelo *Fay Herriot***

Código en RStudio
<pre>## carga de librerías rm(list = ls()) library(sae) library(tidyverse) library(XML) # Carga de base de datos con información de estimaciones puntuales, varianzas e información auxiliar basu <- read.csv("Libro1.csv", sep=";", dec=".") attach(basu) #Ajuste del modelo Fay Herriot mbasu <- mseFH(basura ~ NBIPrecariedadOcupacional + radio, vardir = varianza, method = "REML") ## Extracción del EBLUP, el Error Cuadrático Medio (ECM) y el Error Cuadrado Medio Relatico (ECMR) ECM_basu <- mbasu\$mse Eblup_basu <- mbasu\$est\$eblup ECMR_basu <- sqrt(ECM_basu)/Eblup_basu ## Generación de la base de datos con las estimaciones SAE Estima_SAE_basu <- data.frame(dominio = basu\$dominio, Eblup_basu = Eblup_basu, ECM_basu = ECM_basu, ECMR_basu = ECMR_basu) write.xml(Estima_SAE_basu, "Estimaciones_Recolección_Basura.xls")</pre>

Fuente: elaboración propia con datos de la ENCOVI 2014, programa RStudio.

- Fase 4: Interpretación de información

En esta fase se analizó e interpretó los resultados obtenidos a nivel de departamento en relación con las tasas de privación no censuradas, los niveles de incidencia e intensidad de la pobreza y el IPM-Gt.

Se consideró el criterio de estimaciones aceptables cuando el estimador directo presentó coeficientes de variación inferiores al 15 % y 20 % para el caso de las estimaciones basadas en el EBLUP bajo el modelo *Fay Herriot*.

Por otro lado, se contabilizó el número de veces que la estimación bajo el estimador compuesto presentó una disminución estadísticamente significativa en la cantidad de error relativo estimado.

- Fase 5: Redacción de informe final.

Con base en los resultados obtenidos se procedió a la elaboración del presente informe resumiendo la información en tablas estadísticas y mapas temáticos.

INTRODUCCIÓN

El presente estudio busca sistematizar el uso de metodologías de estimación en áreas pequeñas, específicamente el estimador EBLUP bajo el modelo *Fay Herriot*, para obtener estimaciones desagregadas de los principales indicadores de pobreza multidimensional de la población guatemalteca a nivel de departamento, según la metodología del Índice de Pobreza Multidimensional de Guatemala IPM-Gt.

Según el documento oficial del IPM-Gt (2018), publicado por el Ministerio de Desarrollo Social (MIDES), los resultados principales no consideraron niveles de desagregación departamental, ya que, al emplear el estimador directo basado en el diseño muestral complejo de la ENCOVI 2014, las estimaciones que se obtienen presentan coeficientes de variación superiores al 15 %, umbral estadísticamente aceptable dentro del marco del Sistema Integrado de Encuestas de Hogares (SIEH) del Instituto Nacional de Estadística (INE) de Guatemala. Lo anterior provoca que exista desinformación sobre la situación de los 22 departamentos de la República de Guatemala, con respecto las condiciones de vida y los niveles de pobreza multidimensional.

Derivado de lo anterior, el presente informe describe la implementación de la metodología de estimación en *áreas pequeñas* (término empleado cuando los errores de muestreo de las estimaciones son grandes, generando estimaciones imprecisas de los indicadores de interés) EBLUP bajo el modelo *Fay Herriot*, para obtener estimaciones aceptables sobre los niveles de incidencia e intensidad de la pobreza multidimensional de la población guatemalteca, que contribuya a fortalecer la toma de decisiones a todo nivel.

Los resultados obtenidos en esta investigación servirán para orientar la toma de decisiones informada a nivel técnico-político, ya que coadyuva a decidir estratégicamente sobre dónde focalizar los recursos y maximizar el impacto de los programas sociales en Guatemala. Lo anterior obedece a que la utilización del estimador compuesto representó un 98.4 % de mejora en el proceso inferencial aplicado en comparación con el estimador directo, al observarse una reducción del error relativo de estimación en 433 estimaciones de un total de 440; las restantes estimaciones presentan valores similares entre ambos estimadores.

Para el logro de los objetivos propuestos, en primer lugar, se realizó una revisión de las fuentes bibliográficas, posteriormente, se replicó la metodología del IPM-GT con base en la información contenida en las bases de datos de la ENCOVI 2014.

Tercero, se estimó los indicadores priorizados por el IPM-GT a nivel de departamento utilizado el estimador directo basado en el diseño muestral complejo de la ENCOVI 2014; por lo que se utilizó el paquete estadístico SPSS en su versión 25.

Cuarto, se analizó el uso de información auxiliar proveniente del XII Censo de Población y VII de Vivienda 2018 para Guatemala a través de la plataforma REDATAM, ubicada como opción de uso libre en la página web del Instituto Nacional de Estadística INE.

Quinto, se implementó el estimador EBLUP basado en el modelo *Fay Herriot*, para el cálculo de las estimaciones de los indicadores del IPM-Gt; se complementó el análisis con el uso del paquete estadístico de uso libre Jamovi en su versión 1.8.2., para el análisis del grado de asociación entre las variables utilizadas.

Por último, se evaluó las estimaciones obtenidas a través del análisis de los tamaños de muestra, el uso de información auxiliar y la calidad estadística de las estimaciones.

El estudio fue factible gracias a que se contó con las bases de datos de la Encuesta Nacional de Condiciones de Vida (ENCOVI 2014) y, se pudo extraer información del XII Censo de Población y VII de Vivienda de Guatemala 2018, a través del acceso libre a la plataforma REDATAM en la página web del Instituto Nacional de Estadística (INE).

El informe se integra por cuatro capítulos. El primero presenta el marco referencial, donde se describen las experiencias internacionales sobre mediciones de pobreza multidimensional y el creciente desarrollo de metodologías de estimación en áreas pequeñas para la desagregación de los datos estadísticos.

El segundo, describe el marco teórico donde se abordan los temas principales concernientes a la teoría general del muestreo, las metodologías de estimación en áreas pequeñas y la teoría general para la medición de la pobreza.

En el tercer capítulo, se incluyen los resultados obtenidos en la implementación del estimador del diseño complejo de la ENCOVI 2014 y el estimador EBLUP bajo el modelo *Fay Herriot*, con relación a los 20 indicadores de pobreza multidimensional que componen el IPM-Gt a nivel de departamento, expuestos a través de cuadros y mapas temáticos.

El cuarto capítulo, incluye una discusión de los resultados analizando la validez interna y externa de los mismos, a través de la evaluación de los tamaños de muestra observados, el uso de variables auxiliares, la calidad estadística de

las estimaciones y las experiencias internacionales en la evaluación de la precisión de indicadores.

1. MARCO REFERENCIAL

Con el transcurso del tiempo, han aumentado de manera creciente las necesidades de información desagregada geográficamente sobre las condiciones de vida y los niveles de pobreza, para la ejecución de políticas y programas destinados a atender las necesidades básicas de la población en los países de América Latina y el Caribe.

Yansaneh (2007) argumenta que en las últimas décadas se ha observado un incremento considerable en la “demanda de datos demográficos y socioeconómicos vigentes y detallados sobre los hogares e individuos en los países en desarrollo y en transición” (p.3). Las estadísticas sociales y demográficas resultan esenciales para la planificación y el control de los programas de desarrollo socioeconómico de un país, ya que son un elemento indispensable para la toma de decisiones en todos los niveles.

Según el Departamento de Asuntos Económicos y Sociales de la Secretaría de las Naciones Unidas (DESA, 2009), para la recopilación de datos sobre estadísticas sociales y demográficas pueden llevarse a cabo diferentes tipos de encuestas de hogares. Éstas incluyen las encuestas especializadas, las encuestas en varias etapas, las encuestas sobre varios temas y las encuestas longitudinales.

En este contexto, para atender la demanda de información oportuna sobre indicadores relacionados con los fenómenos sociales en Guatemala, tales como discapacidad, la pobreza, la participación de la población activa, la percepción de violencia e inseguridad, el estado nutricional, entre otros, las autoridades del

Instituto Nacional de Estadística (INE) y de otras instituciones del Sistema Estadístico Nacional (SEN) han recurrido con frecuencia a las encuestas de hogares.

La principal fuente estadística sobre las condiciones socioeconómicas en Guatemala son la Encuesta Nacional Empleo e Ingresos (ENEI), la Encuesta Nacional de Condiciones de Vida (ENCOVI) y la Encuesta Nacional de Ingresos y Gastos Familiares (ENIGFAM), todas a cargo del INE. La ENIGFAM se realiza cada decenio, la ENCOVI con una periodicidad quinquenal, y la ENEI a través de dos mediciones sincrónicas (transversales) por año a partir del 2017.

Romero (2000) afirma que la pobreza: “depende del desarrollo relativo alcanzado por determinada región o país. Así, tanto el carácter como la magnitud de la pobreza en los llamados países en desarrollo difiere cuantitativa y cualitativamente de la misma categoría en los países más avanzados” (p.2). Bajo este argumento, el fenómeno de la pobreza a nivel mundial es uno de los temas priorizados por las encuestas oficiales que realizan los institutos nacionales de estadística, así como los institutos regionales y otros organismos o instituciones a nivel supranacional o internacional.

La Comisión Económica para América Latina y el Caribe (CEPAL, 2018) presenta una actualización de las estimaciones de la pobreza por ingresos en 18 países, con el fin de brindar una panorámica de la pobreza a nivel regional que ofrezca una comparabilidad mayor que la que se obtiene directamente de las mediciones efectuadas en los países analizados. En el documento se describe la metodología de líneas de pobreza que la CEPAL ha utilizado, hasta esa actualización, las cuales se basan en canastas básicas de consumo que datan de la década de 1980.

En lo concerniente a las principales metodologías para la conceptualización y medición de la pobreza, destacan el método de líneas de pobreza, el método de las necesidades básicas insatisfechas y el método multivariado. El más ampliamente utilizado en los países de América Latina y el Caribe es el método tradicional de líneas de pobreza por ingresos. Larrañaga (2007) indica que:

La medición tradicional de la pobreza se ha realizado en base a ingresos. En parte ello responde al desarrollo de un marco conceptual que ha resuelto las preguntas fundamentales para la medición de la pobreza, también denominado paradigma de la pobreza según ingresos (*income poverty paradigm*). En cambio, la medición de la pobreza incorporando otras dimensiones se ha limitado en el pasado a la identificación del déficit en las respectivas variables, como es el caso del enfoque de Necesidades Básicas Insatisfechas, así como la más reciente medición que realiza el PNUD a través del Índice de Pobreza Humana. No obstante, lo anterior, en los últimos años han existido importantes contribuciones conceptuales en la medición de la pobreza multidimensional. (p.19)

Romero (2000) afirma que: “En términos monetarios la pobreza significa la carencia de ingresos suficientes con respecto al umbral de ingreso absoluto, o línea de pobreza” (p.37). Desde la posición de Romero:

La conceptualización de la pobreza a partir de los niveles de ingreso no permite explicar el acceso efectivo a los bienes y servicios fundamentales, ni a la capacidad de elección por parte del consumidor y, por lo tanto, revela sólo parcialmente el impacto de la disponibilidad monetaria sobre el bienestar, aunque se supone que el ingreso permite satisfacer las necesidades fundamentales. (p.37)

En lo referente a la cumbre anual de líderes del G8, celebrada en Okinawa en el año 2000, Romero (2000) enfatiza que:

En contra del enfoque de la pobreza basada solamente en los niveles de ingreso se manifestaron los países más desarrollados en la cumbre de los ocho, celebrada en Okinawa este año. Allí se afirmó que la pobreza "va más allá de la carencia de ingresos", pues esta es de carácter multidimensional e incluye lo económico, lo social y lo gubernamental. (p.37)

Un enfoque más complejo de pobreza es el que propone el premio Nobel de Economía, Amartya Sen, en el año 2002, para quien "la pobreza debe concebirse como la privación de capacidades básicas y no meramente como la falta de ingresos, que es el criterio habitual con el que se identifica la pobreza" (p.114). Desde su perspectiva, la pobreza basada en capacidades no entraña el rechazo de la falta de ingresos como una de las principales causas de pobreza.

De acuerdo con Larrañaga (2007) "En los años recientes ha habido algunos desarrollos teóricos orientados a derivar medidas de pobreza multidimensional. En cierto sentido, se trata de generalizar el marco de la pobreza en una dimensión (ingresos) al plano de diferentes variables" (p.21). Desde su punto de vista, se requiere resolver los problemas de identificación y de agregación, así como poder comparar distintas distribuciones de los atributos del bienestar en términos de una medida de pobreza.

Derivado del enfoque multidimensional de la pobreza, Alkire y Foster (2008) refieren que:

La pobreza multidimensional ha capturado la atención tanto de los investigadores como de quienes desarrollan políticas públicas, debido, en parte, al convincente trabajo conceptual de Amartya Sen y a la disponibilidad sin precedentes de datos relevantes. Una línea fundamental de investigación ha estado orientada al desarrollo de un marco coherente para medir la pobreza en un contexto multidimensional que es análogo a la serie de técnicas desarrolladas en el espacio unidimensional. (p.1)

Alkire y Foster (2008) proponen una nueva metodología para la medición multidimensional de la pobreza a través del Índice de Pobreza Multidimensional (IPM). El proceso incluye doce pasos propuestos por la Iniciativa de Oxford para el Desarrollo Humano y la Reducción de la Pobreza (OPHI). Los primeros seis son comunes a muchas medidas de pobreza multidimensional; el resto son específicos del Método Alkire-Foster.

La metodología de Alkire-Foster de OPHI utiliza dos tipos de línea de corte para identificar a los pobres. La primera es la línea tradicional de pobreza o línea de corte basada en dimensiones específicas, que identifica si una persona sufre privaciones en relación con esa dimensión. La segunda marca cuán amplias deben ser las privaciones que sufre una persona para ser considerada pobre. Alkire y Foster (2008) afirman que:

La metodología satisface una variedad de axiomas deseables, incluyendo la descomponibilidad, una propiedad que facilita la focalización, y un nuevo requisito de monotonía dimensional a través del cual una expansión en el rango de privaciones sufridas por una persona pobre se ve reflejada en el nivel general de pobreza. (p.2)

El aumento en los países de América Latina y el Caribe del uso de las encuestas de hogares, para el abordaje de los fenómenos sociales provocó un incremento en la demanda de datos estadísticos cada vez más desagregados. De acuerdo con Molina (2018):

Existe una demanda creciente de información para grupos específicos de la población o para áreas geográficas menores. A manera de ejemplo, el marco global de indicadores para el seguimiento de los Objetivos del Desarrollo Sostenible plantea que la información debe ser desagregada no solo de manera geográfica (en subregiones de interés como provincias, municipalidades, comunas o localidades), sino también por grupos de ingreso, sexo, edad, raza, origen étnico, estatus migratorio y condición de discapacidad. No obstante, la confiabilidad en la inferencia de los indicadores disminuye a medida que lo hace el tamaño de muestra, por lo que generalmente no es posible lograr los niveles de desagregación deseados con una precisión adecuada. (p.9)

Según refiere Álvarez (2001), los métodos de estimación indirecta para áreas geográficas menores han experimentado un largo desarrollo en América Latina y el Caribe por tres razones principales.

La primera, el fuerte aumento de la demanda de información desagregada, vinculada a aspectos legales de representación política o a ciertas asignaciones presupuestarias, en el marco de procesos de descentralización y la focalización de las políticas sociales en general.

La segunda razón, una mayor frecuencia, periodicidad y sistematización del levantamiento de encuestas de hogares y demográficas con representación nacional, urbana y rural y, por último, la más amplia disponibilidad de medios

computacionales de bajo costo y alta velocidad de procesamiento de bases de datos.

La Comisión Económica para América Latina y el Caribe (CEPAL, 2018) postula que: “Si bien las encuestas de hogares son menos costosas que los censos, pueden llegar a encarecerse notablemente si los resultados han de producirse por separado para cada pequeño territorio administrativo, como las provincias o los distritos” (p.3). La capacidad de la encuesta de producir datos fiables se reduce considerablemente cuanto más pequeña sea la muestra en el área de interés.

En cuanto a una definición de “área pequeña” en el campo de la desagregación de datos de las encuestas de hogares, Molina (2018) la establece de la siguiente manera:

En la literatura, las subdivisiones para las cuales se desea obtener datos estadísticos (o estimaciones) se denominan comúnmente “áreas” o “dominios”, con independencia de si son delimitaciones geográficas, o socioeconómicas, o un cruce de ambos tipos. A la hora de estimar un indicador concreto en una de estas áreas, llamamos estimador directo a un estimador que usa solamente los datos de la encuesta para esa área. Los estimadores directos habituales son insesgados o prácticamente insesgados respecto a la distribución del diseño muestral, e.g., a lo largo de todas las posibles muestras que se pueden extraer de la población mediante el correspondiente diseño muestral. No obstante, si la encuesta no se había planificado para estimar a un nivel tan desagregado, el tamaño muestral en algunas de las áreas puede ser demasiado pequeño, lo cual se traduce en errores de muestreo excesivamente grandes para los estimadores directos de los indicadores de interés en esas áreas. Las

áreas en las que esto ocurre, independientemente de su tamaño poblacional, se conocen en la literatura como “áreas pequeñas”. Por tanto, no es el tamaño poblacional del área el que le confiere el adjetivo de “pequeña” pues, en muchos casos, áreas de gran tamaño poblacional (e.g. los estados de EE. UU.) se consideran áreas pequeñas si no se dispone de estimaciones directas de suficiente calidad. Concretamente, el término “área pequeña” hace referencia a áreas, en las cuales el estimador directo del indicador de interés es ineficiente debido al insuficiente número de observaciones obtenidas (o encuestas realizadas) de esa área. (p.14)

A nivel internacional, ejemplos de estudios realizados utilizando las metodologías de estimación en áreas pequeñas se encuentran en la literatura. De acuerdo con Elbers, Lanjouw y Lan (2002) existen experiencias documentadas sobre la aplicación de estas técnicas estadísticas para la desagregación de datos en Sudáfrica, Brasil, Panamá y Nicaragua. Los autores, con información de fuentes oficiales en Ecuador, describen un procedimiento estadístico que combina el muestreo de encuestas de hogares y la información censal más reciente, para estimar los niveles de bienestar y pobreza en ese país.

Fay y Herriot en el año 1979 (como se citó en Celdrán, 2012) estimaron la renta per cápita en áreas con menos de 1000 habitantes en Estados Unidos, para determinar el reparto de fondos a los gobiernos locales. Su método se utilizó nuevamente en el año 2000 para producir estimaciones por condado de la pobreza infantil en Estados Unidos.

Celdrán (2012) agrega que en el ámbito europeo la Oficina Nacional de Estadística del Reino Unido (*Office for National Statistics*, ONS) generó estimaciones basadas en estas metodologías con datos de desempleo a partir

de su encuesta de fuerza de trabajo (*Labor Force Survey, LFS*) y que recientemente, han sido aceptadas como estadística oficial en el país.

En lo concerniente a la utilización de las metodologías de estimación en áreas pequeñas en el ámbito guatemalteco, Romero y Zapil (2009) refieren que con los datos de la ENIGFAM 1998-1999, efectuaron estimaciones de las condiciones de pobreza empleando el método de estimación en áreas pequeñas desarrollado por Chris Elbers, Jean Lanjouw y Peter Lanjouw en el año 2003. En esa ocasión, el 64.3 % de la población se encontraba en condición de pobreza. Por otro lado, las zonas más atrasadas en lo que se refiere a infraestructura, formación de capacidades humanas en lo educativo, salud y otras dimensiones, fueron la Región del Sur Occidente y la Región Norte que oscilan entre 75.2 % y 79.0 %, respectivamente.

Al explorar las experiencias en Guatemala sobre procesos de estimación de la pobreza multidimensional a un nivel de desagregación determinado, el Ministerio de Desarrollo Social (MIDES, 2018) revela que la combinación de los componentes del IPM-Gt ilustra una realidad hasta hoy desconocida en el país y que, además, el Índice de Pobreza Multidimensional para Guatemala IPM-Gt se concibe en el marco de dos grandes agendas de desarrollo: el Plan Nacional de Desarrollo Nuestra Guatemala K'atun 2032 y la Agenda de Desarrollo Sostenible 2030.

Según el informe del MIDES, se estima que en Guatemala el 61.6 % de las personas con un intervalo de confianza del 95 % de [58.6 %, 64.7 %] viven en pobreza multidimensional, es decir, seis de cada 10 guatemaltecos enfrentan privaciones en el 30 por ciento o más de los indicadores ponderados incluidos en el IPM-Gt. La intensidad de la pobreza estimada fue de 48.6 % con un intervalo de confianza al 95 % de [47.5 %, 49.7 %], y que el IPM-Gt para el año 2014 se

estima en 0.299 (el IPM es el producto de multiplicar la incidencia por la intensidad de pobreza multidimensional).

Como evidencian los párrafos anteriores, el interés por la inferencia de indicadores de pobreza multidimensional, por medio de las metodologías de estimación en áreas pequeñas, ha aumentado enormemente, al grado ser aceptadas como estadística oficial en algunos países.

2. MARCO TEÓRICO

A continuación, se presenta la base conceptual que sirve de guía para comprender la metodología e interpretación de los resultados de la presente investigación.

2.1. Teoría general de la inferencia estadística

Cuando se desea obtener información acerca de las características de una población, pero sólo se dispone de datos sobre una muestra, se necesitan algunos procedimientos para utilizar los datos de la muestra y sacar conclusiones acerca de la población de interés.

2.1.1. Inferencia estadística o estadística inductiva

Citando a Pérez (2010) “Se denomina inferencia estadística o estadística inductiva a la metodología consistente en inferir resultados, predicciones y generalizaciones sobre la población estadística, basada en la información contenida en las muestras representativas previamente elegidas por métodos de muestreo formales” (p.4). La inferencia estadística tiene sus bases en la teoría de la probabilidad, pero tiene un carácter distinto. Desde la perspectiva de Pérez (2010) “En inferencia estadística se consideran fenómenos en los que se manifiesta la regularidad estadística y se construyen modelos probabilísticos para describirlos” (p.4).

Cabe señalar que para la inferencia de poblaciones finitas los estadísticos e investigadores utilizan dos enfoques principales: la inferencia basada en el

diseño y la inferencia basada en el modelo. Sin embargo, en las últimas décadas ha surgido un nuevo enfoque nombrado como la inferencia asistida por un modelo y basada en el diseño.

2.1.1.1. Inferencia basada en el diseño

Särndal, Swensson y Wretman (1992) afirman que este enfoque realiza las estimaciones de los parámetros de interés dependiendo del esquema de muestreo escogido para seleccionar las unidades en la muestra. Este método no toma en cuenta las propiedades de la población. El enfoque basado en el diseño fue consecuencia del desarrollo del método representativo de muestreo.

Según Rao (2005), el estadístico noruego Anders Nicolai Kiær (1897) es quizás el primero en promover la utilización del muestreo representativo (más adelante conocido como teoría de muestreo) en lugar de una enumeración completa de una población. Sin embargo, Hacking (como se citó Rao, 2005) señala que “La referencia más antigua al muestreo se remonta a la gran epopeya india el Mahabharata (Hacking 1975, page 7)” (p. 117). En el método representativo la muestra debe reflejar a la población objetivo, ya sea a través de una selección intencionada, o bien mediante muestreo aleatorio.

Rao (2005) refiere que, para la década de 1920, el muestreo representativo era usado de manera difundida en varios países alrededor del mundo. Por ejemplo, en Rusia fue utilizado a principios de 1900 y en los Estados Unidos se dirigieron encuestas por muestreo alrededor del mismo período utilizando este método. Esta situación trajo consigo la creación, en 1924, del ISI (Instituto Internacional de Estadística por su traducción del inglés al castellano) el cual crea un comité para la evaluación y discusión sobre los aspectos teóricos y prácticos del método de muestreo aleatorio.

Rao (2005) menciona que la contribución de Bowley (1926) a este informe incluyó su trabajo fundamental sobre muestreo estratificado aleatorio con asignación proporcional, lo que conduce a una muestra con iguales probabilidades de inclusión.

De acuerdo con Rao (2005), como consecuencia de estos estudios, Neyman examina las bases de la inferencia en poblaciones finitas y escribe su artículo histórico en 1934, estableciendo así los fundamentos de la teoría del muestreo probabilístico (o inferencia basada en el diseño). Este enfoque se basó en la inferencia a partir de encuestas.

Según Rao (2005) Neyman demostró, tanto teóricamente como con ejemplos prácticos, que el muestreo aleatorio estratificado es preferible al muestreo equilibrado, porque este último puede funcionar mal si no se cumplen los supuestos del modelo. Cabe resaltar que el enfoque propuesto por Neyman es ampliamente utilizado por las agencias de estadística del mundo.

2.1.1.2. Inferencia basada en el modelo

Valliant, Dorfman y Royall (como citó Gutiérrez, 2015) argumentan que “la inferencia basada en el modelo supone el uso de información auxiliar y relaciona la característica de interés con la información auxiliar mediante un modelo de superpoblación ξ ” (p.405). Bajo esta perspectiva no es requisito que los datos provengan de una muestra probabilística (la forma de escogencia de la muestra no es tomada en cuenta en la estimación) y la observación medida en las unidades poblaciones x_i se define como una realización de la variable aleatoria X_i .

2.1.1.3. La inferencia asistida por un modelo y basada en el diseño

Gutiérrez (2015) da a conocer que los estadísticos Cassel, Särndal y Wretman en el año 1976 acuñaron el término de inferencia asistida por un modelo y basada en el diseño (*model assisted design based*, en el inglés original). Como lo hace notar Gutiérrez (2015) en este enfoque el diseño de muestreo es la base de la inferencia estadística, pero la estrategia de muestreo es complementada teniendo en consideración un ξ modelo de estimación para las características de interés.

Desde la posición de Gutiérrez (2015), “la inferencia basada en modelos predictivos y la inferencia basada en el diseño de muestreo no se deben ver como competencia sino como puntos de vista que pueden llegar a ser complementarios” (p. 406). Es así como surge la inferencia basada en el diseño, pero asistida por modelos predictivos. Sin embargo, estos tipos de inferencia, aunque se pueden combinar, no se pueden conciliar porque su filosofía es literalmente distinta.

Citando a Gutiérrez (2015), “La inferencia basada en el diseño difiere radicalmente de la inferencia basada en modelos predictivos y quizás de cualquier otro modelo estadístico, porque está basada exclusivamente en las observaciones muestrales y no hace supuestos a priori” (p. 406). Desde su punto de vista, la inferencia basada en el diseño de muestreo es inversa y la inferencia basada en modelos predictivos es directa.

2.1.2. Teoría general del muestreo estadístico

Pérez (2010) plantea que en toda investigación estadística existe un conjunto de unidades elementales sobre las cuales se toman mediciones. Este conjunto de elementos se denomina con el nombre de población o universo estadístico. Cuando se toma información de todos y cada uno de los elementos de la población se dice que se efectuó un censo.

Sin embargo, en muchas ocasiones, ya sea por el coste que resulta de la toma de los datos, o bien por el factor tiempo, o porque la toma de información implica la destrucción de los elementos en estudio, o que la población tenga infinitos elementos, o por otras causas, la toma de información para las unidades de observación no es posible.

Pérez (2010) enfatiza que esta situación lleva al investigador a tomar datos sólo de una fracción del total de elementos de la población, proceso que recibe el nombre de muestreo. El conjunto finito de elementos de los que se toman mediciones en el proceso de muestreo se llama muestra y el número de unidades elementales que la componen tamaño de la muestra.

El investigador utiliza la información de la muestra para inferir respecto a la población estadística, por lo tanto, es requisito fundamental que las características o variables de interés se reflejen en la muestra de la manera más cercana posible. Para precisar respecto al concepto de una muestra representativa es necesario definir algunos conceptos.

2.1.2.1. Taxonomía de unidades en el muestreo

Empleando las palabras de Kish (1972) “los elementos de una población son las unidades de las que se busca información; son los individuos, las unidades elementales que forman la población acerca de la cual se van a hacer inferencias” (p. 27). Estos elementos son las unidades de análisis, y su naturaleza es determinada mediante los objetivos de investigación.

Agrega Kish (1972) que “La población se define juntamente con los elementos: la población es el agregado de los elementos, y los elementos son las unidades básicas que forman y definen una población” (p.47).

Por otra parte, la población de la muestra que realmente se obtenga tal vez difiera un poco de la población objetivo, las principales diferencias suelen deberse a la no respuesta y a la cobertura deficiente. Es importante señalar que una muestra puede dar información acerca de varias y diversas poblaciones, incluso, la misma encuesta puede abarcar contenidos diferentes.

2.1.2.2. Parámetro poblacional y estadísticas

Desde el punto de vista de la teoría de muestreo, tal y como refiere Kish (1972), un parámetro poblacional se define de la siguiente manera:

Un valor de la población es una expresión numérica que sintetiza los valores de una o varias características de los N elementos de una población completa; es una medida resumida de una cualidad de la distribución de la variable o variables en la población definida. (p. 29)

También refiere que el valor calculado en la muestra, o estadística, es una estimación calculada a partir de los n elementos que integraron la muestra.

Los principales parámetros estimados en función de los valores que la característica cuantitativa X o cualitativa A toma sobre los elementos de la población, son: el total, la media, el total de clase A y la proporción.

2.1.2.3. Métodos de muestreo

De acuerdo con Pérez (2005) se define por métodos de muestreo “al conjunto de técnicas que estudian la forma de seleccionar una muestra lo suficientemente representativa de la población cuya información permita inferir respecto a las propiedades o características de toda la población cometiendo un error medible y acotable” (p.3). Existen varios tipos de muestreo, dependiendo de que la población estadística sea finita o infinita. Para el caso de poblaciones finitas, seleccionar una muestra mediante un determinado método de muestreo probabilístico permite estimar las características de interés (media, total, proporción, etc.) con un error cuantificable y controlable.

2.1.2.4. Muestreo probabilístico

Como afirma Pérez (2005) para medir el grado de representatividad de una muestra es necesario hacer uso del muestreo probabilístico. Pérez (2005) argumenta que “el muestreo es probabilístico cuando pueda establecerse la probabilidad de obtener cada una de las muestras que sea posible seleccionar, esto es, cuando la selección de muestras constituya un fenómeno aleatorio probabilizable” (p. 3).

Por lo tanto, la selección de una muestra representativa de la población estadística debe realizarse en condiciones de azar, lo que permitirá medir los errores cometidos en el proceso de muestreo (a través de varianzas u otras medidas estadísticas).

Algunos ejemplos de muestras probabilísticas son: muestreo aleatorio simple, muestreo estratificado, muestreo por conglomerados, muestreo en dos etapas, entre otros.

2.1.2.5. Diseño de encuestas y diseño de muestras

Kish (1972) postula que el diseño de muestras debe considerar dos aspectos: un proceso de selección, que consiste en definir las reglas y operaciones mediante las cuales se incluirán en la muestra algunos elementos de la población estadística; y un proceso de estimación (o estimador) para estimar los parámetros de interés a partir de la muestra.

Según Kish (1972) el diseño completo de encuestas debe incluir otros aspectos importantes como los objetivos de la encuesta, la definición de las variables de la encuesta, los métodos de observación (medición), el uso o utilización de los resultados, los métodos de análisis, la precisión que se desea para los resultados de la encuesta, entre otros.

2.1.2.6. Distribución de muestreo

Teniendo en cuenta a Kish (1972) si se especifica un diseño de muestreo, que incluya el tamaño de la muestra, así como el procedimiento de selección y de estimación; y posteriormente se aplican estos criterios a una población estadística, seleccionando muestra tras muestra, y calculando el parámetro de

interés de cada muestra posible, y después graficamos estos valores según la distribución de sus frecuencias relativas.

Conforme aumente el número de puntos graficados, la forma de la distribución se hará cada vez más estable y gradualmente se acercará a la distribución verdadera de muestreo. Esto sucederá cuando a medida que las frecuencias relativas para los diferentes valores del parámetro de interés se acerquen a sus probabilidades reales de materialización.

Derivado de lo anterior Kish (1972) indica que “la distribución de muestreo de un estimador (media) es una distribución teórica de todos los valores posibles del estimador (\bar{y}_c), cada uno con su probabilidad de materialización (P_c)” (p. 31).

Kish (1972) agrega que la media de la distribución de muestreo es el valor esperado (o la esperanza matemática) del estimador, por consiguiente,

$$E(\bar{y}) = \sum_c P_c \bar{y}_c \quad (\text{Ec. 01})$$

2.1.2.7. Estimador

De acuerdo con Mirás (1985) “La función $\hat{\theta}$ que asocia a cada muestra, s , el valor numérico $\hat{\theta}(s)$, se denomina estimador y $\hat{\theta}(s)$ es la estimación particular de θ dada s ” (p. 3). Definido un procedimiento de muestreo, un estimador se convierte en una variable aleatoria, al considerar la variabilidad de las muestras. Los errores se cuantifican mediante varianzas, desviaciones típicas o errores cuadráticos medios de los estimadores, que miden la precisión de éstos.

Mirás (1985) señala que para establecer un procedimiento de estimación es necesario considerar dos operaciones fundamentales: una de ellas consiste en diseñar el procedimiento de muestreo y la otra consiste en formar el estimador.

2.1.2.8. Varianzas de los estimadores

Teniendo en cuenta a Cochran (1974) la varianza de y_i en una población finita es usualmente definida como,

$$\sigma^2 = \frac{\sum_1^N (y_i - \bar{Y})^2}{N} \quad (\text{Ec. 02})$$

Por motivos de notación, Cochran (1974) presenta la expresión anterior de una forma ligeramente diferente, en donde el divisor $(N - 1)$ es usado en lugar de N . Es así como se tiene:

$$S^2 = \frac{\sum_1^N (y_i - \bar{Y})^2}{N - 1} \quad (\text{Ec. 03})$$

Esta expresión ha sido utilizada por estadísticos cuyo enfoque a la teoría del muestreo lo hacen por medio del análisis de la varianza. Ahora bien, si se considera la varianza de la media, es decir, $E(\bar{y} - \bar{Y})^2$ tomada sobre todas las muestras posibles ${}_N C_n$, la varianza de una muestra simple aleatoria es,

$$V(\bar{y}) = E(\bar{y} - \bar{Y})^2 = \frac{S^2}{n} \frac{(N - n)}{N} = \frac{S^2}{n} (1 - f) \quad (\text{Ec. 04})$$

En donde $f = n/N$ es la fracción de muestreo.

Por otra parte, el error estándar (error de muestreo) de la \bar{y} se obtienen de la siguiente manera:

$$\sigma_{\bar{y}} = \frac{S}{\sqrt{n}} \sqrt{\frac{(N-n)}{N}} = \frac{S}{\sqrt{n}} \sqrt{1-f} \quad (\text{Ec. 05})$$

El estimador de la varianza de un total $\hat{Y} = N\bar{y}$ es

$$V(\bar{y}) = E(\bar{y} - \bar{Y})^2 = \frac{N^2 S^2 (N-n)}{n N} = \frac{N^2 S^2}{n} (1-f) \quad (\text{Ec. 06})$$

Y el error estándar del total \hat{Y} es

$$\sigma_{\hat{Y}} = \frac{NS}{\sqrt{n}} \sqrt{\frac{(N-n)}{N}} = \frac{NS}{\sqrt{n}} \sqrt{1-f} \quad (\text{Ec. 07})$$

El factor $(1-f)$ para la varianza y $\sqrt{1-f}$ para el error estándar se denominan correcciones debidas a población finita o corrección por finitud.

2.1.2.9. Estimación puntual y por intervalos de confianza

Como lo hace notar Pérez (2005) al estimar las características de interés de la población estadística basándose en la información contenida en una muestra, pueden emplearse los valores puntuales de un estadístico basado en la misma, o bien puede utilizarse un intervalo de valores dentro del cual se tendrá cierta confianza de que esté el valor del parámetro de interés.

En el primer caso se hace referencia al proceso de estimación puntual, en el que se utiliza directamente los valores de un estadístico, denominado estimador puntual. En el segundo caso se emplea la estimación por intervalos de confianza, donde se calcula un intervalo atendiendo a un nivel de confianza prefijado.

Pérez (2005) agrega que al analizar la forma de los estimadores puntuales óptimos de $\hat{\theta} = (X_1, \dots, X_n)$ para los parámetros poblacionales tradicionales, resulta que las mejores propiedades suelen presentarlas los estimadores lineales insesgados de la forma:

$$\hat{\theta} = \sum_{i=1}^n w_i Y_i \quad (\text{Ec. 08})$$

En donde los valores w_i se denominan pesos o factores de expansión (inverso de la probabilidad de selección), y serán los números por lo que hay que multiplicar los valores muestrales para obtener las estimaciones poblacionales.

Para el caso del muestreo sin reposición, el estimador óptimo es el de Horvitz y Thompson:

$$\hat{\theta}_{HT} = \sum_{i=1}^n \frac{Y_i}{\pi_i} \quad (\text{Ec. 09})$$

En donde π_i es la probabilidad que tiene la unidad u_i de la población estadística de pertenecer a la muestra. Los pesos o factores de elevación son en este caso:

$$w_i = \frac{1}{\pi_i} \quad (\text{Ec. 10})$$

Para el muestreo con reposición, el estimador óptimo es el de Hansen y Hurwitz.

$$\hat{\theta}_{HH} = \sum_{i=1}^n \frac{Y_i}{nP_i} \quad (\text{Ec. 11})$$

En donde P_i es la probabilidad unitaria de selección de la unidad u_i de la población estadística para la muestra. Los pesos o factores de elevación son en este caso:

$$w_i = \frac{1}{nP_i} \quad (\text{Ec. 12})$$

Para el cálculo de intervalos de confianza debe evaluarse si el estimador es sesgado o insesgado. En el caso de estimadores insesgados deberá evaluarse si el estimador tiene una distribución normal o no. Si el estimador $\hat{\theta}$ es insesgado y tiene una distribución normal, el intervalo de confianza para el parámetro poblacional θ basado en $\hat{\theta}$ será:

$$[\hat{\theta} - \lambda_{\alpha}\sigma(\hat{\theta}), \hat{\theta} + \lambda_{\alpha}\sigma(\hat{\theta})] \text{ con } \lambda_{\alpha} = F_{N(0,1)}^{-1}\left(1 - \frac{\alpha}{2}\right) \quad (\text{Ec. 13})$$

En donde F es la función de distribución de la normal (0,1), y α es el nivel de confianza. Si se duda de que el estimador insesgado del parámetro $\hat{\theta}$ tenga una distribución normal, puede utilizarse la distribución t de Student con n - 1

grados de libertad para calcular el intervalo de confianza, en este caso el intervalo será:

$$[\hat{\theta} - t_{\alpha}\sigma(\hat{\theta}), \hat{\theta} + t_{\alpha}\sigma(\hat{\theta})] \text{ con } t_{\alpha} = F_{t_{n-1}}^{-1}\left(1 - \frac{\alpha}{2}\right) \quad (\text{Ec. 14})$$

Si el estimador no tiene una distribución normal, Pérez (2005) manifiesta que derivado de la desigualdad de Tchevichev, para el parámetro poblacional θ basado en $\hat{\theta}$ que cubre el valor de θ con una probabilidad $1 - \alpha$ (coeficiente de confianza), será:

$$\left[\hat{\theta} - \frac{\sigma(\hat{\theta})}{\sqrt{\alpha}}, \hat{\theta} + \frac{\sigma(\hat{\theta})}{\sqrt{\alpha}}\right] \quad (\text{Ec. 15})$$

Este intervalo suele ser más ancho que el obtenido cuando la distribución de $\hat{\theta}$ es normal. Por último, el intervalo de confianza para θ basado en el estimador $\hat{\theta}$ en presencia del sesgo no despreciable $B(\hat{\theta}) = E(\hat{\theta}) - \theta$ es el siguiente:

$$[\hat{\theta} - \lambda_{\alpha}\sigma(\hat{\theta}) - |B(\hat{\theta})|, \hat{\theta} + \lambda_{\alpha}\sigma(\hat{\theta}) + |B(\hat{\theta})|] \quad (\text{Ec. 16})$$

2.1.2.10. Propiedades y precisión de los estimadores

Martínez (2012) plantea que un buen estimador debe reunir cuatro condiciones o propiedades: insesgado, es decir, que no tenga sesgo (sin error), lo cual sucede cuando la estimación es igual al parámetro poblacional. El

segundo aspecto para considerar es que el estimador sea consistente, es decir que, al aumentar el tamaño de la muestra, la varianza disminuye siendo menor la diferencia entre el valor real y el estimado. Un tercer aspecto es que sea eficiente, lo cual implica que el estimador que tiene la menor varianza entre todos los estimadores. Por último, cuando se incluye toda la información que la muestra puede proporcionar acerca del parámetro se dice que el estimador es suficiente.

Pérez (2005) refiere que “Para analizar la precisión de un estimador suelen utilizarse los conceptos de error de muestreo (o desviación típica), acuracidad (o error cuadrático medio) y sesgo” (p. 6). A juicio de Pérez (2005) suele llamarse precisión a la acuracidad, lo que no es del todo correcto, ya que, algunas veces el análisis se realiza en función de otras magnitudes. Todas estas magnitudes que influyen en la precisión de un estimador pueden relacionarse a partir de la descomposición del error cuadrático medio de la siguiente forma,

$$ECM(\hat{\theta}) = \sigma(\hat{\theta})^2 + B(\hat{\theta})^2 \quad (\text{Ec. 17})$$

Por lo tanto, la acuracidad (o error cuadrático medio) de un estimador se descompone en la suma del cuadrado del error de muestreo y el cuadrado del sesgo. Cochran (1974) al referirse al sesgo y sus efectos, destaca que en el caso de considerar la probabilidad de cometer un error de más de 1.96σ , el sesgo tiene poco efecto, siempre y cuando sea menor de una décima de la desviación estándar.

$$\left| \frac{B(\hat{\theta})^2}{\sigma(\hat{\theta})} \right| < \frac{1}{10} \quad (\text{Ec. 18})$$

- Cuantificación de la ganancia en precisión de los estimadores

Pérez (2005) postula que “para medir la precisión de los estimadores suele utilizarse el error cuadrático medio, el error relativo (coeficiente de variación) o el error de muestreo (desviación típica)” (p. 8). En cada caso, la ganancia en precisión estará dada por las respectivas tasas de variación:

$$\left(\frac{ECM(\hat{\theta}_1)}{ECM(\hat{\theta}_2)} - 1 \right) \times 100 \quad (\text{Ec. 19})$$

$$\left(\frac{CV(\hat{\theta}_1)}{CV(\hat{\theta}_2)} - 1 \right) \times 100 \quad (\text{Ec. 20})$$

$$\left(\frac{\sigma(\hat{\theta}_1)}{\sigma(\hat{\theta}_2)} - 1 \right) \times 100 \quad (\text{Ec. 21})$$

2.1.2.11. Formas básicas de selección de muestras probabilísticas

Como plantea Pérez (2005) las formas básicas de selección de una muestra se clasifican atendiendo a los siguientes criterios:

- Atendiendo a las probabilidades de selección

Bajo esta modalidad, la selección de la muestra puede darse de dos maneras: con probabilidades iguales y con probabilidades desiguales. Con probabilidades iguales se refiere a que todas las unidades de la población estadística tienen la misma probabilidad de ser incluidas en la muestra en cada

extracción, para el caso de con probabilidades desiguales, al menos dos unidades tienen distintas probabilidades de inclusión en cierta extracción.

- Atendiendo a la mecánica de selección

La mecánica de la selección puede darse de dos maneras: con reposición y sin reposición. Una selección con reposición se efectúa si cada unidad que es extraída para formar parte de la muestra en una extracción particular se repone a la población antes de realizar la siguiente extracción; por lo tanto, la estructura poblacional permanece invariante.

Para el caso del muestreo sin reposición, cada unidad que es extraída para formar parte de la muestra en una extracción no se repone a la población antes de realizar la siguiente extracción, por lo que una unidad elemental podrá aparecer en la muestra a lo sumo una vez y la estructura poblacional cambiará de una extracción a otra.

Al combinar estos cuatro tipos de muestreo resulta:

- Muestreo con reposición y probabilidades iguales.
- Muestreo sin reposición y probabilidades iguales.
- Muestreo con reposición y probabilidades desiguales.
- Muestreo sin reposición y probabilidades desiguales.

2.1.2.12. Marco de muestreo

Por otro lado, como plantea Pérez (2005) “para seleccionar la muestra, es necesario un listado de unidades de muestreo denominado marco que teóricamente debiera coincidir con la población objetivo” (p. 3). Desde su punto

de vista un marco será más adecuado cuanto mejor cubra la población estadística, es decir, cuanto menor sea el error de cobertura. En la práctica, para el caso de los marcos de muestreo, son inevitables las desactualizaciones, las omisiones de algunas unidades de muestreo, las duplicaciones de otras y la presencia de unidades extrañas y otras impurezas que obligan a su depuración.

El marco puede estar constituido por unidades elementales de muestreo o por unidades compuestas. Pérez (2005) refiere que “una unidad elemental (o simple) es la unidad de muestreo más sencilla posible y una unidad compuesta (o primaria) está formada por varias unidades elementales” (p. 3).

Como en la práctica no es fácil disponer de marcos de unidades elementales, una solución es intentar conseguir marcos de unidades compuestas que son más accesibles. Por ejemplo, para estudiar habitantes de una región es más fácil disponer de un listado de manzanas que de un listado de individuos. Se selecciona la muestra de un marco de manzanas (unidades compuestas de varios individuos) y después se estudian las propiedades de los individuos con técnicas adecuadas.

2.1.2.13. Tamaño de la muestra

Como lo hace notar Cochran (1974) en la planeación de una encuesta por muestreo, siempre se alcanza un punto en el cual se debe tomar una decisión acerca del tamaño de la muestra. La decisión es importante. Una muestra muy grande implicará el desperdicio de recursos, mientras que una muestra muy pequeña disminuye la utilidad de los resultados. Por esta razón, la teoría del muestreo proporciona un planteamiento general dentro del cual puede resolverse este problema.

Según Martínez (2012) en términos técnicos, algunos de los componentes que son comunes a las diferentes fórmulas para el cálculo del tamaño de la muestra, dependiendo del método que se aplique, son los siguientes: nivel de confianza, grado de variabilidad, error de muestreo y el tamaño de la población estadística.

El grado de confianza es fijado por el investigador y se refiere al porcentaje de intervalos de confianza que incluirían el parámetro de interés si se tomaran muestras de la misma población repetidamente, una y otra vez bajo el mismo esquema de muestreo. Por lo general, un nivel de confianza de 95 % funciona adecuadamente.

El grado de variabilidad está dado por la varianza. Según Martínez (2012) “Entre más variabilidad presente la característica, mayor será el tamaño de la muestra, necesario para que represente a la población” (p.223).

Otro aspecto que fija el investigador de acuerdo con su experiencia y el conocimiento que tenga de la población estadística que va a medir es el error de muestreo. Desde el punto de vista de Cochran (1974) “El establecimiento de la precisión deseada puede ser hecho indicando la cantidad de error que estamos dispuestos a tolerar en los estimadores muestrales” (p.108). El error de muestreo es el error estándar, o raíz cuadrada de la varianza de la distribución de muestreo; surge a causa de observar solamente una parte de la población completa.

Pérez (2005) argumenta que los tamaños muestrales necesarios para cometer un error absoluto o relativo de muestreo dado en presencia de un coeficiente de confianza adicional son los siguientes:

Tabla V. **Fórmulas para el tamaño de muestra sin reposición**

Tipo de error → Parámetro ↓	Absoluto e	Relativo e_r	Absoluto y coeficiente de confianza adicional e_α	Relativo y confianza $e_{r\alpha}$
Media	$\frac{NS^2}{Ne^2 + S^2}$	$\frac{NC_{1,x}^2}{Ne_r^2 + C_{1,x}^2}$	$\frac{\lambda_\alpha^2 NS^2}{Ne^2 + \lambda_\alpha^2 S^2}$	$\frac{\lambda_\alpha^2 NC_{1,x}^2}{Ne_{r\alpha}^2 + \lambda_\alpha^2 C_{1,x}^2}$
Total	$\frac{N^2 S^2}{e^2 + NS^2}$	$\frac{NC_{1,x}^2}{Ne_r^2 + C_{1,x}^2}$	$\frac{\lambda_\alpha^2 N^2 S^2}{e^2 + \lambda_\alpha^2 NS^2}$	$\frac{\lambda_\alpha^2 NC_{1,x}^2}{Ne_{r\alpha}^2 + \lambda_\alpha^2 C_{1,x}^2}$
Proporción	$\frac{NPQ}{e^2(N-1) + PQ}$	$\frac{NQ}{P(N-1)e_r^2 + Q}$	$\frac{\lambda_\alpha^2 NPQ}{e^2(N-1) + \lambda_\alpha^2 PQ}$	$\frac{NQ\lambda_\alpha^2}{e_{r\alpha}^2(N-1)P + \lambda_\alpha^2 Q}$
Total de clase	$\frac{N^3 PQ}{e^2(N-1) + N^2 PQ}$	$\frac{NQ}{P(N-1)e_r^2 + Q}$	$\frac{\lambda_\alpha^2 N^3 PQ}{e^2(N-1) + \lambda_\alpha^2 N^2 PQ}$	$\frac{NQ\lambda_\alpha^2}{e_{r\alpha}^2(N-1)P + \lambda_\alpha^2 Q}$

Fuente: Pérez. (2005) *Técnicas de muestreo estadístico*, p.118.

En donde en todas las fórmulas S^2 es la cuasivarianza poblacional y $C_{1,x}^2 = (S/\bar{X})^2$. Por otra parte, λ_α es el valor crítico de la normal unitaria a nivel α . Para el caso de la proporción debe tenerse en cuenta que para poblaciones grandes o fracción de muestreo pequeña ($N \rightarrow \infty$), el valor máximo de n se obtiene para $P = Q = 1/2$. Agrega Pérez (2005) que para el caso del muestreo con reposición las fórmulas son las siguientes:

Tabla VI. **Fórmulas para el tamaño de muestra con reposición**

Tipo de error → Parámetro ↓	Absoluto e	Relativo e_r	Absoluto y coeficiente de confianza adicional e_α	Relativo y confianza $e_{r\alpha}$
Media	$\frac{\sigma^2}{e^2}$	$\frac{C_x^2}{e_r^2}$	$\frac{\lambda_\alpha^2 \sigma^2}{e^2}$	$\frac{\lambda_\alpha^2 C_x^2}{e_{r\alpha}^2}$
Total	$\frac{N^2 \sigma^2}{e^2}$	$\frac{C_x^2}{e_r^2}$	$\frac{\lambda_\alpha^2 N^2 \sigma^2}{e^2}$	$\frac{\lambda_\alpha^2 C_x^2}{e_{r\alpha}^2}$
Proporción	$\frac{PQ}{e^2}$	$\frac{Q}{Pe_r^2}$	$\frac{\lambda_\alpha^2 PQ}{e^2}$	$\frac{\lambda_\alpha^2 Q}{Pe_{r\alpha}^2}$
Total de clase	$\frac{N^2 PQ}{e^2}$	$\frac{Q}{Pe_r^2}$	$\frac{\lambda_\alpha^2 N^2 PQ}{e^2}$	$\frac{\lambda_\alpha^2 Q}{Pe_{r\alpha}^2}$

Fuente: Pérez. (2005). *Técnicas de muestreo estadístico*, p.120.

En donde en todas las fórmulas σ^2 es la varianza poblacional y $C_x^2 = (\sigma/\bar{X})^2$. Por otra parte, λ_α es el valor crítico de la normal unitaria a nivel α .

2.1.2.14. Muestras complejas

El análisis y evaluación de los diseños muestrales resulta de particular importancia en el caso de los grandes programas nacionales de encuestas, como los desarrollados por institutos nacionales de estadística, que con frecuencia se diseñan para que sean la fuente única de información precisa sobre una gran variedad de temas de interés. En este sentido, la obtención de información precisa y oportuna sobre algunos indicadores sociodemográficos demanda el diseño y puesta en marcha de muestras complejas.

Según Lohr (2000) los diseños de encuestas complejas difieren de los muestreos aleatorios simples en aspectos como la probabilidad desigual del muestreo, el muestreo polietápico y la estratificación. La mayoría de las encuestas grandes comprende varias de estas características, por lo que es necesario recurrir a análisis ponderados para obtener estimaciones de los parámetros de población no sesgadas (o casi sin sesgo). Para los estimadores, la estimación de varianzas depende de los detalles específicos del plan de muestreo y requiere de métodos aproximados, por lo general la linealización mediante el desarrollo en serie de Taylor o las técnicas de replicación.

Un ejemplo de una muestra compleja se observa en el diseño estadístico de la Encuesta Nacional de Condiciones de Vida ENCOVI 2014, a cargo del Instituto Nacional de Estadística INE, cuyo esquema adoptado es de tipo probabilístico, bietápico, estratificado y por conglomerados, donde la unidad última de selección está constituida por doce viviendas ocupadas seleccionadas con igual probabilidad al interior de cada unidad primaria de muestreo (UPM).

Cada UPM fue seleccionada de manera sistemática y proporcional al tamaño (PPT), con afijación proporcional en las áreas urbanas y rurales y al interior de ellas en 4 estratos urbanos y 4 rurales. El diseño muestral contempló además para sus estimaciones y proceso inferencial, 22 dominios de estudio. Los 22 dominios de estudio lo comprenden los 22 departamentos de la república, pudiendo integrarse si se desea por región, total país urbano, total país rural y total del país. La cobertura de la encuesta es entonces nacional y abarca la división político-administrativa del país

2.1.2.15. Estimación en dominios

De acuerdo con Lohr (2000) “Con frecuencia queremos separar las estimaciones de alguna subpoblación; las subpoblaciones se llaman dominios o subdominios” (p. 77). Al suponer que existen D dominios. Sea U_d el conjunto de unidades en la población estadística que están en el dominio d y sea s_d el conjunto de elementos en la muestra que están en el dominio d , para $d = 1, 2, \dots, D$. Sea N_d el número de unidades de la población en U_d y n_d el número de unidades en la muestra s_d . Si se desea estimar la media del ingreso de los jóvenes para el d -ésimo dominio de estudio, se tiene:

$$\bar{y}_{U_d} = \frac{\sum_{i \in s_d} y_i}{n_d} \quad (\text{Ec. 22})$$

2.1.3. Estimación en áreas pequeñas

En el contexto del muestreo de encuestas, Rao (2005) argumenta que “Un dominio (área) se considera grande (o principal) si la muestra especificada para el dominio es lo suficientemente grande como para producir estimaciones

directas con la precisión adecuada” (p. 1). En este contexto, agrega Rao (2005), que se utiliza el término “área pequeña” para denotar dominios de estudio en los que el tamaño de la muestra no es lo suficientemente grande para que una estimación directa basada en el diseño sea lo suficientemente precisa.

2.1.3.1. Demanda de estimaciones en áreas pequeñas

Como refiere Molina (2019) “Existe una demanda creciente de información para grupos específicos de la población o para áreas geográficas menores” (p. 9). Es así como en años recientes ha tenido auge el concepto de la desagregación de datos, referida como aquella información que ha sido recopilada de diferentes fuentes, o medida a través de múltiples variables o incluso diferentes unidades de observación, y que se compila en forma agregada y resumida. Agrega Molina (2019) que:

El propósito de esta agregación es presentar a la sociedad estimaciones de interés que gocen de buenas propiedades estadísticas para que sea posible que, a partir de ellas, se pueda extraer información e incluso formular políticas públicas en cada uno de los subgrupos de interés. (p. 9)

2.1.3.2. Límites para la desagregación de datos estadísticos

De acuerdo con Molina (2019) aunque no existe una definición formal, “un área se denomina “pequeña” cuando el error de muestreo del estimador directo considerado para el indicador de interés no es aceptable” (p.14). Sin embargo, no existe una regla de carácter universal para establecer el límite superior para este error de muestreo por encima del cual el área donde se estima se considere

“pequeña”. En la práctica, cada agencia de estadística establece su propio criterio al definir el límite para el error relativo de muestreo (coeficiente de variación), a partir del cual las estimaciones se consideran no fiables y por lo tanto no se publican. Algunas veces este tipo de estimaciones se publican con la indicación de que carecen de la calidad estadística deseada.

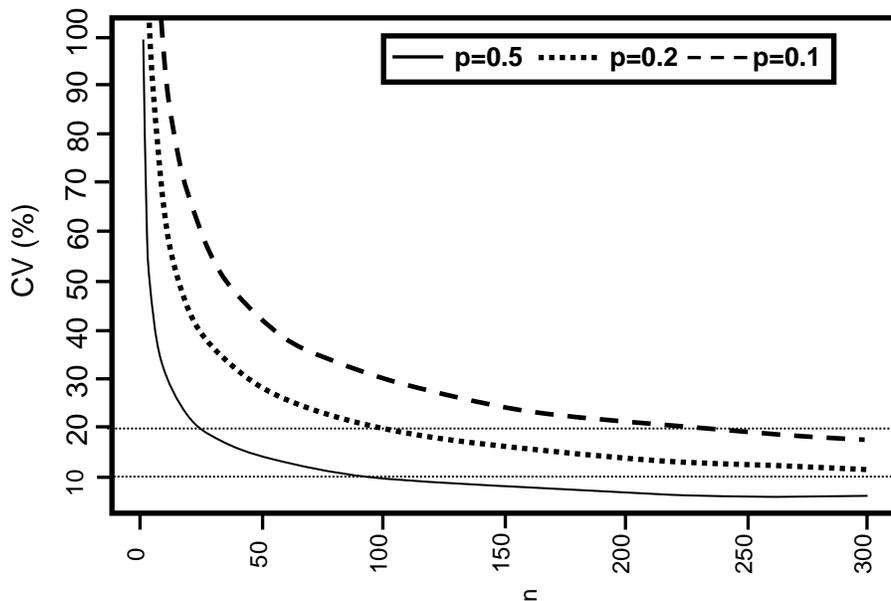
Existe otro aspecto que debe ser considerado al establecer el límite para la desagregación de datos estadísticos. Según Molina (2019):

Tampoco existe un tamaño muestral concreto, por debajo del cual el área se considere pequeña, pues el error de muestreo varía no solo dependiendo del tamaño muestral, sino también del indicador que se estima y del estimador concreto que se utiliza. (p. 14)

A pesar de no existir límites superiores universales para los errores de muestreo, ni inferiores para los tamaños muestrales, a partir de los cuales se considera que una estimación no tiene la suficiente calidad, algunos Institutos Nacionales de Estadística coinciden en establecer como no preciso cuando el error relativo de muestreo o CV supera un 20 %. Para el caso de Guatemala, se considera un 15 %.

Debido a que algunos indicadores de pobreza habituales son proporciones, el siguiente gráfico ilustra el tamaño muestral mínimo necesario para obtener un CV máximo concreto para la proporción muestral bajo muestreo aleatorio simple.

Figura 1. CV de la proporción muestral \hat{p} en función del tamaño muestral n , para cada valor de la verdadera proporción p



Fuente: Molina. (2019). *Serie estudios estadísticos 97: Desagregación de datos en encuestas de hogares. Metodologías de estimación en áreas pequeñas*, p. 15.

2.1.3.3. Metodologías para superar los límites de la desagregación

Molina (2019) plantea que, si se desea evitar incrementos al tamaño de la muestra debido a los costes, o si la demanda de datos a un nivel más desagregado se ha producido al concluir una encuesta, una forma efectiva en términos de costes es obtener estimaciones para las áreas de interés mediante métodos indirectos. Estos métodos utilizan, no sólo los datos del área correspondiente, sino también datos de otras áreas asumiendo algún tipo de similitud con dominio de interés. Esta similitud se suele presentar a través de un modelo matemático (que representa un conjunto de hipótesis). Ejemplos de

estimadores indirectos son los estimadores sintéticos y los estimadores compuestos.

Molina (2019) agrega que “Estimadores indirectos algo más sofisticados, que consideran la existencia de heterogeneidad entre las áreas, son los basados en modelos de regresión” (p. 17). Dos grandes grupos de modelos de regresión son utilizados para la desagregación de datos: los modelos a nivel de áreas y los modelos a nivel de individuo, aunque también es posible establecer modelos a niveles de agregación intermedio (por ejemplo, a nivel de grupos de edad y sexo dentro de un área).

2.1.4. Métodos directos para la desagregación de datos

Como se ha indicado, un estimador de un parámetro poblacional en un área determinada se califica como “directo” si se calcula utilizando los datos del diseño muestral para esta área, sin hacer uso de datos de ninguna otra. Molina (2019) refiere que este tipo de estimadores son utilizados por defecto en los institutos nacionales de estadística, debido a sus buenas propiedades respecto al diseño muestral (como insesgadez) en áreas con suficiente tamaño muestral.

2.1.4.1. Estimador de Horvitz-Thompson (HT)

Según refiere Molina (2019) el estimador insesgado con respecto al diseño muestral de la media del área d , \bar{Y}_d , es conocido como estimador de Horvitz-Thompson (HT). Este estimador requiere conocer el verdadero tamaño del área N_d y los pesos muestrales $w_{di} = \pi_{di}^{-1}$ para los elementos de la muestra en el área d . Asumiendo que éstos son conocidos, el estimador HT de \bar{Y}_d es:

$$\hat{Y}_d = N_d^{-1} \sum_{i \in S_d} w_{di} Y_{di} \quad (\text{Ec. 23})$$

Para el caso del total del área d , $Y_d = \sum_{i=1}^{N_d} Y_{di}$, el estimador de HT es simplemente $\sum_{i \in S_d} w_{di} Y_{di}$, por lo tanto, no requiere conocer el tamaño poblacional del área N_d . El estimador de la varianza, que no depende de las probabilidades de inclusión de segundo orden:

$$\widehat{var}_{\pi}(\hat{Y}_d) = N_d^{-2} \sum_{i \in S_d} w_{di} (w_{di} - 1) Y_{di}^2 \quad (\text{Ec. 24})$$

Si no se dispone de la información sobre el diseño muestral (solamente de los pesos muestrales), el estimador de la varianza de la ecuación 23 es utilizado. El programa estadístico R, incluye este procedimiento de cálculo a través de la función `direct()` del paquete R *sae*. R es un entorno de software libre para gráficos y computación estadística. Se compila y se ejecuta en una amplia variedad de plataformas UNIX, Windows y MacOS. En el caso de disponer de información muestral, existen paquetes en R más adecuados como *survey* o *sampling*.

Por último, Molina (2019) argumenta que el estimador de HT tiene la desventaja de ser muy ineficiente para áreas pequeñas cuando el tamaño de muestra es pequeño. Otra desventaja es que no se pueden calcular estimaciones para áreas o dominios no muestreados; es decir, con tamaño muestral n_d igual a cero.

2.1.4.2. Estimador de Hájek (HA)

Molina (2019) postula que: “Un estimador ligeramente sesgado para n_d pequeño, pero con varianza algo menor, y que no requiere el conocimiento del

tamaño del área N_d para estimar la media \bar{Y}_d , es el estimador Hájek” (p.23). Este estimador es igual a la media ponderada en las observaciones del área, usando como ponderaciones los pesos muestrales, es decir:

$$\hat{Y}_d^{HA} = \hat{N}_d^{-1} \sum_{i \in S_d} w_{di} Y_{di}, \text{ donde } \hat{N}_d = \sum_{i \in S_d} w_{di} \quad (\text{Ec. 25})$$

Para el total $Y_d = \sum_{i=1}^{N_d} Y_{di}$, el estimador de Hájek es $\hat{Y}_d^{HA} = N_d \hat{Y}_d^{HA}$ que si requiere conocer el tamaño poblacional N_d .

2.1.4.3. Estimadores GREG y de calibración

Molina (2019) alude que un estimador más sofisticado que el estimador de HT, en el sentido de que utiliza información auxiliar, es el estimador generalizado de regresión (en inglés, *generalized regression*, GREG). Este estimador requiere conocer el total $X_d = \sum_{i=1}^{N_d} x_{di}$, o la media $\bar{X}_d = N_d^{-1} \sum_{i=1}^{N_d} x_{di}$, para el área d de un vector x_{di} de valores de p variables auxiliares relacionadas con Y_{di} , para el individuo i dentro del área d . Si $\hat{X}_d = N_d^{-1} \sum_{i \in S_d} w_{di} x_{di}$ es el estimador de HT de \bar{X}_d , el estimador GREG de \bar{Y}_d , está dado por:

$$\hat{Y}_d^{GREG} = \hat{Y}_d + (\bar{X}_d - \hat{X}_d)' \hat{B}_d \quad (\text{Ec. 26})$$

Aquí, $\hat{B}_d = (\sum_{i \in S_d} w_{di} x_{di} x_{di}' / c_{di})^{-1} \sum_{i \in S_d} w_{di} x_{di} Y_{di} / c_{di}$ es el estimador de mínimos cuadrados ponderados (que utiliza los pesos del diseño muestral) del vector de coeficientes de la siguiente regresión lineal asumida para las unidades del área d .

$$Y_{di} = \mathbf{x}_{di}'\boldsymbol{\beta}_d + \epsilon_{di}, \quad i = 1, \dots, N_d \quad (\text{Ec. 27})$$

En donde los errores del modelo ϵ_{di} son independientes, con esperanza cero y varianza $\sigma^2 c_{di}$, siendo $c_{di} > 0$ constantes que representan la posible heteroscedasticidad, $i = 1, \dots, N_d$. Las constantes c_{di} se determinan estudiando los residuos del modelo lineal sin heteroscedasticidad, es decir, con $c_{di} = 1, i = 1, \dots, N_d$. Según refiere Molina (2019) observando un diagrama de dispersión de los residuos frente a cada una de las variables auxiliares se puede observar gráficamente si la varianza de los residuos aumenta con alguna de las variables. En este caso, se tomaría como constantes c_{di} , los valores de esta variable en las unidades del área o, más generalmente, una función $c_{di} = f(x_{di}) > 0$, de los valores de dicha variable auxiliar.

Molina (2019) postula que un estimador consistente (cuando n_d crece) para la varianza del estimador GREG se obtiene usando el método de linealización de Taylor. El estimador resultante se obtiene de reemplazar Y_{di} por $\tilde{e}_{di} = Y_{di} - \mathbf{x}_{di}'\hat{\mathbf{B}}_d$ en la varianza estimada del estimador de HT.

$$\begin{aligned} \widehat{var}_{\pi}(\hat{Y}_d^{GREG}) = N_d^{-2} \left\{ \sum_{i \in S_d} \frac{\tilde{e}_{di}^2}{\pi_{di}^2} (1 - \pi_{di}) + \right. \\ \left. 2 \sum_{i \in S_d} \sum_{\substack{j \in S_d \\ j > i}} \frac{\tilde{e}_{di}\tilde{e}_{dj}}{\pi_{di}\pi_{dj}} \left(\frac{\pi_{d,ij} - \pi_{di}\pi_{dj}}{\pi_{d,ij}} \right) \right\} \end{aligned} \quad (\text{Ec. 28})$$

Por otra parte, Molina (2019) destaca que los estimadores de calibración son muy utilizados por los institutos nacionales de estadística para estimar medias o totales en dominios con suficiente tamaño muestral. Deville y Särndal en 1992 (como citó Molina, 2019) propusieron el método de calibración para estimar el total de una variable de interés usando información auxiliar de p variables relacionadas con ella. Molina (2019) expresa:

El método de calibración consiste en encontrar unos nuevos pesos h_{di} , lo más cercanos posible a los pesos muestrales originales w_{di} en virtud de una distancia $G_{di}(h_{di}, w_{di})$, tales que el total X_d de las variables auxiliares se estima con estos pesos de forma exacta; es decir, sin error. Si la variable de interés está linealmente relacionada con estas variables auxiliares y los totales de éstas se estiman de forma exacta, se espera que los totales de la variable de interés también se estimen con poco error. (p. 28)

Formalmente, para estimar la media \bar{Y}_d , se buscan unos nuevos pesos para las unidades de la muestra, $h_{di}, i \in s_d$ que son la solución al problema. Molina (2019) deduce que utilizando la pseudo distancia ji-cuadrado dada por $G_{di}(h_{di}, w_{di}) = c_{di}(h_{di} - w_{di})^2/w_{di}$, que es probablemente la más popular, y resolviendo el problema mediante el método de multiplicadores de Lagrange, los pesos resultantes son:

$$h_{di} = w_{di} \left\{ 1 + \mathbf{x}_{di}' \left(\sum_{i \in s_d} w_{di} \mathbf{x}_{di} \mathbf{x}_{di}' / c_{di} \right)^{-1} \left(\mathbf{X}_d - \sum_{i \in s_d} w_{di} \mathbf{x}_{di} / c_{di} \right) \right\}, i \in s_d \quad (\text{Ec. 29})$$

Los pesos calibrados h_{di} son el resultado de realizar un ajuste a los pesos originales, $h_{di} = w_{di} g_{di}$, donde el factor de ajuste g_{di} viene dado por el término entre llaves en la ecuación (Ec. 29). El estimador de calibración \bar{Y}_d se obtiene entonces simplemente igual que el estimador de HT, pero usando los pesos calibrados en lugar de los originales, de la forma:

$$\hat{Y}_d^{CAL} = N_d^{-1} \sum_{i \in s_d} h_{di} Y_{di} \quad (\text{Ec. 30})$$

Como afirma Molina (2019) estos estimadores tienen la ventaja de ser insesgados (y consistentes cuando n_d crece) respecto al diseño muestral, independientemente de si el modelo se verifica o no (son no paramétricos).

2.1.5. Metodologías de estimación en áreas pequeñas

Como plantea Mancho (2002) la estimación en áreas pequeñas es un tema que ha despertado gran interés en los últimos años debido, fundamentalmente, a la importancia que tanto para el sector público como para el privado tiene la obtención de información precisa y confiable acerca de los dominios específicos de interés.

Algunos métodos indirectos básicos incluyen el estimador post-estratificado sintético, los estimadores sintéticos de regresión a nivel de área y a nivel de individuo, y estimadores compuestos. Métodos indirectos más complejos incluyen: EBLUP bajo el modelo *Fay-Harriot*, EBLUP bajo el modelo con errores anidados, método ELL, el mejor predictor empírico bajo el modelo con errores anidados, el método jerárquico Bayes bajo el modelo con errores anidados y los métodos basados en modelos lineales generalizados mixtos.

2.1.5.1. Estimador post-estratificado sintético

En la opinión de Molina (2019) este estimador es escasamente utilizado en aplicaciones reales de estimación en áreas pequeñas debido a que se basa en hipótesis poco realistas.

Según refiere Molina (2019) al disponer de una variable cualitativa relacionada con la variable Y_{di} . Esta variable cualitativa tiene J categorías posibles, las cuales dividen una población U en J grupos, U^1, \dots, U^J de tamaños

N^1, \dots, N^J , llamados post-estratos, que se entrecruzan con las áreas. Por lo tanto, el área U_d de la población queda igualmente dividida en J trozos de post-estratos, U_d^1, \dots, U_d^J de tamaños poblacionales N_d^1, \dots, N_d^J y con medias $\bar{Y}_d^1, \dots, \bar{Y}_d^J$, donde $\bar{Y}_d^j = \sum_{i \in U_d^j} Y_{di} / N_d^j, j = 1, \dots, J$. Dado que las medias son indicadores aditivos, pueden descomponerse en sumas por los J estratos.

$$\bar{Y}_d = \frac{1}{N_d} \sum_{i=1}^{N_d} Y_{di} = \frac{1}{N_d} \sum_{j=1}^J N_d^j \bar{Y}_d^j \quad (\text{Ec. 31})$$

En este método se asume que los individuos dentro de cada estrato se comportan de forma homogénea, independientemente del área al que pertenecen; concretamente, se asume:

$$\bar{Y}_d^j = \bar{Y}^j, \quad j = 1, \dots, J \quad (\text{Ec. 32})$$

En donde $\bar{Y}^j = \sum_{i \in U^j} Y_{di} / N^j$ es la media del estrato j . Entonces, al aprovechar esta homogeneidad dentro de los estratos para estimar la media de cada área estimando las medidas de los estratos (que deben tener tamaños de muestra grandes). Es decir, sustituyendo la ecuación (Ec. 32) en (Ec. 31).

$$\bar{Y}_d = \frac{1}{N_d} \sum_{j=1}^J N_d^j \bar{Y}^j \quad (\text{Ec. 33})$$

El estimador post-estratificado sintético (PS-SYN) de \bar{Y}_d se obtiene estimando las medias de cada estrato en (Ec. 32) mediante los estimadores de Hájek, es decir:

$$\hat{Y}_d^{PS-SYN} = \frac{1}{N_d} \sum_{j=1}^J N_d^j \hat{Y}_{j,HA} \quad (\text{Ec. 34})$$

Rao y Molina (2015) argumentan que, dado que el sesgo de un estimador sintético no es despreciable, en lugar de su varianza, la cual podría dar una imagen errónea de la calidad del estimador, es de interés obtener su error cuadrático medio (ECM) que refleja ambas cosas. Un estimador del ECM bajo el diseño está dado por:

$$\widehat{MSE}_\pi(\hat{Y}_d^{SYN}) = (\hat{Y}_d^{SYN} - \hat{Y}_d^{DIR})^2 - \widehat{var}_\pi(\hat{Y}_d^{DIR}) \quad (\text{Ec. 35})$$

Molina (2019) menciona que, entre las ventajas de este método está que, si los estratos tienen suficientes observaciones en la muestra, la varianza se disminuye considerablemente en comparación con el estimador directo. Una desventaja es la hipótesis de homogeneidad considerada para las variables Y_{di} al ser considerada poco realista en la práctica. Si no se verifica este supuesto, los estimadores resultantes pueden tener un sesgo considerable y, por lo tanto, no reflejar la realidad.

2.1.5.2. Estimador sintético de regresión a nivel de área

Molina (2019) postula que “Los estimadores sintéticos de regresión asumen un modelo de regresión lineal que se puede plantear, bien a nivel de área o bien a nivel de individuo, dependiendo de la información auxiliar disponible” (p.37). Para el caso del estimador sintético de regresión a nivel de

área, si x_d es el vector disponible de p variables auxiliares a nivel de área, se asume que el indicador que se desea estimar θ_d varía respecto de estos datos agregados x_d de forma constante para todas las áreas, según un modelo de regresión lineal. Debido a que los verdaderos parámetros son desconocidos se consideran los estimadores directos $\hat{\theta}_d, d = 1, \dots, D$. Así, el modelo a nivel de área asume:

$$\hat{\theta}_d = x_d' \alpha + \varepsilon_d, \quad d = 1, \dots, D \quad (\text{Ec. 36})$$

En donde los términos de error ε_d se asumen independientes, con esperanza cero y varianza ψ_d conocida, $d = 1, \dots, D$. Debe observarse que, dado que x_d es el valor poblacional y por lo tanto tiene varianza cero, ψ_d es la varianza del estimador directo $\hat{\theta}_d$, es decir, $\psi_d = \text{var}(\hat{\theta}_d)$. En la práctica estas varianzas se estiman con los microdatos de la encuesta. El estimador sintético de regresión (REG1-SYN) para el indicador del área d viene entonces dado por la predicción del estimador a través del modelo, es decir, $\hat{\alpha} = (\sum_{d=1}^D \psi_d^{-1} x_d x_d')^{-1} \sum_{d=1}^D \psi_d^{-1} x_d \hat{\theta}_d$ es el estimador de α obtenido por mínimos cuadrados ponderados, el estimador viene dado por:

$$\hat{\theta}_d^{\text{REG1-SYN}} = x_d' \hat{\alpha} \quad (\text{Ec. 37})$$

En el modelo (Ec. 35), ε_d es el error debido a que se utiliza un estimador directo $\hat{\theta}_d$ en lugar del verdadero valor del estimador θ_d , ya que este es desconocido, y se está asumiendo que el verdadero valor θ_d es exactamente igual al término de la regresión, $\theta_d = x_d' \alpha$, sin dejar ningún grado de heterogeneidad. Por esta razón se denominan modelos sintéticos. De hecho, el

sesgo bajo el diseño de $\hat{\theta}_d^{REG1-SYN}$ para α conocido, viene dado por $\mathbf{x}'_d \alpha - \theta_d$, que no depende del tamaño muestral del área n_d ; por lo tanto, este sesgo no disminuye al aumentar el tamaño muestral del área.

De acuerdo con Molina (2019) “Una ventaja de los estimadores basados en modelos es que permiten estimar en áreas no muestreadas; es decir, con tamaño muestral igual a cero, si se dispone de la correspondiente información auxiliar” (p. 37).

2.1.5.3. Estimador sintético de regresión a nivel de individuo

Para el caso del estimador sintético de regresión a nivel de individuo, Molina (2019) refiere que se debe disponer de datos a nivel de individuo (o microdatos) de las p variables auxiliares en la encuesta, \mathbf{x}_{di} , $i \in s_d$, $d = 1, \dots, D$. En esta situación, se puede obtener un estimador sintético de regresión para el indicador de interés asumiendo un modelo de regresión lineal a nivel de individuo para Y_{di} . Entonces $\mathbf{y}_d = (Y_{d1}, \dots, Y_{dN_d})'$ será el vector de valores de la variable de interés para los individuos del área d . El indicador para estimar en el área d es una función de este vector, es decir, $\theta_d = \theta_d(\mathbf{y}_d)$. El modelo de regresión sintético básico considera que las variables Y_{di} para todos los elementos de la población siguen el modelo de regresión lineal:

$$Y_{di} = \mathbf{x}'_{di} \boldsymbol{\beta} + \varepsilon_{di}, \quad i = 1, \dots, N_d, \quad d = 1, \dots, D. \quad (\text{Ec. 38})$$

En donde los errores ε_{di} son independientes, con esperanza cero y varianza $\sigma^2 k_{di}^2$, y k_{di} son constantes conocidas que representan la posible

heteroscedasticidad en el modelo ($k_{di} = 1$ para todo i y d si no existe heteroscedasticidad). Al estimar $\boldsymbol{\beta}$ mediante el estimador $\hat{\boldsymbol{\beta}} = (\sum_{d=1}^D \sum_{i \in s_d} a_{di} \mathbf{x}_{di} \mathbf{x}_{di}')^{-1} \sum_{d=1}^D \sum_{i \in s_d} a_{di} \mathbf{x}_{di} Y_{di}$ de mínimos cuadrados ponderados, siendo $a_{di} = k_{di}^{-2}$, se obtienen predicciones, a través del modelo, para cada individuo del área, $\hat{Y}_{di} = \mathbf{x}_{di}' \hat{\boldsymbol{\beta}}$, $i = 1, \dots, N_d$. El vector de predicciones para el área d es entonces $\hat{\mathbf{y}}_d = (\hat{Y}_{d1}, \dots, \hat{Y}_{dN_d})'$. Utilizando este vector en lugar de \mathbf{y}_d para calcular el indicador de interés, se obtiene el estimador sintético de regresión de θ_d , es decir:

$$\hat{\theta}_d^{REG2-SYN} = \theta_d(\hat{\mathbf{y}}_d) \quad (\text{Ec. 39})$$

Molina (2019) considera que para estimar la media del área d , $\theta_d = \bar{Y}_d$, si $\bar{\mathbf{X}}_d$ es el vector de medias poblacionales de las p variables auxiliares consideradas en esa área, el estimador sintético basado en el modelo sería:

$$\hat{\bar{Y}}_d^{REG2-SYN} = \bar{\mathbf{X}}_d' \hat{\boldsymbol{\beta}} \quad (\text{Ec. 40})$$

Agrega Molina (2019) que, para un área no muestreada, este estimador se obtiene de la misma forma. Para $\boldsymbol{\beta}$ conocido, el sesgo bajo el diseño del estimador de la media es $\bar{\mathbf{X}}_d' \boldsymbol{\beta} - \bar{Y}_d$, que no depende del tamaño de la muestra del área n_d ; por lo tanto, este sesgo no disminuye al aumentar el número de observaciones en la muestra.

Entre las ventajas mencionadas por Molina (2019) al emplear este método de estimación se encuentran: la reducción de la varianza de los estimadores directos y de los estimadores a partir de un modelo a nivel de área. Entre las desventajas describe que el método no representa los casos en los que no se dispone de todas las variables auxiliares que expliquen la heterogeneidad entre las áreas. Otro aspecto para considerar es que debe evaluarse profundamente el modelo, pues el sesgo de estos estimadores depende de la bondad de ajuste del modelo. Por último, estos modelos requieren la verificación de que la suma de los totales estimados en las áreas de una región mayor coincida con el estimador directo para la región (propiedad de *benchmarking*).

2.1.5.4. Estimadores compuestos

Molina (2019) indica que “los estimadores compuestos nacen con el objeto de disminuir la varianza del estimador directo a cambio de una porción del sesgo de un estimador sintético” (p. 40). El uso de estimadores compuestos pretende aumentar la eficiencia del estimador directo y reducir el sesgo del estimador sintético. Sea \hat{Y}_d^{DIR} un estimador directo genérico de \bar{Y}_d y \hat{Y}_d^{SYN} un estimador sintético, un estimador compuesto de \bar{Y}_d tiene la forma:

$$\hat{Y}_d^C = \phi_d \hat{Y}_d^{DIR} + (1 - \phi_d) \hat{Y}_d^{SYN}, \quad 0 \leq \phi_d \leq 1 \quad (\text{Ec. 41})$$

Molina (2019) expresa que: “El peso ϕ_d otorgado al estimador directo se puede establecer de forma semi-óptima minimizando una aproximación del error cuadrático medio (ECM) bajo el diseño muestral o fijándolo de forma arbitraria” (p. 40). Drew, Singh y Choudhry en el año 1982 (como cito Molina, 2019) propusieron un peso ϕ_d que depende del tamaño muestral del área, dando lugar un estimador dependiente del tamaño de la muestra (en inglés, *sample-size*

dependent, SSD) que, tomando un valor de $\delta > 0$ preestablecido (por defecto puede ser 1), el peso propuesto tiene la forma:

$$\phi_d = \begin{cases} 1, & \text{si } \hat{N}_d \geq \delta N_d; \\ \hat{N}_d(\delta N_d) & \text{si } \hat{N}_d < \delta N_d, \end{cases} \quad (\text{Ec. 42})$$

En donde $\hat{N}_d = \sum_{i \in S_d} w_{di}$. Un ejemplo para comprender la idea intuitiva de este estimador puede darse para el caso de un muestreo aleatorio simple (MAS) de la población. En este caso el tamaño del área n_d es aleatorio y se obtiene de la siguiente fórmula:

$$\hat{N}_d = \sum_{i \in S_d} w_{di} = \sum_{i \in S_d} \frac{N}{n} = N \frac{n_d}{n} \quad (\text{Ec. 43})$$

En donde \hat{N}_d es insesgado, su esperanza bajo el diseño es igual a $NE_\pi(n_d) = N_d$, con lo cual $E_\pi(n_d) = nN_d/N$ y el peso resulta ser:

$$\phi_d = \begin{cases} 1, & \text{si } n_d \geq \delta E_\pi(n_d); \\ n_d / \{\delta E_\pi(n_d)\} & \text{si } n_d < \delta E_\pi(n_d) \end{cases} \quad (\text{Ec. 44})$$

Molina (2019) deduce que, si se fija $\delta = 1$, entonces el estimador dependiente del tamaño muestral (SSD) le da una ponderación de 1 al estimador directo cuando el tamaño de la muestra del área es mayor o igual al tamaño muestral esperado, y le da un peso menor que 1 en caso contrario. Sin embargo, un área determinada puede tener tamaño muestral n_d pequeño, pero puede darse el caso que éste supere el tamaño de muestra esperado, con lo cual se le

daría un peso igual a 1 al estimador directo y por lo tanto no habría mejora en la eficiencia respecto al estimador directo considerado.

Molina (2019) postula que es posible obtener estimadores compuestos aproximadamente óptimos respecto al diseño de muestreo, tomando el peso ϕ_d que minimiza (de forma aproximada) el error cuadrático medio bajo el diseño del estimador compuesto, $MSE_\pi(\hat{Y}_d^C)$. Al considerar que la covarianza entre el estimador directo y el sintético es despreciable, y minimizando se obtiene:

$$MSE_\pi(\hat{Y}_d^C) \approx \phi_d^2 var_\pi(\hat{Y}_d^{DIR}) + (1 - \phi_d)^2 MSE_\pi(\hat{Y}_d^{SYN}) \quad (\text{Ec. 45})$$

Se obtiene el peso óptimo a través de la siguiente expresión:

$$\phi_d^* = \frac{MSE_\pi(\hat{Y}_d^{SYN})}{var_\pi(\hat{Y}_d^{DIR}) + MSE_\pi(\hat{Y}_d^{SYN})} \quad (\text{Ec. 46})$$

El estimador de $\widehat{MSE}_d(\hat{Y}_d^{SYN})$ es:

$$\widehat{MSE}_d(\hat{Y}_d^{SYN}) = (\hat{Y}_d^{SYN} - \hat{Y}_d^{DIR})^2 - var_\pi(\hat{Y}_d^{DIR}) \quad (\text{Ec. 47})$$

Reemplazando el estimador del error cuadrático medio de (Ec. 46) en el estimador de peso óptimo de (Ec. 45), se obtiene un estimador de este peso óptimo dado por:

$$\phi_d^* = \frac{MSE_{\pi}(\hat{Y}_d^{SYN})}{(\hat{Y}_d^{SYN} - \hat{Y}_d^{DIR})^2} = 1 - \frac{var_{\pi}(\hat{Y}_d^{DIR})}{(\hat{Y}_d^{SYN} - \hat{Y}_d^{DIR})^2} \quad (\text{Ec. 48})$$

Según refiere Molina (2019) este peso depende del estimador directo \hat{Y}_d^{DIR} , que es muy volátil. Esto significa que el peso óptimo estimado $\hat{\phi}_d^*$ también es volátil. Se obtiene un peso estimado más estable promediando para todas las áreas de la forma:

$$\phi^* = \frac{\sum_{\ell=1}^D \widehat{MSE}_{\pi}(\hat{Y}_{\ell}^{SYN})}{\sum_{\ell=1}^D (\hat{Y}_{\ell}^{SYN} - \hat{Y}_{\ell}^{DIR})^2} = 1 - \frac{var_{\pi}(\hat{Y}_{\ell}^{DIR})}{(\hat{Y}_{\ell}^{SYN} - \hat{Y}_{\ell}^{DIR})^2} \quad (\text{Ec. 49})$$

El peso ϕ^* , es muy estable, pero no depende del área d ; es decir, es constante para todas las áreas, sin depender siquiera de su tamaño muestral. Por este tipo de inconvenientes, los estimadores compuestos óptimos son menos utilizados en la práctica que los basados en modelos.

Molina (2019) manifiesta que una ventaja de este estimador es que está diseñado para reducir el sesgo del estimador sintético y la varianza del estimador directo. Entre las desventajas de este método se encuentra que, para un área con tamaño de muestra pequeño, se espera que este tamaño no sea inferior al tamaño muestral esperado. Otra desventaja es que el peso que se da al estimador sintético no depende de lo bien explicada que esté la variable de interés por las variables auxiliares, además, no se puede calcular para áreas o dominios no muestreados y, por último, requieren de un ajuste para verificar que la suma de los totales estimados en las áreas de una región mayor coincida con el estimador directo para dicha región.

2.1.5.5. EBLUP basado en el modelo *Fay-Herriot*

Como afirma Molina (2019) el modelo *Fay-Herriot* (FH) es un modelo muy popular a nivel de área que fue introducido por Fay y Herriot en el año 1979 para estimar niveles de ingreso e indicadores de pobreza en áreas pequeñas de Estados Unidos de Norteamérica, como parte de las actividades de la Oficina del Censo de EE. UU. (*U.S. Census Bureau*).

Según refiere Molina (2019) el modelo enlaza los indicadores de interés para todas las áreas $\theta_d, d = 1, \dots, D$, asumiendo que éstos varían respecto de un vector con valores de p variables auxiliares x_d de forma constante para todas las áreas, siguiendo un modelo de regresión lineal.

$$\theta_d = x_d' \boldsymbol{\beta} + u_d, \quad d = 1, \dots, D. \quad (\text{Ec. 50})$$

En donde $\boldsymbol{\beta}$ es el vector de coeficientes común a todas las áreas y u_d es el término de error de la regresión, también conocido como efecto aleatorio del área d , el cual es diferente para cada área. Estos efectos aleatorios u_d representan la heterogeneidad de los indicadores θ_d a través de las áreas de interés, no debida a (o no explicada por) las variables auxiliares consideradas en el modelo.

Dado que los verdaderos valores de los parámetros de interés θ_d no son observables, el modelo (Ec. 50) no puede ajustarse. En este caso, utilizando el estimador directo $\hat{\theta}_d^{DIR}$ de θ_d , debe tenerse en cuenta que este estimador tiene un error debido al muestreo. El modelo FH considera que este estimador directo

$\hat{\theta}_d^{DIR}$ es insesgado bajo el diseño. Por lo tanto, el error debido al muestreo de este estimador se representa mediante:

$$\hat{\theta}_d^{DIR} = \theta_d + e_d, \quad d = 1, \dots, D. \quad (\text{Ec. 51})$$

En donde e_d es el error de muestreo correspondiente al área d . Por otra parte, este método asume que los errores de muestreo e_d son independientes entre sí y también son independientes de los efectos aleatorios de las áreas, u_d , además, tienen media cero y varianzas conocidas ψ_d ; es decir, $e_d \sim^{ind}(0, \psi_d)$. Según refiere Molina (2019) en la práctica, dichas varianzas, $\psi_d = \text{var}_\pi(\hat{\theta}_d^{DIR} | \theta_d)$, $d = 1, \dots, D$, se estiman con los microdatos de la encuesta. Combinando los modelos (Ec. 50) y (Ec. 51) se tiene:

$$\hat{\theta}_d^{DIR} = \mathbf{x}'_d \boldsymbol{\beta} + u_d + e_d, \quad d = 1, \dots, D. \quad (\text{Ec. 52})$$

Calculando por medio del método de multiplicadores de Lagrange el estimador lineal en los datos $\hat{\theta}_d^{DIR}$, $d = 1, \dots, D$, que es insesgado bajo el modelo (Ec. 52) y que minimiza el error cuadrático medio (ECM) bajo el modelo, se obtiene el mejor predictor lineal insesgado (en inglés, *best linear unbiased predictor*, BLUP) de $\theta_d = \mathbf{x}'_d \boldsymbol{\beta} + u_d$. El estimador resultante se obtiene simplemente ajustando el modelo mixto (Ec. 52); es decir, el BLUP bajo el modelo FH de θ_d viene dado por:

$$\tilde{\theta}_d^{FH} = \mathbf{x}_d' \boldsymbol{\beta} + \tilde{u}_d \quad (\text{Ec. 53})$$

En donde $\tilde{u}_d = \gamma_d(\hat{\theta}_d^{DIR} - \mathbf{x}_d' \tilde{\boldsymbol{\beta}})$ es el BLUP de u_d , siendo $\gamma_d = \sigma_u^2 / (\sigma_u^2 + \psi_d)$ y $\tilde{\boldsymbol{\beta}}$ es el estimador de mínimos cuadrados ponderados de $\boldsymbol{\beta}$ bajo el modelo (Ec. 53), dado por:

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{d=1}^D \gamma_d \mathbf{x}_d \mathbf{x}_d' \right)^{-1} \sum_{d=1}^D \gamma_d \mathbf{x}_d \hat{\theta}_d^{DIR} \quad (\text{Ec. 54})$$

Al sustituir $\tilde{u}_d = \gamma_d(\hat{\theta}_d^{DIR} - \mathbf{x}_d' \tilde{\boldsymbol{\beta}})$ en el BLUP bajo el modelo FH dado en (Ec. 53), se puede expresar el BLUP como una combinación lineal convexa del estimador directo y del estimador sintético de regresión, es decir:

$$\tilde{\theta}_d^{FH} = \gamma_d \hat{\theta}_d^{DIR} + (1 - \gamma_d) \mathbf{x}_d' \tilde{\boldsymbol{\beta}} \quad (\text{Ec. 55})$$

Con un peso para el estimador directo dado por $\gamma_d = \sigma_u^2 / (\sigma_u^2 + \psi_d) \in (0, 1)$. Este peso depende del tamaño de la muestra en el área a través de la varianza ψ_d del estimador directo y de la bondad de ajuste del modelo sintético medido por σ_u^2 (es decir, la heterogeneidad no explicada entre las áreas). Por lo tanto, para un área d en la que el estimador directo $\hat{\theta}_d^{DIR}$ sea eficiente debido a disponer de suficientes observaciones en la muestra; es decir, con una varianza muestral ψ_d pequeña comparada con la heterogeneidad no explicada σ_u^2 , $\gamma_d = \sigma_u^2 / (\sigma_u^2 + \psi_d)$ es cercano a uno y por lo que $\tilde{\theta}_d^{FH}$ le da más peso al estimador directo.

Por otro lado, en las áreas d en las que el estimador directo carece de calidad debido al pequeño tamaño de la muestra, donde su varianza muestral ψ_d sea mayor que la heterogeneidad no explicada σ_u^2 , entonces Υ_d se acerca a cero y por lo que se le da más peso al estimador sintético de regresión $x_d' \tilde{\beta}$, el cual utiliza los datos de todas las áreas para estimar el parámetro común β . Es decir, este estimador toma prestada información de las otras áreas a través del estimador sintético de regresión $x_d' \tilde{\beta}$ en la medida en que es necesario, dependiendo de la eficiencia del estimador directo.

Molina (2019) indica que, el hecho de que BLUP $\tilde{\theta}_d^{FH}$ se acerque al estimador directo cuando el tamaño muestral del área es grande (ψ_d pequeña) es una propiedad muy deseable, pues no es necesario saber cuándo un área es lo suficientemente “pequeña” para utilizar este estimador en lugar del estimador directo, ya que tiende al estimador directo cuando el tamaño de la muestra en el área d crece, y al mismo tiempo mejora el estimador directo en las áreas con tamaño muestral pequeño. En principio, este estimador se puede utilizar para todas las áreas mientras exista alguna “pequeña” (si no existe alguna no es necesario su uso).

Según postula Molina (2019) el BLUP de θ_d depende del verdadero valor de la varianza σ_d^2 de los efectos aleatorios u_d . En la práctica, esta varianza es desconocida y debe estimarse. Algunos métodos para su estimación son el de máxima verosimilitud (en inglés, *maximum likelihood*, ML) y máxima verosimilitud restringida o residual (en inglés, *restricted residual ML*, REML). Fay y Herriot en el año 1979 (como citó Molina, 2019) propusieron un método de ajuste basado en los momentos (método FH). Sea $\hat{\sigma}_u^2$ un estimador consistente de σ_d^2 , al reemplazar σ_d^2 por $\hat{\sigma}_u^2$ en (Ec. 53), se obtiene el BLUP empírico (en inglés, *empirical BLUP*, EBLUP) de θ_d .

$$\hat{\theta}_d^{FH} = \gamma_d \hat{\theta}_d^{DIR} + (1 - \gamma_d) \mathbf{x}_d' \hat{\boldsymbol{\beta}} \quad (\text{Ec. 56})$$

Donde $\hat{\gamma}_d = \hat{\sigma}_u^2 / (\hat{\sigma}_u^2 + \psi_d)$ y $\hat{\boldsymbol{\beta}} = (\sum_{d=1}^D \hat{\gamma}_d \mathbf{x}_d \mathbf{x}_d')^{-1} \sum_{d=1}^D \hat{\gamma}_d \mathbf{x}_d \hat{\theta}_d^{DIR}$. El modelo resultante es el llamado EBLUP basado en el modelo *Fay-Herriot*.

Si los parámetros del modelo $\boldsymbol{\beta}$ y σ_u^2 son conocidos, el error cuadrático medio del BLUP, $\tilde{\theta}_d^{FH}$, basado en el modelo (Ec. 52) viene dado por:

$$MSE(\tilde{\theta}_d^{FH}) = \gamma_d \psi_d \leq \psi_d = \text{var}_{\pi}(\hat{\theta}_d^{DIR} | \theta_d) \quad (\text{Ec. 57})$$

Por lo anterior, dado el verdadero valor del indicador θ_d , si σ_u^2 y $\boldsymbol{\beta}$ son conocidos, el BLUP bajo el modelo FH, $\tilde{\theta}_d^{FH}$, no puede ser menos eficiente que el estimador directo. Molina (2019) señala que, “En la práctica, σ_u^2 y $\boldsymbol{\beta}$ son estimados y el error debido a la estimación de estos dos parámetros se añade al ECM del estimador FH” (p. 47). Sin embargo, estos dos términos que se añaden al error cuadrático medio tienden a cero cuando el número de áreas D tiende al infinito. Por lo tanto, para un número de áreas D suficiente, es probable que el estimador FH incluso mejore al estimador directo en cuanto a ECM.

Si existe desconocimiento de los parámetros del modelo $\boldsymbol{\beta}$ y σ_u^2 , Molina (2019) señala que un estimador insesgado de segundo orden del ECM del estimador FH, llamado estimador Prasad-Rao, viene dado por la siguiente expresión:

$$MSE(\tilde{\theta}_d^{FH}) = g_{d1}(\hat{\sigma}_u^2) + g_{d2}(\hat{\sigma}_u^2) + 2g_{d3}(\hat{\sigma}_u^2) \quad (\text{Ec. 58})$$

En donde:

$$g_{d1}(\hat{\sigma}_u^2) = \gamma_d \psi_d \quad (\text{Ec. 59})$$

$$g_{d2}(\hat{\sigma}_u^2) = (1 - \gamma_d)^2 \mathbf{x}_d' \left(\sum_{d=1}^D (\sigma_u^2 + \psi_d)^{-1} \mathbf{x}_d \mathbf{x}_d' \right)^{-1} \mathbf{x}_d \quad (\text{Ec. 60})$$

$$2g_{d3}(\hat{\sigma}_u^2) = (1 - \gamma_d)(\sigma_u^2 + \psi_d)^{-1} \overline{\text{var}}(\hat{\sigma}_u^2) \quad (\text{Ec. 61})$$

A juicio de Molina (2019) los modelos a nivel de individuo, basados en el tamaño muestral total n , pueden ganar mucha más eficiencia que los modelos a nivel de área, mientras existan variables auxiliares a este nivel que sean lo suficientemente informativas sobre la variable respuesta. Sin embargo, una ventaja del estimador EBLUP basado en el modelo FH es que utiliza los factores de expansión (pesos) correspondientes al diseño muestral a través del estimador directo y es consistente bajo el diseño cuando el tamaño de la muestra del área n_d aumenta, mientras el peso del estimador directo es $\gamma_d > 0$.

Molina (2019) agrega que, uno de los requisitos de este modelo es disponer de datos agregados para las p variables auxiliares que se consideren para las áreas, $\mathbf{x}_d, d = 1, \dots, D$. Entre las ventajas de este método destaca el hecho de que suele mejorar la eficiencia del estimador directo, por otro lado, el

modelo de regresión considerado incorpora heterogeneidad no explicada entre las áreas, además, es un estimador compuesto que toma información del resto de áreas a medida que es necesario. Una desventaja del método es que se basa en un modelo; por lo tanto, es necesario analizar dicho modelo a través de los residuos (para parámetros no lineales, podría existir algún problema de linealidad).

Otra desventaja, es que el número de observaciones usadas para ajustar el modelo es igual al número de áreas muestreadas, el cual es mucho menor que el utilizado para justar modelos a nivel de individuo. Por último, al igual que en otros modelos, requiere un reajuste para verificar que la suma de los totales estimados en las áreas de una región mayor coincida con el estimador directo para dicha región.

2.1.5.6. EBLUP basado en el modelo con errores anidados

El modelo con errores anidados fue propuesto por Battese, Harter y Fuller en el año 1977 (como citó Molina, 2019) para estimar la producción de maíz y soya en condados de EE. UU. En esencia, este modelo relaciona de forma lineal los valores de una variable de interés Y_{di} para el individuo i dentro del área d , con los valores de p variables auxiliares para este mismo individuo, de la forma siguiente:

$$Y_{di} = \mathbf{x}'_{di}\boldsymbol{\beta} + u_d + e_{di}, \quad i = 1, \dots, N_d, d = 1, \dots, D \quad (\text{Ec. 62})$$

En donde $\boldsymbol{\beta}$ es el vector de coeficientes de las variables auxiliares, común para todas las áreas, u_d es el efecto aleatorio del área y e_{di} es el error a nivel de

individuo. Los efectos aleatorios representan la heterogeneidad no explicada de los valores Y_{di} a través de las áreas. En este caso, los efectos aleatorios se consideran independientes de los errores, con $u_d \sim iid(0, \sigma_d^2)$ y $e_{di} \sim ind(0, \sigma_e^2 k_{di}^2)$, siendo k_{di} constantes conocidas que representan la posible heteroscedasticidad.

En relación con la media del área d , ésta puede descomponerse en la suma de los valores observados en la muestra y los no muestreados, de la forma siguiente:

$$\bar{Y}_d = N_d^{-1} \left(\sum_{i \in s_d} Y_{di} + \sum_{i \in r_d} Y_{di} \right) \quad (\text{Ec. 63})$$

En este sentido, no es necesario predecir los valores muestrales puesto que vienen dados. El BLUP de \bar{Y}_d bajo el modelo con errores anidados (Ec. 62) se obtiene ajustando el modelo a los datos de la muestra y prediciendo los valores de las variables Y_{di} fuera de la muestra del área d , es decir:

$$\tilde{Y}_d^{EBLUP} = N_d^{-1} \left(\sum_{i \in s_d} Y_{di} + \sum_{i \in r_d} \tilde{Y}_{di}^{BLUP} \right) \quad (\text{Ec. 64})$$

En donde tomando el estimador de mínimos cuadrados ponderados $\tilde{\beta}$ de β bajo el modelo (Ec. 62), los valores predichos son:

$$\tilde{Y}_{di}^{EBLUP} = \mathbf{x}_{di}' \tilde{\beta} + \tilde{u}_d \quad (\text{Ec. 65})$$

En donde:

$$\tilde{u}_d = \gamma_d(\bar{y}_{da} - \bar{x}_{da}'\tilde{\beta}) \quad (\text{Ec. 66})$$

$$\gamma_d = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2/a_d). \quad (\text{Ec. 67})$$

Las medias muestrales ponderadas de la variable respuesta:

$$\bar{y}_{da} = a_d^{-1} \sum_{i \in s_d} a_{di} Y_{di}. \quad (\text{Ec. 68})$$

$$\bar{x}_{da} = a_d^{-1} \sum_{i \in s_d} a_{di} \mathbf{x}_{di}. \quad (\text{Ec. 69})$$

Con los pesos

$$a_{di} = k_{di}^{-2} \quad (\text{Ec. 70})$$

En donde:

$$a_d = \sum_{i \in s_d} a_{di} \quad (\text{Ec. 71})$$

De nuevo \tilde{u}_d es BLUP de u_d y los valores predichos, \tilde{Y}_{di}^{BLUP} son BLUPs de las variables $Y_{di}, i \in r_d$ bajo el modelo (EC. 62).

Construyendo el vector de variables de respuesta para el área d , $y_d = (Y_{d1}, \dots, Y_{dN_d})'$ y la matriz correspondiente de variables auxiliares, $\mathbf{X}_d =$

$(x_{d1}, \dots, x_{dN_d})'$ bajo el modelo de errores anidados (EC. 62), se tiene que $\mathbf{y}_d \sim \text{ind}N(\mathbf{X}_d\boldsymbol{\beta}, \mathbf{V}_d)$, $d = 1, \dots, D$, en donde:

$$\mathbf{V}_d = \sigma_u^2 \mathbf{1}_{N_d} \mathbf{1}'_{N_d} + \sigma_e^2 \mathbf{A}_d \quad (\text{Ec. 72})$$

En donde $\mathbf{A}_d = \text{diag}(k_{di}^2; i = 1, \dots, N_d)$. Ahora se descompone el vector \mathbf{y}_d del área d en los subvectores para las unidades de la muestra y para las unidades fuera de la muestra de la forma $\mathbf{y}_d = (\mathbf{y}_{ds}', \mathbf{y}_{dr}')$ y las matrices \mathbf{X}_d y \mathbf{V}_d de la misma forma:

$$\mathbf{X}_d = \begin{pmatrix} \mathbf{X}_{ds} \\ \mathbf{X}_{dr} \end{pmatrix} \quad (\text{Ec. 73})$$

$$\mathbf{V}_d = \begin{pmatrix} \mathbf{V}_{ds} & \mathbf{V}_{dsr} \\ \mathbf{V}_{drs} & \mathbf{V}_{dr} \end{pmatrix} \quad (\text{Ec. 74})$$

Con esta notación el estimador de mínimos cuadrados ponderados de $\boldsymbol{\beta}$ está dado por:

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{d=1}^D \mathbf{X}_{ds} \mathbf{V}_{ds}^{-1} \mathbf{X}_{ds}' \right)^{-1} \sum_{d=1}^D \mathbf{X}_{ds} \mathbf{V}_{ds}^{-1} \mathbf{y}_{ds} \quad (\text{Ec. 75})$$

Para áreas con fracción muestral despreciable, es decir, donde $n_d/N_d \approx 0$, el BLUP de la media \bar{Y}_d se puede escribir de la forma:

$$\hat{Y}_d^{BLUP} \approx \gamma_d \{ \bar{y}_d + (\bar{\mathbf{X}}_d - \bar{\mathbf{x}}_{da})' \tilde{\boldsymbol{\beta}} \} + (1 - \gamma_d) \bar{\mathbf{X}}_d' \tilde{\boldsymbol{\beta}} \quad (\text{Ec. 76})$$

Molina (2019) refiere que como $\gamma_d \in (0,1)$, el BLUP es una media ponderada entre el estimador $\bar{y}_{da} + (\bar{X}_d - \bar{x}_{da})'\tilde{\beta}$, conocido como estimador “*survey regression*” y el estimador sintético de regresión, $\bar{X}_d'\tilde{\beta}$. Este estimador se obtiene ajustando el modelo (Ec. 62) pero tomando los efectos de las áreas u_d como fijos en lugar de aleatorios.

El BLUP dado en el modelo (Ec. 64) depende de los verdaderos valores de las componentes de la varianza del modelo (Ec. 62), $\theta = (\sigma_u^2, \sigma_e^2)'$. Sustituyendo el verdadero θ por un estimador consistente $\hat{\theta} = (\hat{\sigma}_u^2, \hat{\sigma}_e^2)'$ en BLUP (Ec. 64).

$$\hat{Y}_{di}^{EBLUP} = x_{di}'\hat{\beta} + \hat{u}_d \quad (\text{Ec. 77})$$

En donde:

$$\hat{u}_d = \hat{\gamma}_d(\bar{y}_{da} - \bar{x}_{da}'\hat{\beta}) \quad (\text{Ec. 78})$$

$$\hat{\gamma}_d = \hat{\sigma}_u^2 / (\hat{\sigma}_u^2 + \hat{\sigma}_e^2 / a_d) \quad (\text{Ec. 79})$$

Al sustituir θ por el estimador $\hat{\theta}$, el EBLUP sigue siendo insesgado bajo el modelo (Ec. 62) y bajo ciertas condiciones sobre el estimador $\hat{\theta}$.

2.1.5.7. Modelo ELL

El método de Elbers, Lanjouw y Lanjow es llamado método ELL (como citó Molina (2019), es un método tradicionalmente utilizado por el Banco Mundial para

construir mapas de pobreza o desigualdad. Según refiere Molina (2019) este método fue el primero que apareció en la literatura que permite estimar indicadores más complejos que simplemente medias y totales, siempre que sean función de una variable que mide el poder adquisitivo de los individuos (generalmente ingresos netos). El método asume el método de errores anidados (Ec. 62) para la transformación por el logaritmo de esta variable, en donde los efectos aleatorios son para las unidades primarias de muestreo del diseño muestral (diseño por conglomerados) en lugar de para las áreas de interés.

Para ejemplificar este método y al mismo tiempo simplificar la notación, se considera que las unidades de primera etapa equivalen a las áreas. En este caso, si E_{di} es la variable que mide el poder adquisitivo del individuo i en el área d , tomando $Y_{di} = \log(E_{di} + c)$, donde $c > 0$ es constante (tradicionalmente este método emplea un valor de $c = 0$). El modelo ELL es el siguiente:

$$Y_{di} = \mathbf{x}'_{di}\boldsymbol{\beta} + u_d + e_{di} \quad i = 1, \dots, N_d, d = 1, \dots, D. \quad (\text{Ec. 80})$$

En donde $u_d \sim^{iid}(0, \sigma_u^2)$ y $e_{di} \sim^{ind}(0, k_{di}^2)$, siendo u_d y e_{di} independientes, y k_{di} constantes conocidas que representan la posible heteroscedasticidad.

De acuerdo con Molina (2019) el estimador ELL de un parámetro general $\theta_d = \theta_d(y_d)$ bajo el modelo (Ec. 80) se obtiene mediante un procedimiento *Bootstrap*. Este procedimiento *Bootstrap* proporciona una aproximación numérica del estimador ELL teórico, que viene dado por la esperanza marginal $\hat{\theta}_d^{ELL} = E[\theta_d]$, a diferencia del método EBLUP basado en el modelo con errores anidados que está condicionado a la muestra y_s . Para estimar el ECM del estimador ELL se utiliza el mismo procedimiento *Bootstrap*.

Dicho con palabras de Molina (2019) el procedimiento *Bootstrap* funciona de la siguiente manera: En primer lugar, a partir de los residuos del modelo (Ec. 80) ajustado a los datos, se generan efectos aleatorios u_d^* para cada área $d = 1, \dots, D$ y errores e_{di}^* para cada individuo $i = 1, \dots, N_d, d = 1, \dots, D$. A partir de ellos, del estimador $\widehat{\boldsymbol{\beta}}$ del parámetro de regresión $\boldsymbol{\beta}$, y utilizando los valores de las variables auxiliares para los elementos dentro y fuera de la muestra, se generan valores *Bootstrap* de la variable respuesta para todos los elementos de la población de la forma siguiente:

$$Y_{di}^* = \mathbf{x}'_{di}\boldsymbol{\beta} + u_d^* + e_{di}^*, \quad i = 1, \dots, N_d, d = 1, \dots, D \quad (\text{Ec. 81})$$

Este procedimiento proporciona un censo de la variable respuesta, con el que se puede calcular cualquier tipo de parámetro, lo que resulta ventajoso mientras estén definidos como una función de las variables de respuesta Y_{di} . Esta mecánica de generación se repite para $a = 1, \dots, A^*$, obteniendo así censos completos. Para cada censo a , se calcula el indicador de interés $\theta_d^{*(a)} = \theta_d(\mathbf{y}_d^{*(a)})$, donde $\mathbf{y}_d^{*(a)} = (y_d^{*(a)}, \dots, y_{dN_d}^{*(a)})'$ son los valores de la variable respuesta en el área d en el censo *Bootstrap* a . Finalmente, el estimador ELL se obtiene promediando para los A censos:

$$\widehat{\theta}_d^{ELL} = \frac{1}{A^*} \sum_{a=1}^{A^*} \theta_d^{*(a)} \quad (\text{Ec. 82})$$

Por otro lado, en este método, el ECM se estima mediante:

$$mse_{ELL}(\widehat{\theta}_d^{ELL}) = \frac{1}{A^*} \sum_{a=1}^{A^*} (\theta_d^{*(a)} - \widehat{\theta}_d^{ELL})^2 \quad (\text{Ec. 83})$$

Este método está basado en datos a nivel de individuo, que proporciona información con mayor detalle que los datos a nivel de área. Rao y Molina (2015) refieren que una de las desventajas de este método es que los estimadores ELL pueden presentar un alto ECM, pudiendo incluso comportarse peor que los estimadores directos basados en el diseño si la heterogeneidad entre áreas no explicada es significativa. Otra desventaja es que pueden verse afectados por valores atípicos.

2.1.5.8. Mejor predictor empírico bajo el modelo con errores anidados

El mejor predictor (en inglés, *best/Bayes predictor*, BP) basado en el modelo con errores anidados fue propuesto por Molina y Rao en el año 2010 (como citó Molina, 2019) para estimar indicadores generales de tipo no lineal. El método asume que las variables $Y_{di} = \log(E_{di} + c)$ siguen el modelo (Ec. 62) con normalidad para los efectos aleatorios de las áreas u_d y para los errores e_{di} . Como lo hace notar Molina (2019), bajo este modelo, los vectores de variables para cada área, $\mathbf{y}_d = (Y_{d1}, \dots, Y_{dN_d})'$, $d = 1, \dots, D$, son independientes y verifican $\mathbf{y}_d \sim^{ind} N(\boldsymbol{\mu}_d, \mathbf{V}_d)$, con valor de medias $\boldsymbol{\mu}_d = \mathbf{X}_d \boldsymbol{\beta}$, siendo $\mathbf{X}_d = (\mathbf{x}_{d1}, \dots, \mathbf{x}_{dN_d})'$ y matriz de covarianzas $\mathbf{V}_d = \sigma_u^2 \mathbf{1}_{N_d} \mathbf{1}_{N_d}' + \sigma_e^2 \mathbf{A}_d$, donde $\mathbf{A}_d = \text{diag}(k_{di}^2; i = 1, \dots, N_d)$. Para un indicador general definido como una función de \mathbf{y}_d , es decir, $\theta_d = \theta_d(\mathbf{y}_d)$, el mejor predictor es aquel que minimiza el ECM y viene dado por:

$$\tilde{\theta}_d^B(\boldsymbol{\theta}) = E_{y_{dr}}[\theta_d(\mathbf{y}_d) | \mathbf{y}_{ds}; \boldsymbol{\theta}] \quad (\text{Ec. 84})$$

En donde la esperanza se toma respecto de la distribución del vector de valores fuera de la muestra y_{dr} del dominio d dados los valores de la muestra y_{ds} . Molina (2019) hace notar que esta distribución condicionada depende del valor de los parámetros del modelo para θ y que al reemplazar θ por un estimador

consistente $\hat{\theta}$ en el mejor predictor (Ec. 80), se obtiene el llamado mejor predictor empírico (en inglés, *empirical best/Bayes*, EB) $\hat{\theta}_d^{EB} = \hat{\theta}_d^B(\hat{\theta})$.

2.1.5.9. Método jerárquico Bayes bajo el modelo con errores anidados

El cálculo de los estimadores descritos en los párrafos anteriores junto con sus errores cuadráticos medios requiere computación intensiva y puede ser inviable en algunas situaciones si la población de interés es muy grande o para indicadores muy complejos (por ejemplo, los que requieren ordenaciones). Con el propósito de desarrollar un método computacionalmente más eficiente, Molina, Nandram y Rao en año 2014 (como citó Molina, 2019) propusieron el método jerárquico Bayes (en inglés, *hierarchical Bayes*, HB) para la estimación de parámetros generales. Este método no requiere el uso del método *Bootstrap* en la estimación del error cuadrático medio, ya que proporciona muestras de la distribución posterior del parámetro de interés, a partir de las cuales se obtienen las varianzas posteriores que juegan el papel de ECM, o cualquier otra medida resumen.

Molina (2019) argumenta que el método HB se basa en reparametrizar el modelo con errores anidados (Ec. 62) en términos del coeficiente de correlación intraclase $\rho = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$ y considerando distribuciones previas para los parámetros del modelo $(\boldsymbol{\beta}, \rho, \sigma_e^2)$ que reflejan la falta de información previa sobre ellos. Concretamente, el modelo es el siguiente:

$$(i) Y_{di} | u_d, \boldsymbol{\beta}, \sigma_e^2 \stackrel{ind}{\sim} N(\mathbf{x}'_{di} \boldsymbol{\beta} + u_d, \sigma_e^2 k_{di}^2), \quad i = 1, \dots, N_d \quad (\text{Ec. 85})$$

$$(ii) u_d | \rho, \sigma_e^2 \stackrel{iid}{\sim} N\left(0, \frac{\rho}{1-\rho} \sigma_e^2\right), \quad d = 1, \dots, D \quad (\text{Ec. 86})$$

$$(iii) \pi(\boldsymbol{\beta}, \rho, \sigma_e^2) \propto \frac{1}{\sigma_e^2}, \quad \epsilon \leq \rho \leq 1 - \epsilon, \sigma_e^2 > 0, \boldsymbol{\beta} \in R^p \quad (\text{Ec. 87})$$

En donde $\epsilon > 0$ se selecciona muy pequeño para reflejar la falta de información previa.

2.1.5.10. Métodos basados en modelos lineales generalizados mixtos

Debido a que el acceso a determinados servicios educativos o de salud, o la disponibilidad de ciertas condiciones en la vivienda se miden habitualmente mediante proporciones de unidades elementales que tienen o no acceso al servicio o indicador en cuestión. Para obtener estimaciones en el espacio natural $[0, 1]$ en el que están las proporciones, Molina (2019) postula que es habitual utilizar modelos lineales generalizados mixtos (en inglés, *generalized linear mixed models*, GLMM). Si $Y_{di} \in \{0, 1\}$ es la variable binaria que mide la presencia o no del atributo de interés, el modelo de estimación en áreas pequeñas más habitual es el GLMM con efectos aleatorios de las áreas, dado por:

$$\begin{aligned} Y_{di} | v_d &\sim \text{Bern}(p_{di}), g(p_{di}) = \mathbf{x}'_{di} \boldsymbol{\alpha} + v_d, v_d \stackrel{iid}{\sim} N(0, \sigma_v^2), \\ i &= 1, \dots, N_d, \\ d &= 1, \dots, D \end{aligned} \quad (\text{Ec. 88})$$

En donde v_d es el efecto del área d , $\boldsymbol{\alpha}$ es el vector de coeficientes de regresión y $g: (0, 1) \rightarrow R$ es la función vínculo (biyectiva, con derivada común).

En particular, el vínculo logístico dado por $g(p) = \log(p/(1-p))$ es posiblemente el más utilizado en la práctica. El mejor predictor bajo el modelo (que minimiza el ECM bajo el modelo) de la proporción $P_d = \bar{Y}_d$ viene dado por:

$$\bar{P}_d^B(\boldsymbol{\theta}) = E(P_d | \mathbf{y}_{sd}; \boldsymbol{\theta}) = \frac{1}{N_d} \left\{ \sum_{i \in s_d} y_{di} + \sum_{i \in r_d} E(Y_{di} | \mathbf{y}_{ds}; \boldsymbol{\theta}) \right\} \quad (\text{Ec. 89})$$

La distribución de $Y_{di} | \mathbf{y}_{ds}$ depende del vector $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \sigma_v^2)'$ de parámetros del modelo. En la práctica, se obtiene el predictor EB reemplazando $\boldsymbol{\theta}$ por un estimador consistente $\hat{\boldsymbol{\theta}}$ en el mejor predictor, es decir, $\hat{P}_d^{EB} = \bar{P}_d^B(\hat{\boldsymbol{\theta}})$.

Molina (2019) refiere que el estimador $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}', \sigma_v^2)'$ de $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \sigma_v^2)'$ se obtiene ajustando el modelo GLMM dado en (Ec. 88) a los datos de la muestra $\mathbf{y}_s = (\mathbf{y}_{1s}', \dots, \mathbf{y}_{Ds}')$. Si se desea ajustar el modelo mediante el método de máxima verosimilitud, es necesario maximizar la verosimilitud dada por $f(\mathbf{y}_s) = \int R^D f(\mathbf{y}_s | \mathbf{v}) f(\mathbf{v}) d\mathbf{v}$, donde $\mathbf{v} = (v_1, \dots, v_D)'$. Una vez ajustado el modelo se calculan las esperanzas $E(Y_{di} | \mathbf{y}_{ds}; \hat{\boldsymbol{\theta}})$ que definen el predictor EB. Dicha esperanza se puede expresar de la forma:

$$E(Y_{di} | \mathbf{y}_{ds}; \hat{\boldsymbol{\theta}}) = \frac{E\{h(\mathbf{x}_{di}\boldsymbol{\alpha} + v_d) f(\mathbf{Y}_{ds} | v_d; \hat{\boldsymbol{\theta}})\}}{E\{f(\mathbf{y}_{ds} | v_d; \hat{\boldsymbol{\theta}})\}} \quad (\text{Ec. 90})$$

En donde $h = g^{-1}$ es el vínculo inverso y el vínculo logístico es igual a:

$$\begin{aligned}
f(\mathbf{y}_{ds}|v_d) &= \prod_{i \in s_d} p_{di}^{Y_{di}} (1 - p_{di})^{(1-Y_{di})} \\
&= \prod_{i \in s_d} h(\mathbf{x}_{di}'\boldsymbol{\alpha} + v_d)^{Y_{di}} \{1 - h(\mathbf{x}_{di}'\boldsymbol{\alpha} + v_d)\}^{(1-Y_{di})}
\end{aligned} \tag{Ec. 91}$$

El vínculo inverso es $h(\mathbf{x}_{di}'\boldsymbol{\alpha} + v_d) = \exp(\mathbf{x}_{di}'\boldsymbol{\alpha} + v_d) / \{1 + \exp(\mathbf{x}_{di}'\boldsymbol{\alpha} + v_d)\}$. Utilizando (Ec. 91) puede aproximarse las dos esperanzas que aparecen en (Ec. 90) mediante simulación Monte Carlo, generando $v_d^{(r)} \sim N(0, \hat{\sigma}_v^2)$, $r = 1, \dots, R$, y después calculando:

$$E(Y_{di}|\mathbf{y}_{ds}; \hat{\boldsymbol{\theta}}) \approx \frac{R^{-1} \sum_{r=1}^R h(\mathbf{x}_{di}'\hat{\boldsymbol{\alpha}} + v_d^{(r)}) \hat{f}(\mathbf{y}_{ds}|v_d^{(r)})}{R^{-1} \hat{f}(\mathbf{y}_{ds}|v_d^{(r)})}, \quad i \in r_d \tag{Ec. 92}$$

En donde \hat{f} es la densidad condicionada $f(\mathbf{y}_{ds}|v_d)$, con $\boldsymbol{\alpha}$ reemplazado por $\hat{\boldsymbol{\alpha}}$. Sin embargo, ajustar el modelo GLMM y ajustar la aproximación Monte Carlo, requiere de un tiempo computacional considerable. Un estimador simple cuando se estima una proporción, si $\hat{\boldsymbol{\alpha}}$ y \hat{v}_d son los estimadores $\boldsymbol{\alpha}$ y v_d que devuelve el software, se puede calcular por el método de analogía prediciendo los valores fuera de la muestra a través del modelo, es decir, tomando:

$$\hat{p}_d^{PI} = \frac{1}{N_d} \left(\sum_{i \in s_d} Y_{di} + \sum_{i \in r_d} \hat{p}_{di} \right) \tag{Ec. 93}$$

En donde $\hat{p}_{di} = h(\mathbf{x}_{di}'\hat{\boldsymbol{\alpha}} + \hat{v}_d)$ es el valor predicho de la observación de la muestra Y_{di} , $i \in r_d$.

2.2. Teoría general para la medición de la pobreza

Empleando las palabras de Romero (2000) “Nada es más difícil que definir el concepto de pobreza, pues esta involucra múltiples factores determinantes, los cuales varían dependiendo de las circunstancias de cada país, región o época” (p. 36). Desde su perspectiva, para entender mejor la pobreza es necesario conocer los diferentes enfoques que existen sobre la misma y que reflejan, de una u otra manera, determinados intereses que son respaldados por planteamientos teóricos y técnicos.

2.2.1. Privaciones y teorías normativas del bienestar

De acuerdo con Larrañaga (2007) “Los enfoques normativos del bienestar constituyen macro-visiones del ordenamiento social, representando el marco natural de referencia para la discusión de privaciones socioeconómicas” (p. 13). Agrega que en la actualidad tres son los enfoques más relevantes: el utilitarismo, la justicia como imparcialidad de Rawls y el enfoque de funcionamientos y capacidades de Sen.

2.2.1.1. Utilitarismo

Larrañaga (2007) declara que este enfoque origina la mayor parte de los postulados de la economía. Empleando sus palabras “El utilitarismo postula que el bien del individuo reside en la satisfacción de sus preferencias individuales; mientras que el bien social consiste en maximizar una función de bienestar social construida a partir de la agregación de las preferencias individuales” (p.13). En el contexto, agrega que en el enfoque utilitarista la pobreza se define en términos de un nivel de bienestar mínimo a alcanzar por parte de cada persona, cuyo equivalente en términos de ingresos es la línea de pobreza.

2.2.1.2. La justicia como imparcialidad de Rawls

La justicia como imparcialidad fue abordada por Rawls en el año 1971 en su libro "*A Theory of Justice*". Rawls (1971) postula que una sociedad es una asociación cooperativa, cuyos miembros tienen intereses comunes en los frutos que se deriven de la cooperación social, además que, existe un conflicto sobre la distribución de los beneficios y responsabilidades asociados a la vida en sociedad. La resolución del conflicto compete a los principios de justicia. Los principios de justicia considerados determinan de este modo la estructura básica de la sociedad. Por lo tanto, en palabras de Rawls (1971), "La justicia es la primera virtud de las instituciones sociales, como la verdad lo es de los sistemas de pensamiento" (p.3).

El sujeto de la justicia al que hace referencia el trabajo de Rawls (1971), son las instituciones básicas de la sociedad, aquellas que tienen efectos fundamentales y perdurables sobre la distribución de los derechos y beneficios de la cooperación social. El logro de la justicia se define por tanto en el diseño de estas instituciones, puesto que ellas determinan los términos de la cooperación social entre las personas.

2.2.1.3. El enfoque de funcionamientos y capacidades de SEN

Larrañaga (2007) afirma que los enfoques anteriores definen el bienestar de las personas como un logro que se alcanza a través de medios como el ingreso. Larrañaga (2007) agrega que para Sen existe una categoría intermedia, denominada funcionamientos, que sirve de enlace entre los medios y los logros. Los funcionamientos para Sen (como citó Larrañaga, 2007) "son los constituyentes del bienestar de las personas: estar saludable, bien nutrido, tener

educación, participar de la cultura de la sociedad, estar integrado a la sociedad, etc.” (p.16). Desde esta perspectiva, debe hacerse énfasis en las actividades de la persona (ser y hacer) antes que en tenencias.

Teniendo en cuenta a Larrañaga (2007), las capacidades para Sen “constituyen vectores de funcionamientos, que entregan posibilidades para vivir uno u otro tipo de vida de acuerdo con los proyectos de las personas” (p.16). Las capacidades, por lo tanto, entregan la capacidad de elegir el tipo de vida que las personas desean vivir, según sus razones.

2.2.2. Metodologías utilizadas para medir la pobreza

En términos generales, la pobreza es entendida como una situación de privación que sufren los hogares y las personas. Según refiere el Instituto Nacional de Estadística INE, Proyecto MECOVI (2002):

La pobreza se relaciona no sólo con las oportunidades de acceso a bienes materiales y servicios, sino también se asocia con la vulnerabilidad y las limitaciones a la participación y decisión y con la probabilidad que tienen muchas personas de salir desfavorecidas por las crisis económicas, los desastres, las enfermedades y la violencia personal. (p. 3)

En este sentido, dada la complejidad que conlleva la conceptualización de la pobreza, su identificación y medición es un fenómeno complejo que comprende múltiples dimensiones. Para su evaluación, se suele utilizar dos vertientes principales: el método directo y el método indirecto.

Las mediciones directas se apoyan en factores sobre los cuales se identifica el nivel de bienestar, por ejemplo, una nutrición adecuada, un lugar decente para vivir, educación básica, entre otros.

Al concebirse la pobreza como compleja y multidimensional, recurrir a una medida que indique el nivel de bienestar o calidad de vida de las personas engloba una cantidad numerosa de factores, por lo que en la práctica la medición directa resulta casi imposible. Por otra parte, la medición indirecta de la pobreza se apoya en la construcción de mediciones y líneas monetarias.

Sumado a esos dos métodos, en la actualidad, para el abordaje de la pobreza se sumó una nueva línea de trabajo que considera o toma parte de los dos anteriores. Entre los métodos más utilizados a nivel internacional figuran: el método de las necesidades básicas insatisfechas, el método de la línea de pobreza, el método de la medición integrada y el método multidimensional.

2.2.2.1. Método de necesidades básicas insatisfechas NBI

Feres y Mancero (2001) refieren que el método directo de las Necesidades Básicas Insatisfechas (NBI) fue propuesto en la década de 1960 para la identificación de ciertas carencias críticas de la población y la caracterización de la pobreza. La Comisión Económica para América Latina y el Caribe (CEPAL) recomendó su utilización a partir de la década de 1980, por lo que es un método muy utilizado en América Latina y el Caribe. La metodología de las NBI considera un conjunto mínimo de indicadores relacionados con las necesidades básicas estructurales (vivienda, salud, educación, entre otros) para evaluar el bienestar individual de las personas.

Este conjunto de indicadores puede variar de país en país. Para el caso de Guatemala, el Instituto Nacional de Estadística INE (2006) da a conocer la incorporación en la medición seis necesidades básicas consideradas como el mínimo vital en la sociedad guatemalteca. Los criterios adoptados de satisfacción fueron especificados de formas distintas para el área urbana y el área rural del país. Las seis necesidades básicas utilizadas para la identificación de insatisfacción a nivel de hogares son:

- Calidad de la vivienda

El criterio de insatisfacción utilizado para el área urbana es que las viviendas estén construidas con: pared de bajareque, lepa, palo, caña u otro, o bien, pared de ladrillo, block, adobe, madera, lámina y con piso de tierra. Para el caso del área rural se requiere que las viviendas sean construidas con pared de lámina, bajareque, lepa, palo, caña u otro y piso de tierra.

- Hacinamiento

Para ambas áreas geográficas, se consideran hogares con necesidad insatisfecha cuando residen más de tres personas por cuarto (excluyendo baño y cocina).

- Origen y abastecimiento de agua

Para el caso del área urbana el criterio de insatisfacción utilizado considera a los hogares sin conexión a una red de distribución de agua y con servicio de abastecimiento de chorro público, pozo, camión, tonel, río, lago o manantial. A nivel rural, los hogares con servicio de abastecimiento de agua de río, lago, manantial u otra forma.

- Acceso a servicio sanitario

En áreas urbanas se considera insatisfechos a hogares sin conexión a una red de captación de desagüe y con tipo de servicio sanitario igual a excusado lavable, letrina, pozo ciego o sin servicio sanitario. En áreas rurales se emplea el criterio de insatisfacción a hogares que no cuentan con ningún sistema de eliminación de excretas o que no disponen de servicio sanitario.

- Asistencia escolar

Para el caso del área urbana, bajo este criterio, se considera insatisfecho a hogares con niños de 7 a 12 años que no asisten a establecimientos de educación formal. En el área rural aplica para hogares con niños de 7 a 10 años que no asisten a establecimientos de educación formal.

- Precariedad ocupacional

Este indicador tampoco hace diferencia entre el área urbana y rural, por lo que se considera insatisfacción cuando el jefe del hogar no tiene ningún nivel de instrucción y hay más de cuatro personas por miembro del hogar ocupado.

Las NBI anteriores permiten la estimación de un Índice de Necesidades Básicas Insatisfechas (INBI). El INBI permite clasificar a los hogares según su nivel socioeconómico como “pobres” y “no pobres”, bajo este criterio aquellos hogares que tienen al menos una necesidad básica insatisfecha se consideran como pobres. Derivado de la clasificación en pobreza, el INE en el año 2002 incorporó además el criterio de “pobreza extrema” para aquellos hogares con dos o más NBI.

2.2.2.2. Método de línea de pobreza

A juicio de Larrañaga (2007), el método más utilizado internacionalmente, a pesar de sus limitaciones es el método de la línea de pobreza (LP), el cual utiliza el ingreso o el gasto de consumo como medidas del bienestar, estableciéndose un valor per cápita de una canasta mínima de consumo necesario para la sobrevivencia, el cual permite la diferenciación de los niveles de pobreza.

Guatemala adoptó la medición por el método de líneas de pobreza a partir del año 2000, el cuál define como línea de pobreza extrema el costo que representa adquirir las 2,172 calorías mínimas recomendadas para Guatemala por el Instituto de Nutrición de Centro América y Panamá (INCAP). Según el Instituto Nacional de Estadística INE (2016) la línea de pobreza extrema se fijó en Q 5,750 per cápita anual. La línea de pobreza general incluye, además del costo alimenticio, un monto adicional que corresponde al porcentaje de consumo de bienes y servicios, éste monto para el año 2014 fue de Q10,218.

Teniendo en cuenta a Molina (2019) “El indicador de pobreza más utilizado es la incidencia o tasa de pobreza, también llamada tasa de riesgo de pobreza, que se calcula como la proporción de individuos con ingresos (netos equivalentes) por debajo del umbral o línea de pobreza” (p. 19). Otro indicador referido por Molina (2019) es el de brecha de la pobreza “que mide la magnitud de la pobreza en lugar de la frecuencia de individuos en riesgo de pobreza” (p. 19). Estos indicadores forman parte de una familia de indicadores más amplia definida por Foster, Greer y Thorbecke en el año 1984, llamada familia de indicadores FGT, y que poseen la ventaja de ser aditivos en los elementos de la población.

2.2.2.3. Método de medición integrada

De acuerdo con Feres y Mancero (2001) un tercer método, denominado método integrado de medición de la pobreza, combina el método de la línea de pobreza y el de necesidades básicas insatisfechas para clasificar a la población en cuatro niveles de bienestar (pobres crónicos, recientes, inerciales e integrados socialmente). Esto como consecuencia de que tienen características propias que se complementan de manera muy útil para ciertos propósitos. La principal limitante de este método es que el porcentaje total de pobres siempre resulta mayor que el que proporciona cualquiera de los dos métodos que integra.

2.2.2.4. Método multidimensional

Como lo hace notar Larrañaga (2007) “En los años recientes ha habido algunos desarrollos teóricos orientados a derivar medidas de pobreza multidimensional” (p.21). Desde su perspectiva, se trata de generalizar el marco de la pobreza en una dimensión (ingresos) al plano de diferentes variables. En este sentido, agrega Larrañaga (2007) “Se requiere por tanto resolver los problemas de identificación y de agregación, así como poder comparar distintas distribuciones de los atributos del bienestar en términos de la medida de pobreza” (p. 21).

A nivel de la identificación, Larrañaga (2007) señala que: “(a) identificar las dimensiones del bienestar que serán evaluadas en el indicador de pobreza, y (b) definir umbrales mínimos para cada una de las dimensiones” (p. 21). La selección de las dimensiones de la pobreza es un tema fundamental, ya que supone la identificación de aquellos aspectos del bienestar que son centrales para las personas vivan una “vida mínima decente”. La identificación y agregación

conlleva problemas para los diferentes marcos conceptuales que se empleen para la evaluación social del bienestar.

2.2.3. Índice de Pobreza Multidimensional de Guatemala IPM-Gt

Según el Ministerio de Desarrollo Social (2018), el Índice de Pobreza Multidimensional de Guatemala IPM-Gt se concibe en el marco de dos grandes agendas de desarrollo: el Plan Nacional de Desarrollo Nuestra Guatemala K'atun 2032 y la Agenda de Desarrollo Sostenible 2030. En este contexto, el IPM-Gt define el siguiente objetivo:

Sustentar la toma de decisiones en la gestión técnico-política que permita la adecuada planificación, programación y asignación de recursos del Estado, así como el seguimiento y evaluación en función de las prioridades nacionales de desarrollo (definidas en el Plan Nacional de Desarrollo Nuestra Guatemala K'atun 2032 y la Agenda de Desarrollo Sostenible 2030), para el bienestar e igualdad de oportunidades con énfasis en la población en pobreza multidimensional. (p. 4)

2.2.3.1. Metodología del IPM-Gt

De acuerdo con la documentación del Índice de Pobreza Multidimensional para Guatemala, el IPM-Gt fue construido a partir del método Alkire-Foster propuesto por James Foster y Sabina Alkire. El IPM-Gt está conformado por cinco dimensiones y 17 indicadores. Las dimensiones son: salud y seguridad alimentaria y nutricional, educación, empleo digno, vivienda y servicios básicos. Los 17 indicadores se seleccionan a partir de aspectos relevantes para la sociedad guatemalteca.

Según refiere el Ministerio de Desarrollo Social (2018) para llegar al cálculo del IPM-Gt es necesario estimar dos componentes: la tasa de incidencia (H) y la tasa de intensidad (A). La tasa de incidencia (H) es el porcentaje de hogares que se encuentran en pobreza multidimensional y la tasa de intensidad (A) es el porcentaje promedio de privaciones que sufren las personas multidimensionalmente pobres. Esta medida indica qué tan pobres son los pobres y el grado o severidad de su pobreza en promedio, pudiéndose determinar con claridad en qué indicadores y en qué dimensiones se encuentra privadas las personas multidimensionalmente pobres. El IPM-Gt (M_o) entonces es el resultado de la multiplicación de la intensidad y la incidencia de la pobreza multidimensional, como se muestra en la siguiente fórmula.

$$\text{IPM-Gt } (M_o) = H * A \quad (\text{Ec. 94})$$

Con base a los datos provenientes de la Encuesta Nacional de Condiciones de Vida ENCOVI 2014, se estimó que 61.6 % de la población guatemalteca es multidimensionalmente pobre con un intervalo de estimación al 95% de confianza de [58.6 %, 64.7 %]. La intensidad de la pobreza se estimó en 48.6% con el mismo nivel de confianza [47.5 %, 49.7 %], esto refleja que las personas multidimensionalmente pobres en promedio están privadas en casi la mitad de los indicadores ponderados.

El IPM-Gt para el año 2014 se estimó en 0.299 con un intervalo de confianza al 95 % de [0.28, 0.318 %]. Esto significa que las personas multidimensionalmente pobres experimentan el 29.9 % de privaciones ponderadas respecto al máximo posible de privaciones que tendrían si todas las personas en Guatemala fueran pobres y privados en todas las dimensiones.

2.2.3.2. Diseño de la medida de pobreza multidimensional IPM-Gt

Los distintos indicadores de pobreza pueden ser evaluados según cumplan o no los axiomas deseables. En este sentido, la elección de una medida de pobreza multidimensional para Guatemala, asociada a la unidad de identificación hogar, giró en torno a la descripción de tres componentes: dimensiones e indicadores, ponderaciones y el método de umbrales duales (privación y pobreza).

- Dimensiones e indicadores

Como lo hace notar el informe del Ministerio de Desarrollo Social (2018), la selección de las dimensiones e indicadores se fundamentó en el análisis de los nodos críticos que arrojó las metas estratégicas y prioridades de desarrollo definidas por la Secretaría de Planificación de la Presidencia de Guatemala (SEGEPLAN), con el acompañamiento de la CEPAL. La siguiente tabla, detalla las dimensiones y los indicadores del IPM-Gt con sus ponderaciones, así como los umbrales de privación establecidos para cada indicador.

Tabla VII. **IPM-Gt: Matriz de dimensiones, indicadores, umbrales y pesos**

Dimensión	Peso	Tema	Peso	Umbrales de privación
1. Salud y Seguridad Alimentaria y Nutricional	20%	Acceso a servicios de salud	5%	<p>El hogar está privado en acceso a servicios de salud (uso efectivo de servicios de salud) si una persona mayor de 5 años se enfermó y no tuvo acceso a servicios de salud por las siguientes causas:</p> <ul style="list-style-type: none"> -el lugar de atención está lejos -falta de dinero -no hay medio de transporte -no hay médicos/enfermeras -no hablan en el idioma de la persona enferma -la espera es muy larga -no alcanzó número o cupo -otro.
		Seguridad alimentaria y nutricional	5%	<p>El hogar se considera privado si se encuentra en los niveles de inseguridad alimentaria y nutricional moderada o severa definidos por la <i>Food Insecurity Experience Scale</i> (Escala Latinoamericana y Caribeña de Seguridad Alimentaria -ELCSA), en la cual:</p> <ul style="list-style-type: none"> -Inseguridad alimentaria moderada: el hogar reportó entre 6 y 10 experiencias de inseguridad alimentaria, en hogares con menores de 18 años, y entre 4 y 6 en hogares sin menores de 18 años. -Inseguridad alimentaria severa: el hogar reportó entre 11 y 15 experiencias de inseguridad alimentaria, en hogares con menores de 18 años, y entre 7 y 8 en hogares sin menores de 18 años.
1. Salud y Seguridad Alimentaria y Nutricional	20%	Embarazo en adolescentes	5%	El hogar se considera privado si hay alguna niña o adolescente menor de 20 años embarazada o que ha estado embarazada.
		Cuidado prenatal	5%	El hogar se considera privado si hay al menos una mujer de 12 a 49 años que ha estado embarazada en los cinco años anteriores a la realización de la encuesta, que en su último embarazo no asistió a por lo menos 4 controles de embarazo.
2. Educación	20%	Asistencia escolar	5%	El hogar es privado cuando hay al menos un niño o adolescente de 7 a 15 años que no asiste a la escuela, o hay al menos un adolescente de 16 a 17 años que no asiste a la escuela y no ha completado 9 años de estudios.
		Años de escolaridad	5%	<p>El hogar se considera privado si hay al menos un adulto (18 años o más) con menos de los años de escolaridad según las edades siguientes:</p> <ul style="list-style-type: none"> -18 a 32 años: 9 años de escolaridad -33 a 64 años: 6 años de escolaridad -Mayores de 65 años: no saber leer y escribir.
		Rezago educativo	5%	El hogar se considera privado cuando hay al menos un niño o adolescente de 8 a menos de 18 años que presenta más de dos años de rezago escolar, estimado según su edad y el grado en el que se inscribió.

Continúa tabla VII

		Cuidado infantil	5%	El hogar se considera privado si hay al menos un niño menor de 7 años que permanece con el padre o la madre en su lugar de trabajo, o con vecinos o amigos, o permaneció la mayor parte del tiempo solo.
3. Empleo Digno	20%	Empleo informal	10%	El hogar se considera privado cuando hay al menos una persona ocupada en alguna de las siguientes categorías: -Empleadores, empleados y obreros de empresas de menos de 6 personas. -Todos los trabajadores por cuenta propia o autónoma (excluyendo profesionales y técnicos, medido como 12 años de escolaridad o más). -Todos los familiares no remunerados. -Ocupados en servicio doméstico.
		Trabajo infantil	10%	El hogar se considera privado si existe al menos un niño de 15 años o menos que trabaja, o cuando existe al menos un adolescente de 16 a 17 años que trabaja, no estudia y no tiene al menos 9 años de educación completos.
4. Vivienda	20%	Materiales de la vivienda	6.67%	El hogar se considera privado si la vivienda que habita está construida con materiales inapropiados, bajo los siguientes criterios: -se consideran privadas las viviendas con piso de tierra -se consideran privadas las viviendas, sin importar el material del piso, cuando el material de pared está hecho de adobe, lámina metálica, bajareque, lepa, palo o caña, u otro.
4. Vivienda	20%	Hacinamiento	6.67%	El hogar se encuentra privado por hacinamiento cuando habitan más de tres personas en promedio por cuarto (sin considerar los cuartos utilizados para cocina, baños, pasillos, garajes y los dedicados a negocios).
		Combustible para cocinar	6.67%	El hogar se considera privado si el combustible utilizado para cocinar es leña y no cuenta con una chimenea o una salida de escape para el humo.
5. Servicios básicos	20%	Acceso al agua	5%	El hogar se considera privado cuando no tiene acceso al agua según los siguientes criterios: -Área urbana: el agua es obtenida principalmente de chorro público, pozo perforado público o privado, río, lago, manantial, camión cisterna, agua de lluvia y otro. -Área rural: el agua es obtenida principalmente de río, lago, manantial, camión cisterna, agua de lluvia y otro. -Para ambas áreas: el hogar recibe agua por tubería, pero no tuvo servicio durante 15 días o más en el mes anterior a la encuesta.
		Energía eléctrica	5%	El hogar se considera privado cuando la vivienda que habita no está conectada a servicio de energía eléctrica, o cuando sí está conectada a red de distribución de energía eléctrica, pero no contó con el servicio durante más de un día en el mes anterior a la encuesta.
		Recolección de basura	5%	El hogar se considera privado si no tiene acceso a servicio de recolección de basura y la mayor parte de la basura la queman, la entierran, la tiran en cualquier lugar o respondieron "otra" forma de eliminación de la basura.

Continúa tabla VII

		Saneamiento	5%	El hogar se considera privado si no cuenta con saneamiento mejorado, según los siguientes criterios: -Área urbana: la vivienda no está conectada a una red de drenajes o si está conectada a una red de drenaje, pero su servicio sanitario es un excusado lavable, letrina o pozo ciego, o no tiene. -Área rural: la vivienda cuenta con servicio sanitario tipo letrina o pozo ciego, o no tiene.
--	--	-------------	----	---

Fuente: Ministerio de Desarrollo Social, *Índice de Pobreza Multidimensional IPM-Gt*, 2018, p. 17.

- Ponderaciones

Como plantea el documento del IPM-Gt, el criterio de ponderación utilizado es conocido como ponderación anidada. En consecuencia, las cinco dimensiones del IPM-Gt tienen ponderaciones iguales, es decir cada dimensión tiene un peso de 20 por ciento (un quinto), mientras cada indicador dentro de una misma dimensión tiene el mismo peso. Por ejemplo, en el caso de la dimensión de salud y seguridad alimentaria y nutricional cada indicador tiene un peso de 5 por ciento. Este criterio de ponderación fue adoptado debido a que, al asignar un mismo peso a las dimensiones priorizadas, se busca reflejar la misma importancia relativa que éstas tienen en la vida de las personas.

- El método de umbrales duales: privación y pobreza

Según refiere el Ministerio de Desarrollo Social (2018), al utilizar el método Alkire-Foster en la creación de la medida de pobreza multidimensional para Guatemala, se aplicaron dos tipos de umbrales: 1) Umbrales de privación, que determinan si un hogar se encuentra privado en un indicador específico; 2) Umbral de pobreza multidimensional, que identifica a los individuos como

multidimensionalmente pobres de acuerdo al número de privaciones (ponderadas) que un hogar registró, según la ENCOVI 2014.

Como lo hace notar el informe del IPM-Gt, el umbral de privación establece el valor de referencia para determinar si una persona enfrenta carencia o no en un determinado indicador. Por ejemplo, en el indicador de trabajo infantil, el hogar se considera privado cuando existe al menos un niño de 15 años o menos que trabaja, o cuando existe al menos un adolescente de 16 a 17 años que trabaja, no estudia y no tiene al menos 9 años de educación completos. El detalle de umbrales de privación aplicados en el IPM-Gt se aprecia en la matriz de dimensiones, indicadores y umbrales presentados en la Tabla VII.

Por otro lado, la metodología del IPM-Gt refiere que el umbral de pobreza multidimensional establece la cantidad mínima de privaciones que una persona debe tener para ser considerada como multidimensionalmente pobre. En el caso del IPM-Gt se definió un umbral de pobreza multidimensional igual al 30 % que equivale a tener privaciones en cinco de los 17 indicadores incluidos en el IPM-Gt (es decir, un número de indicadores mayor a una dimensión). En consecuencia, todos los hogares cuyo conteo ponderado de privaciones sea igual o mayor al 30 %, serán considerados multidimensionalmente pobres.

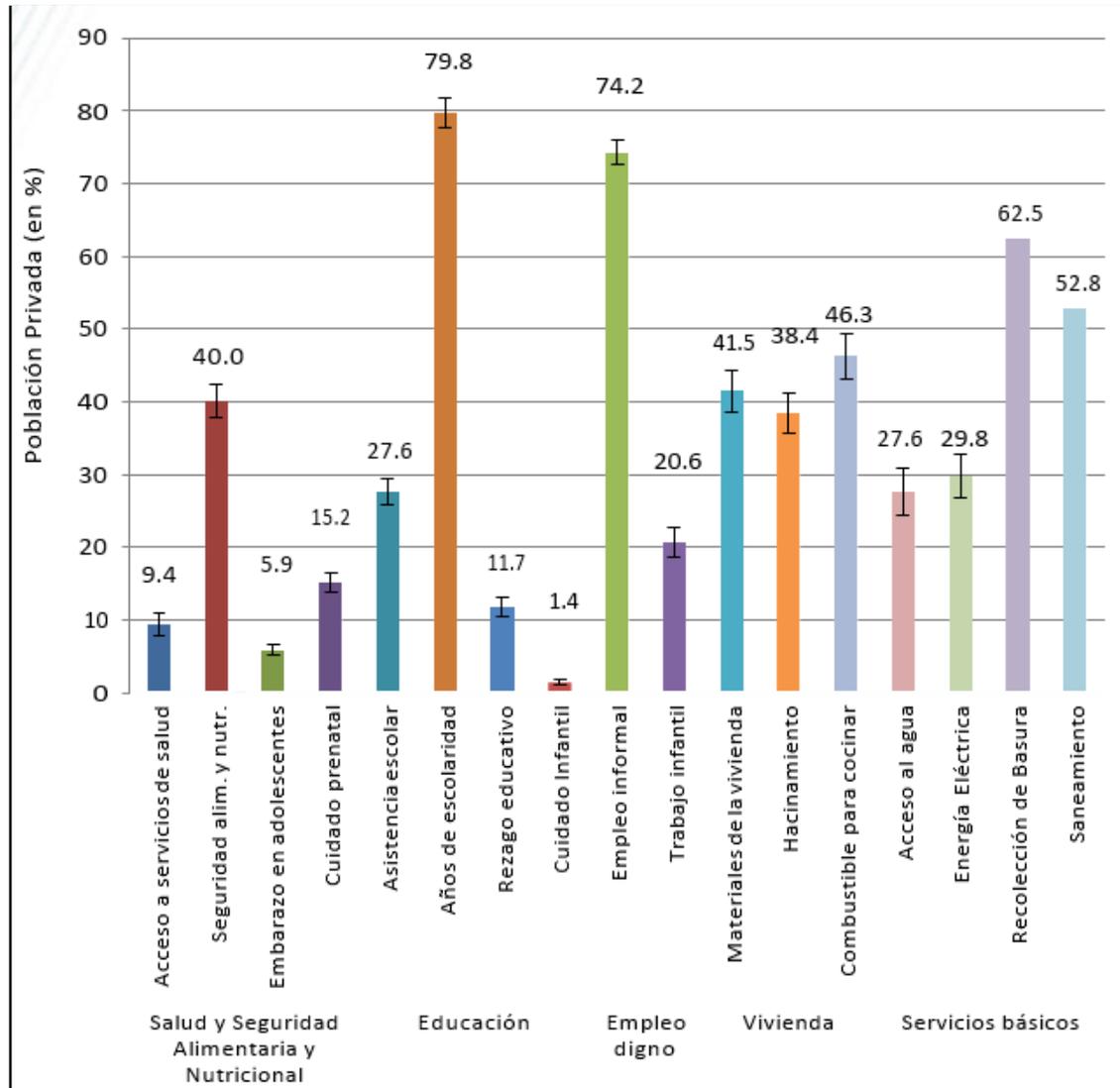
2.2.3.3. Tasas de privaciones no censuradas de los indicadores del IPM-Gt

De acuerdo con el Ministerio de Desarrollo Social (2018) “Las tasas de privación no censuradas se definen como el porcentaje de personas privadas en el respectivo indicador sobre el total de la población” (p. 21). Para su estimación, no se considera aún la clasificación de hogares por su nivel de pobreza multidimensional.

Para el caso de Guatemala, según los resultados del IPM-Gt ocho de cada diez personas pertenecen a hogares privados en el indicador de años de escolaridad, tal como fue definido en la matriz de dimensiones, indicadores y umbrales presentada anteriormente.

El informe del IPM-Gt refiere que este indicador registra la tasa de privación no censurada más alta entre los 17 indicadores. La siguiente gráfica muestra los resultados estimados a nivel nacional.

Figura 2. Tasas de privación no censuradas por indicador al 2014



Fuente: Ministerio de Desarrollo Social. (2018). Índice de Pobreza Multidimensional de Guatemala IPM-Gt, p. 23.

3. PRESENTACIÓN DE RESULTADOS

De acuerdo con los objetivos propuestos, se presentan los resultados obtenidos con relación a los niveles de pobreza multidimensional de Guatemala.

3.1. Objetivo 1: Tasas de privación no censuradas de los indicadores del IPM-Gt a nivel de departamento

A continuación, se exponen las estimaciones de los 17 indicadores priorizados por el IPM-Gt a nivel nacional y departamental.

3.1.1. Acceso a servicios de salud

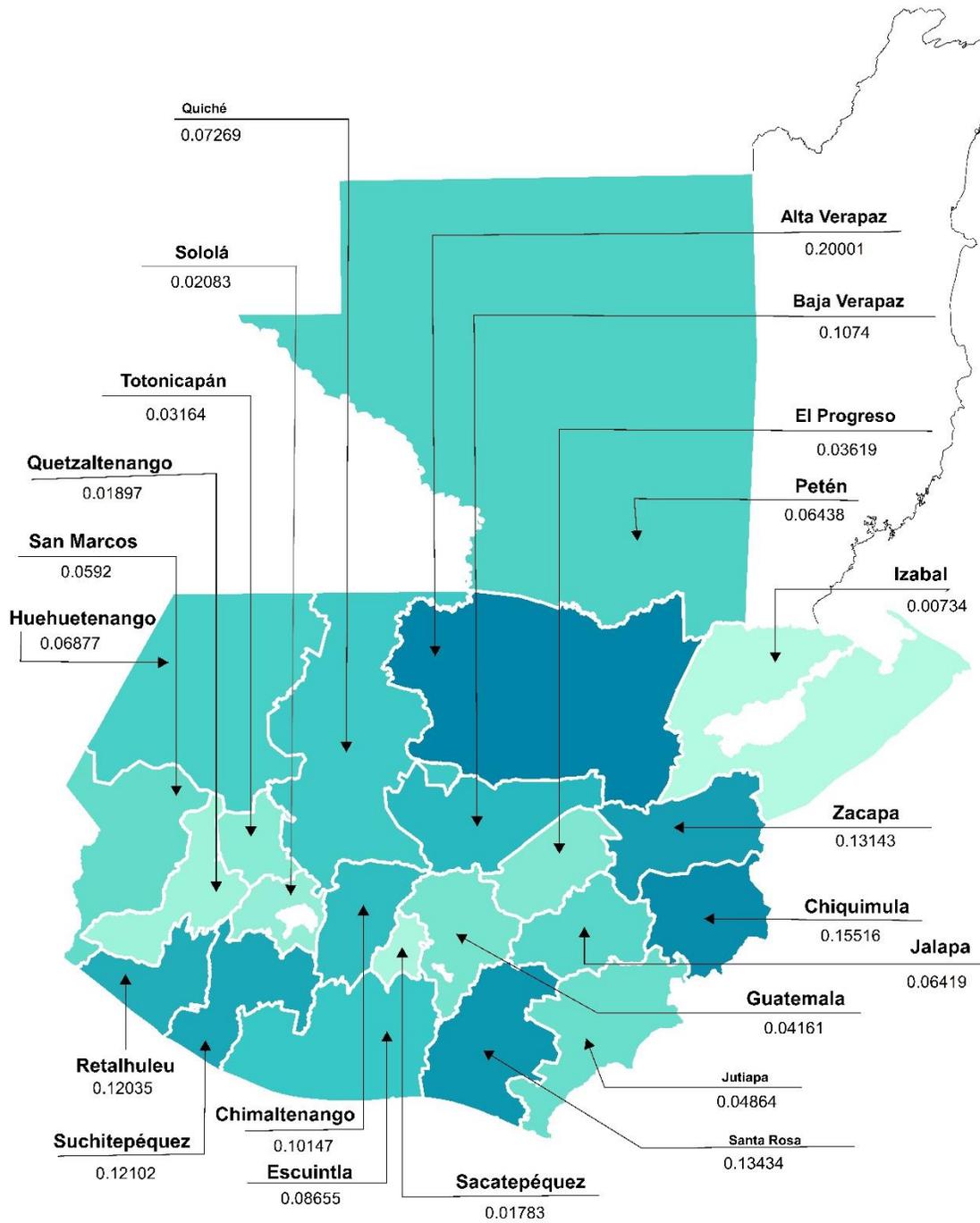
A nivel nacional, el 8.1 % de los guatemaltecos viven en hogares privados en acceso a servicios de salud, con un intervalo de confianza al 95 % de [6.5 %, 9.6 %]. A nivel departamental se observaron los siguientes resultados:

Tabla VIII. Acceso a servicios de salud, según ENCOVI 2014

Departamento	Estimador directo	EE	CV	EBLUP Fay Herriot	ECM	ECMR
Guatemala	0.0426	0.0158232	0.3713	0.0416	0.0002324	0.3664
El Progreso	0.0356	0.0089436	0.2513	0.0362	0.0000779	0.2439
Sacatepéquez	0.0177	0.0054427	0.3079	0.0178	0.0000294	0.3039
Chimaltenango	0.1073	0.0202408	0.1887	0.1015	0.0003559	0.1859
Escuintla	0.0812	0.0178337	0.2197	0.0866	0.0002901	0.1968
Santa Rosa	0.1477	0.0212849	0.1441	0.1343	0.0003878	0.1466
Sololá	0.0207	0.0068866	0.3330	0.0208	0.0000468	0.3283
Totonicapán	0.0276	0.0125578	0.4544	0.0316	0.0001494	0.3863
Quetzaltenango	0.0183	0.0059029	0.3228	0.0190	0.0000344	0.3092
Suchitepéquez	0.1286	0.0208692	0.1622	0.1210	0.0003755	0.1601
Retalhuleu	0.1164	0.0256254	0.2201	0.1203	0.0005516	0.1952
San Marcos	0.0564	0.0169095	0.2999	0.0592	0.0002591	0.2719
Huehuetenango	0.0648	0.0226500	0.3493	0.0688	0.0004309	0.3018
Quiché	0.0707	0.0179289	0.2536	0.0727	0.0002880	0.2335
Baja Verapaz	0.1294	0.0350263	0.2706	0.1074	0.0008250	0.2674
Alta Verapaz	0.2723	0.0546048	0.2006	0.2000	0.0015875	0.1992
Petén	0.0607	0.0133646	0.2202	0.0644	0.0001688	0.2018
Izabal	0.0067	0.0041812	0.6268	0.0073	0.0000174	0.5681
Zacapa	0.1790	0.0351931	0.1966	0.1314	0.0008382	0.2203
Chiquimula	0.1834	0.0307375	0.1676	0.1552	0.0006957	0.1700
Jalapa	0.0631	0.0206840	0.3280	0.0642	0.0003695	0.2995
Jutiapa	0.0480	0.0096490	0.2009	0.0486	0.0000902	0.1953

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 3. Acceso a servicios de salud, estimador EBLUP bajo el modelo *Fay Herriot*



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.2. Seguridad alimentaria y nutricional

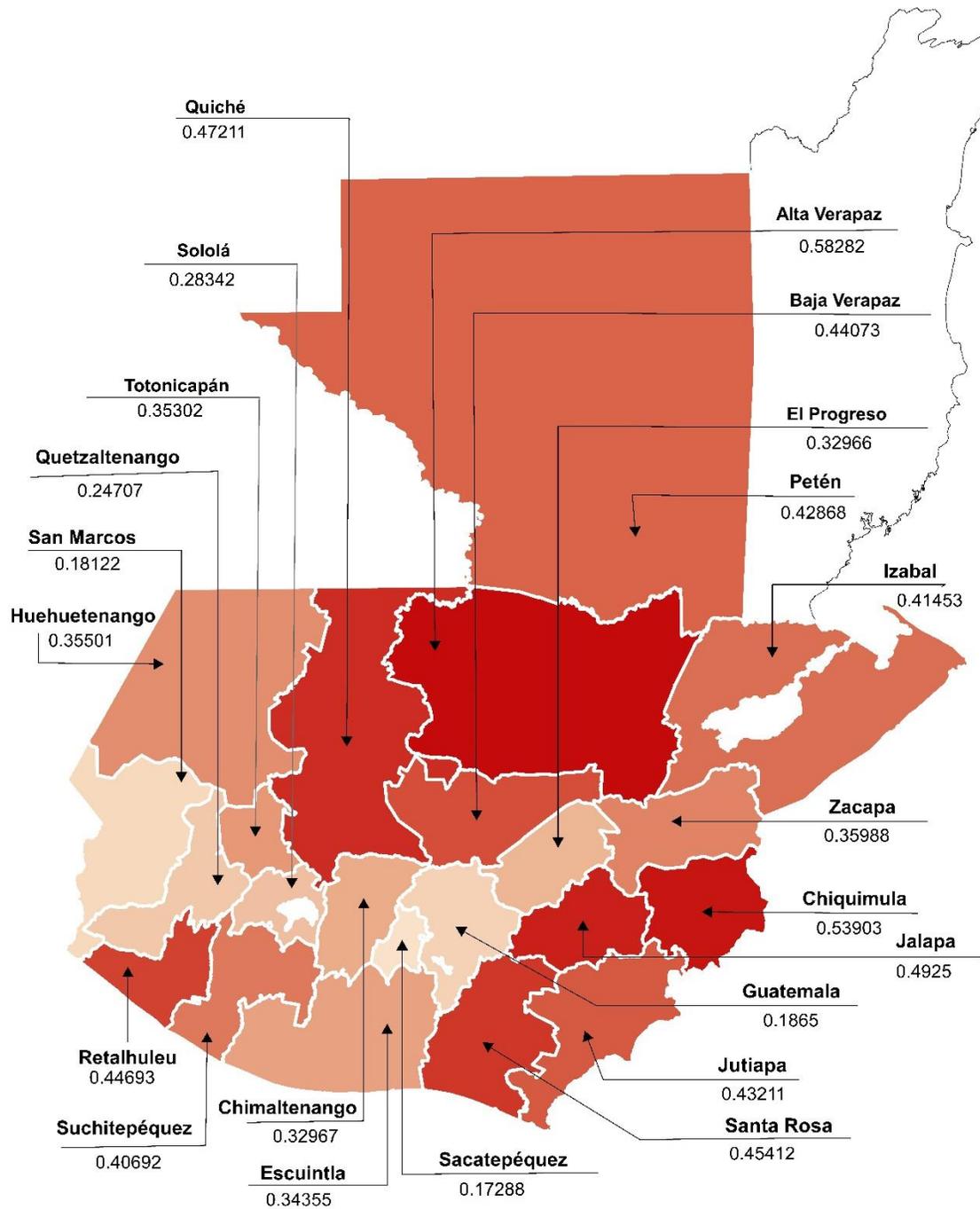
Para el caso del indicador de seguridad alimentaria y nutricional, a nivel nacional se estima que el 34.0 % de los guatemaltecos viven en hogares que se encuentran en los niveles de inseguridad alimentaria y nutricional moderada o severa definidos por la *Food Insecurity Experience Scale* (Escala Latinoamericana y Caribeña de Seguridad Alimentaria -ELCSA). El intervalo de estimación para el indicador es de [31.8 %, 36.2 %] con un 95 % de confianza. A nivel de departamento se obtuvieron las siguientes estimaciones:

Tabla IX. Seguridad alimentaria y nutricional, según ENCOVI 2014

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.1758	0.0277693	0.1579	0.1865	0.0007104	0.1429
El Progreso	0.3423	0.0428079	0.1250	0.3297	0.0014813	0.1167
Sacatepéquez	0.1639	0.0191948	0.1171	0.1729	0.0003567	0.1092
Chimaltenango	0.3291	0.0289396	0.0879	0.3297	0.0007581	0.0835
Escuintla	0.3416	0.0257809	0.0755	0.3436	0.0006157	0.0722
Santa Rosa	0.4858	0.0394225	0.0811	0.4541	0.0012874	0.0790
Sololá	0.2824	0.0434233	0.1538	0.2834	0.0015473	0.1388
Totonicapán	0.3370	0.0406665	0.1207	0.3530	0.0013558	0.1043
Quetzaltenango	0.2427	0.0392243	0.1616	0.2471	0.0012994	0.1459
Suchitepéquez	0.4263	0.0396328	0.0930	0.4069	0.0013034	0.0887
Retalhuleu	0.4549	0.0377128	0.0829	0.4469	0.0012003	0.0775
San Marcos	0.1609	0.0316460	0.1967	0.1812	0.0009017	0.1657
Huehuetenango	0.3507	0.0490499	0.1399	0.3550	0.0019054	0.1230
Quiché	0.5000	0.0463221	0.0926	0.4721	0.0016595	0.0863
Baja Verapaz	0.4643	0.0517121	0.1114	0.4407	0.0019499	0.1002
Alta Verapaz	0.5449	0.0532344	0.0977	0.5828	0.0025261	0.0862
Petén	0.4514	0.0397710	0.0881	0.4287	0.0013105	0.0844
Izabal	0.4476	0.1117906	0.2497	0.4145	0.0044932	0.1617
Zacapa	0.3602	0.0346842	0.0963	0.3599	0.0010422	0.0897
Chiquimula	0.5322	0.0576990	0.1084	0.5390	0.0025075	0.0929
Jalapa	0.4901	0.0450405	0.0919	0.4925	0.0017999	0.0861
Jutiapa	0.4427	0.0339490	0.0767	0.4321	0.0010015	0.0732

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 4. Seguridad alimentaria y nutricional, estimador EBLUP bajo el modelo *Fay Herriot*



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.3. Embarazo en adolescentes

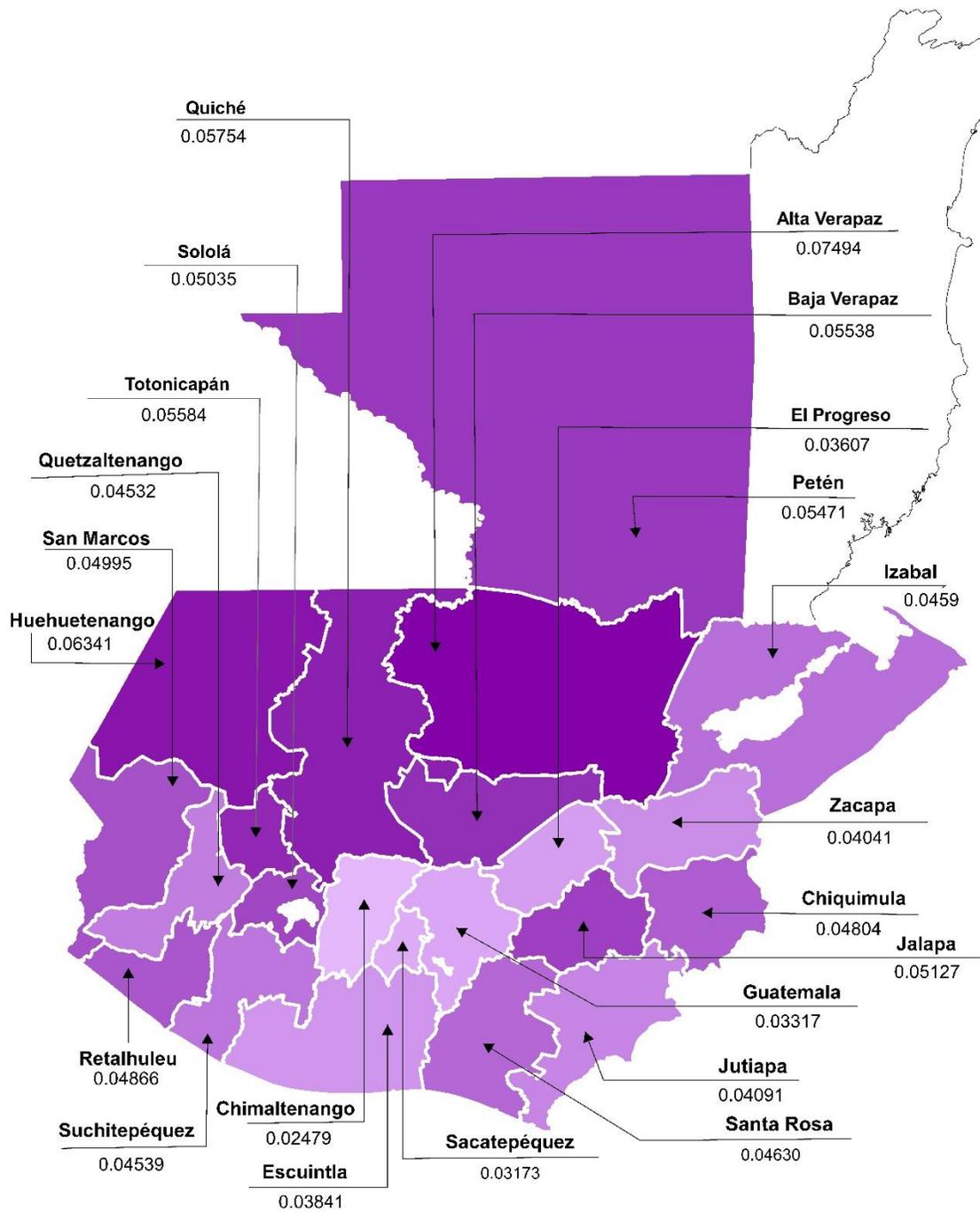
De acuerdo con la ENCOVI 2014, en Guatemala el 5.2 % de la población viven en hogares en donde hay alguna niña o adolescente menor de 20 años embarazada o que ha estado embarazada, lo que hace que se consideren privados en este indicador. El intervalo al 95 % de confianza obtenido fue de [4.4 %, 5.9 %]. La siguiente tabla detalla los resultados a nivel departamental:

Tabla X. **Embarazo en adolescentes, según ENCOVI 2014**

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.0341	0.0085830	0.2514	0.0332	0.0000598	0.2332
El Progreso	0.0327	0.0113708	0.3473	0.0361	0.0000740	0.2386
Sacatepéquez	0.0311	0.0070031	0.2255	0.0317	0.0000450	0.2113
Chimaltenango	0.0171	0.0068061	0.3970	0.0248	0.0000414	0.2595
Escuintla	0.0399	0.0114372	0.2866	0.0384	0.0000749	0.2253
Santa Rosa	0.0754	0.0182784	0.2425	0.0463	0.0000922	0.2073
Sololá	0.0444	0.0130326	0.2937	0.0504	0.0000819	0.1798
Totonicapán	0.0582	0.0148709	0.2557	0.0558	0.0000895	0.1694
Quetzaltenango	0.0510	0.0126462	0.2480	0.0453	0.0000775	0.1943
Suchitepéquez	0.0467	0.0158277	0.3386	0.0454	0.0000855	0.2037
Retalhuleu	0.0680	0.0163602	0.2406	0.0487	0.0000870	0.1916
San Marcos	0.0585	0.0182876	0.3126	0.0500	0.0000896	0.1895
Huehuetenango	0.0863	0.0196100	0.2273	0.0634	0.0001046	0.1613
Quiché	0.0424	0.0114117	0.2690	0.0575	0.0000937	0.1682
Baja Verapaz	0.0728	0.0157344	0.2161	0.0554	0.0000865	0.1680
Alta Verapaz	0.1033	0.0215158	0.2083	0.0749	0.0001441	0.1602
Petén	0.0596	0.0127923	0.2145	0.0547	0.0000801	0.1636
Izabal	0.0390	0.0142771	0.3665	0.0459	0.0000828	0.1983
Zacapa	0.0343	0.0154073	0.4486	0.0404	0.0000852	0.2284
Chiquimula	0.0383	0.0120003	0.3135	0.0480	0.0000780	0.1839
Jalapa	0.0558	0.0121871	0.2184	0.0513	0.0000759	0.1700
Jutiapa	0.0374	0.0109246	0.2919	0.0409	0.0000698	0.2043

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 5. Embarazo en adolescentes, estimador EBLUP bajo el modelo *Fay Herriot*



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.4. Cuidado prenatal

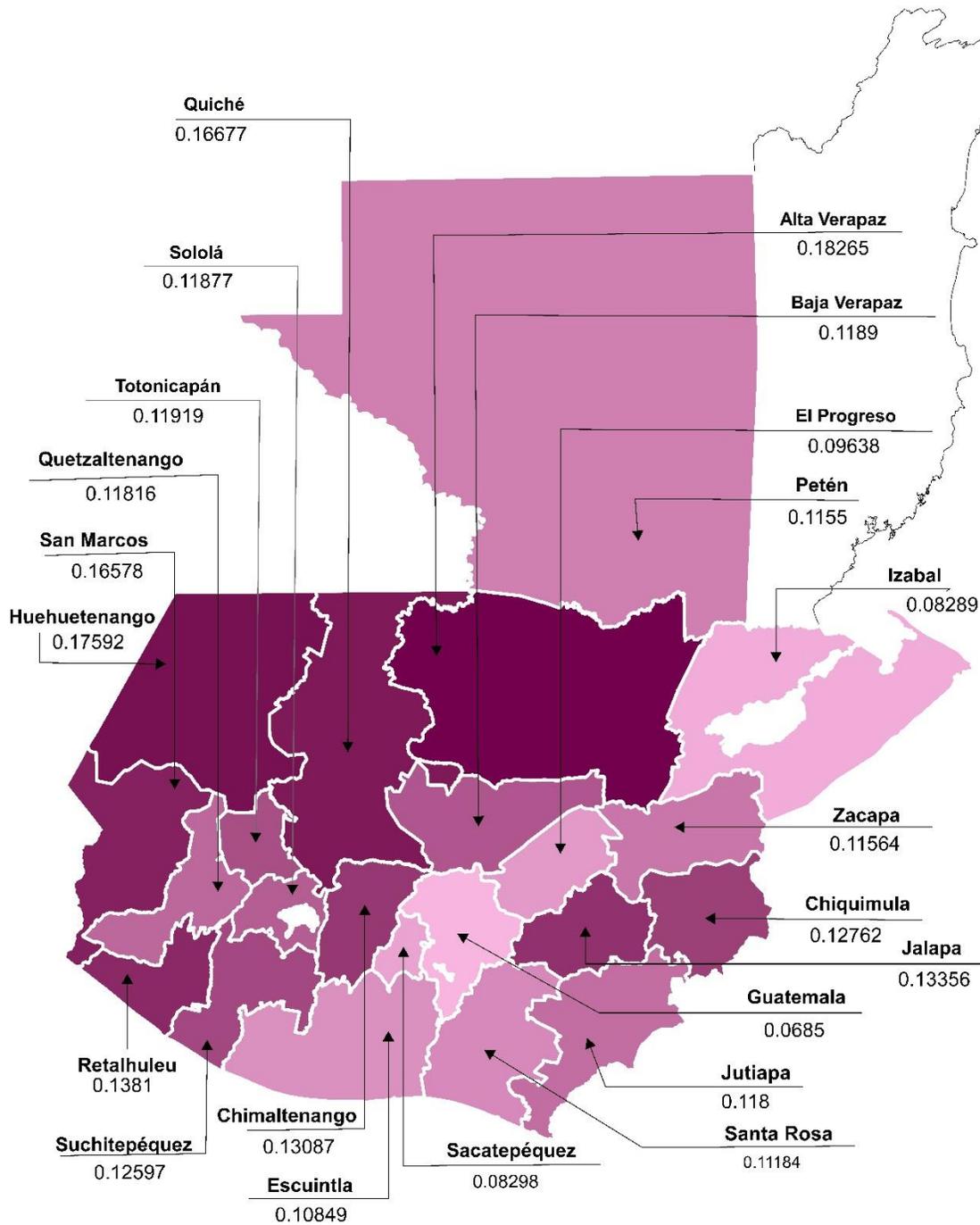
Con relación al cuidado prenatal, se estima que a nivel nacional el 12.9 % de la población viven en hogares en donde hay al menos una mujer de 12 a 49 años que ha estado embarazada en los cinco años anteriores a la realización de la encuesta, que en su último embarazo no asistió a por lo menos cuatro controles de embarazo. El intervalo de confianza calculado al 95 % para este indicador fue de [11.6 %, 14.1 %]. En la tabla siguiente se muestran los resultados obtenidos a nivel departamental:

Tabla XI. Cuidado prenatal, según ENCOVI 2014

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.0667	0.0157780	0.2365	0.0685	0.0002178	0.2154
El Progreso	0.0939	0.0183448	0.1953	0.0964	0.0002486	0.1636
Sacatepéquez	0.0839	0.0163834	0.1953	0.0830	0.0002233	0.1801
Chimaltenango	0.1381	0.0203715	0.1475	0.1309	0.0002764	0.1270
Escuintla	0.1082	0.0222505	0.2056	0.1085	0.0003073	0.1616
Santa Rosa	0.1124	0.0217269	0.1933	0.1118	0.0002987	0.1545
Sololá	0.1372	0.0292764	0.2134	0.1188	0.0003916	0.1666
Totonicapán	0.1139	0.0212164	0.1862	0.1192	0.0002896	0.1428
Quetzaltenango	0.1346	0.0261038	0.1939	0.1182	0.0003594	0.1604
Suchitepéquez	0.1244	0.0180885	0.1454	0.1260	0.0002394	0.1228
Retalhuleu	0.1551	0.0238507	0.1538	0.1381	0.0003265	0.1308
San Marcos	0.2066	0.0336980	0.1631	0.1658	0.0004485	0.1277
Huehuetenango	0.1937	0.0302428	0.1562	0.1759	0.0004597	0.1219
Quiché	0.1811	0.0285193	0.1574	0.1668	0.0004188	0.1227
Baja Verapaz	0.1029	0.0203255	0.1975	0.1189	0.0002802	0.1408
Alta Verapaz	0.2032	0.0236927	0.1166	0.1827	0.0003640	0.1045
Petén	0.0919	0.0198331	0.2158	0.1155	0.0002765	0.1440
Izabal	0.0513	0.0194055	0.3782	0.0829	0.0002611	0.1949
Zacapa	0.1327	0.0280788	0.2117	0.1156	0.0003824	0.1691
Chiquimula	0.1076	0.0300534	0.2792	0.1276	0.0004031	0.1573
Jalapa	0.1260	0.0223812	0.1776	0.1336	0.0003140	0.1327
Jutiapa	0.1267	0.0196287	0.1549	0.1180	0.0002673	0.1386

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 6. **Cuidado prenatal, estimador EBLUP bajo el modelo *Fay Herriot***



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.5. Asistencia escolar

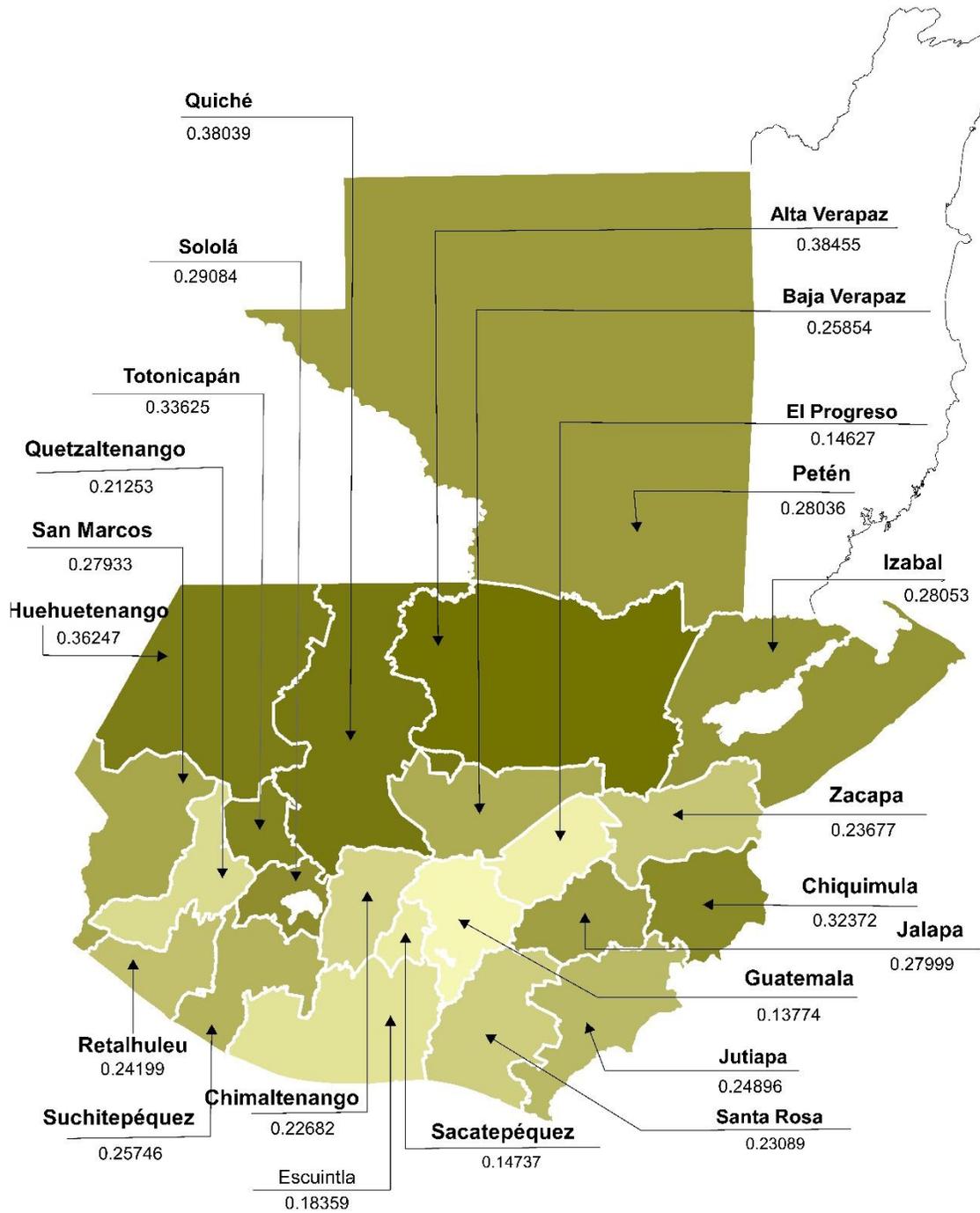
En lo que respecta a la asistencia escolar, de acuerdo con la última encuesta de condiciones de vida en el país (ENCOVI 2014) el 25.9 % de los guatemaltecos viven en hogares privados de este indicador, es decir, hogares en donde hay al menos un niño o adolescente de 7 a 15 años que no asiste a la escuela, o hay al menos un adolescente de 16 a 17 años que no asiste a la escuela y no ha completado 9 años de estudios. El intervalo de estimación obtenido al 95 % de confianza es de [24.1 %, 27.7 %]. En la tabla siguiente se presentan las estimaciones a nivel de departamento:

Tabla XII. **Asistencia escolar, según ENCOVI 2014**

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.1292	0.0226983	0.1757	0.1377	0.0004330	0.1511
El Progreso	0.1258	0.0206121	0.1638	0.1463	0.0003654	0.1307
Sacatepéquez	0.1385	0.0199431	0.1440	0.1474	0.0003578	0.1284
Chimaltenango	0.2449	0.0291072	0.1189	0.2268	0.0005796	0.1061
Escuintla	0.1774	0.0200219	0.1128	0.1836	0.0003480	0.1016
Santa Rosa	0.3365	0.0591132	0.1757	0.2309	0.0009837	0.1358
Sololá	0.2896	0.0280039	0.0967	0.2908	0.0005994	0.0842
Totonicapán	0.3779	0.0393930	0.1043	0.3363	0.0007913	0.0837
Quetzaltenango	0.2019	0.0305338	0.1512	0.2125	0.0006037	0.1156
Suchitepéquez	0.2625	0.0259307	0.0988	0.2575	0.0005009	0.0869
Retalhuleu	0.2567	0.0337855	0.1316	0.2420	0.0006586	0.1061
San Marcos	0.2996	0.0466523	0.1557	0.2793	0.0008837	0.1064
Huehuetenango	0.3945	0.0415913	0.1054	0.3625	0.0009133	0.0834
Quiché	0.3493	0.0343744	0.0984	0.3804	0.0008921	0.0785
Baja Verapaz	0.2414	0.0398902	0.1652	0.2585	0.0008077	0.1099
Alta Verapaz	0.3481	0.0387447	0.1113	0.3846	0.0009736	0.0811
Petén	0.2691	0.0302922	0.1126	0.2804	0.0005981	0.0872
Izabal	0.3364	0.0789499	0.2347	0.2805	0.0010101	0.1133
Zacapa	0.2385	0.0357240	0.1498	0.2368	0.0007363	0.1146
Chiquimula	0.3760	0.0522674	0.1390	0.3237	0.0009184	0.0936
Jalapa	0.2878	0.0305807	0.1063	0.2800	0.0005961	0.0872
Jutiapa	0.2739	0.0318179	0.1162	0.2490	0.0006922	0.1057

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 7. **Asistencia escolar, estimador EBLUP bajo el modelo *Fay Herriot***



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.6. Años de escolaridad

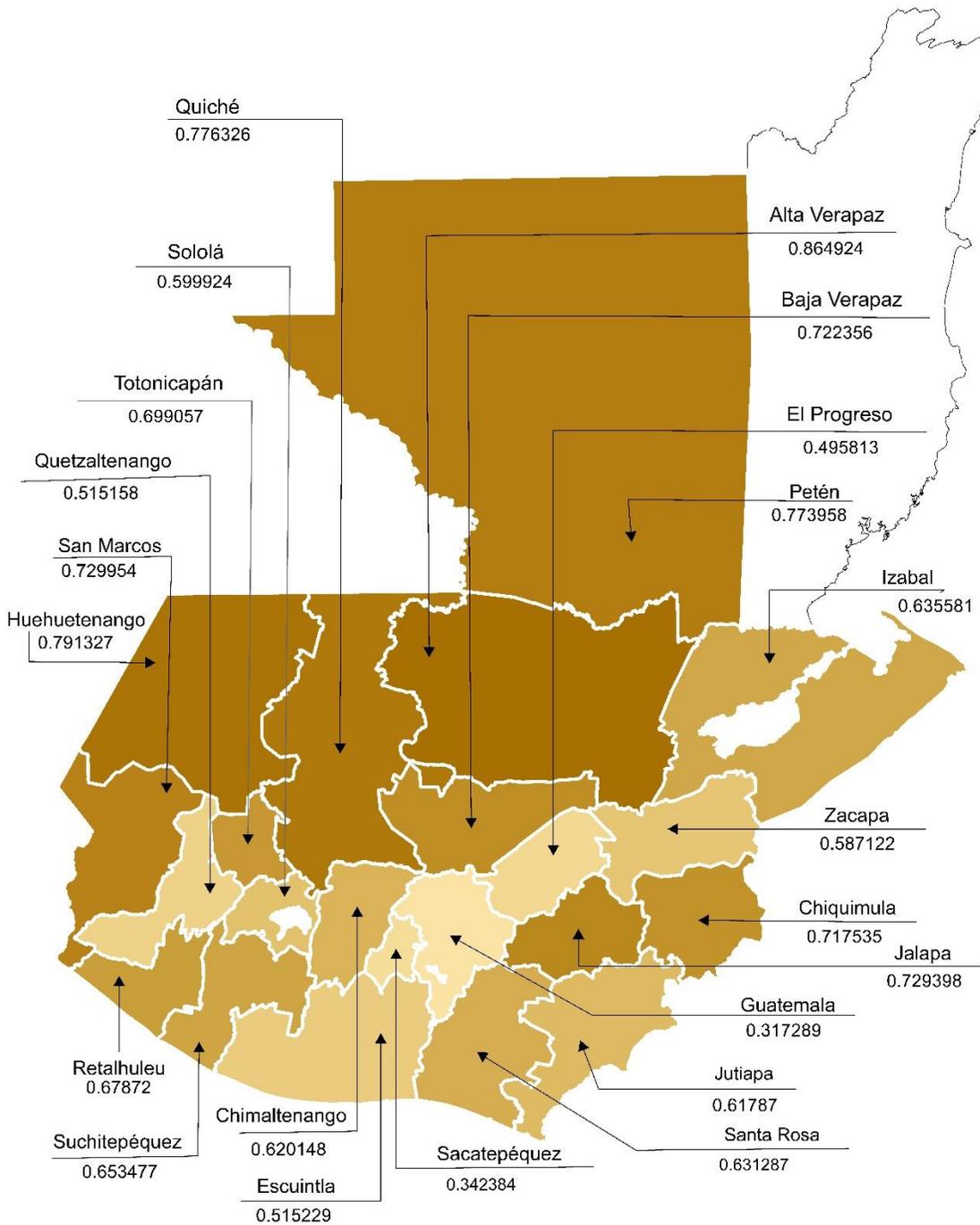
El nivel de escolaridad de hombres y mujeres es uno de los principales determinantes en el desarrollo de un país. A nivel nacional, se estima que el 59.5 % de los guatemaltecos viven en hogares en donde hay al menos un adulto (18 años o más) con menos de los años de escolaridad según las edades siguientes: 18 a 32 años, 9 años de escolaridad; 33 a 64 años, 6 años de escolaridad y mayores de 65 años que no saben leer y escribir. El intervalo de estimación obtenido al 95 % de confianza es de [56.7 %, 62.2 %]. La siguiente tabla detalla las estimaciones departamentales:

Tabla XIII. Años de escolaridad, según ENCOVI 2014

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.2913	0.0406852	0.1397	0.3173	0.0013583	0.1162
El Progreso	0.4797	0.0547443	0.1141	0.4958	0.0018370	0.0864
Sacatepéquez	0.3160	0.0338457	0.1071	0.3424	0.0009885	0.0918
Chimaltenango	0.6157	0.0403162	0.0655	0.6201	0.0012245	0.0564
Escuintla	0.4973	0.0359912	0.0724	0.5152	0.0010437	0.0627
Santa Rosa	0.6614	0.0399277	0.0604	0.6313	0.0012113	0.0551
Sololá	0.6347	0.0494166	0.0779	0.5999	0.0016098	0.0669
Totonicapán	0.7434	0.0502839	0.0676	0.6991	0.0016275	0.0577
Quetzaltenango	0.4807	0.0606420	0.1262	0.5152	0.0020141	0.0871
Suchitepéquez	0.6499	0.0376293	0.0579	0.6535	0.0011071	0.0509
Retalhuleu	0.7025	0.0411011	0.0585	0.6787	0.0012560	0.0522
San Marcos	0.6960	0.0567350	0.0815	0.7300	0.0019209	0.0600
Huehuetenango	0.7298	0.0535715	0.0734	0.7913	0.0019175	0.0553
Quiché	0.7457	0.0494599	0.0663	0.7763	0.0016891	0.0529
Baja Verapaz	0.7267	0.0375611	0.0517	0.7224	0.0011120	0.0462
Alta Verapaz	0.8757	0.0465330	0.0531	0.8649	0.0015943	0.0462
Petén	0.7821	0.0330892	0.0423	0.7740	0.0009208	0.0392
Izabal	0.6130	0.1032212	0.1684	0.6356	0.0027982	0.0832
Zacapa	0.6283	0.0516139	0.0821	0.5871	0.0017027	0.0703
Chiquimula	0.7214	0.0693024	0.0961	0.7175	0.0022599	0.0663
Jalapa	0.7306	0.0467193	0.0639	0.7294	0.0015104	0.0533
Jutiapa	0.6422	0.0332250	0.0517	0.6179	0.0009173	0.0490

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 8. **Años de escolaridad, estimador EBLUP bajo el modelo *Fay Herriot***



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.7. Rezago educativo

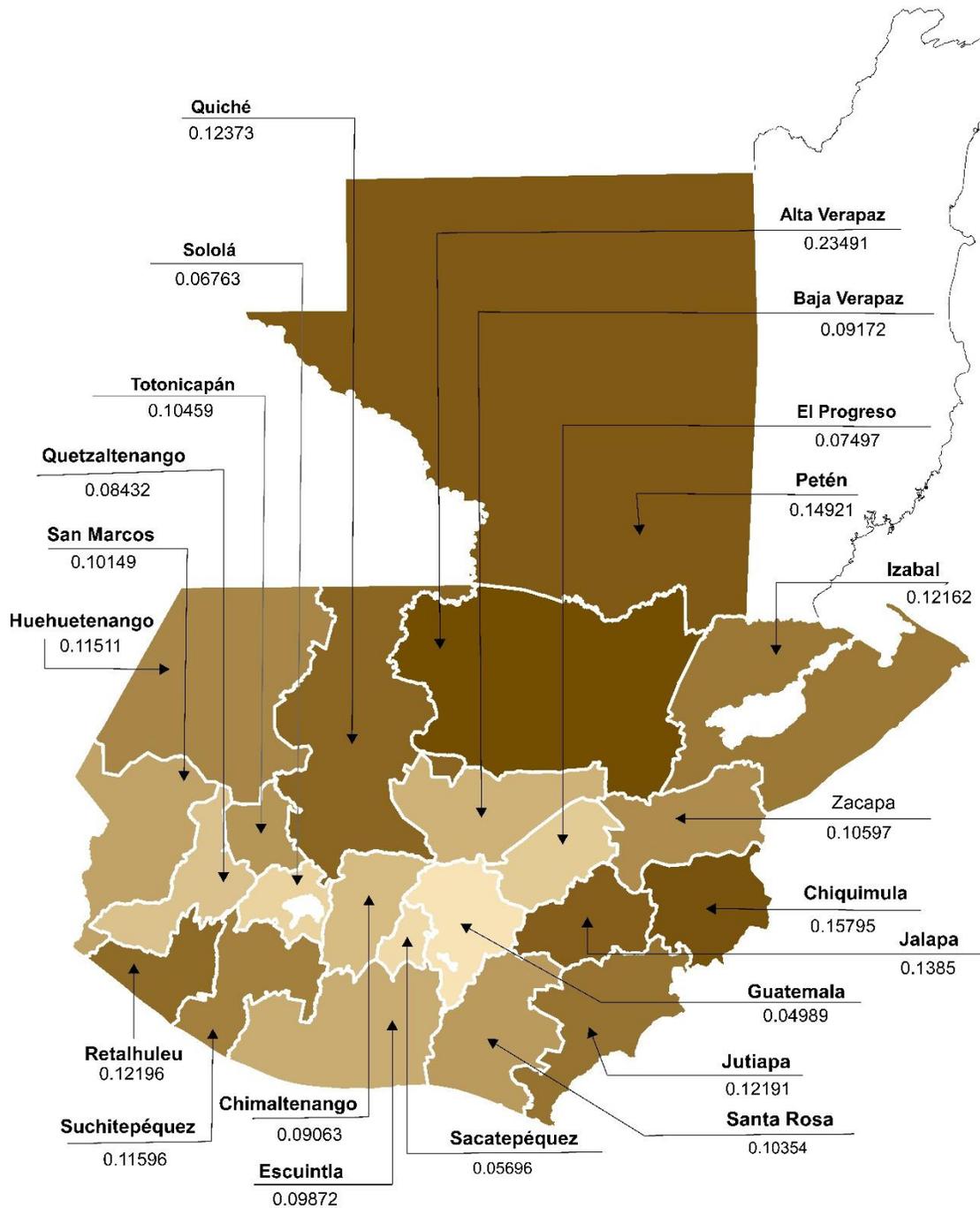
En referencia al indicador de rezago educativo, se estima que en el país el 11.0 % de los guatemaltecos residen en hogares en donde hay al menos un niño o adolescente de 8 a menos de 18 años que presenta más de dos años de rezago escolar, estimado según su edad y el grado en el que se inscribió. El intervalo de confianza al 95 % para este indicador fue de [9.8 %, 12.3 %]. Los resultados obtenidos a nivel de departamento se muestran a continuación:

Tabla XIV. **Rezago educativo, según ENCOVI 2014**

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.0451	0.0100632	0.2229	0.0499	0.0000941	0.1945
El Progreso	0.0691	0.0234537	0.3392	0.0750	0.0003280	0.2416
Sacatepéquez	0.0552	0.0111376	0.2017	0.0570	0.0001142	0.1876
Chimaltenango	0.0847	0.0181744	0.2146	0.0906	0.0002382	0.1703
Escuintla	0.0912	0.0157020	0.1721	0.0987	0.0001951	0.1415
Santa Rosa	0.1128	0.0268171	0.2377	0.1035	0.0003637	0.1842
Sololá	0.0692	0.0162503	0.2347	0.0676	0.0002131	0.2159
Totonicapán	0.0958	0.0192920	0.2013	0.1046	0.0002586	0.1538
Quetzaltenango	0.0900	0.0172918	0.1920	0.0843	0.0002287	0.1794
Suchitepéquez	0.1406	0.0228035	0.1622	0.1160	0.0003136	0.1527
Retalhuleu	0.1170	0.0225998	0.1932	0.1220	0.0003153	0.1456
San Marcos	0.1098	0.0234428	0.2135	0.1015	0.0003220	0.1768
Huehuetenango	0.1045	0.0196385	0.1880	0.1151	0.0002690	0.1425
Quiché	0.1252	0.0229419	0.1832	0.1237	0.0003177	0.1441
Baja Verapaz	0.0776	0.0156902	0.2021	0.0917	0.0001957	0.1525
Alta Verapaz	0.2580	0.0403507	0.1564	0.2349	0.0009132	0.1286
Petén	0.2103	0.0284089	0.1351	0.1492	0.0003807	0.1308
Izabal	0.1004	0.0268086	0.2670	0.1216	0.0003815	0.1606
Zacapa	0.1227	0.0309883	0.2526	0.1060	0.0004058	0.1901
Chiquimula	0.1631	0.0412354	0.2527	0.1580	0.0005528	0.1488
Jalapa	0.1625	0.0263155	0.1619	0.1385	0.0003615	0.1373
Jutiapa	0.1295	0.0199902	0.1544	0.1219	0.0002694	0.1346

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 9. **Rezago educativo, estimador EBLUP bajo el modelo *Fay Herriot***



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.8. Cuidado infantil

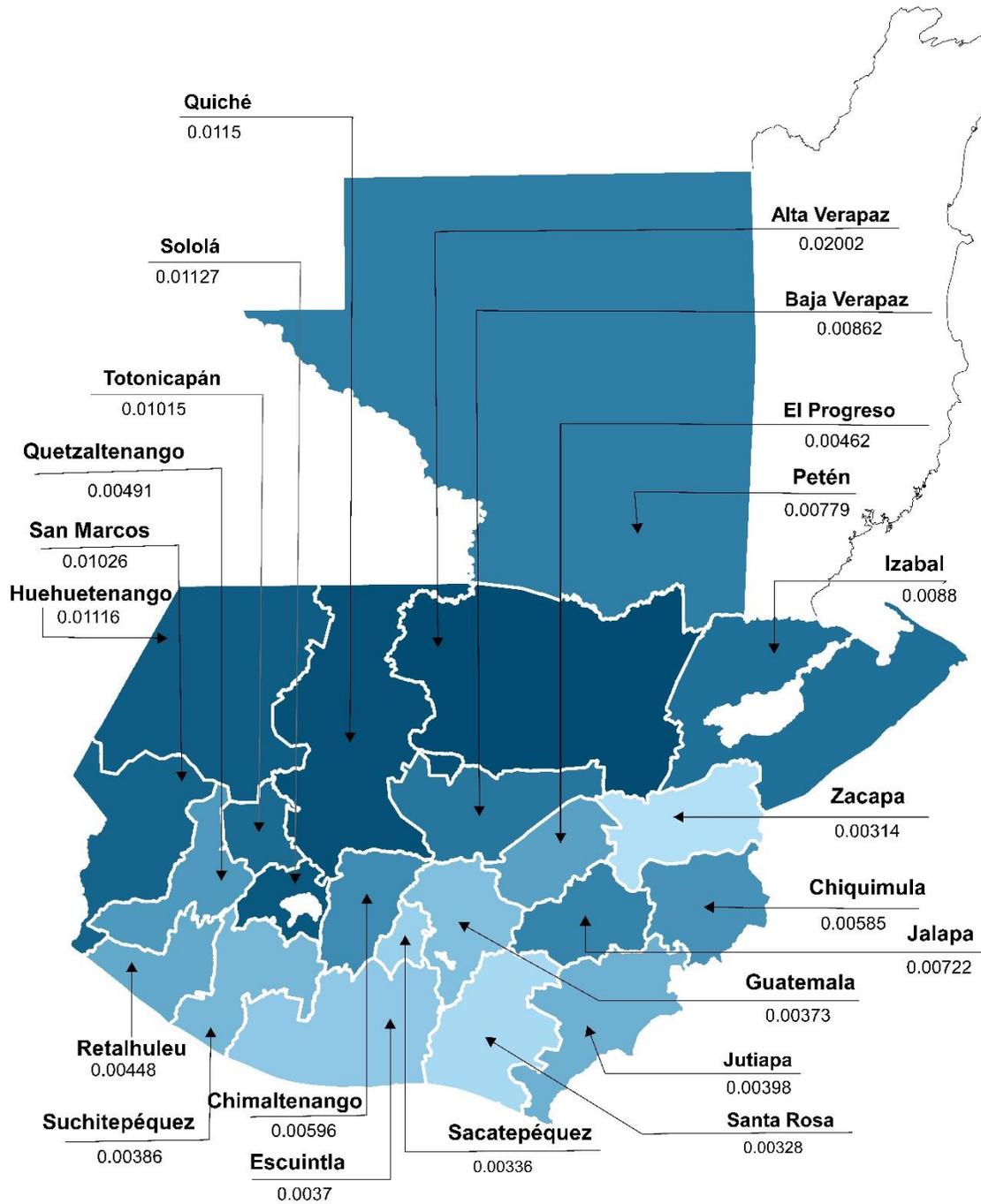
De acuerdo con los resultados obtenidos, el 1.1 % de la población guatemalteca pertenece a hogares en los que hay al menos un niño menor de 7 años que permanece con el padre o la madre en su lugar de trabajo, o con vecinos o amigos, o permaneció la mayor parte del tiempo solo. Con un 95 % de confianza se estima que el intervalo para este indicador fue de [0.7 %, 1.4 %]. Para el caso de los departamentos se obtuvieron las siguientes estimaciones:

Tabla XV. **Cuidado infantil, según ENCOVI 2014**

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.0094	0.0041640	0.4443	0.0037	0.0000004	0.1692
El Progreso	0.0111	0.0064443	0.5802	0.0046	0.0000002	0.0938
Sacatepéquez	0.0021	0.0015499	0.7356	0.0034	0.0000017	0.3932
Chimaltenango	0.0104	0.0050281	0.4830	0.0060	0.0000002	0.0820
Escuintla	0.0057	0.0030734	0.5436	0.0037	0.0000006	0.2084
Santa Rosa	0.0068	0.0049183	0.7222	0.0033	0.0000004	0.1904
Sololá	0.0189	0.0090412	0.4775	0.0113	0.0000007	0.0723
Totonicapán	0.0120	0.0070286	0.5866	0.0102	0.0000005	0.0709
Quetzaltenango	0.0111	0.0042211	0.3815	0.0049	0.0000003	0.1129
Suchitepéquez	0.0041	0.0033634	0.8130	0.0039	0.0000005	0.1842
Retalhuleu	0.0143	0.0082014	0.5733	0.0045	0.0000001	0.0863
San Marcos	0.0066	0.0043396	0.6531	0.0103	0.0000008	0.0852
Huehuetenango	0.0127	0.0056578	0.4463	0.0112	0.0000008	0.0812
Quiché	0.0118	0.0068042	0.5755	0.0115	0.0000008	0.0787
Baja Verapaz	0.0057	0.0031812	0.5610	0.0086	0.0000007	0.0963
Alta Verapaz	0.0167	0.0078083	0.4674	0.0200	0.0000061	0.1231
Petén	0.0131	0.0051859	0.3951	0.0078	0.0000003	0.0727
Izabal	0.0440	0.0182652	0.4152	0.0088	0.0000002	0.0490
Zacapa	0.0037	0.0021688	0.5785	0.0031	0.0000011	0.3373
Chiquimula	0.0025	0.0025057	0.9964	0.0059	0.0000007	0.1380
Jalapa	0.0044	0.0033699	0.7714	0.0072	0.0000005	0.0962
Jutiapa	0.0011	0.0010552	0.9968	0.0040	0.0000030	0.4372

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 10. **Cuidado infantil, estimador EBLUP bajo el modelo *Fay Herriot***



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.9. Empleo informal

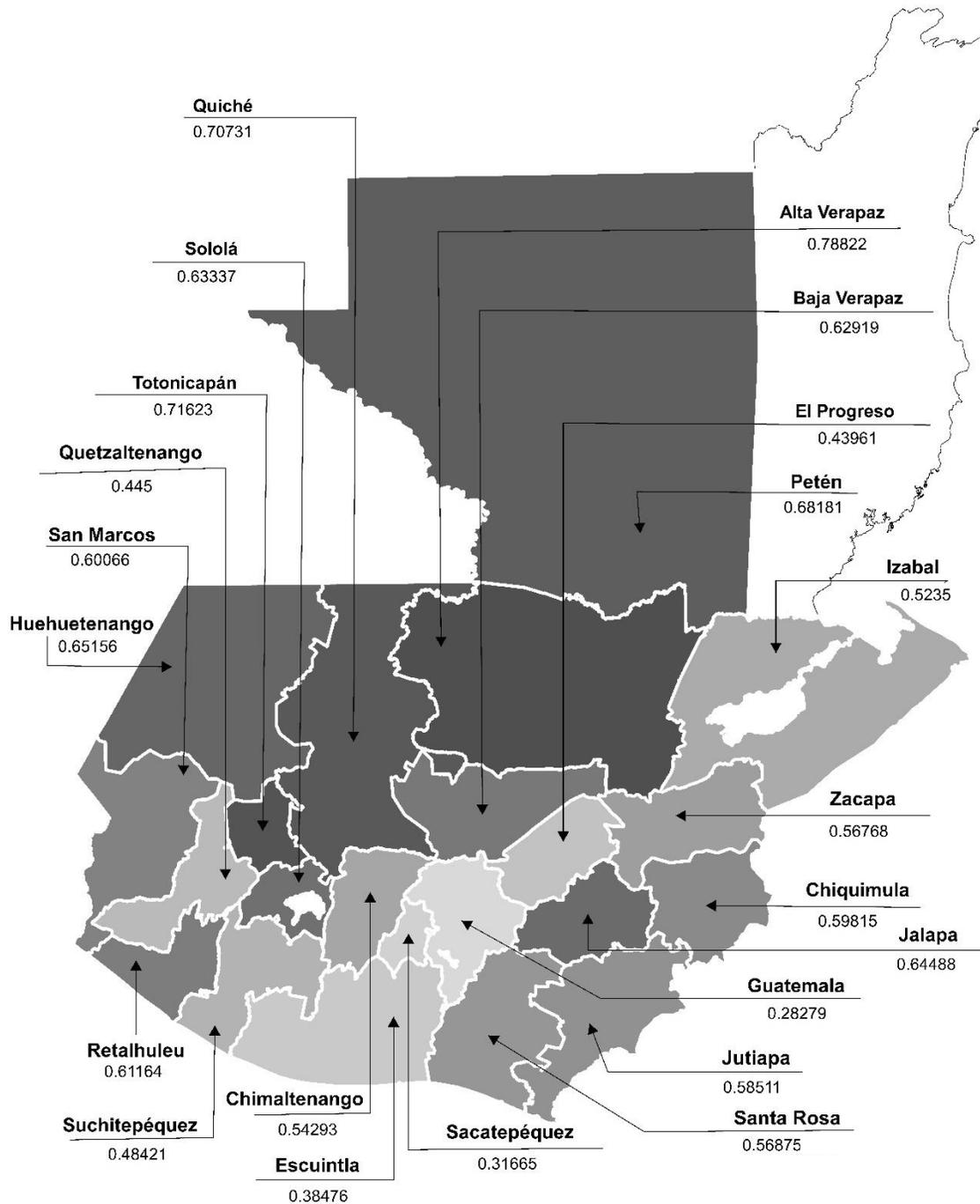
A nivel de la República de Guatemala se estima que el 52.7 % de la población reside en hogares en donde hay al menos una persona ocupada en alguna de las siguientes categorías: empleadores, empleados y obreros de empresas de menos de 6 personas; todos los trabajadores por cuenta propia o autónoma (excluyendo profesionales y técnicos, medido como 12 años de escolaridad o más), todos los familiares no remunerados y, ocupados en servicio doméstico. El intervalo al 95 % de confianza obtenido fue de [50.1 %, 55.3 %]. La siguiente tabla detalla los resultados a nivel departamental:

Tabla XVI. **Empleo informal, según ENCOVI 2014**

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.2673	0.0353533	0.1323	0.2828	0.0011245	0.1186
El Progreso	0.4334	0.0458755	0.1059	0.4396	0.0016780	0.0932
Sacatepéquez	0.2994	0.0331861	0.1108	0.3167	0.0009891	0.0993
Chimaltenango	0.5568	0.0389347	0.0699	0.5429	0.0012850	0.0660
Escuintla	0.3733	0.0305010	0.0817	0.3848	0.0008435	0.0755
Santa Rosa	0.6033	0.0447908	0.0742	0.5688	0.0016160	0.0707
Sololá	0.6148	0.0507176	0.0825	0.6334	0.0019608	0.0699
Totonicapán	0.7334	0.0487675	0.0665	0.7162	0.0018525	0.0601
Quetzaltenango	0.4155	0.0545174	0.1312	0.4450	0.0021441	0.1041
Suchitepéquez	0.4683	0.0384586	0.0821	0.4842	0.0012512	0.0731
Retalhuleu	0.6289	0.0300722	0.0478	0.6116	0.0008180	0.0468
San Marcos	0.6116	0.0504026	0.0824	0.6007	0.0019095	0.0727
Huehuetenango	0.6337	0.0484528	0.0765	0.6516	0.0018466	0.0660
Quiché	0.6818	0.0432194	0.0634	0.7073	0.0016002	0.0566
Baja Verapaz	0.6251	0.0351796	0.0563	0.6292	0.0010842	0.0523
Alta Verapaz	0.7984	0.0448358	0.0562	0.7882	0.0016776	0.0520
Petén	0.6990	0.0312059	0.0446	0.6818	0.0008727	0.0433
Izabal	0.5119	0.0846099	0.1653	0.5235	0.0035308	0.1135
Zacapa	0.5870	0.0513497	0.0875	0.5677	0.0019661	0.0781
Chiquimula	0.5617	0.0595836	0.1061	0.5981	0.0024405	0.0826
Jalapa	0.6694	0.0453546	0.0677	0.6449	0.0016319	0.0626
Jutiapa	0.5965	0.0360978	0.0605	0.5851	0.0011222	0.0573

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 11. Empleo informal, estimador EBLUP bajo el modelo *Fay Herriot*



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.10. Trabajo infantil

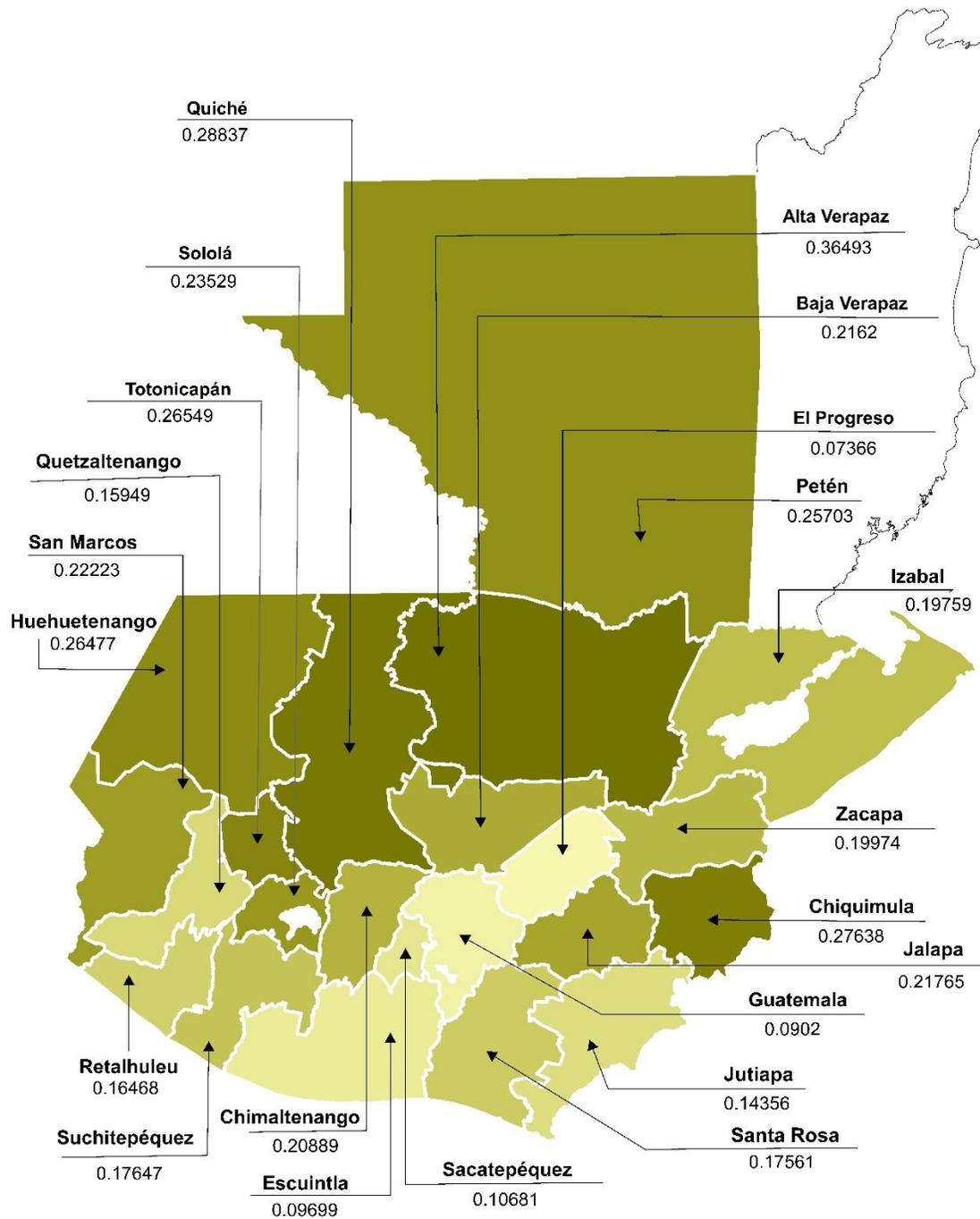
Con base en los resultados obtenidos, se estima que en Guatemala el 20.1 % de la población vive en hogares en los que existe al menos un niño de 15 años o menos que trabaja, o cuando existe al menos un adolescente de 16 a 17 años que trabaja, no estudia y no tiene al menos 9 años de educación completos. El intervalo de estimación al 95 % de confianza fue de [18.2 %, 22.1 %]. A nivel departamental se obtuvieron los siguientes valores estimados:

Tabla XVII. Trabajo infantil, según ENCOVI 2014

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.0881	0.0187638	0.2131	0.0902	0.0003166	0.1973
El Progreso	0.0650	0.0138618	0.2134	0.0737	0.0001799	0.1821
Sacatepéquez	0.1074	0.0199464	0.1856	0.1068	0.0003510	0.1754
Chimaltenango	0.2701	0.0369382	0.1368	0.2089	0.0008235	0.1374
Escuintla	0.0907	0.0155363	0.1712	0.0970	0.0002220	0.1536
Santa Rosa	0.2088	0.0401580	0.1923	0.1756	0.0008976	0.1706
Sololá	0.2315	0.0314915	0.1360	0.2353	0.0006848	0.1112
Totonicapán	0.2756	0.0330607	0.1200	0.2655	0.0007332	0.1020
Quetzaltenango	0.1550	0.0283930	0.1832	0.1595	0.0005882	0.1521
Suchitepéquez	0.1729	0.0249753	0.1445	0.1765	0.0004880	0.1252
Retalhuleu	0.1635	0.0189382	0.1158	0.1647	0.0003132	0.1075
San Marcos	0.2345	0.0314717	0.1342	0.2222	0.0006725	0.1167
Huehuetenango	0.2583	0.0303100	0.1173	0.2648	0.0006719	0.0979
Quiché	0.2463	0.0367848	0.1494	0.2884	0.0009462	0.1067
Baja Verapaz	0.2170	0.0331133	0.1526	0.2162	0.0007200	0.1241
Alta Verapaz	0.4002	0.0610001	0.1524	0.3649	0.0015953	0.1094
Petén	0.2677	0.0256684	0.0959	0.2570	0.0005135	0.0882
Izabal	0.1751	0.0505982	0.2889	0.1976	0.0010766	0.1661
Zacapa	0.2404	0.0440396	0.1832	0.1997	0.0009648	0.1555
Chiquimula	0.3483	0.0567337	0.1629	0.2764	0.0011946	0.1251
Jalapa	0.2203	0.0224984	0.1021	0.2177	0.0004159	0.0937
Jutiapa	0.1362	0.0187041	0.1373	0.1436	0.0003063	0.1219

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 12. Trabajo infantil, estimador EBLUP bajo el modelo *Fay Herriot*



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.11. Materiales de la vivienda

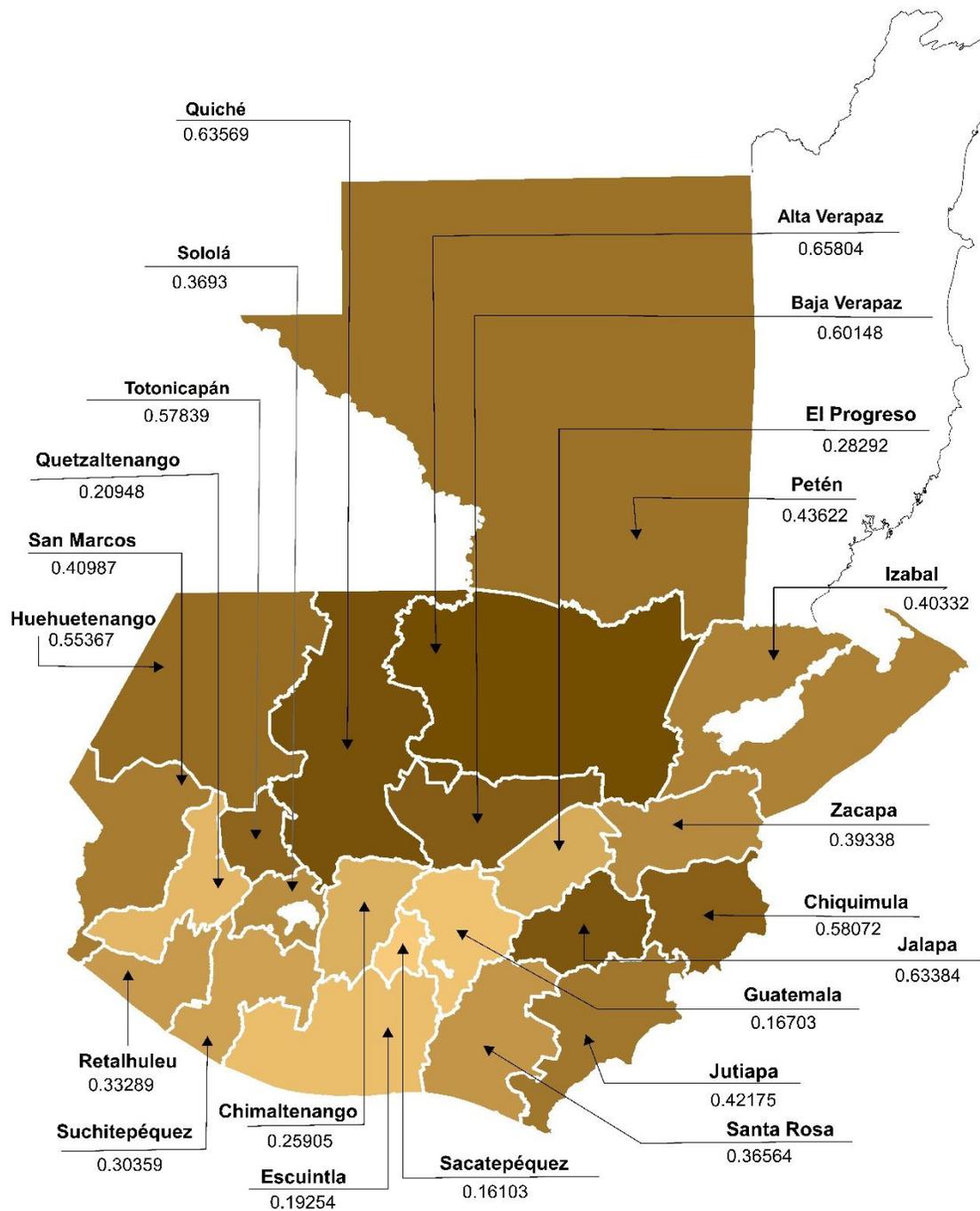
En esta parte se presentan los resultados relacionados con los materiales de la vivienda. En este sentido, se estimó que el 37.7 % de la población guatemalteca habita en viviendas construidas con materiales inapropiados (con piso de tierra, o bien, paredes de adobe, lámina metálica, bajareque, lepa, palo o caña, u otro). Este indicador presenta un intervalo de confianza al 95% de [34.9 %, 40.5 %]. Para el caso de los departamentos, la siguiente tabla detalla las estimaciones obtenidas con el estimador directo de la encuesta y con el EBLUP bajo la modelo *Fay Herriot*.

Tabla XVIII. **Materiales de la vivienda, según ENCOVI 2014**

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.1584	0.0345725	0.2182	0.1670	0.0009639	0.1859
El Progreso	0.2639	0.0385158	0.1459	0.2829	0.0010797	0.1161
Sacatepéquez	0.1517	0.0224462	0.1479	0.1610	0.0004657	0.1340
Chimaltenango	0.2684	0.0336712	0.1254	0.2590	0.0009056	0.1162
Escuintla	0.1890	0.0197929	0.1047	0.1925	0.0003695	0.0998
Santa Rosa	0.3984	0.0487894	0.1225	0.3656	0.0014304	0.1034
Sololá	0.3861	0.0486676	0.1260	0.3693	0.0014247	0.1022
Totonicapán	0.6329	0.0566642	0.0895	0.5784	0.0017225	0.0718
Quetzaltenango	0.2030	0.0304081	0.1498	0.2095	0.0007783	0.1332
Suchitepéquez	0.3096	0.0358227	0.1157	0.3036	0.0009802	0.1031
Retalhuleu	0.3608	0.0419054	0.1162	0.3329	0.0012112	0.1045
San Marcos	0.3872	0.0614984	0.1588	0.4099	0.0017666	0.1025
Huehuetenango	0.5659	0.0599942	0.1060	0.5537	0.0018185	0.0770
Quiché	0.5703	0.0597008	0.1047	0.6357	0.0020110	0.0705
Baja Verapaz	0.5890	0.0580376	0.0985	0.6015	0.0018473	0.0715
Alta Verapaz	0.6497	0.0582851	0.0897	0.6580	0.0019395	0.0669
Petén	0.3691	0.0416394	0.1128	0.4362	0.0012266	0.0803
Izabal	0.3790	0.1202264	0.3172	0.4033	0.0023794	0.1209
Zacapa	0.4303	0.0654054	0.1520	0.3934	0.0018391	0.1090
Chiquimula	0.6388	0.0785191	0.1229	0.5807	0.0022054	0.0809
Jalapa	0.6823	0.0492890	0.0722	0.6338	0.0015422	0.0620
Jutiapa	0.4436	0.0354170	0.0798	0.4218	0.0009572	0.0734

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 13. **Materiales de la vivienda, estimador EBLUP bajo el modelo Fay Herriot**



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.12. Hacinamiento

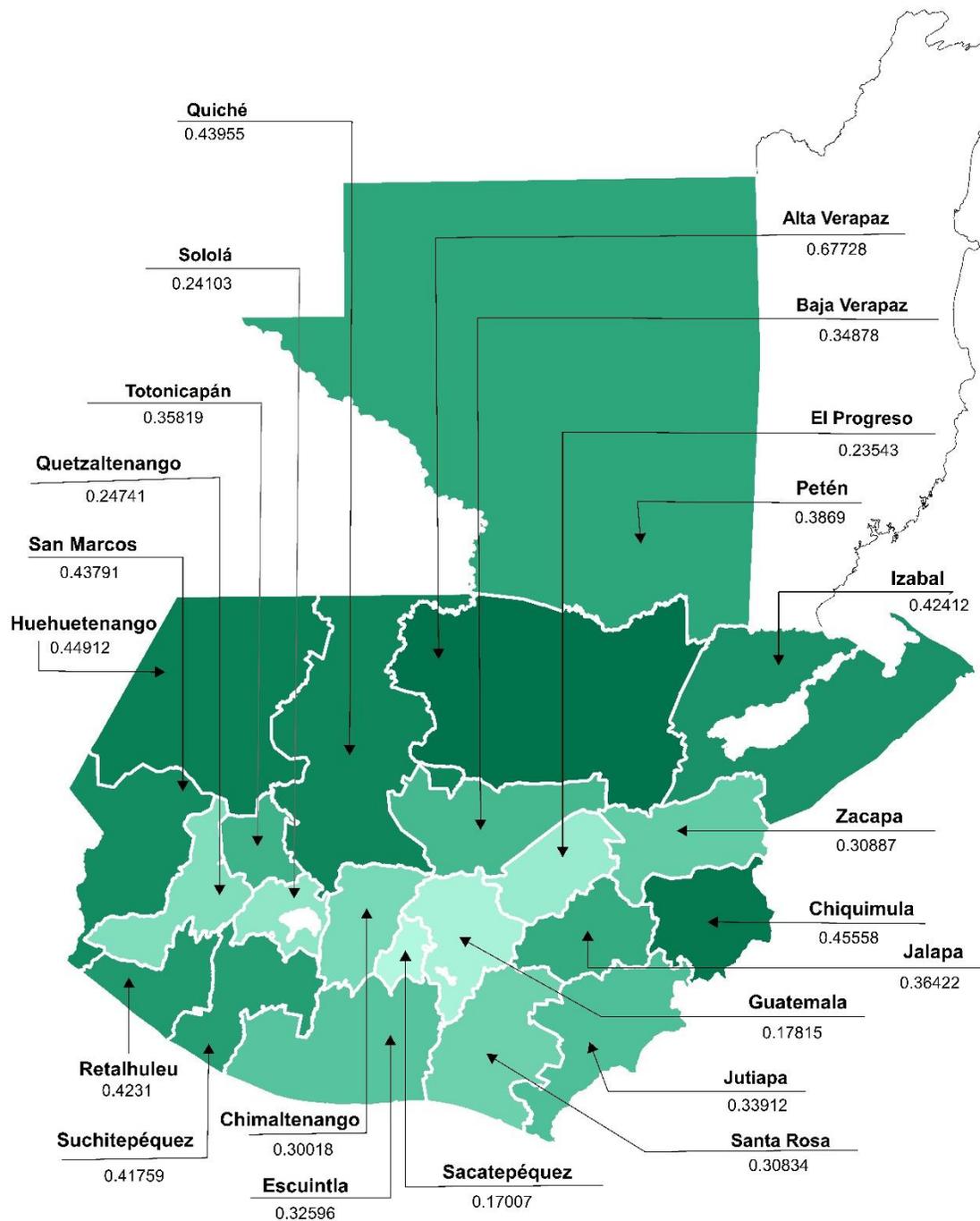
De acuerdo con los cálculos realizados, a nivel nacional se estimó que el 35.0 % de los guatemaltecos se encuentra privado por hacinamiento cuando habitan más de tres personas en promedio por cuarto (sin considerar los cuartos utilizados para cocina, baños, pasillos, garajes y los dedicados a negocios). El intervalo de confianza al 95 % obtenido fue de [32.4 %, 37.7 %]. La tabla siguiente muestra los resultados obtenidos a nivel departamental:

Tabla XIX. Hacinamiento, según ENCOVI 2014

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.1717	0.0334140	0.1946	0.1782	0.0007792	0.1567
El Progreso	0.2336	0.0361888	0.1549	0.2354	0.0007583	0.1170
Sacatepéquez	0.1530	0.0214798	0.1404	0.1701	0.0004225	0.1209
Chimaltenango	0.2813	0.0300580	0.1069	0.3002	0.0006095	0.0822
Escuintla	0.3163	0.0296038	0.0936	0.3260	0.0006151	0.0761
Santa Rosa	0.3554	0.0428398	0.1205	0.3083	0.0008232	0.0931
Sololá	0.2109	0.0329766	0.1563	0.2410	0.0007549	0.1140
Totonicapán	0.3420	0.0380232	0.1112	0.3582	0.0007541	0.0767
Quetzaltenango	0.2586	0.0401018	0.1550	0.2474	0.0008150	0.1154
Suchitepéquez	0.4499	0.0356932	0.0793	0.4176	0.0008950	0.0716
Retalhuleu	0.4265	0.0338593	0.0794	0.4231	0.0007095	0.0630
San Marcos	0.4951	0.0526748	0.1064	0.4379	0.0012916	0.0821
Huehuetenango	0.4732	0.0617374	0.1305	0.4491	0.0010129	0.0709
Quiché	0.3869	0.0405447	0.1048	0.4395	0.0009390	0.0697
Baja Verapaz	0.3459	0.0474819	0.1373	0.3488	0.0008993	0.0860
Alta Verapaz	0.6340	0.0577367	0.0911	0.6773	0.0020567	0.0670
Petén	0.3983	0.0327658	0.0823	0.3869	0.0006736	0.0671
Izabal	0.4337	0.1463763	0.3375	0.4241	0.0011075	0.0785
Zacapa	0.3355	0.0456584	0.1361	0.3089	0.0008368	0.0937
Chiquimula	0.4385	0.0626719	0.1429	0.4556	0.0011690	0.0750
Jalapa	0.3869	0.0359271	0.0929	0.3642	0.0007467	0.0750
Jutiapa	0.3718	0.0300009	0.0807	0.3391	0.0006448	0.0749

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 14. **Hacinamiento, estimador EBLUP bajo el modelo *Fay Herriot***



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.13. Combustible para cocinar

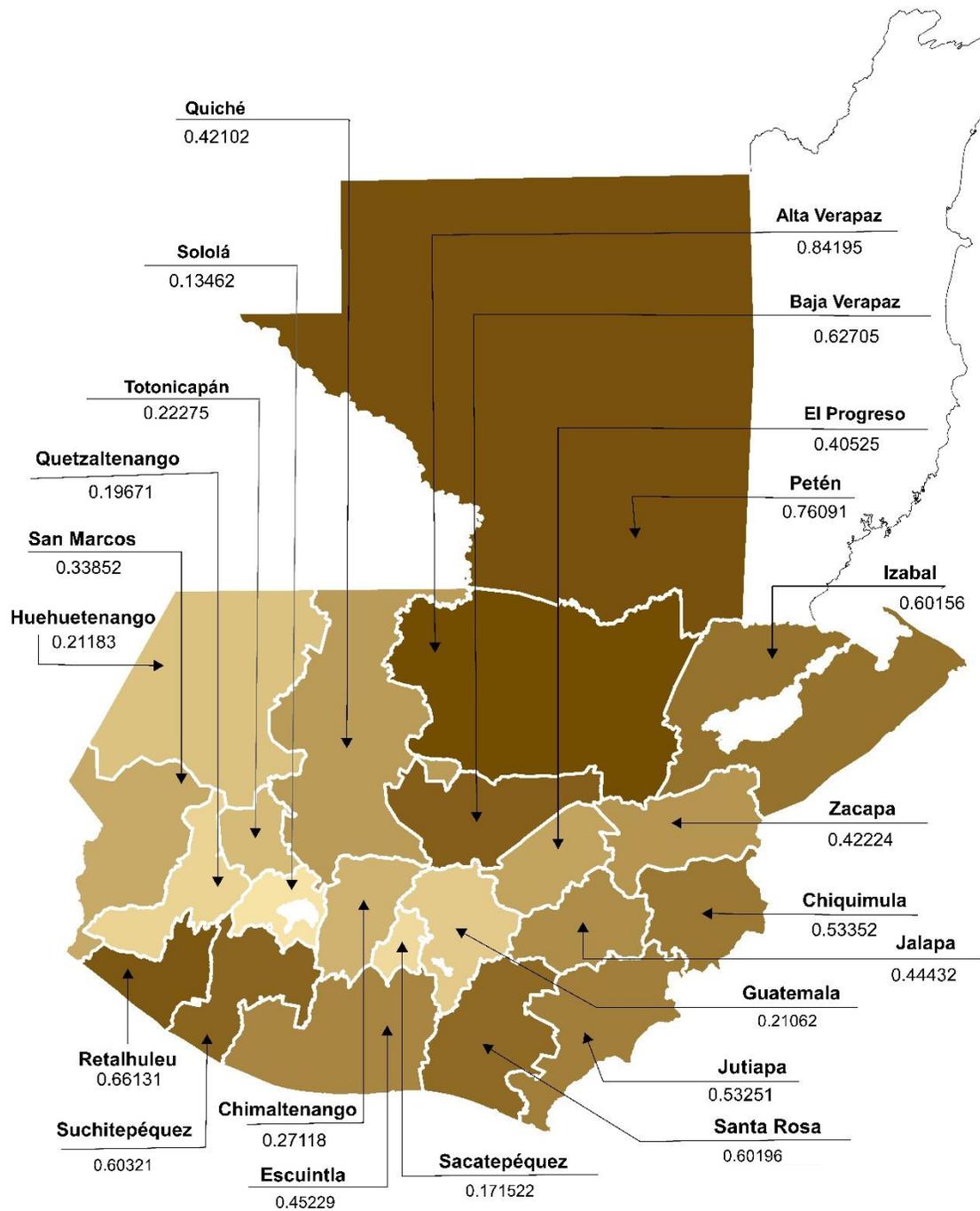
Al considerar el combustible para cocinar como indicador de privación de pobreza multidimensional para Guatemala, se estimó que el 39.4 % de la población habita en hogares en donde se utiliza para cocinar la leña y no se cuenta con una chimenea o una salida de escape para el humo. El resultado tiene un intervalo de confianza al 95 % de [36.4 %, 42.4 %]. Para el caso de los departamentos, las estimaciones se muestran en la siguiente tabla:

Tabla XX. **Combustible para cocinar, según ENCOVI 2014**

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.2172	0.0344835	0.1587	0.2106	0.0011347	0.1599
El Progreso	0.4087	0.0494114	0.1209	0.4052	0.0021710	0.1150
Sacatepéquez	0.1734	0.0275489	0.1589	0.1715	0.0007360	0.1582
Chimaltenango	0.2713	0.0315749	0.1164	0.2712	0.0009520	0.1138
Escuintla	0.4525	0.0401306	0.0887	0.4523	0.0014832	0.0851
Santa Rosa	0.6352	0.0542561	0.0854	0.6020	0.0025420	0.0838
Sololá	0.1222	0.0300017	0.2456	0.1346	0.0008649	0.2185
Totonicapán	0.2257	0.0420441	0.1863	0.2228	0.0016682	0.1834
Quetzaltenango	0.1973	0.0358646	0.1817	0.1967	0.0012184	0.1774
Suchitepéquez	0.6083	0.0353884	0.0582	0.6032	0.0011871	0.0571
Retalhuleu	0.6769	0.0390088	0.0576	0.6613	0.0014086	0.0568
San Marcos	0.3311	0.0446479	0.1348	0.3385	0.0018006	0.1253
Huehuetenango	0.1780	0.0416066	0.2338	0.2118	0.0015873	0.1881
Quiché	0.4077	0.0457531	0.1122	0.4210	0.0018815	0.1030
Baja Verapaz	0.6390	0.0442517	0.0692	0.6270	0.0017786	0.0673
Alta Verapaz	0.8481	0.0469024	0.0553	0.8419	0.0021311	0.0548
Petén	0.7717	0.0357875	0.0464	0.7609	0.0012173	0.0459
Izabal	0.5719	0.1179386	0.2062	0.6016	0.0078074	0.1469
Zacapa	0.4078	0.0563753	0.1383	0.4222	0.0027307	0.1238
Chiquimula	0.5168	0.0723343	0.1400	0.5335	0.0041415	0.1206
Jalapa	0.4340	0.0356206	0.0821	0.4443	0.0011936	0.0778
Jutiapa	0.5421	0.0381196	0.0703	0.5325	0.0013466	0.0689

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 15. **Combustible para cocinar, estimador EBLUP bajo el modelo Fay Herriot**



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.14. Acceso al agua

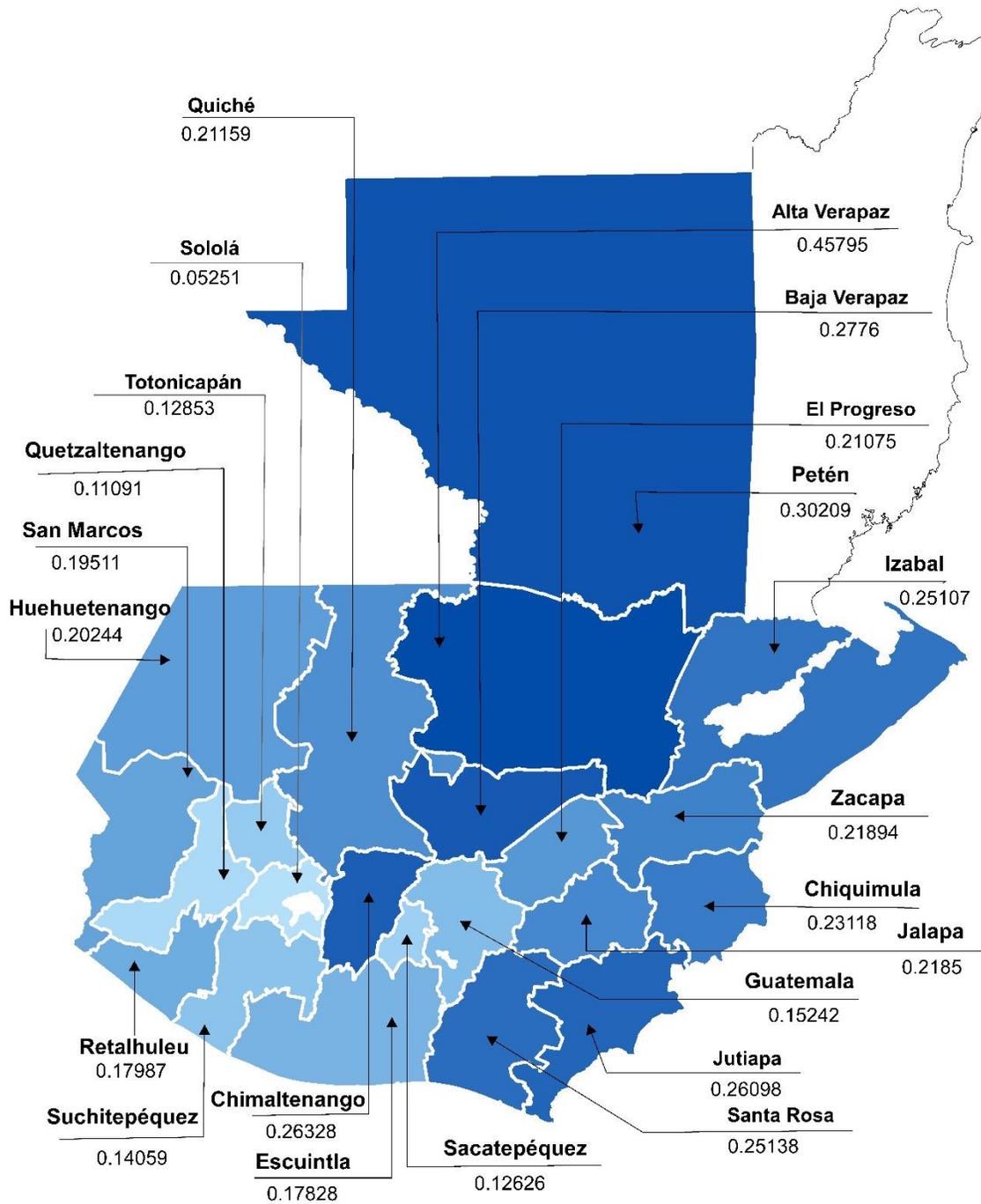
De acuerdo con los resultados obtenidos, el 21.2 % de los guatemaltecos no tienen acceso al agua según los siguientes criterios: área urbana, el agua es obtenida principalmente de chorro público, pozo perforado público o privado, río, lago, manantial, camión cisterna, agua de lluvia y otro. Área rural, el agua es obtenida principalmente de río, lago, manantial, camión cisterna, agua de lluvia y otro. El intervalo de estimación al 95 % de confianza fue de [18.2 %, 24.1 %]. A nivel departamental se obtuvieron las siguientes estimaciones:

Tabla XXI. Acceso al agua, según ENCOVI 2014

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.1583	0.0386905	0.2444	0.1524	0.0011710	0.2245
El Progreso	0.2520	0.0500294	0.1985	0.2108	0.0012238	0.1660
Sacatepéquez	0.1118	0.0223287	0.1997	0.1263	0.0004694	0.1716
Chimaltenango	0.2838	0.0409627	0.1443	0.2633	0.0013536	0.1397
Escuintla	0.1793	0.0389102	0.2171	0.1783	0.0010220	0.1793
Santa Rosa	0.2938	0.0543570	0.1850	0.2514	0.0013218	0.1446
Sololá	0.0475	0.0172310	0.3631	0.0525	0.0002950	0.3271
Totonicapán	0.1658	0.0409063	0.2467	0.1285	0.0012995	0.2805
Quetzaltenango	0.1080	0.0232916	0.2156	0.1109	0.0005004	0.2017
Suchitepéquez	0.1340	0.0420871	0.3140	0.1406	0.0012382	0.2503
Retalhuleu	0.1441	0.0377398	0.2619	0.1799	0.0010093	0.1766
San Marcos	0.1975	0.0359427	0.1820	0.1951	0.0009034	0.1541
Huehuetenango	0.1916	0.0337700	0.1763	0.2024	0.0008564	0.1446
Quiché	0.1539	0.0426437	0.2771	0.2116	0.0011740	0.1619
Baja Verapaz	0.2762	0.0555867	0.2013	0.2776	0.0014291	0.1362
Alta Verapaz	0.5196	0.0951068	0.1830	0.4580	0.0048222	0.1516
Petén	0.2924	0.0571546	0.1955	0.3021	0.0016550	0.1347
Izabal	0.1919	0.0723293	0.3769	0.2511	0.0017580	0.1670
Zacapa	0.2577	0.0495649	0.1924	0.2189	0.0012116	0.1590
Chiquimula	0.2426	0.0394784	0.1627	0.2312	0.0009973	0.1366
Jalapa	0.2663	0.0456869	0.1716	0.2185	0.0011706	0.1566
Jutiapa	0.2780	0.0383994	0.1381	0.2610	0.0009919	0.1207

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 16. Acceso al agua, estimador EBLUP bajo el modelo *Fay Herriot*



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.15. Energía eléctrica

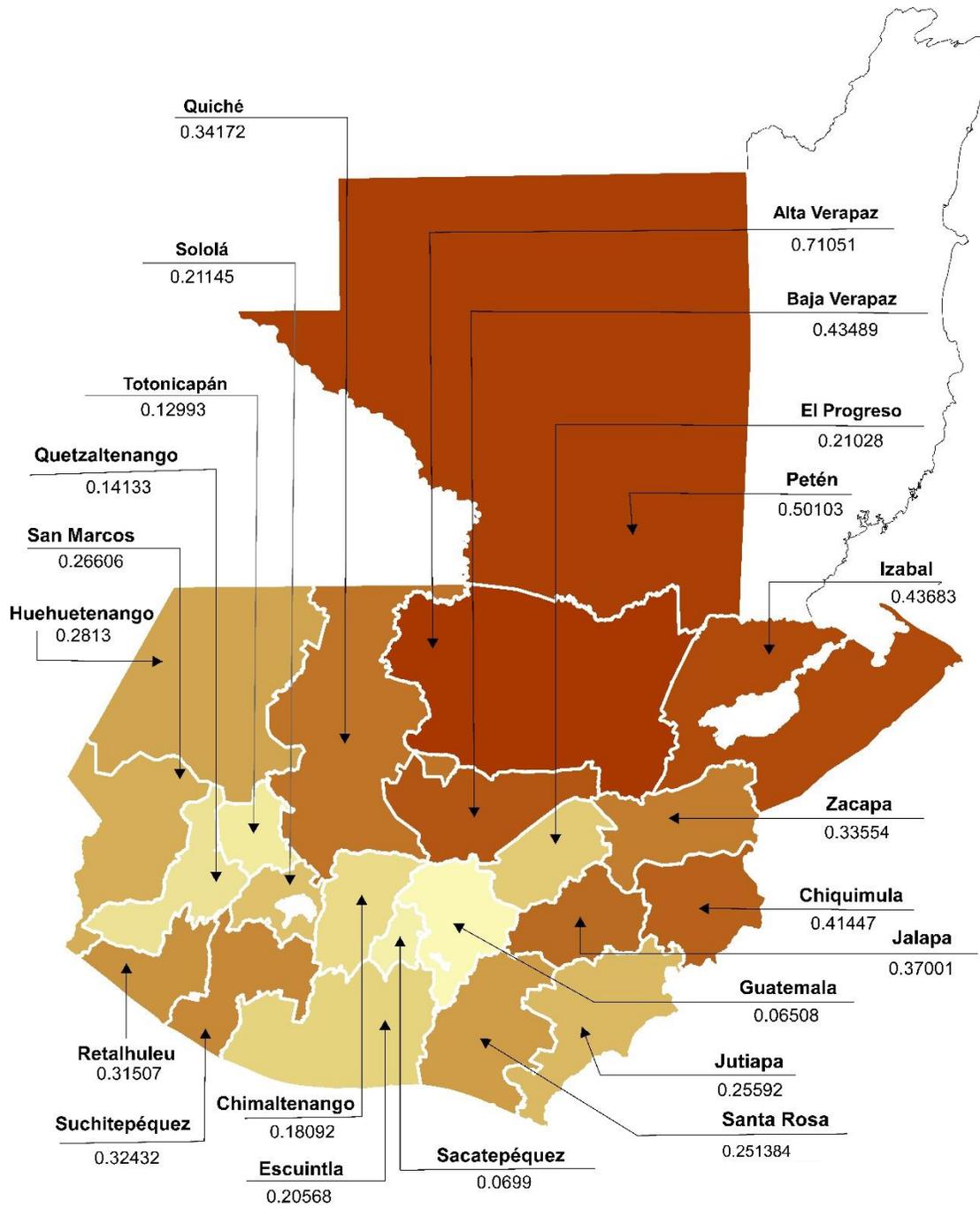
Con base en los resultados alcanzados, se estima que el 27.0 % de los guatemaltecos habitan en viviendas sin conexión a servicio de energía eléctrica, o bien, viviendas conectas a una red de distribución de energía eléctrica, pero que no contaron con el servicio durante más de un día en el mes anterior a la encuesta. Este indicador obtuvo un intervalo de estimación al 95 % de confianza de [24.1 %, 29.9 %]. La siguiente tabla refleja los resultados obtenidos a nivel departamental:

Tabla XXII. **Energía eléctrica, según ENCOVI 2014**

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.0638	0.0173156	0.2714	0.0651	0.0002858	0.2598
El Progreso	0.1855	0.0321189	0.1731	0.2103	0.0007483	0.1301
Sacatepéquez	0.0686	0.0135071	0.1968	0.0699	0.0001779	0.1908
Chimaltenango	0.1979	0.0301070	0.1522	0.1809	0.0006910	0.1453
Escuintla	0.1862	0.0315361	0.1694	0.2057	0.0007641	0.1344
Santa Rosa	0.3661	0.0511202	0.1396	0.3132	0.0012159	0.1113
Sololá	0.2128	0.0343375	0.1613	0.2114	0.0008156	0.1351
Totonicapán	0.1353	0.0230066	0.1700	0.1299	0.0004747	0.1677
Quetzaltenango	0.1524	0.0323170	0.2121	0.1413	0.0007777	0.1973
Suchitepéquez	0.3275	0.0389400	0.1189	0.3243	0.0010823	0.1014
Retalhuleu	0.3465	0.0437307	0.1262	0.3151	0.0011294	0.1067
San Marcos	0.2761	0.0449618	0.1628	0.2661	0.0010709	0.1230
Huehuetenango	0.2235	0.0362370	0.1621	0.2813	0.0008854	0.1058
Quiché	0.2871	0.0442576	0.1542	0.3417	0.0011163	0.0978
Baja Verapaz	0.4626	0.0595804	0.1288	0.4349	0.0014556	0.0877
Alta Verapaz	0.7541	0.0611597	0.0811	0.7105	0.0028035	0.0745
Petén	0.4855	0.0571292	0.1177	0.5010	0.0015171	0.0777
Izabal	0.4106	0.1537704	0.3745	0.4368	0.0019167	0.1002
Zacapa	0.3768	0.0597393	0.1585	0.3355	0.0013685	0.1102
Chiquimula	0.5246	0.0782133	0.1491	0.4145	0.0015628	0.0954
Jalapa	0.3921	0.0494359	0.1261	0.3700	0.0012353	0.0950
Jutiapa	0.2473	0.0297065	0.1201	0.2559	0.0006732	0.1014

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 17. **Energía eléctrica, estimador EBLUP bajo el modelo *Fay Herriot***



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.16. Recolección de basura

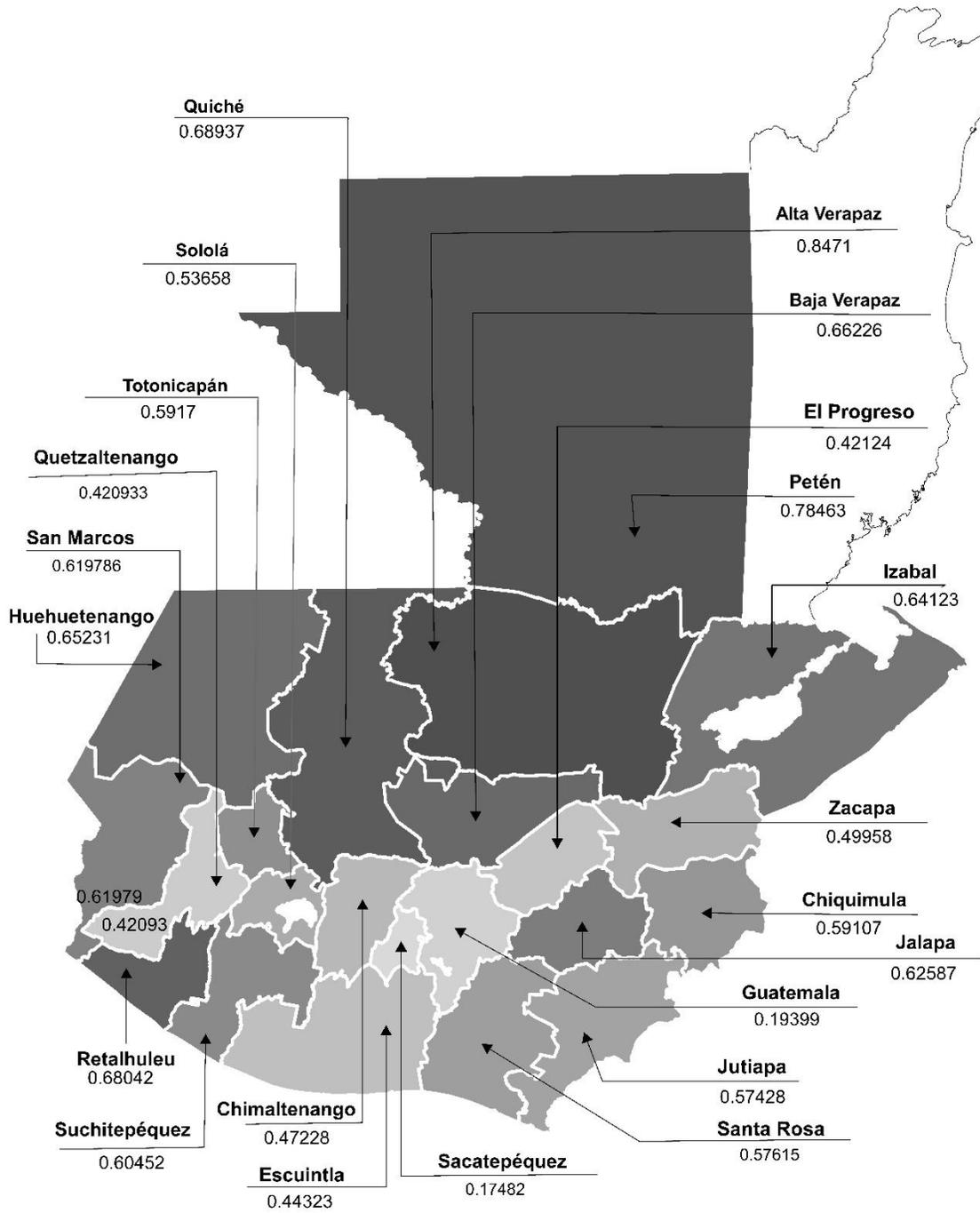
Los resultados muestran que el 51.6 % de los guatemaltecos habitan en hogares que no tienen acceso a servicio de recolección de basura y la mayor parte de la basura la queman, la entierran, la tiran en cualquier lugar o respondieron "otra" forma de eliminación de la basura. Los límites inferior y superior al 95 % de confianza para este estimador fueron de [48.6 %, 54.6 %]. Por departamento, las estimaciones se distribuyeron de la siguiente manera:

Tabla XXIII. **Recolección de basura, según ENCOVI 2014**

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.1771	0.0381230	0.2152	0.1940	0.0013233	0.1875
El Progreso	0.3999	0.0546504	0.1367	0.4212	0.0023363	0.1147
Sacatepéquez	0.1616	0.0283472	0.1754	0.1748	0.0007651	0.1582
Chimaltenango	0.4960	0.0465711	0.0939	0.4723	0.0018213	0.0904
Escuintla	0.4329	0.0430556	0.0995	0.4432	0.0015998	0.0902
Santa Rosa	0.6161	0.0623515	0.1012	0.5761	0.0028386	0.0925
Sololá	0.5164	0.0636302	0.1232	0.5366	0.0029262	0.1008
Totonicapán	0.6440	0.0571308	0.0887	0.5917	0.0027679	0.0889
Quetzaltenango	0.4350	0.0614346	0.1412	0.4209	0.0028291	0.1264
Suchitepéquez	0.6013	0.0450507	0.0749	0.6045	0.0017312	0.0688
Retalhuleu	0.7084	0.0402680	0.0568	0.6804	0.0014189	0.0554
San Marcos	0.6613	0.0631108	0.0954	0.6198	0.0028209	0.0857
Huehuetenango	0.6446	0.0624527	0.0969	0.6523	0.0028744	0.0822
Quiché	0.6608	0.0540179	0.0817	0.6894	0.0024556	0.0719
Baja Verapaz	0.6716	0.0463307	0.0690	0.6623	0.0017895	0.0639
Alta Verapaz	0.8305	0.0539089	0.0649	0.8471	0.0024711	0.0587
Petén	0.7961	0.0357826	0.0449	0.7846	0.0011673	0.0435
Izabal	0.5955	0.1087445	0.1826	0.6412	0.0052645	0.1132
Zacapa	0.4539	0.0754522	0.1662	0.4996	0.0036013	0.1201
Chiquimula	0.5473	0.0641548	0.1172	0.5911	0.0029185	0.0914
Jalapa	0.6218	0.0564093	0.0907	0.6259	0.0024342	0.0788
Jutiapa	0.5848	0.0397724	0.0680	0.5743	0.0013782	0.0646

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 18. **Recolección de basura, estimador EBLUP bajo el modelo Fay Herriot**



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.1.17. Saneamiento

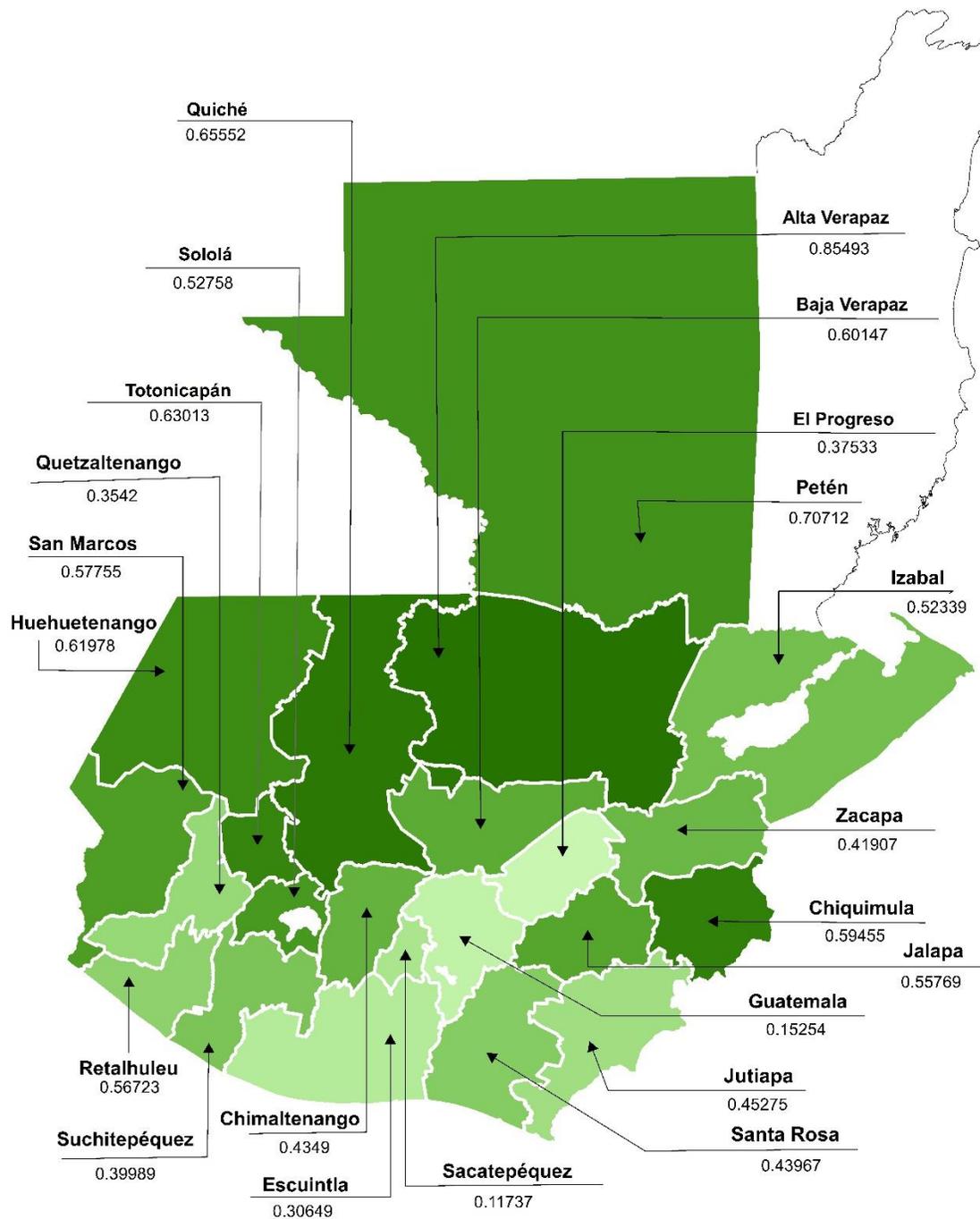
Conforme a la metodología adoptada para las estimaciones, el 45.9 % de la población guatemalteca reside en hogares en donde no cuenta con saneamiento mejorado, según los siguientes criterios: Área urbana, la vivienda no está conectada a una red de drenajes o sí está conectado a una red de drenaje, pero su servicio sanitario es un excusado lavable, letrina o pozo ciego, o no tiene. Área rural, la vivienda cuenta con servicio sanitario tipo letrina o pozo ciego, o no tiene. El intervalo de estimación calculado al 95 % de confianza fue de [42.7 %, 49.2 %]. La siguiente tabla detalla las estimaciones a nivel de departamento:

Tabla XXIV. Saneamiento, según ENCOVI 2014

Departamento	Estimador directo	EE	CV	EBLUP Fay Herriot	ECM	ECMR
Guatemala	0.1426	0.0410842	0.2881	0.1525	0.0015032	0.2542
El Progreso	0.3806	0.0568270	0.1493	0.3753	0.0023808	0.1300
Sacatepéquez	0.1053	0.0224268	0.2129	0.1174	0.0004860	0.1878
Chimaltenango	0.4354	0.0552401	0.1269	0.4349	0.0026932	0.1193
Escuintla	0.2999	0.0379409	0.1265	0.3065	0.0012649	0.1160
Santa Rosa	0.4500	0.0607442	0.1350	0.4397	0.0025786	0.1155
Sololá	0.5550	0.0580859	0.1047	0.5276	0.0024847	0.0945
Totonicapán	0.6786	0.0589298	0.0868	0.6301	0.0024984	0.0793
Quetzaltenango	0.3661	0.0577511	0.1577	0.3542	0.0024701	0.1403
Suchitepéquez	0.3940	0.0448789	0.1139	0.3999	0.0017087	0.1034
Retalhuleu	0.6268	0.0507528	0.0810	0.5672	0.0020101	0.0790
San Marcos	0.6082	0.0696324	0.1145	0.5775	0.0031049	0.0965
Huehuetenango	0.5753	0.0753718	0.1310	0.6198	0.0034408	0.0946
Quiché	0.6166	0.0672594	0.1091	0.6555	0.0031369	0.0854
Baja Verapaz	0.5690	0.0652679	0.1147	0.6015	0.0028758	0.0892
Alta Verapaz	0.8240	0.0573664	0.0696	0.8549	0.0027549	0.0614
Petén	0.7261	0.0380839	0.0524	0.7071	0.0012845	0.0507
Izabal	0.5247	0.1261129	0.2404	0.5234	0.0051232	0.1368
Zacapa	0.3836	0.0763535	0.1991	0.4191	0.0033922	0.1390
Chiquimula	0.5605	0.0847424	0.1512	0.5946	0.0038080	0.1038
Jalapa	0.5720	0.0564588	0.0987	0.5577	0.0023638	0.0872
Jutiapa	0.4606	0.0428218	0.0930	0.4528	0.0015497	0.0869

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 19. Saneamiento, estimador EBLUP bajo el modelo *Fay Herriot*



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.2. Objetivo 2: Niveles de incidencia e intensidad de la pobreza de la población guatemalteca a nivel de departamento

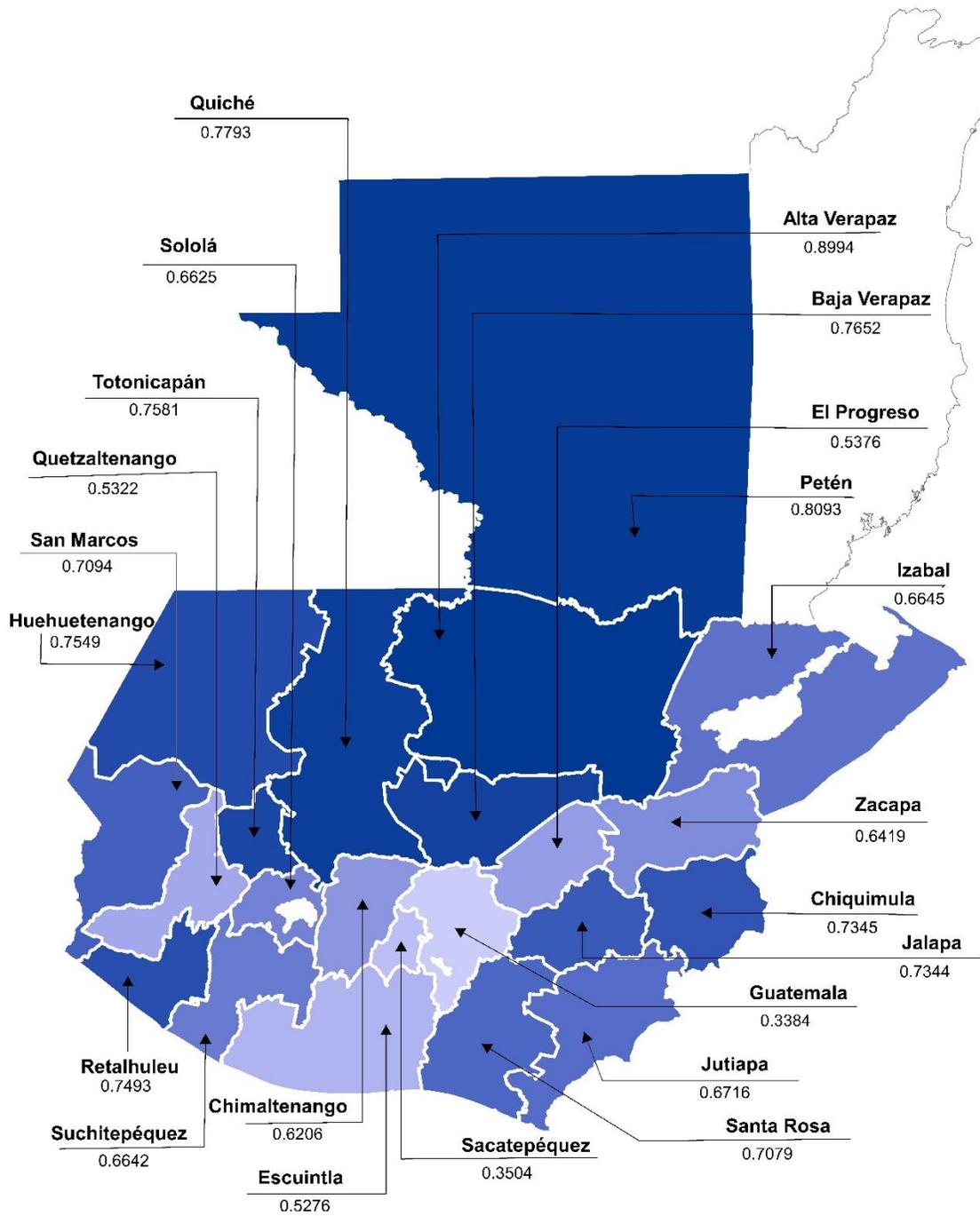
De acuerdo con los resultados obtenidos la incidencia de pobreza multidimensional de la población guatemalteca es de 61.6 % con un intervalo de confianza al 95 por ciento de [58.6 %, 64.7 %], es decir, se estima que seis de cada 10 guatemaltecos enfrentan privaciones en el 30 por ciento o más de los indicadores ponderados incluidos en el IPM-Gt. A nivel de departamento se obtuvieron los siguientes resultados:

Tabla XXV. **Incidencia de pobreza multidimensional departamental, según ENCOVI 2014**

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.3072	0.0416367	0.1355	0.3384	0.0015468	0.1162
El Progreso	0.5284	0.0557837	0.1056	0.5376	0.0024763	0.0926
Sacatepéquez	0.3295	0.0344373	0.1045	0.3504	0.0010964	0.0945
Chimaltenango	0.6308	0.0408151	0.0647	0.6206	0.0014726	0.0618
Escuintla	0.5232	0.0378579	0.0724	0.5276	0.0012914	0.0681
Santa Rosa	0.7360	0.0421397	0.0573	0.7079	0.0015560	0.0557
Sololá	0.6427	0.0491344	0.0765	0.6625	0.0020114	0.0677
Totonicapán	0.7599	0.0493946	0.0650	0.7581	0.0020348	0.0595
Quetzaltenango	0.4973	0.0623908	0.1255	0.5322	0.0029043	0.1013
Suchitepéquez	0.6681	0.0359998	0.0539	0.6642	0.0011720	0.0515
Retalhuleu	0.7676	0.0338441	0.0441	0.7493	0.0010491	0.0432
San Marcos	0.7235	0.0563708	0.0779	0.7094	0.0024882	0.0703
Huehuetenango	0.7433	0.0527826	0.0710	0.7549	0.0022858	0.0633
Quiché	0.7476	0.0496203	0.0664	0.7793	0.0021381	0.0593
Baja Verapaz	0.7768	0.0365238	0.0470	0.7652	0.0012029	0.0453
Alta Verapaz	0.8960	0.0380798	0.0425	0.8994	0.0013506	0.0409
Petén	0.8198	0.0319752	0.0390	0.8093	0.0009458	0.0380
Izabal	0.6458	0.0968718	0.1500	0.6645	0.0049508	0.1059
Zacapa	0.6510	0.0508975	0.0782	0.6419	0.0021248	0.0718
Chiquimula	0.7326	0.0689742	0.0941	0.7345	0.0033576	0.0789
Jalapa	0.7474	0.0459384	0.0615	0.7344	0.0017914	0.0576
Jutiapa	0.6789	0.0350847	0.0517	0.6716	0.0011195	0.0498

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 20. **Incidencia de pobreza multidimensional departamental, estimador EBLUP bajo el modelo *Fay Herriot***



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

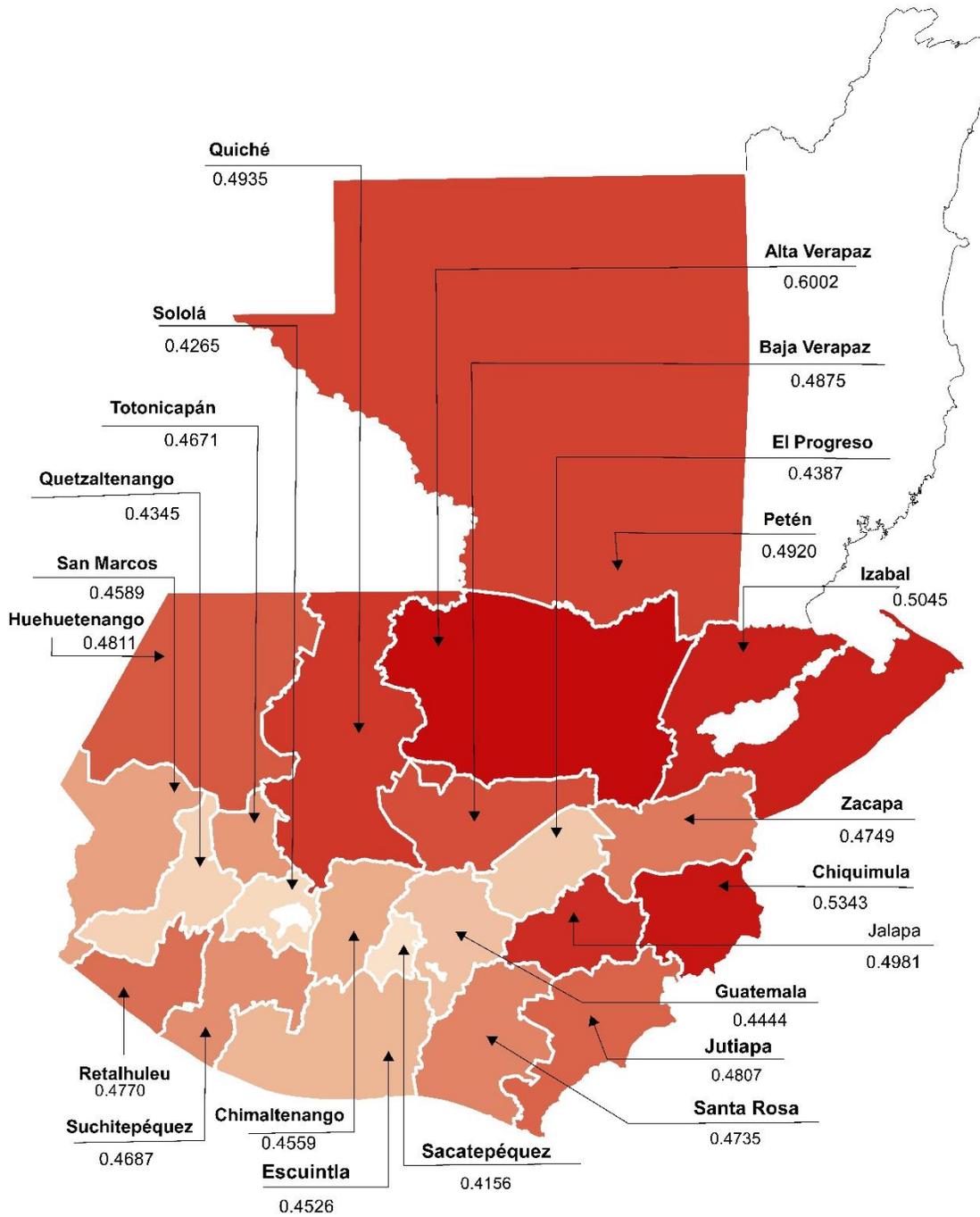
Para el caso de la intensidad de la pobreza de la población guatemalteca a nivel nacional, se estima que fue de 48.6 % con un intervalo al 95 % de confianza de [47.5 %, 49.7 %]. Este indicador refleja que las personas multidimensionalmente pobres en promedio están privadas en casi la mitad de los indicadores ponderados del IPM-GT. En la siguiente tabla se presentan los resultados obtenidos a nivel departamental.

Tabla XXVI. Intensidad de la pobreza multidimensional departamental, según ENCOVI 2014

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.4519	0.0158597	0.0351	0.4444	0.0001457	0.0272
El Progreso	0.4365	0.0100835	0.0231	0.4387	0.0000807	0.0205
Sacatepéquez	0.4108	0.0100695	0.0245	0.4156	0.0000825	0.0219
Chimaltenango	0.4525	0.0097429	0.0215	0.4559	0.0000756	0.0191
Escuintla	0.4440	0.0086278	0.0194	0.4526	0.0000624	0.0175
Santa Rosa	0.4852	0.0147400	0.0304	0.4735	0.0001306	0.0241
Sololá	0.4264	0.0112309	0.0263	0.4265	0.0000971	0.0231
Totonicapán	0.4611	0.0095938	0.0208	0.4671	0.0000739	0.0184
Quetzaltenango	0.4335	0.0091728	0.0212	0.4345	0.0000701	0.0193
Suchitepéquez	0.4749	0.0103733	0.0218	0.4687	0.0000834	0.0195
Retalhuleu	0.4721	0.0085086	0.0180	0.4770	0.0000614	0.0164
San Marcos	0.4594	0.0095530	0.0208	0.4589	0.0000736	0.0187
Huehuetenango	0.4694	0.0137348	0.0293	0.4811	0.0001222	0.0230
Quiché	0.4963	0.0109208	0.0220	0.4935	0.0000899	0.0192
Baja Verapaz	0.4877	0.0155730	0.0319	0.4875	0.0001390	0.0242
Alta Verapaz	0.6021	0.0206939	0.0344	0.6002	0.0002904	0.0284
Petén	0.5016	0.0112916	0.0225	0.4920	0.0000937	0.0197
Izabal	0.5095	0.0366285	0.0719	0.5045	0.0002378	0.0306
Zacapa	0.4909	0.0164044	0.0334	0.4749	0.0001457	0.0254
Chiquimula	0.5422	0.0133644	0.0246	0.5343	0.0001251	0.0209
Jalapa	0.5054	0.0118194	0.0234	0.4981	0.0001002	0.0201
Jutiapa	0.4822	0.0099735	0.0207	0.4807	0.0000783	0.0184

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 21. **Intensidad de la pobreza multidimensional departamental, estimador EBLUP bajo el modelo *Fay Herriot***



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

3.3. Objetivo 3: Índice de Pobreza Multidimensional IPM-Gt a nivel de departamento

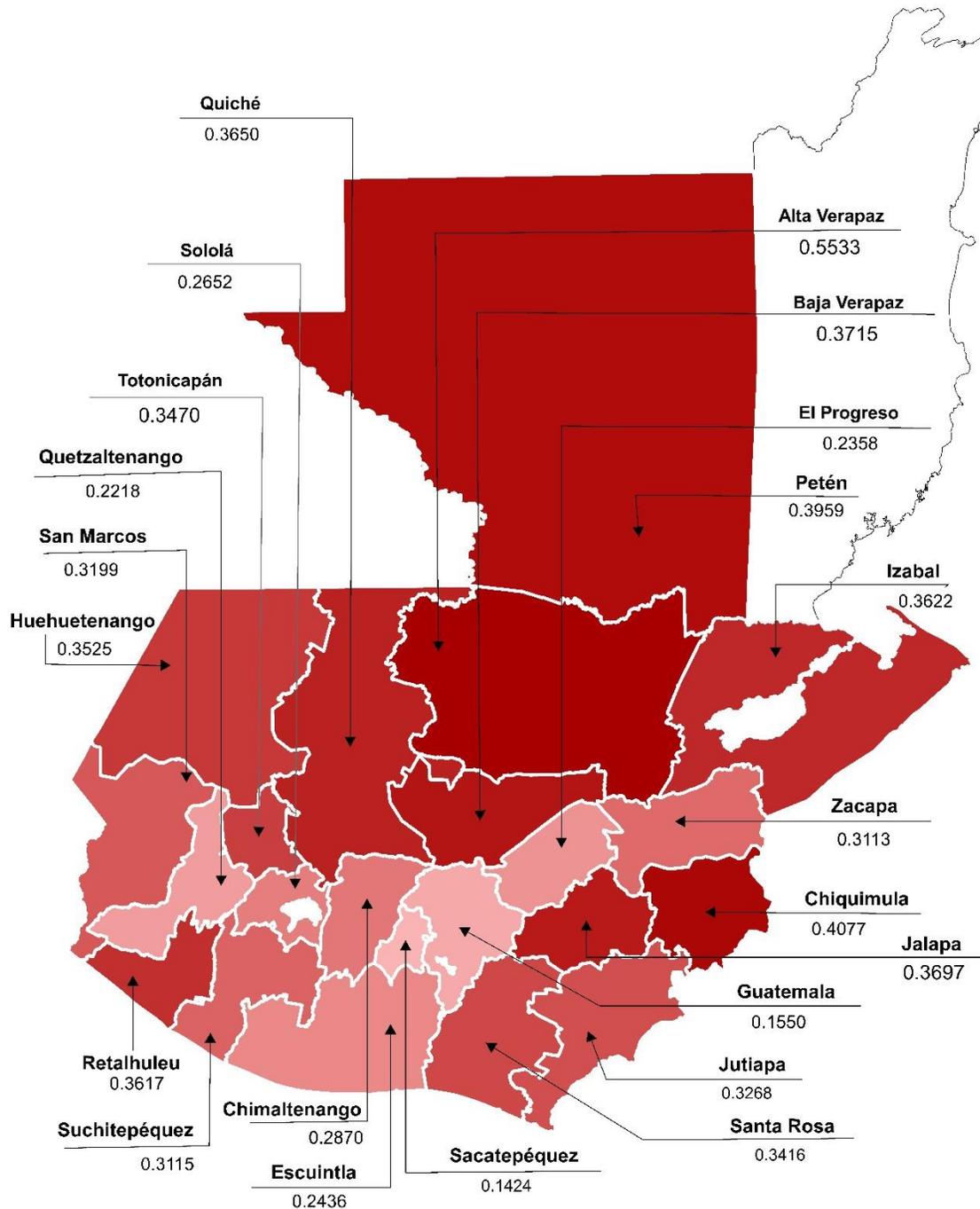
A nivel nacional se estimó en 0.299 (producto de la incidencia por la intensidad de la pobreza) el IPM-Gt con un intervalo de confianza al 95 % de [0.282, 0.317]. Este resultado indica que las personas multidimensionalmente pobres experimentan el 29.9 por ciento de privaciones ponderadas respecto al máximo posible de privaciones que tendrían si todas las personas en Guatemala fueran pobres y privados en todas las dimensiones. A nivel departamental se obtuvieron los siguientes resultados:

Tabla XXVII. Índice de Pobreza Multidimensional IPM-Gt a nivel de departamento, según ENCOVI 2014

Departamento	Estimador directo	EE	CV	EBLUP <i>Fay Herriot</i>	ECM	ECMR
Guatemala	0.1388	0.0216245	0.1558	0.1550	0.0004168	0.1317
El Progreso	0.2307	0.0263174	0.1141	0.2358	0.0005813	0.1022
Sacatepéquez	0.1353	0.0151342	0.1118	0.1424	0.0002175	0.1036
Chimaltenango	0.2854	0.0196806	0.0689	0.2870	0.0003495	0.0651
Escuintla	0.2323	0.0181149	0.0780	0.2436	0.0003009	0.0712
Santa Rosa	0.3571	0.0273564	0.0766	0.3416	0.0006149	0.0726
Sololá	0.2740	0.0218083	0.0796	0.2652	0.0004258	0.0778
Totonicapán	0.3504	0.0263900	0.0753	0.3470	0.0005794	0.0694
Quetzaltenango	0.2156	0.0285806	0.1326	0.2218	0.0006678	0.1165
Suchitepéquez	0.3173	0.0207510	0.0654	0.3115	0.0003850	0.0630
Retalhuleu	0.3624	0.0187028	0.0516	0.3617	0.0003198	0.0494
San Marcos	0.3324	0.0290051	0.0873	0.3199	0.0006754	0.0812
Huehuetenango	0.3489	0.0299634	0.0859	0.3525	0.0007134	0.0758
Quiché	0.3710	0.0281556	0.0759	0.3650	0.0006447	0.0696
Baja Verapaz	0.3789	0.0264388	0.0698	0.3715	0.0005822	0.0650
Alta Verapaz	0.5395	0.0375073	0.0695	0.5533	0.0012352	0.0635
Petén	0.4112	0.0221246	0.0538	0.3959	0.0004299	0.0524
Izabal	0.3291	0.0704523	0.2141	0.3622	0.0018870	0.1199
Zacapa	0.3195	0.0312706	0.0979	0.3113	0.0007578	0.0884
Chiquimula	0.3973	0.0439621	0.1107	0.4077	0.0012584	0.0870
Jalapa	0.3778	0.0270024	0.0715	0.3697	0.0006023	0.0664
Jutiapa	0.3274	0.0201248	0.0615	0.3268	0.0003639	0.0584

Fuente: elaboración propia con datos de la ENCOVI 2014.

Figura 22. Índice de Pobreza Multidimensional IPM-Gt a nivel de departamento, estimador EBLUP bajo el modelo *Fay Herriot*



Fuente: elaboración propia con datos de la ENCOVI 2014, programa QGIS.

4. DISCUSIÓN DE RESULTADOS

4.1. Tamaños de muestra departamentales

La evaluación de la validez interna de las estimaciones obtenidas a través del estimador directo del diseño complejo de la ENCOVI 2014 y el EBLUP bajo el modelo *Fay Herriot*, fue dimensionado, inicialmente, a través del análisis de los tamaños de muestra obtenidos por indicador a nivel de departamento. Los resultados obtenidos fueron calculados a partir de 53,083 observaciones muestrales, las cuales se desglosan en las tablas siguientes:

Tabla XXVIII. Tamaños de muestra parte 1, según ENCOVI 2014

Departamento	Tamaño de la muestra	Acceso a salud		Seguridad alimentaria		Embarazo en adolescentes	
		No privado	Privación censurada	No privado	Privación censurada	No privado	Privación censurada
Guatemala	4,204	4,037	167	3,526	678	4,056	148
El Progreso	1,992	1,913	79	1,255	737	1,925	67
Sacatepéquez	3,747	3,667	80	3,113	634	3,630	117
Chimaltenango	2,523	2,213	310	1,674	849	2,467	56
Escuintla	3,140	2,862	278	2,005	1,135	3,010	130
Santa Rosa	1,687	1,435	252	876	811	1,581	106
Sololá	2,283	2,220	63	1,582	701	2,140	143
Totonicapán	2,233	2,189	44	1,418	815	2,109	124
Quetzaltenango	3,322	3,248	74	2,488	834	3,132	190
Suchitepéquez	3,020	2,537	483	1,703	1,317	2,884	136
Retalhuleu	1,959	1,762	197	1,041	918	1,817	142
San Marcos	2,167	2,004	163	1,809	358	2,038	129
Huehuetenango	2,435	2,276	159	1,598	837	2,224	211
Quiché	2,494	2,264	230	1,109	1,385	2,362	132
Baja Verapaz	1,691	1,461	230	845	846	1,560	131
Alta Verapaz	2,153	1,620	533	906	1,247	1,944	209
Petén	2,739	2,577	162	1,524	1,215	2,551	188
Izabal	1,672	1,655	17	1,137	535	1,570	102
Zacapa	1,548	1,280	268	989	559	1,492	56
Chiquimula	1,789	1,431	358	897	892	1,711	78
Jalapa	2,292	2,146	146	1,102	1,190	2,166	126
Jutiapa	2,993	2,845	148	1,662	1,331	2,882	111

Fuente: elaboración propia con datos obtenidos de la ENCOVI 2014.

En la tabla anterior se observa que las tasas no censuradas “acceso a salud” y “embarazo en adolescentes” presentan tamaños de muestra promedio de 202 y 109 observaciones por departamento, respectivamente, llegando a ser de 17 individuos para el departamento de Izabal en la columna de la privación censurada “acceso a salud”.

Tabla XXIX. Tamaños de muestra parte 2, según ENCOVI 2014

Departamento	Tamaño de la muestra	Cuidado prenatal		Asistencia escolar		Escolaridad en adultos	
		No privado	Privación censurada	No privado	Privación censurada	No privado	Privación censurada
Guatemala	4,204	3,929	275	3,725	479	3,111	1,093
El Progreso	1,992	1,800	192	1,704	288	958	1,034
Sacatepéquez	3,747	3,463	284	3,216	531	2,571	1,176
Chimaltenango	2,523	2,180	343	1,857	666	942	1,581
Escuintla	3,140	2,793	347	2,573	567	1,504	1,636
Santa Rosa	1,687	1,495	192	1,195	492	591	1,096
Sololá	2,283	1,956	327	1,598	685	808	1,475
Totonicapán	2,233	1,951	282	1,385	848	543	1,690
Quetzaltenango	3,322	2,813	509	2,599	723	1,549	1,773
Suchitepéquez	3,020	2,620	400	2,239	781	1,051	1,969
Retalhuleu	1,959	1,649	310	1,457	502	563	1,396
San Marcos	2,167	1,659	508	1,477	690	531	1,636
Huehuetenango	2,435	1,951	484	1,387	1,048	571	1,864
Quiché	2,494	1,942	552	1,567	927	448	2,046
Baja Verapaz	1,691	1,493	198	1,262	429	414	1,277
Alta Verapaz	2,153	1,716	437	1,415	738	256	1,897
Petén	2,739	2,488	251	1,987	752	619	2,120
Izabal	1,672	1,521	151	1,181	491	761	911
Zacapa	1,548	1,337	211	1,202	346	596	952
Chiquimula	1,789	1,620	169	1,106	683	499	1,290
Jalapa	2,292	1,999	293	1,566	726	527	1,765
Jutiapa	2,993	2,609	384	2,077	916	963	2,030

Fuente: elaboración propia con datos obtenidos de la ENCOVI 2014.

En la tabla anterior, se observa que el menor número de observaciones se observó en el indicador “cuidado prenatal”, el cual presentó un promedio de 323 observaciones por departamento. Los indicadores “asistencia escolar” y

“escolaridad en adultos” presentaron tamaños de muestra promedio por dominio de 650 y 1,532 observaciones respectivamente, con relación al número de individuos que presentan la privación no censurada del IPM-Gt.

Tabla XXX. **Tamaños de muestra parte 3, según ENCOVI 2014**

Departamento	Tamaño de la muestra	Rezago educativo		Cuidado infantil		Empleo informal	
		No privado	Privación censurada	No privado	Privación censurada	No privado	Privación censurada
Guatemala	4,204	4,023	181	4,166	38	3,185	1,019
El Progreso	1,992	1,871	121	1,972	20	1,057	935
Sacatepéquez	3,747	3,550	197	3,739	8	2,633	1,114
Chimaltenango	2,523	2,310	213	2,498	25	1,111	1,412
Escuintla	3,140	2,830	310	3,122	18	1,919	1,221
Santa Rosa	1,687	1,508	179	1,672	15	719	968
Sololá	2,283	2,114	169	2,239	44	833	1,450
Totonicapán	2,233	1,990	243	2,204	29	598	1,635
Quetzaltenango	3,322	2,950	372	3,285	37	1,833	1,489
Suchitepéquez	3,020	2,614	406	3,007	13	1,588	1,432
Retalhuleu	1,959	1,733	226	1,931	28	732	1,227
San Marcos	2,167	1,932	235	2,146	21	716	1,451
Huehuetenango	2,435	2,187	248	2,403	32	835	1,600
Quiché	2,494	2,152	342	2,465	29	640	1,854
Baja Verapaz	1,691	1,553	138	1,691	4	637	1,054
Alta Verapaz	2,153	1,577	576	2,101	52	461	1,692
Petén	2,739	2,220	519	2,689	50	831	1,908
Izabal	1,672	1,519	153	1,642	30	954	718
Zacapa	1,548	1,389	159	1,539	9	665	883
Chiquimula	1,789	1,498	291	1,775	14	758	1,031
Jalapa	2,292	1,912	380	2,276	16	688	1,604
Jutiapa	2,993	2,597	396	2,989	4	1,140	1,853

Fuente: elaboración propia con datos obtenidos de la ENCOVI 2014.

Para el caso de la tasa no censurada “cuidado infantil” se observó un promedio de 24 observaciones por departamento. Este indicador fue el que menor número tamaño de muestra registró a nivel de campo de estimación. El “rezago educativo” tuvo un promedio de 275 observaciones, mientras que el

indicador de “empleo informal” presentó 1,343 elementos con la privación censurada.

Tabla XXXI. Tamaños de muestra parte 4, según ENCOVI 2014

Departamento	Tamaño de la muestra	Trabajo infantil		Materiales de la vivienda		Hacinamiento	
		No privado	Privación censurada	No privado	Privación censurada	No privado	Privación censurada
Guatemala	4,204	3,878	326	3,643	561	3,590	614
El Progreso	1,992	1,837	155	1,407	585	1,495	497
Sacatepéquez	3,747	3,351	396	3,213	534	3,180	567
Chimaltenango	2,523	1,788	735	1,840	683	1,782	741
Escuintla	3,140	2,861	279	2,495	645	2,073	1,067
Santa Rosa	1,687	1,388	299	1,028	659	1,086	601
Sololá	2,283	1,762	521	1,376	907	1,752	531
Totonicapán	2,233	1,638	595	775	1,458	1,411	822
Quetzaltenango	3,322	2,775	547	2,549	773	2,344	978
Suchitepéquez	3,020	2,524	496	2,106	914	1,654	1,366
Retalhuleu	1,959	1,637	322	1,227	732	1,085	874
San Marcos	2,167	1,655	512	1,304	863	1,030	1,137
Huehuetenango	2,435	1,767	668	940	1,495	1,159	1,276
Quiché	2,494	1,802	692	863	1,631	1,378	1,116
Baja Verapaz	1,691	1,297	394	592	1,099	1,046	645
Alta Verapaz	2,153	1,361	792	803	1,350	818	1,335
Petén	2,739	2,015	724	1,706	1,033	1,644	1,095
Izabal	1,672	1,461	211	1,158	514	1,117	555
Zacapa	1,548	1,200	348	871	677	1,064	484
Chiquimula	1,789	1,213	576	663	1,126	1,047	742
Jalapa	2,292	1,744	548	652	1,640	1,326	966
Jutiapa	2,993	2,530	463	1,531	1,462	1,828	1,165

Fuente: elaboración propia con datos obtenidos de la ENCOVI 2014.

En lo respecta a la privación censurada del indicador de pobreza multidimensional “trabajo infantil” se obtuvieron 482 observaciones en promedio, 970 unidades de observación promedio en la privación censurada “materiales de la vivienda” y 872 observaciones promedio en la privación censurada “hacinamiento”. Estos valores tienen implicaciones importantes en el cálculo de las varianzas del diseño complejo de la ENCOVI 2014. En esencia, a menor

número de observaciones mayor grado de imprecisión de las estimaciones. La siguiente tabla detalla los tamaños de muestra observados para las privaciones no censuradas del IPM-Gt: combustible para cocinar, acceso a agua y energía eléctrica.

Tabla XXXII. Tamaños de muestra parte 5, según ENCOVI 2014

Departamento	Tamaño de la muestra	Combustible para cocinar		Acceso a agua		Energía eléctrica	
		No privado	Privación censurada	No privado	Privación censurada	No privado	Privación censurada
Guatemala	4,204	3,408	796	3,650	554	3,972	232
El Progreso	1,992	1,087	905	1,430	562	1,585	407
Sacatepéquez	3,747	3,125	622	3,331	416	3,500	247
Chimaltenango	2,523	1,868	655	1,870	653	2,024	499
Escuintla	3,140	1,652	1,488	2,610	530	2,502	638
Santa Rosa	1,687	621	1,066	1,227	460	1,086	601
Sololá	2,283	1,943	340	2,169	114	1,750	533
Totonicapán	2,233	1,697	536	1,913	320	1,918	315
Quetzaltenango	3,322	2,523	799	2,904	418	2,767	555
Suchitepéquez	3,020	1,204	1,816	2,715	305	1,991	1,029
Retalhuleu	1,959	620	1,339	1,598	361	1,240	719
San Marcos	2,167	1,346	821	1,680	487	1,447	720
Huehuetenango	2,435	1,936	499	1,967	468	1,793	642
Quiché	2,494	1,364	1,130	2,112	382	1,698	796
Baja Verapaz	1,691	574	1,117	1,208	483	867	824
Alta Verapaz	2,153	332	1,821	1,143	1,010	566	1,587
Petén	2,739	666	2,073	1,991	748	1,445	1,294
Izabal	1,672	812	860	1,337	335	1,137	535
Zacapa	1,548	970	578	1,103	445	982	566
Chiquimula	1,789	859	930	1,359	430	952	837
Jalapa	2,292	1,233	1,059	1,605	687	1,332	960
Jutiapa	2,993	1,289	1,704	2,172	821	2,180	813

Fuente: elaboración propia con datos obtenidos de la ENCOVI 2014.

En la tabla anterior se observa un promedio de 871 observaciones para las privaciones censuradas de los tres indicadores considerados. El menor promedio lo obtuvo el indicador “acceso a agua” (500 unidades en la muestra) y el mayor número de unidades en la muestra se observó en el indicador “combustible para

cocinar” (1,415 observaciones). Por último, se presentan los tamaños de muestra obtenidos en la ENCOVI 2014 para los indicadores “recolección de basura” y “saneamiento”.

Tabla XXXIII. Tamaños de muestra parte 6, según ENCOVI 2014

Departamento	Tamaño de la muestra	Recolección de basura		Saneamiento	
		No privado	Privación censurada	No privado	Privación censurada
Guatemala	4,204	3,588	616	3,734	470
El Progreso	1,992	1,081	911	1,162	830
Sacatepéquez	3,747	3,152	595	3,340	407
Chimaltenango	2,523	1,282	1,241	1,375	1,148
Escuintla	3,140	1,706	1,434	2,183	957
Santa Rosa	1,687	679	1,008	986	701
Sololá	2,283	1,044	1,239	1,013	1,270
Totonicapán	2,233	739	1,494	659	1,574
Quetzaltenango	3,322	1,687	1,635	1,958	1,364
Suchitepéquez	3,020	1,231	1,789	1,898	1,122
Retalhuleu	1,959	552	1,407	721	1,238
San Marcos	2,167	568	1,599	697	1,470
Huehuetenango	2,435	778	1,657	851	1,584
Quiché	2,494	679	1,815	745	1,749
Baja Verapaz	1,691	507	1,184	627	1,064
Alta Verapaz	2,153	343	1,810	400	1,753
Petén	2,739	590	2,149	752	1,987
Izabal	1,672	769	903	1,002	670
Zacapa	1,548	790	758	921	627
Chiquimula	1,789	748	1,041	851	938
Jalapa	2,292	759	1,533	878	1,414
Jutiapa	2,993	1,094	1,899	1,525	1,468

Fuente: elaboración propia con datos obtenidos de la ENCOVI 2014.

De acuerdo con la distribución de los tamaños de muestra observados en las tablas anteriores, para los distintos indicadores del IPM-Gt, una de las principales limitaciones del estudio se presentó en aquellos campos de estimación en los que se observaron tamaños relativamente bajos para la privación censurada en consideración. Los coeficientes de variación superiores

al 15% obtenidos están explicados por el esquema de muestreo adoptado y los tamaños de muestra observados. Otro aspecto importante en la evaluación de la consistencia de los resultados obtenidos se dio a través del análisis de la información auxiliar que se utilizó para la aplicación del EBLUP bajo el modelo *Fay Herriot*.

4.2. Uso de variables auxiliares, estimador EBLUP bajo el modelo *Fay Herriot*

El EBLUP bajo el modelo *Fay Herriot* enlazó los indicadores de interés a nivel departamental, asumiendo que éstos varían respecto de un vector con valores de p variables auxiliares X_d de forma constante para todas las áreas o dominios, siguiendo un modelo de regresión lineal y considerando el término de error de la regresión de cada departamento como efecto aleatorio. Este efecto aleatorio representó la heterogeneidad de los indicadores de pobreza multidimensional δ_d a través de los departamentos, no debida (o no explicada) por las variables auxiliares consideradas.

La búsqueda de variables útiles para la exploración del método de estimación EBLUP basado en el modelo *Fay Herriot* se efectuó a través del análisis de la información contenida en las bases del XII Censo Nacional de Población y VII de Vivienda realizado en el año 2018 por el Instituto Nacional de Estadística INE. Los datos fueron organizados a nivel de vivienda, hogar y personas de acuerdo con las necesidades de información de cada indicador estimado bajo el modelo. La siguiente tabla resume las distintas variables auxiliares utilizadas para la aplicación del modelo, esto como resultado de una evaluación exhaustiva del mejor ajuste para la estimación de los parámetros de pobreza multidimensional de la población guatemalteca.

Tabla XXXIV. **Variables auxiliares utilizadas, estimador EBLUP bajo el modelo *Fay Herriot***

No.	Indicador	Variable auxiliar utilizada
1	Acceso a salud	Proporción de viviendas con acceso a agua potable
2	Seguridad alimentaria y nutricional	NBI calidad de la vivienda y NBI Origen y abastecimiento de agua
3	Embarazo en adolescentes	NBI Precariedad ocupacional
4	Cuidado prenatal	Proporción de niños de 0 a 6 años con relación al total de la población
5	Asistencia escolar	NBI Asistencia escolar y NBI Precariedad ocupacional
6	Años de escolaridad	Proporción de niños de 0 a 6 años con relación al total de la población
7	Rezago educativo	NBI Origen y abastecimiento de agua
8	Cuidado infantil	NBI Hacinamiento
9	Empleo informal	Proporción de personas 15 años y más analfabetas
10	Trabajo infantil	NBI Precariedad ocupacional
11	Materiales de la vivienda	Proporción de viviendas con paredes adecuadas
12	Hacinamiento	NBI Origen y abastecimiento de agua y proporción de personas en hacinamiento
13	Combustible para cocinar	Proporción de hogares con radio
14	Acceso al agua	Proporción de viviendas con acceso a energía eléctrica
15	Energía eléctrica	Proporción de hogares con radio y acceso a energía eléctrica
16	Recolección de basura	NBI Precariedad ocupacional y proporción de hogares con radio
17	Saneamiento	Proporción de hogares con televisión y servicio de cable
18	Incidencia	NBI Precariedad ocupacional
19	Intensidad	NBI Origen y abastecimiento de agua
20	IPM-Gt	NBI Origen y abastecimiento de agua

Fuente: elaboración propia con datos obtenidos del CENSO 2018.

El uso de la información anterior permitió que se mejorara la eficiencia de las estimaciones realizadas con el estimador EBLUP bajo el modelo *Fay Herriot*, ya que, al tratarse de un estimador compuesto, tomó prestada información del resto de departamentos dando mayor ponderación al estimador sintético de regresión en los casos en los que fue necesario, es decir, cuando el estimador directo presentó una mayor varianza, o menor tamaño muestral. Esta consideración técnica resultó siendo una ventaja significativa en la estimación de los indicadores de interés.

Cuando se evaluó el uso de información auxiliar para la modelación, los indicadores de pobreza multidimensional que presentaron el mayor grado de dificultar fueron: acceso a salud y embarazo en adolescentes. Sin embargo, el estimador compuesto utilizado en esta investigación tiene la ventaja de combinar el estimador directo del diseño complejo y el estimador sintético de regresión a nivel de área o dominio de estudio d , por lo que las estimaciones obtenidas son estadísticamente aceptables.

De acuerdo con los resultados obtenidos, la metodología de estimación en área pequeña utilizada demandó el uso de información auxiliar correlacionada con el fenómeno de estudio para la desagregación de los resultados del IPM-Gt. En este sentido, es importante señalar que el fortalecimiento de los sistemas de registros administrativos de un país permite optimizar los procesos inferenciales de encuestas de hogares cuando se desea desagregar resultados.

Otra consideración técnica, para el caso de Guatemala, es que la metodología empleada puede generalizarse al momento que el Instituto Nacional de Estadística (INE) incluya en su Sistema Integrado de Encuestas de Hogares (SIEH) la medición de los indicadores del IPM-Gt con una mayor periodicidad y genere desagregaciones con metodologías de estimación en áreas pequeñas. Lo que sentará las bases para obtención de estimaciones desagregadas y oficiales en el país.

Por otro lado, es estadísticamente relevante que el estimador utilizado fue diseñado para reducir al mismo tiempo el sesgo derivado del estimador sintético y la varianza del estimador basado en el diseño de muestreo. Por lo tanto, no puede tener peor eficiencia que el estimador directo ni mayor sesgo que el estimador sintético.

Si para un departamento de interés, la ponderación dada al estimador directo fue estrictamente positiva ($\gamma_d > 0$) el uso de los pesos muestrales a través del estimador directo generó estimaciones consistentes bajo el esquema de una muestra compleja. Esto significa que el estimador se vería menos afectado por diseños de tipo informativo (probabilidades de selección según la variable de interés).

Una ventaja observada en la aplicación del EBLUP bajo el modelo *Fay Herriot* fue la utilización de información auxiliar agregada, lo que evitó problemas de confidencialidad de los microdatos obtenidos del XII Censo Nacional de Población y VII de Vivienda 2018.

La estimación de los errores cuadráticos medios ECM del EBLUP bajo el modelo *Fay Herriot* utilizó el estimador de *Prasad-Rao*, el cual es estable e insesgado bajo el diseño cuando se promedia a lo largo de muchas áreas (en este caso, 22 departamentos).

La utilización de la librería SAE de Isabel Molina para el paquete RStudio permitió reajustar las sumas de los totales por departamento para que coincidieran con los totales obtenidos bajo el diseño muestral complejo de la ENCOVI 2014 (*propiedad de benchmarking*).

4.3. Calidad estadística de las estimaciones

La siguiente tabla permite cuantificar el número de estimaciones estadísticamente aceptables, con base en las 440 estimaciones realizadas por estimador.

Tabla XXXV. Evaluación del número de estimaciones aceptables

Indicador	Estimación directa con CV menor a 15%	Estimación directa con CV superior a 15%	EBLUP <i>Fay Herriot</i> con ECMR menor a 20%	EBLUP <i>Fay Herriot</i> con ECMR mayor a 20%
Acceso a salud	1	21	8	14
Seguridad alimentaria y nutricional	17	5	22	0
Embarazo en adolescentes	0	22	13	9
Cuidado prenatal	3	19	21	1
Asistencia escolar	15	7	22	0
Años de escolaridad	21	1	22	0
Rezago educativo	1	21	20	2
Cuidado infantil	0	22	18	4
Empleo informal	21	1	22	0
Trabajo infantil	11	11	22	0
Materiales de la vivienda	18	4	22	0
Hacinamiento	17	5	22	0
Combustible para cocinar	15	7	21	1
Acceso al agua	2	20	17	5
Energía eléctrica	9	13	21	1
Recolección de basura	18	4	22	0
Saneamiento	16	6	21	1
Incidencia	21	1	22	0
Intensidad	22	0	22	0
IPM-Gt	20	2	22	0
Totales	248	192	402	38
Porcentajes	56.4%	43.6%	91.4%	8.6%

Fuente: elaboración propia con datos obtenidos de la ENCOVI 2014.

La evaluación de los niveles de precisión de las estimaciones obtenidas por ambos métodos se definió con base en dos criterios. Las estimaciones directas que presentaron coeficientes de variación superiores al 15 % se consideraron no estadísticamente aceptables, para el caso de las estimaciones basadas en el EBLUP bajo el modelo *Fay Herriot* se consideró un máximo de 20 % del error cuadrático medio relativo o estándar obtenido bajo el estimador de *Prasad-Rao*; este se obtiene del cociente entre la raíz cuadrada del error cuadrático medio ECM estimado y el estimador puntual bajo el modelo.

Con base en la evaluación del número de estimaciones aceptables bajo los criterios técnicos definidos, se observó que bajo el estimador EBLUP bajo el modelo *Fay Herriot* se obtuvieron 402 estimaciones con ECMR inferior al 20 %, representando un 91.4 % de estimaciones aceptables. Sin embargo, al comparar el error relativo de muestreo (CV) obtenido bajo el estimador directo del diseño de muestreo complejo de la ENCOVI 2014 y el error cuadrático medio estándar o relativo calculado bajo el estimador EBLUP bajo el modelo *Fay Herriot*, se observó que 433 estimaciones presentaron un menor error relativo al utilizar el estimador compuesto, lo que representó un 98.4 % de mejora en proceso inferencial aplicado.

Para el caso del estimador directo se observó que el 46.3 % de las estimaciones realizadas presentaron coeficientes de variación superiores al 15 %, observándose incluso un 99.7 % de CV en el indicador de cuidado infantil correspondiente a Jutiapa. Este indicador fue reducido a 43.7 % con método de estimación en áreas pequeñas utilizado.

Por tratarse del primer ejercicio de desagregación de indicadores del IPM-Gt a nivel de departamento no fue posible comparar los resultados con algún estudio previo en Guatemala. Sin embargo, la validez externa del proceso estadístico aplicado de llevó a cabo a través de la evaluación de las experiencias internacionales en la desagregación de indicadores de pobreza multidimensional. En ese sentido, se verificó que la metodología expuesta en el presente informe cumplió con el marco de referencia propuesto por la Comisión Económica para América Latina y el Caribe (CEPAL) para la desagregación de datos estadísticos mediante las metodologías de estimación en áreas pequeñas.

Básicamente el marco de referencia para la producción de estadísticas oficiales en áreas pequeñas plantea la necesidad de especificar las variables de

interés, definir los objetivos de estimación, evaluar el tipo de indicador y los niveles de desagregación priorizados, determinar la disponibilidad de datos y cobertura, generar estimaciones iniciales bajo el estimador directo, evaluación de los errores relativos de las estimaciones, la utilización del principio de parsimonia en la selección de la información auxiliar utilizada para la aplicación del método y, finalmente, la evaluación de la incertidumbre de las estimaciones obtenidas a través del análisis de los errores cuadráticos medios relativos obtenidos.

Por otro lado, se revisó la experiencia que tienen otros institutos nacionales de estadística con relación al tema y se determinó, por ejemplo, que en Instituto Nacional de Estadística y Geografía (INEGI) de México para la evaluación de la calidad de las estimaciones utiliza como criterio que si la estimación presenta un error relativo de 0 a 15 % se considera buena, al igual que el INE de Guatemala, de 15 a 25 % aceptable y mayor de 25 % con reserva. Para el caso de CHILE, el Instituto Nacional de Estadística agrega en la evaluación el número de observaciones muestrales, definiendo los errores relativos de 0 a 15 % como estimaciones confiables (sin importar el tamaño de muestra observado), de 15 % a 20 % como descriptivo cuando es alto el número de observaciones en la muestra y estimador no confiable si tiene pocas observaciones. Estimaciones superiores al 20 % son consideradas no confiables independientemente del número de observaciones en la muestra.

Con base en los anteriores criterios, el estimador EBLUP bajo el modelo *Fay Herriot* presentó un 95.7 % de estimaciones precisas y aceptables, por lo que se garantiza la integridad y confiabilidad de los resultados obtenidos.

CONCLUSIONES

1. Se estimaron las tasas de privación no censuradas de los indicadores del IPM-Gt a nivel de departamento, obteniendo un 50 % de estimaciones aceptables bajo el estimador directo basado en el diseño muestral complejo de la ENCOVI 2014 y el 90 % mediante la metodología de estimación en área pequeña EBLUP basada en el modelo *Fay Herriot*.
2. Se estimaron los niveles de incidencia e intensidad de la pobreza multidimensional a nivel de departamento, obteniendo un 98 % de estimaciones aceptables bajo el estimador directo basado en el diseño muestral complejo de la ENCOVI 2014 y el 100 % mediante la metodología de estimación en área pequeña EBLUP basada en el modelo *Fay Herriot*.
3. Se estimó el Índice de Pobreza Multidimensional IPM-Gt a nivel de departamento, obteniendo un 91 % de estimaciones aceptables bajo el estimador directo basado en el diseño muestral complejo de la ENCOVI 2014 y el 100 % mediante la metodología de estimación en área pequeña EBLUP basado en el modelo *Fay Herriot*.
4. Se estimaron los principales indicadores de pobreza multidimensional de la población guatemalteca a nivel de departamento, obteniendo un 56 % de estimaciones aceptables bajo el estimador directo basado en el diseño muestral complejo de la ENCOVI 2014 y el 91 % mediante la metodología de estimación en área pequeña EBLUP basado en el modelo *Fay Herriot*.

RECOMENDACIONES

1. Utilizar las estimaciones obtenidas mediante la metodología de estimación en áreas pequeñas, estimador EBLUP bajo el modelo *Fay Herriot*, como estadísticas oficiales de las tasas de privación no censuradas del IPM-Gt a nivel de departamento.
2. Considerar las estimaciones obtenidas mediante la metodología de estimación en áreas pequeñas, estimador EBLUP bajo el modelo *Fay Herriot*, como estadísticas oficiales de los niveles de incidencia e intensidad de la pobreza multidimensional de la población guatemalteca a nivel de departamento.
3. Hacer uso de las estimaciones obtenidas mediante la metodología de estimación en áreas pequeñas, estimador EBLUP bajo el modelo *Fay Herriot*, como estadísticas oficiales del Índice de Pobreza Multidimensional IPM-Gt a nivel de departamento
4. Emplear las estimaciones obtenidas mediante la metodología de estimación en áreas pequeñas, estimador EBLUP bajo el modelo *Fay Herriot*, como estadísticas oficiales de los niveles de pobreza multidimensional de la población guatemalteca a nivel de departamento.

REFERENCIAS

1. Alkire, S. y Foster, J. (2008). *Recuento y medición multidimensional de la pobreza*. Recuperado de <https://www.ophi.org.uk/wp-content/uploads/ophi-wp7-es.pdf>.
2. Álvarez, G. (2001). *Estimación de población en áreas menores mediante variables sintomáticas: una aplicación para los departamentos de la República Argentina (1991 y 1996)*. Santiago, Chile: Naciones Unidas.
3. Celdrán, D. (2012). *Estimación Trimestral en Áreas Pequeñas de los Ocupados de la Encuesta de Población Activa Según su Actividad Económica* (tesis de maestría). Universidad de Santiago de Compostela. Galicia, España.
4. Cochran, W. G. (1974). *Técnicas de Muestreo (Trad. E. Casas)*. En Compañía Editorial Continental, S. A., México D.F., México (Reimpreso de *Sampling Techniques*, Ed. 1971, New York, Estados Unidos de Norte América: JOHN WILEY & SONS, INC.).
5. Comisión Económica para América Latina y el Caribe CEPAL (2018). *Medición de la pobreza por ingresos: Actualización metodológica y resultados*. Santiago, Chile: Naciones Unidas.

6. Consejo Nacional de Desarrollo Urbano y Rural (2014). *Plan Nacional de Desarrollo K'atun: nuestra Guatemala 2032*. Guatemala, Guatemala: Conadur/Segeplán.
7. Departamento de Asuntos Económicos y Sociales (DESA 2009). *Estudios de métodos Serie F No. 98: Diseño de muestras para encuestas de hogares: directrices prácticas*. New York, Estados Unidos de Norte América: Naciones Unidas.
8. Barneche, P., Bugallo, A., Ferrea, H., Ilarregui, M., Monterde, C., Pérez, V., Santa María, T., ..., Angeletti, K. (agosto de 2010). *Métodos de Medición de la Pobreza*. Conceptos y aplicaciones en América Latina. *Entrelíneas de la Política Económica*. Volumen 4 (Nº 26). 31-41.
9. Elbers, C., Lanjouw, J. y Lan, P. (2002). Policy Research Working Paper 2911: *Micro-Level Estimation of Welfare*. Washington, DC, Estados Unidos de Norte América: The World Bank. Development Research Group. Poverty Team.
10. Feres, J. y Mancero, X. (2001). *El Método de las Necesidades Básicas Insatisfechas (NBI) y sus aplicaciones en América Latina*. Santiago de Chile, Chile: Naciones Unidas.
11. Instituto Nacional de Estadística INE, Proyecto MECOVI (2002). *Encuesta Nacional de Condiciones de Vida ENCOVI 2000: Perfil de la Pobreza en Guatemala*. Ciudad de Guatemala, Guatemala: Editorial Serviprensa, S. A.

12. Instituto Nacional de Estadística INE (2006). *Necesidades Básicas Insatisfechas al 2002*. Ciudad de Guatemala, Guatemala: Instituto Nacional de Estadística INE.
13. Instituto Nacional de Estadística INE (2016). *Encuesta Nacional de Condiciones de Vida 2014*. Ciudad de Guatemala, Guatemala: Instituto Nacional de Estadística INE.
14. Gutiérrez, A. (2015). *Estrategias de Muestreo, Diseño de Encuestas y Estimación de Parámetros*. Bogotá, Colombia: Universidad Santo Tomas.
15. Kish, L. (1972). *Muestreo de encuestas* (Trad. R. V. Cruz). México D.F., México (Reimpreso de Survey Sampling, Ed. 1965, New York, Estados Unidos de Norte América: JOHN WILEY & SONS, INC.).
16. Larrañaga, O. (2007). *Serie 58 estudios estadísticos y prospectivos: la medición de la pobreza en dimensiones distintas al ingreso*. Santiago, Chile: División de Estadística y Proyecciones Económicas CEPAL.
17. Lohr, S. (2000) *Muestreo: Diseño y Análisis* (Trad. O. A. Palmas). México D.F., México (Reimpreso de *Sampling: Design and Analysis*, Ed. 1999, Arizona, Estados Unidos de Norte América: Duxbury Press.)
18. Mancho, J. (2002). *Técnicas de Estimación en Áreas Pequeñas*. Euskadi, España: Eustat.

19. Martínez, C. (2012). *Estadística y Muestreo*. Bogotá, Colombia: Imagen Editorial Impresores.
20. Ministerio de Desarrollo Social (2018). *Índice de Pobreza Multidimensional de Guatemala (IPM-Gt)*. Recuperado de https://mppn.org/wp-content/uploads/2019/10/Guatemala-Report-IPM-gt_29jul19-v1.1.pdf
21. Mirás, J. (1985). *Elementos de Muestreo para Poblaciones Finitas*. Madrid, España: Instituto Nacional de Estadística.
22. Molina, I. (2019). *Serie estudios estadísticos 97: Desagregación de datos en encuestas de hogares. Metodologías de estimación en áreas pequeñas*. Santiago, Chile: Naciones Unidas.
23. Naciones Unidas (2018), *La Agenda 2030 y los Objetivos de Desarrollo Sostenible: una oportunidad para América Latina y el Caribe (LC/G.2681-P/Rev.3)*, Santiago, Chile: Naciones Unidas.
24. Pérez, C. (2010). *Técnicas de muestreo estadístico*. Madrid, España: IBERGARCETA PUBLICACIONES, S.L.
25. Pérez, C. (2005). *Muestreo Estadístico, Conceptos y Problemas Resueltos*. Madrid, España: PEARSON EDUCACIÓN, S.A.
26. Rao, J. N. K. (2005). *Small Area Estimation*. New Jersey, Estados Unidos de Norte América: JOHN WILEY & SONS, INC.

27. Rao, J. N. K. (2005). *Survey Methodology*. A Journal Published by Statistics Canada, Volumen 31, (No. 2). p. 117-138.
28. Rao, J. N. K. y Molina, I. (2015). *Small Area Estimation*. New Jersey, Estados Unidos de Norte América: JOHN WILEY & SONS, INC.
29. Rawls, J. (1971). *A Theory of Justice*. Massachusetts, Estados Unidos de Norte América: Harvard University.
30. Romero, A. (2000). *El mundo de la pobreza (primera parte)*. Revista TENDENCIAS, Volumen 1, (No.2). p. 35-59.
31. Romero, W. y Zapil, P., (2009). *Dinámica Territorial del consumo, la pobreza y la desigualdad en Guatemala: 1998-2006. Documento de Trabajo N° 51. Programa Dinámicas Territoriales Rurales*. Santiago, Chile: Rimisp, Centro Latinoamericano para el Desarrollo Rural.
32. Särndal, C., Swensson, B. y Wretman, J. (1992) *Model Assisted Survey Sampling*. New York, Estados Unidos de Norte America: Springer Series in Statistics.
33. Sen, A. (2000), *Desarrollo y Libertad*, Buenos Aires, Argentina: Editorial Planeta Argentina, S. A. I. C.
34. Yansaneh, I. S., (2007). Introducción. En I. S. Yansaneh, (Ed.). *Estudios de métodos Serie F No. 96: Encuestas de hogares en los países en desarrollo y en transición*. (pp. 3). New York, Estados Unidos: Naciones Unidas.

APÉNDICES

A continuación, se muestra la sintaxis utilizada para replicar el Índice de Pobreza Multidimensional de Guatemala IPM-Gt en el paquete estadístico SPSS.

Apéndice 1. Sintaxis para replicar el IPM-Gt

Código en SPSS
<p>*Consideraciones técnicas: ***Todos los cálculos se han efectuado a partir de las bases de datos de la Encuesta Nacional de Condiciones de Vida -ENCOVI- 2014. ***Las bases de datos están disponibles en la página del Instituto Nacional de Estadística ***disponibles en el siguiente link: https://www.ine.gob.gt/index.php/encuestas-de-hogares-y-personas/condiciones-de-vida ***Base Personas 2014: https://www.ine.gob.gt/sistema/uploads/2015/12/11/JD6quWN8hqDw0KFpPOGuPZLc8d1Wih7l.sav. ***Base Hogares 2014: https://www.ine.gob.gt/sistema/uploads/2015/12/11/ZYdzkwlx6qvAqZ05Uimcw1131bm04lh.sav. *****</p> <p>INTEGRACION DE DATOS PARA EL PROCESO DEL IPM-GT *****</p> <p>*Abrir la base de datos "Personas" de la Encuesta y ordenarla por la variable llave llamada "NUMHOG". GET FILE='C:/IPM-Gt/JD6quWN8hqDw0KFpPOGuPZLc8d1Wih7l.sav'. DATASET NAME Personas WINDOW=FRONT. SORT CASES BY NUMHOG (A).</p> <p>*Abrir la base de datos "Hogares" de la Encuesta y ordenarla por la variable llave llamada "NUMHOG". GET FILE='C:/IPM-Gt/ZYdzkwlx6qvAqZ05Uimcw1131bm04lh.sav'. DATASET NAME Hogares WINDOW=FRONT. SORT CASES BY NUMHOG (A).</p> <p>***Para facilitar el cálculo de los indicadores, se integran ambas bases de datos a nivel de personas. ****Debe recordarse que la unidad de identificación de la pobreza multidimensional es el hogar y la unidad de análisis son las personas.</p> <p>DATASET ACTIVATE Personas. STAR JOIN /SELECT t0.REGION, t0.DEPTO, t0.AREA, t0.UPM, t0.FACTOR, t0.POBREZA, t0.THOGAR, t0.ID, t0.PPA02, t0.PPA03, t0.PPA04A, t0.PPA04B, t0.PPA04C, t0.PPA04C, t0.PPA05, t0.PPA06, t0.PPA07, t0.PPA08, t0.PPA09, t0.P03A01, t0.P03A02, t0.P03A03B, t0.P03A04, t0.P03B01, t0.P03B02, t0.P04A01A, t0.P04A01B, t0.P04A02, t0.P04A03A, t0.P04A04A, t0.P04A04B, t0.P04A05, t0.P04A06A, t0.P04A07A, t0.P04A08A, t0.P04A09A, t0.P04A10, t0.P04A11A, t0.P05A01, t0.P05A02, t0.P05A03, t0.P05A04, t0.P05A05, t0.P05A06, t0.P05B01A, t0.P05B01B, t0.P05B02, t0.P05B03A, t0.P05B03B, t0.P05B04, t0.P05B05A, t0.P05B05B, t0.P05B06A, t0.P05B06B, t0.P05B06C, t0.P05B07, t0.P05C01, t0.P05C02, t0.P05C03, t0.P05C04, t0.P05C05, t0.P05C06, t0.P05C07A, t0.P05C07B, t0.P05C08A, t0.P05C08B, t0.P05C09A, t0.P05C09B, t0.P05C09C, t0.P05C09D, t0.P05C09E, t0.P05C09F, t0.P05C09G, t0.P05C10, t0.P05D01, t0.P05D02, t0.P05D03A, t0.P05D03B, t0.P05D04A, t0.P05D04B, t0.P05D05, t0.P05D06A, t0.P05D06B, t0.P05D07, t0.P05D08A, t0.P05D08B, t0.P05D09A, t0.P05D09B, t0.P05D10A, t0.P05D10B, t0.P05D11A, t0.P05D11B, t0.P05D12A, t0.P05D12B, t0.P05D13A, t0.P05D13B, t0.P05D14A, t0.P05D14B, t0.P05D15A, t0.P05D15B, t0.P05D16A, t0.P05D16B, t0.P05D17A, t0.P05D17B, t0.P05D18A, t0.P05D18B, t0.P05D18C, t0.P05D18D, t0.P05D18E, t0.P05D18F, t0.P05D18G, t0.P05D18H, t0.P05D19, t0.P05E01, t0.P05E02, t0.P05E03, t0.P05E04, t0.P05E05, t0.P05E06A, t0.P05E06B, t0.P05E07, t0.P05E08, t0.P05E09, t0.P05E10, t0.P05E11A, t0.P05E11B, t0.P05E12, t0.P05E13A, t0.P05E13B, t0.P05E14, t0.P05E15, t0.P05E16A, t0.P05E16B, t0.P05E17, t0.P05E18, t0.P05E19, t0.P05E20, t0.P05E21A, t0.P05E21B, t0.P05E22, t0.P05E23A, t0.P05E23B, t0.P05E23C, t0.P05E23D, t0.P05E23E, t0.P05E23F, t0.P05E23G, t0.P05E23H, t0.P05E23I, t0.P05E23J, t0.P05E24A, t0.P05E24B, t0.P05E24C, t0.P05E24D, t0.P05E24E, t0.P05E24F, t0.P05E24G, t0.P05E24H, t0.P05E24I, t0.P05E24J, t0.P06A01, t0.P06A02, t0.P06A03A,</p>

Continúa apéndice 1.

Código en SPSS
<p>t0.P06A03B, t0.P06A04A, t0.P06A04B, t0.P06A05A, t0.P06A05B, t0.P06A06A, t0.P06A06B, t0.P06A07A, t0.P06A07B, t0.P06A08A, t0.P06A08B, t0.P06A09A, t0.P06A09B, t0.P06B01, t0.P06B02A, t0.P06B03C, t0.P06B03, t0.P06B04, t0.P06B05, t0.P06B06A, t0.P06B06B, t0.P06B07A, t0.P06B07C, t0.P06B08A, t0.P06B08B, t0.P06B09, t0.P06B10A, t0.P06B10B, t0.P06B11A, t0.P06B11B, t0.P06B12A, t0.P06B12B, t0.P06B13A, t0.P06B13B, t0.P06B14A, t0.P06B14B, t0.P06B15, t0.P06B16A, t0.P06B16B, t0.P06B17, t0.P06B18A, t0.P06B18B, t0.P06B19A, t0.P06B19B, t0.P06B20A, t0.P06B20B, t0.P06B21A, t0.P06B21B, t0.P06B22, t0.P06B23, t0.P06B24, t0.P06B25A, t0.P06B25B, t0.P06B26, t0.P06B27A, t0.P07A01, t0.P07A02A, t0.P07A02B, t0.P07A03, t0.P07A04, t0.P07A05, t0.P07A06A, t0.P07A06C, t0.P08A01A, t0.P08A01B, t0.P08A01C, t0.P08A01D, t0.P08A01E, t0.P08A02, t0.P08A03A, t0.P08A03B, t0.P08A03C, t0.P08A03D, t0.P08A03E, t0.P08A04, t0.P09A01A, t0.P09A01B, t0.P09A01C, t0.P09A02A, t0.P09A02B, t0.P09A02C, t0.P09A03A, t0.P09A03B, t0.P09A03C, t0.P09A04A, t0.P09A04B, t0.P09A04C, t0.P09B01A, t0.P09B01B, t0.P09B01C, t0.P09B02A, t0.P09B02B, t0.P09B02C, t0.P09C01A, t0.P09C01B, t0.P09C01C, t0.P09C02A, t0.P09C02B, t0.P09C02C, t0.P09C03A, t0.P09C03B, t0.P09C03C, t0.P09C04A, t0.P09C04B, t0.P09C04C, t0.P09C04D, t0.P09C05A, t0.P09C05B, t0.P09C05C, t0.P09C06A, t0.P09C06B, t0.P09C06C, t0.P09C07A, t0.P09C07B, t0.P09C07C, t0.P09C08A, t0.P09C08B, t0.P09C08C, t0.P09C09A, t0.P09C09B, t0.P09C09C, t0.P09D01A, t0.P09D01B, t0.P09D01C, t0.P09D02A, t0.P09D02B, t0.P09D02C, t0.P09D03A, t0.P09D03B, t0.P09D03C, t0.P09E01A, t0.P09E01B, t0.P09E01C, t0.P09E02A, t0.P09E02B, t0.P09E02C, t0.P09F01A, t0.P09F01B, t0.P09F01C, t0.P09F02A, t0.P09F02B, t0.P09F03A, t0.P09F03B, t0.P09F03C, t0.P09F04A, t0.P09F04B, t0.P09F04C, t0.P09F05A, t0.P09F05B, t0.P09F05C, t0.P09F06A, t0.P09F06B, t0.P09F06C, t0.P09F07A, t0.P09F07B, t0.P09F07C, t0.P09F08A, t0.P09F08B, t0.P09F08C, t0.P09F09A, t0.P09F09B, t0.P09F09C, t0.P09G01A, t0.P09G01B, t0.P09H01A, t0.P09H01B, t0.P09H01C, t0.P09H01D, t0.P09I01, t0.P09I02, t0.P09I03, t0.P09I04, t0.P09I05, t0.P10A01, t0.P10A02, t0.P10A03, t0.P10A04, t0.P10A05, t0.P10A06, t0.P10A07, t0.P10B01, t0.P10B02B, t0.P10B02B_1D, t0.P10B03B, t0.P10B03B_1D, t0.P10B04, t0.P10B05, t0.P10B06, t0.P10B07A, t0.P10B07B, t0.P10B08, t0.P10B09A, t0.P10B09B, t0.P10B09C, t0.P10B10A, t0.P10B10B, t0.P10B11A, t0.P10B11B, t0.P10B12A, t0.P10B12B, t0.P10B13A, t0.P10B13B, t0.P10B14A, t0.P10B14B, t0.P10B15A, t0.P10B15B, t0.P10B16A, t0.P10B16B, t0.P10B17, t0.P10B18A, t0.P10B18B, t0.P10B19A, t0.P10B19B, t0.P10B20A, t0.P10B20B, t0.P10B21A, t0.P10B21B, t0.P10B22, t0.P10B23, t0.P10B24A, t0.P10B24B, t0.P10B25, t0.P10B26A, t0.P10B26B, t0.P10B27A, t0.P10B27B, t0.P10B27C, t0.P10B27D, t0.P10B27E, t0.P10B27F, t0.P10B27G, t0.P10C01, t0.P10C02B, t0.P10C02B_1D, t0.P10C03B, t0.P10C03B_1D, t0.P10C04, t0.P10C05, t0.P10C06A, t0.P10C06B, t0.P10C07A, t0.P10C07B, t0.P10C08A, t0.P10C08B, t0.P10C09A, t0.P10C09B, t0.P10C10A, t0.P10C10B, t0.P10C11, t0.P10C12, t0.P10C13A, t0.P10C13B, t0.P10C14, t0.P10C15, t0.P10D01, t0.P10D02A, t0.P10D02B, t0.P10D02C, t0.P10D03, t0.P10D04, t0.P10D05, t0.P10E01, t0.P10E02B, t0.P10E02B_1D, t0.P10E03, t0.P10E04, t0.P10E05, t0.P10F01, t0.P10F02, t0.P10F03, t0.P10F04, t0.P10F05, t0.P10F06, t0.P10F07A, t0.P10F07B, t0.P11A01A, t0.P11A01B, t0.P11A02A, t0.P11A02B, t0.P11A03A, t0.P11A03B, t0.P11A04A, t0.P11A04B, t0.P11A05A, t0.P11A05B, t0.P11A06A, t0.P11A06B, t0.P11A07A, t0.P11A07B, t0.P11A08A, t0.P11A08B, t0.P11A09A, t0.P11A09B, t0.P11A10A, t0.P11A10B, t0.P11B01A, t0.P11B01B, t0.P11B02A, t0.P11B02B, t0.P11B03A, t0.P11B03B, t0.P11B04A, t0.P11B04B, t1.FACTOR3, t1.PPB01, t1.PPB02, t1.PPB03, t1.PPB04, t1.PPD04, t1.PPD04A, t1.PPD04B, t1.PPD04C, t1.PPD06, t1.PPD07, t1.PPD08, t1.PPA11A, t1.PPA11B, t1.P01A01, t1.P01A02, t1.P01A03, t1.P01A04, t1.P01A05A, t1.P01A05B, t1.P01A05C, t1.P01A05D, t1.P01A05E, t1.P01A05F, t1.P01A06, t1.P01B01, t1.P01B02, t1.P01B03, t1.P01C01, t1.P01C02A, t1.P01C02B, t1.P01C02C, t1.P01C02D, t1.P01C02E, t1.P01C02F, t1.P01C02G, t1.P01C02H, t1.P01C02I, t1.P01C02J, t1.P01C02K, t1.P01C02L, t1.P01C02M, t1.P01C03A, t1.P01C03B, t1.P01C03C, t1.P01C03D, t1.P01C03E, t1.P01C04, t1.P01D01, t1.P01D02, t1.P01D03, t1.P01D04, t1.P01D05, t1.P01D06, t1.P01D07, t1.P01D08, t1.P01D09, t1.P01D10, t1.P01D11, t1.P01D12, t1.P01D13A, t1.P01D13B, t1.P01D14, t1.P01D15A, t1.P01D15B, t1.P01D16, t1.P01D17, t1.P01D18, t1.P01D19A, t1.P01D19B, t1.P01D19C, t1.P01D19D, t1.P01D20A, t1.P01D20B, t1.P01D20C, t1.P01D20D, t1.P01D21, t1.P01D22A, t1.P01D22B, t1.P01D23A, t1.P01D24, t1.P01D25, t1.P01D26, t1.P01D27, t1.P01D28, t1.P01D29A, t1.P01D29B, t1.P01D29C, t1.P01D30, t1.P01D31, t1.ID_FUENTES_1, t1.P01E01_1, t1.P01E02_1, t1.UNIDAD_FUENTES_1, t1.P01E03_1, t1.ID_FUENTES_2, t1.P01E01_2, t1.P01E02_2, t1.UNIDAD_FUENTES_2, t1.P01E03_2, t1.ID_FUENTES_3, t1.P01E01_3, t1.P01E02_3, t1.UNIDAD_FUENTES_3, t1.P01E03_3, t1.ID_FUENTES_4, t1.P01E01_4, t1.P01E02_4, t1.UNIDAD_FUENTES_4, t1.P01E03_4, t1.ID_FUENTES_5, t1.P01E01_5, t1.P01E02_5, t1.UNIDAD_FUENTES_5, t1.P01E03_5, t1.ID_FUENTES_6, t1.P01E01_6, t1.P01E02_6, t1.UNIDAD_FUENTES_6, t1.P01E03_6, t1.ID_FUENTES_7, t1.P01E01_7, t1.P01E02_7, t1.UNIDAD_FUENTES_7, t1.P01E03_7, t1.ID_FUENTES_8, t1.P01E01_8, t1.P01E02_8, t1.UNIDAD_FUENTES_8, t1.P01E03_8, t1.ID_FUENTES_9, t1.P01E01_9, t1.P01E02_9, t1.UNIDAD_FUENTES_9, t1.P01E03_9, t1.P01G01, t1.P01H01, t1.P01H02, t1.P01H03, t1.P01H04, t1.P01H05, t1.P01H06, t1.P01H07, t1.P01H08, t1.P01H09, t1.P01H10, t1.P01H11, t1.P01H12, t1.P01H13, t1.P01H14, t1.P01H15, t1.P01H16, t1.DIA_ENC, t1.MES_ENC, t1.A_ENC</p>
<p>/FROM * AS t0 /JOIN 'Hogares' AS t1 ON t0.NUMHOG=t1.NUMHOG /OUTFILE FILE=*</p>

Continúa apéndice 1.

Código en SPSS

```
*Conservar únicamente la base de datos integrada para trabajar sobre ella todos los cálculos.
DATASET CLOSE Hogares.
SAVE OUTFILE='C:\IPM-Gt\IPM-Gt - Datos de Encovi integrada 2014.sav'
/COMPRESSED.

*****
CALCULO DE INDICADORES PRINCIPALES
*****

*Observaciones: los indicadores fueron construidos siguiendo dos procedimientos:
**En el caso de indicadores que impliquen características de personas se procedió de la siguiente forma (pasos):
***1) Identificar a la población de referencia del indicador (prefijo PR).
***2) Identificar a los hogares con población de referencia entre sus integrantes (prefijo PR).
***3) Identificar la privación a nivel de personas, dentro de la población de referencia (prefijo Priv).
***4) Identificar a los hogares con privación en el indicador (prefijo hh_d_).
**En el caso de indicadores que aborden características a nivel de hogar o vivienda se procedió a identificar la privación según el paso
4.

*DIMENSION: SALUD Y SEGURIDAD ALIMENTARIA Y NUTRICIONAL.
***Dimensión Salud y SAN: Indicador Acceso a salud.
*Paso 1: calcular variable de población de Referencia.
IF (ppa03 <= 5) PR_salud_acceso=0.
IF (ppa03 > 5) PR_salud_acceso=1.
EXECUTE.

*Paso 2: identificar hogares con población de referencia.
AGGREGATE
/OUTFILE=* MODE=ADDVARIABLES
/BREAK=numhog
/PRHOG_salud_acceso=MAX(PR_salud_acceso).
EXECUTE.

*Paso 3: privación en personas.
IF (p05d05 = 3 | p05d05 = 4 | p05d05 = 5 | p05d05 = 7 | p05d05 = 8 | p05d05 = 9 | p05d05 = 10 | p05d05 = 98) priv_salud_acceso=1.
IF (p05d05 = 1 | p05d05 = 2 | p05d05 = 6 ) priv_salud_acceso=0.
IF p05d01 = 2 priv_salud_acceso=0.
IF p05d02 = 2 priv_salud_acceso=0.
IF p05d04a <=5 priv_salud_acceso=0.
IF ppa03 <=5 priv_salud_acceso=$SYSMIS.

*Paso 4: identificar hogares con privación.
AGGREGATE
/OUTFILE=* MODE=ADDVARIABLES
/BREAK=numhog
/hh_d_salud =MAX(priv_salud_acceso).
EXECUTE.

VALUE LABELS hh_d_salud priv_salud_acceso
1 'Privado'
0 'No privado'.

VARIABLE LABELS
PR_salud_acceso 'Indicador Acceso a Salud: población de referencia'
PRHOG_salud_acceso 'Indicador Acceso a Salud: hogares con población de referencia'
priv_salud_acceso 'Indicador Acceso a Salud: personas privadas'
hh_d_salud 'Indicador Acceso a Salud: hogares privados'.

***Dimensión Salud y SAN: Indicador Embarazo en Adolescentes.
*Paso 1: calcular variable de población de Referencia.
IF (ppa03 >= 12 & ppa03 <=19 & ppa02=2) PR_embarazo=1.
IF (ppa03 < 12 | ppa03 > 19) | ppa02=1 PR_embarazo=0.
EXECUTE.
```

Continúa apéndice 1.

Código en SPSS
<p>*Paso 2: identificar hogares con poblacion de referencia. AGGREGATE /OUTFILE=* MODE=ADDVARIABLES /BREAK=numhog /PRHOG_embarazo=MAX(PR_embarazo). EXECUTE.</p> <p>*Paso 3: privación en personas. IF (PR_embarazo=1 & p05e04>0) priv_embarazo=1. IF (PR_embarazo=1 & p05e04=0) priv_embarazo=0.</p> <p>*Paso 4: identificar hogares con privación. AGGREGATE /OUTFILE=* MODE=ADDVARIABLES /BREAK=numhog /hh_d_emba=MAX(priv_embarazo). EXECUTE.</p> <p>IF PRHOG_embarazo=0 hh_d_emba=0.</p> <p>VALUE LABELS hh_d_emba priv_embarazo 0 'No privado' 1 'Privado'.</p> <p>VARIABLE LABELS PR_embarazo 'Indicador Embarazo en adolescentes: población de referencia' PRHOG_embarazo 'Indicador Embarazo en adolescentes: hogares con población de referencia' priv_embarazo 'Indicador Embarazo en adolescentes: personas privadas' hh_d_emba 'Indicador Embarazo en adolescentes: hogares privados'.</p> <p>***Dimensión Salud y SAN: Indicador Inseguridad Alimentaria y Nutricional -ELCSA. *Calcular el total de experiencias de InSAN con respuesta afirmativa, segun la presencia o no de menores de 18 años. *Recodificar los NS/NR como valores perdidos, lo exige la metodologia de la ELCSA. RECODE p01h01 p01h02 p01h03 p01h04 p01h05 p01h06 p01h07 p01h08 p01h10 p01h11 p01h12 p01h13 p01h14 p01h15 p01h16 (9 , 98, 99 =SYSMIS). *Para el computo se suman las dicotómicas, el código "2=No" se convierte ahora en 0. RECODE p01h01 p01h02 p01h03 p01h04 p01h05 p01h06 p01h07 p01h08 p01h10 p01h11 p01h12 p01h13 p01h14 p01h15 p01h16 (2=0). *Calcular la escala para hogares que si tienen menores de 18 años. IF (p01h09=1) puntajeELCSA= p01h01+p01h02+p01h03+p01h04+ p01h05+p01h06+p01h07+p01h08+p01h10+p01h11+p01h12+p01h13+p01h14+p01h15+p01h16. *Calcular la escala para hogares que no tienen menores de 18 años. IF (p01h09=2) puntajeELCSA= p01h01+p01h02+p01h03+p01h04+p01h05+p01h06+p01h07+p01h08. VARIABLE LABELS puntajeELCSA 'Puntaje ELCSA experiencia en SAN'. EXECUTE.</p> <p>**Categorizar al hogar segun la experiencia de InSAN, de acuerdo a la presencia o no de menores de 18 años. IF (p01h09=1 & (puntajeELCSA>=1 & puntajeELCSA<=5)) categoriaELCSA=1. IF (p01h09=1 & (puntajeELCSA>=6 & puntajeELCSA<=10)) categoriaELCSA=2. IF (p01h09=1 & (puntajeELCSA>=11 & puntajeELCSA<=15)) categoriaELCSA=3. IF (p01h09=2 & (puntajeELCSA>=1 & puntajeELCSA<=3)) categoriaELCSA=1. IF (p01h09=2 & (puntajeELCSA>=4 & puntajeELCSA<=6)) categoriaELCSA=2. IF (p01h09=2 & (puntajeELCSA>=7 & puntajeELCSA<=8)) categoriaELCSA=3. IF (puntajeELCSA=0) categoriaELCSA=0.</p> <p>VARIABLE LABELS categoriaELCSA 'Categoria SAN del hogar según la ELCSA'. VALUE LABELS categoriaELCSA 0 'Hogar seguro alimentariamente ELCSA' 1 'Inseguridad alimentaria leve ELCSA' 2 'Inseguridad alimentaria moderada ELCSA' 3 'Inseguridad alimentaria severa ELCSA'.</p> <p>*Identificar los hogares privados a partir de Inseguridad alimentaria moderada y severa. COMPUTE hh_d_segal = categoriaELCSA>=2. IF categoriaELCSA = \$SYSMIS hh_d_segal=\$SYSMIS. VALUE LABELS hh_d_segal 1 'Privado' 0 'No privado'. VARIABLE LABELS hh_d_segal 'Indicador Inseguridad Alimentaria y Nutricional: hogares privados'.</p>

Continúa apéndice 1.

Código en SPSS
<pre> ***Dimensión Salud y SAN: Indicador Cuidado Prenatal. *Paso 1: calcular variable de población de Referencia: Mujeres en edad fértil (12-49 años). *Identificar embarazos en los últimos 5 años. IF (p05e06b< 2009) ene_9=0. IF (p05e06b>=2009) ene_9=1. *Seleccionar a la población de referencia: mujeres en edad fértil (12-49 años). IF (ppa02=2 & (ppa03>=12 & ppa03<=49)) PR_prenatal=1. IF (ppa02=2 & (ppa03< 12 ppa03> 49)) PR_prenatal=0. IF ppa02=1 PR_prenatal = \$SYSMIS. *Identificar hogares en los cuales hay al menos una mujer. COMPUTE mujer = ppa02=2. AGGREGATE /OUTFILE=* MODE=ADDVARIABLES /BREAK=numhog /hh_mujer=sum(mujer). EXECUTE. *Paso 2: identificar hogares con población de referencia. AGGREGATE /OUTFILE=* MODE=ADDVARIABLES /BREAK=numhog /PRHOG_prenatal=MAX(PR_prenatal). EXECUTE. if hh_mujer= 0 PRHOG_prenatal=0. *Paso 3: privación en personas. DO IF (PR_prenatal=1 & p05e04>0). COMPUTE priv_prenatal = 0. END IF. DO IF (PR_prenatal=1 & p05e04>0). IF (p05e07=95 p05e07=96) priv_prenatal=1. END IF. DO IF (PR_prenatal=1 & p05e04>0). IF (p05e08<4) priv_prenatal=1. END IF. DO IF (PR_prenatal=1 & p05e04>0). IF (p05e07=\$SYSMIS & p05e08=\$SYSMIS) priv_prenatal=\$SYSMIS. END IF. IF (PR_prenatal=1 & p05e04=0) priv_prenatal=0. IF (p05e06b<2009 & p05e04>0 & PR_prenatal=1) priv_prenatal=0. *Paso 4: identificar hogares con privación. AGGREGATE /OUTFILE=* MODE=ADDVARIABLES /BREAK=numhog /hh_d_prenatal=MAX(priv_prenatal). EXECUTE. IF PRHOG_prenatal=0 hh_d_prenatal=0. VALUE LABELS hh_d_prenatal 0 'No privado' 1 'Privado'. VARIABLE LABELS PR_prenatal 'Indicador Cuidado Prenatal: poblacion de referencia' PRHOG_prenatal 'Indicador Cuidado Prenatal: hogares con poblacion de referencia' priv_prenatal 'Indicador Cuidado Prenatal: personas privadas' hh_d_prenatal 'Indicador Cuidado Prenatal: hogares privados'. </pre>

Continúa apéndice 1.

Código en SPSS

```
*Paso 3: privación en personas.
IF (ppa03 > 17 & ppa03 <=32 & anios_escolar <9) priv_aniosesco=1.
IF (ppa03 > 17 & ppa03 <=32 & anios_escolar >=9) priv_aniosesco=0.
IF (ppa03 > 32 & ppa03 <=64 & anios_escolar <6) priv_aniosesco=1.
IF (ppa03 > 32 & ppa03 <=64 & anios_escolar >=6) priv_aniosesco=0.
IF (ppa03 > 64 & p06b01=2) priv_aniosesco=1.
IF (ppa03 > 64 & p06b01=1) priv_aniosesco=0.
IF (ppa03>17 & ppa03 <=64) & anios_escolar=$SYSMIS priv_aniosesco=$SYSMIS.
IF (ppa03>64) & p06b01=$SYSMIS priv_aniosesco=$SYSMIS.

*Paso 4: identificar hogares con privación.
AGGREGATE
  /OUTFILE=* MODE=ADDVARIABLES
  /BREAK=numhog
  /hh_d_escol=MAX(priv_aniosesco).
EXECUTE.

IF PRHOG_aniosesco=0 hh_d_escol=0.
VALUE LABELS hh_d_escol
0 'No privado'
1 'Privado'.

VARIABLE LABELS
PR_aniosesco 'Indicador Escolaridad en Adultos: poblacion de referencia'
PRHOG_aniosesco 'Indicador Escolaridad en Adultos: hogares con poblacion de referencia'
priv_aniosesco 'Indicador Escolaridad en Adultos: personas privadas'
hh_d_escol 'Indicador Escolaridad en Adultos: hogares privados'.

***dimensión educación: Indicador Rezago educativo.
*Paso 1: calcular variable de poblacion de Referencia.
IF ppa03 >= 8 & ppa03 < 18 PR_rezago=1.
IF ppa03 < 8 | ppa03 >= 18 PR_rezago=0.
EXECUTE.

*Paso 2: identificar hogares con población de referencia.
AGGREGATE
  /OUTFILE=* MODE=ADDVARIABLES
  /BREAK=numhog
  /PRHOG_rezago=MAX(PR_rezago).
EXECUTE.

*Variable auxiliar para cálculo del indicador: años de estudios grado actual.
*bajo el supuesto que diversificado = 12 años, y a partir de universitario se coloca constante 13.
COMPUTE anios_escoactual=$SYSMIS.
IF p06b06a = 0 | p06b06a=1 anios_escoactual=0.
IF p06b06a=2 anios_escoactual=p06b06b.
IF p06b06a=3 anios_escoactual=6+p06b06b.
IF p06b06a=4 anios_escoactual=6+p06b06b.
IF p06b06a>=5 anios_escoactual=13.

*Paso 3: privación en personas.
IF PR_rezago = 1 & (ppa03 - 6 - anios_escoactual) > 2 priv_rezago=1.
IF PR_rezago = 1 & (ppa03 - 6 - anios_escoactual) <= 2 priv_rezago=0.
IF PR_rezago = 1 & p06b05=2 priv_rezago=0.
IF PR_rezago = 0 priv_rezago= $sysmis.
IF (ppa03 - 6 - anios_escoactual)=$sysmis priv_rezago=$sysmis.

*Paso 4: identificar hogares con privación.
AGGREGATE
  /OUTFILE=* MODE=ADDVARIABLES
  /BREAK=numhog
  /hh_d_reza=MAX(priv_rezago).
EXECUTE.

IF PRHOG_rezago = 0 hh_d_reza =0.

VALUE LABELS hh_d_reza
0 'No privado'
1 'Privado'.
```

Continúa apéndice 1.

Código en SPSS
<p>VARIABLE LABELS PR_rezago 'Indicador Rezago Educativo: población de referencia' PRHOG_rezago 'Indicador Rezago Educativo: hogares con poblacion de referencia' priv_rezago 'Indicador Rezago Educativo: personas privadas' hh_d_reza 'Indicador Rezago Educativo: hogares privados'.</p> <p>***Dimensión Educación: Indicador Cuidado infantil. *Paso 1: calcular variable de población de Referencia. IF (ppa03<6) PR_cuinfa=1. IF (ppa03>=6) PR_cuinfa=0. IF (ppa03 = \$SYSMIS) PR_cuinfa=\$SYSMIS. EXECUTE.</p> <p>*Paso 2: identificar hogares con población de referencia. AGGREGATE /OUTFILE=* MODE=ADDVARIABLES /BREAK=numhog /PRHOG_cuinfa=MAX(PR_cuinfa). EXECUTE.</p> <p>*Paso 3: privación en personas. If PR_cuinfa=1 & (p05a01 = 3 p05a01= 4 p05a01= 9 p05a01=13) priv_cinfantil = 1. IF PR_cuinfa=1 & (p05a01 = 1 p05a01 = 2 p05a01 = 5 p05a01 = 6 p05a01 = 7 p05a01 = 8 p05a01 =10 p05a01 = 11 p05a01 =12 p05a01 = 98) priv_cinfantil = 0. IF p05a01=\$SYSMIS priv_cinfantil=\$SYSMIS.</p> <p>*Paso 4: identificar hogares con privación. AGGREGATE /OUTFILE=* MODE=ADDVARIABLES /BREAK=numhog /hh_d_cinfantil=MAX(priv_cinfantil). EXECUTE. IF PRHOG_cuinfa = 0 hh_d_cinfantil=0.</p> <p>VALUE LABELS hh_d_cinfantil 0 'No privado' 1 'Privado'.</p> <p>VARIABLE LABELS PR_cuinfa 'Indicador Cuidado Infantil: población de referencia' PRHOG_cuinfa 'Indicador Cuidado Infantil: hogares con población de referencia' priv_cinfantil 'Indicador Cuidado Infantil: personas privadas' hh_d_cinfantil 'Indicador Cuidado Infantil: hogares privados'.</p> <p>*DIMENSION: EMPLEO DIGNO. ***Dimensión Empleo: Indicador Empleo informal. *Paso 1: calcular variable de población de Referencia. IF p10a02= 1 p10a03= 1 p10a04= 1 PR_informal=1. IF p10a02> 1 & p10a03= 2 & p10a04= 2 PR_informal=0. EXECUTE.</p> <p>*Paso 2: identificar hogares con población de referencia. AGGREGATE /OUTFILE=* MODE=ADDVARIABLES /BREAK=numhog /PRHOG_informal=MAX(PR_informal). EXECUTE.</p> <p>*Paso 3: privación en personas. IF (p10b04 = 4 p10b04 = 5 p10b04 = 7 p10b04 = 9) priv_informal=1. IF (p10b04 = 1 p10b04 = 2 p10b04 = 3 p10b04 = 6 p10b04 = 8) priv_informal=0. IF (p10b04 = 5 p10b04 = 7) & anios_escolar > 12 & priv_informal = 1 priv_informal=0. IF p10b25<6 & priv_informal =0 priv_informal=1.</p> <p>IF PR_informal = 0 priv_informal= \$SYSMIS.</p>

Continúa apéndice 1.

Código en SPSS

```
*Paso 4: identificar hogares con privación.
AGGREGATE
  /OUTFILE=* MODE=ADDVARIABLES
  /BREAK=numhog
  /hh_d_inf=MAX(priv_informal).
EXECUTE.

IF PRHOG_informal = 0 hh_d_inf=0.

VALUE LABELS hh_d_inf
0 'No privado'
1 'Privado'.

VARIABLE LABELS
PR_informal 'Indicador Empleo Informal: población de referencia'
PRHOG_informal 'Indicador Empleo Informal: hogares con población de referencia'
priv_informal 'Indicador Empleo Informal: personas privadas'
hh_d_inf 'Indicador Empleo Informal: hogares privados'.

***Dimensión Empleo: Indicador Trabajo Infantil.
*Paso 1: calcular variable de población de Referencia.
IF ppa03 >6 & ppa03 < 18 PR_trainfantil=1.
IF ppa03 <=6 | ppa03 >= 18 PR_trainfantil=0.
IF ppa03=$SYSMIS PR_trainfantil=$SYSMIS.
EXECUTE.

*Paso 2: identificar hogares con población de referencia.
AGGREGATE
  /OUTFILE=* MODE=ADDVARIABLES
  /BREAK=numhog
  /PRHOG_trainfantil=MAX(PR_trainfantil).
EXECUTE.

*Paso 3: privación en personas.
IF (PR_trainfantil = 1 & (p10a02= 1 | p10a03= 1 | p10a04= 1)) ninios_ocupados=1.
IF (ppa03<16 & ninios_ocupados=1) priv_trainfantil=1.
IF (ppa03<16 & ninios_ocupados=0) priv_trainfantil=0.
IF ((ppa03=16 | ppa03=17) & ninios_ocupados=1 & priv_escolar=1 & anios_escolar <=9) priv_trainfantil=1.
IF ((ppa03=16 | ppa03=17) & ninios_ocupados=1 & priv_escolar=1 & anios_escolar >9) priv_trainfantil=0.
IF (PR_trainfantil=1 & ninios_ocupados=0) priv_trainfantil=0.
IF PR_trainfantil=0 priv_trainfantil=0.
IF ppa03=$SYSMIS priv_trainfantil=$SYSMIS.

*Paso 4: identificar hogares con privación.
AGGREGATE
  /OUTFILE=* MODE=ADDVARIABLES
  /BREAK=numhog
  /hh_d_trabainf=MAX(priv_trainfantil).
EXECUTE.
IF PRHOG_trainfantil = 0 hh_d_trabainf=0.

VALUE LABELS hh_d_trabainf
0 'No privado'
1 'Privado'.

VARIABLE LABELS
PR_trainfantil 'Indicador Trabajo Infantil: población de referencia'
PRHOG_trainfantil 'Indicador Trabajo Infantil: hogares con población de referencia'
priv_trainfantil 'Indicador Trabajo Infantil: personas privadas'
hh_d_trabainf 'Indicador Trabajo Infantil: hogares privados'.
frequencies PR_trainfantil prhog_trainfantil priv_trainfantil hh_d_trabainf.
```

Continúa apéndice 1.

Código en SPSS
<pre> ***DIMENSION: VIVIENDA. ***Dimensión Vivienda: Indicador Materiales de la vivienda. *Paso 4: identificar hogares con privación. COMPUTE hh_d_mat=0. IF (p01a04 = 7) hh_d_mat = 1. RECODE p01a02 (4=1) (6=1) (7=1) (8=1) (98=1) INTO hh_d_mat. VALUE LABELS hh_d_mat 0 'No privado' 1 'Privado'. VARIABLE LABELS hh_d_mat 'Indicador Materiales de la vivienda: hogares privados'. ***Dimensión Vivienda: Indicador Hacinamiento. *Paso 4: identificar hogares con privación. COMPUTE personas_hogar= thogar/p01a06. COMPUTE hh_d_hacina =0. IF (personas_hogar > 3) hh_d_hacina = 1. VALUE LABELS hh_d_hacina 0 'No privado' 1 'Privado'. VARIABLE LABELS hh_d_hacina 'Indicador Hacinamiento: hogares privados'. ***Dimensión Vivienda: Indicador Combustible para cocinar. *Paso 4: identificar hogares con privación. RECODE p01d27 (1=1) (2=0) INTO hh_d_coci. IF (p01d27 = 1 & p01d31 = 1) hh_d_coci=0. EXECUTE. VALUE LABELS hh_d_coci 0 'No privado' 1 'Privado'. VARIABLE LABELS hh_d_coci 'Indicador Combustible para cocinar: hogares privados'. ***DIMENSION: SERVICIOS BASICOS. ***Dimensión Servicios Básicos: Indicador Acceso al agua. *Paso 4: identificar hogares con privación. COMPUTE hh_d_agua=0. IF (p01d06 >= 3 & area = 1) hh_d_agua = 1. IF (p01d06 >= 5 & area = 2) hh_d_agua = 1. IF (p01d10 >=15 & p01d10 <98 & hh_d_agua = 0) hh_d_agua = 1. VALUE LABELS hh_d_agua 0 'No privado' 1 'Privado'. ***Dimensión Servicios Básicos: Indicador Energía eléctrica. *Paso 4: identificar hogares con privación. RECODE p01a05c (1=0) (2=1) INTO hh_d_elec. EXECUTE. IF (p01d25>1 & hh_d_elec=0) hh_d_elec=1. VALUE LABELS hh_d_elec 0 'No privado' 1 'Privado'. ***Dimensión Servicios Básicos: Indicador Recolección de Basura. *Paso 4: identificar hogares con privación. RECODE p01d21 (3=1) (4=1) (5=1) (98=1) (1=0) (2=0) (6=0) INTO hh_d_basu. VALUE LABELS hh_d_basu 0 'No Privado' 1 'Privado'. </pre>

Continúa apéndice 1.

```

                                Código en SPSS
*
***Dimensión Servicios Básicos: Indicador Saneamiento.
*Paso 4: identificar hogares con privación.
COMPUTE hh_d_san=0.
IF (p01a05b = 2 & area = 1) hh_d_san = 1.
IF (p01d17 >= 3 & area = 1) hh_d_san = 1.
IF (p01d17 =4 | p01d17 =5 & area= 2) hh_d_san = 1.

VALUE LABELS hh_d_san
0 'No privado'
1 'Privado'.

VARIABLE LABELS
hh_d_agua 'Indicador Acceso al agua: hogares privados'
hh_d_elec 'Indicador Energia Electrica: hogares privados'
hh_d_basu 'Indicador Recoleccion de Basura: hogares privados'
hh_d_san 'Indicador Saneamiento: hogares privados'.

*Generamos estadísticos descriptivos de los indicadores contruidos.
FRECUENCIAS
  hh_d_salud hh_d_emba hh_d_segal hh_d_prenatal
  hh_d_asis hh_d_escol hh_d_reza hh_d_cinfantil
  hh_d_inf hh_d_trabainf
  hh_d_mat hh_d_hacina hh_d_coci
  hh_d_agua hh_d_elec hh_d_basu hh_d_san.

SET WORKSPACE 346921.

MVA VARIABLES=
  hh_d_salud hh_d_emba hh_d_segal hh_d_prenatal
  hh_d_asis hh_d_escol hh_d_reza hh_d_cinfantil
  hh_d_inf hh_d_trabainf
  hh_d_mat hh_d_hacina hh_d_coci
  hh_d_agua hh_d_elec hh_d_basu hh_d_san
/MAXCAT=25
/CATEGORICAL=
  hh_d_salud hh_d_emba hh_d_segal hh_d_prenatal
  hh_d_asis hh_d_escol hh_d_reza hh_d_cinfantil
  hh_d_inf hh_d_trabainf
  hh_d_mat hh_d_hacina hh_d_coci
  hh_d_agua hh_d_elec hh_d_basu hh_d_san
/CROSSTAB PERCENT=1
/MISMATCH PERCENT=1.

COMPUTE missing = 0.

IF NMISS
(hh_d_salud, hh_d_emba ,hh_d_segal, hh_d_prenatal,
hh_d_asis, hh_d_escol, hh_d_reza, hh_d_cinfantil,
hh_d_inf, hh_d_trabainf,
hh_d_mat, hh_d_hacina, hh_d_coci,
hh_d_agua, hh_d_elec, hh_d_basu, hh_d_san) >0 MIS=1.

IF SYSMIS(MIS) mis=0.

FRECUENCIAS MIS.
SELECT IF (MIS=0).

*****
***Estimar el Vector del conteo ponderado de privaciones.
COMPUTE c_privaciones=
(hh_d_salud *1/20 + hh_d_emba *1/20 + hh_d_segal *1/20 + hh_d_prenatal *1/20) +
(hh_d_asis *1/20 + hh_d_escol *1/20 + hh_d_reza *1/20 + hh_d_cinfantil*1/20 ) +
(hh_d_inf * 1/10 + hh_d_trabainf*1/10 ) +
(hh_d_mat *1/15 + hh_d_hacina *1/15 + hh_d_coci *1/15 ) +
(hh_d_agua *1/20 + hh_d_elec *1/20 + hh_d_basu *1/20 + hh_d_san*1/20).

```

Continúa apéndice 1.

Código en SPSS
<pre> VARIABLE LABELS c_privaciones 'Cuento ponderado de privaciones'. EXECUTE. FRECUENCIES c_privaciones. ***Identificar a los hogares pobres multidimensionalmente donde c>=k, donde k=35%. COMPUTE multi_poor=0. IF (c_privaciones >= 0.3) multi_poor =1. VARIABLE LABELS multi_poor 'Hogar Pobre Multidimensionalmente'. VALUE LABELS multi_poor 0 'Hogar no pobre' 1 'Hogar en pobreza multidimensional'. EXECUTE. ***Estimar el Vector de Privación Censurada por indicador. *Indicadores de la Dimensión Salud. IF (hh_d_salud = 1 & multi_poor = 1) hh_d_cens_salud = 1. IF (hh_d_salud = 1 & multi_poor = 0) hh_d_cens_salud = 0. IF (hh_d_salud = 0) hh_d_cens_salud = 0. IF (hh_d_emba = 1 & multi_poor = 1) hh_d_cens_emba = 1. IF (hh_d_emba = 1 & multi_poor = 0) hh_d_cens_emba = 0. IF (hh_d_emba = 0) hh_d_cens_emba = 0. IF (hh_d_segal = 1 & multi_poor = 1) hh_d_cens_segal = 1. IF (hh_d_segal = 1 & multi_poor = 0) hh_d_cens_segal = 0. IF (hh_d_segal = 0) hh_d_cens_segal = 0. IF (hh_d_prenatal = 1 & multi_poor = 1) hh_d_cens_prenatal = 1. IF (hh_d_prenatal = 1 & multi_poor = 0) hh_d_cens_prenatal = 0. IF (hh_d_prenatal = 0) hh_d_cens_prenatal = 0. VARIABLE LABELS hh_d_cens_salud 'Acceso a Salud: Hogar pobre multidimensional y privado en indicador' hh_d_cens_emba 'Embarazo en Adolescentes: Hogar pobre multidimensional y privado en indicador' hh_d_cens_segal 'Seguridad Alimentaria: Hogar pobre multidimensional y privado en indicador' hh_d_cens_prenatal 'Cuidado Prenatal: Hogar pobre multidimensional y privado en indicador'. *Indicadores de la Dimensión Educación. IF (hh_d_asis = 1 & multi_poor = 1) hh_d_cens_asis = 1. IF (hh_d_asis = 1 & multi_poor = 0) hh_d_cens_asis = 0. IF (hh_d_asis = 0) hh_d_cens_asis = 0. IF (hh_d_escol = 1 & multi_poor = 1) hh_d_cens_escol = 1. IF (hh_d_escol = 1 & multi_poor = 0) hh_d_cens_escol = 0. IF (hh_d_escol = 0) hh_d_cens_escol = 0. IF (hh_d_reza = 1 & multi_poor = 1) hh_d_cens_reza = 1. IF (hh_d_reza = 1 & multi_poor = 0) hh_d_cens_reza = 0. IF (hh_d_reza = 0) hh_d_cens_reza = 0. IF (hh_d_cinfantil = 1 & multi_poor = 1) hh_d_cens_cinfantil = 1. IF (hh_d_cinfantil = 1 & multi_poor = 0) hh_d_cens_cinfantil = 0. IF (hh_d_cinfantil = 0) hh_d_cens_cinfantil = 0. VARIABLE LABELS hh_d_cens_asis 'Asistencia Escolar: Hogar pobre multidimensional y privado en indicador' hh_d_cens_escol 'Escolaridad en Adultos: Hogar pobre multidimensional y privado en indicador' hh_d_cens_reza 'Rezago Educativo: Hogar pobre multidimensional y privado en indicador' hh_d_cens_cinfantil 'Cuidado Infantil: Hogar pobre multidimensional y privado en indicador'. *Indicadores de la Dimensión Empleo Digno. IF (hh_d_inf = 1 & multi_poor = 1) hh_d_cens_inf = 1. IF (hh_d_inf = 1 & multi_poor = 0) hh_d_cens_inf = 0. IF (hh_d_inf = 0) hh_d_cens_inf = 0. IF (hh_d_trabainf = 1 & multi_poor = 1) hh_d_cens_trabainf = 1. IF (hh_d_trabainf = 1 & multi_poor = 0) hh_d_cens_trabainf = 0. IF (hh_d_trabainf = 0) hh_d_cens_trabainf = 0. </pre>

Continúa apéndice 1.

Código en SPSS
<p>VARIABLE LABELS hh_d_cens_inf 'Empleo Informal: Hogar pobre multidimensional y privado en indicador' hh_d_cens_trabainf 'Trabajo Infantil: Hogar pobre multidimensional y privado en indicador'.</p> <p>*Indicadores de la Dimensión Vivienda. IF (hh_d_mat = 1 & multi_poor = 1) hh_d_cens_mat = 1. IF (hh_d_mat = 1 & multi_poor = 0) hh_d_cens_mat = 0. IF (hh_d_mat = 0) hh_d_cens_mat = 0.</p> <p>IF (hh_d_hacina = 1 & multi_poor = 1) hh_d_cens_hacina = 1. IF (hh_d_hacina = 1 & multi_poor = 0) hh_d_cens_hacina = 0. IF (hh_d_hacina = 0) hh_d_cens_hacina = 0.</p> <p>IF (hh_d_coci = 1 & multi_poor = 1) hh_d_cens_coci = 1. IF (hh_d_coci = 1 & multi_poor = 0) hh_d_cens_coci = 0. IF (hh_d_coci = 0) hh_d_cens_coci = 0.</p> <p>VARIABLE LABELS hh_d_cens_mat 'Materiales de la Vivienda: Hogar pobre multidimensional y privado en indicador' hh_d_cens_hacina 'Hacinamiento: Hogar pobre multidimensional y privado en indicador' hh_d_cens_coci 'Combustible para cocinar: Hogar pobre multidimensional y privado en indicador'.</p> <p>*Indicadores de la Dimensión Servicios Básicos. IF (hh_d_agua = 1 & multi_poor = 1) hh_d_cens_agua = 1. IF (hh_d_agua = 1 & multi_poor = 0) hh_d_cens_agua = 0. IF (hh_d_agua = 0) hh_d_cens_agua = 0.</p> <p>IF (hh_d_elec = 1 & multi_poor = 1) hh_d_cens_elec = 1. IF (hh_d_elec = 1 & multi_poor = 0) hh_d_cens_elec = 0. IF (hh_d_elec = 0) hh_d_cens_elec = 0.</p> <p>IF (hh_d_basu = 1 & multi_poor = 1) hh_d_cens_basu = 1. IF (hh_d_basu = 1 & multi_poor = 0) hh_d_cens_basu = 0. IF (hh_d_basu = 0) hh_d_cens_basu = 0.</p> <p>IF (hh_d_san = 1 & multi_poor = 1) hh_d_cens_san = 1. IF (hh_d_san = 1 & multi_poor = 0) hh_d_cens_san = 0. IF (hh_d_san = 0) hh_d_cens_san = 0.</p> <p>VARIABLE LABELS hh_d_cens_agua 'Acceso a Agua: Hogar pobre multidimensional y privado en indicador' hh_d_cens_elec 'Energía Eléctrica: Hogar pobre multidimensional y privado en indicador' hh_d_cens_basu 'Recolección de basura: Hogar pobre multidimensional y privado en indicador' hh_d_cens_san 'Saneamiento: Hogar pobre multidimensional y privado en indicador'.</p> <p>VALUE LABELS hh_d_cens_salud hh_d_cens_emba hh_d_cens_segal hh_d_cens_prenatal hh_d_cens_asis hh_d_cens_escol hh_d_cens_reza hh_d_cens_cinfantil hh_d_cens_inf hh_d_cens_trabainf hh_d_cens_mat hh_d_cens_hacina hh_d_cens_coci hh_d_cens_agua hh_d_cens_elec hh_d_cens_basu hh_d_cens_san 0 'No privado' 1 'Privación censurada'.</p> <p>***M0: Variable que permite estimar el IPM-Gt. COMPUTE M0_c_privaciones = c_privaciones. IF multi_poor = 0 M0_c_privaciones = 0. VARIABLE LABELS M0_c_privaciones 'M0: Índice de Pobreza Multidimensional'.</p> <p>***A: Variable que permite estimar la intensidad. IF (multi_poor = 1) A_intensidad=c_privaciones. EXECUTE. VARIABLE LABELS A_intensidad=c_privaciones 'A: intensidad de la pobreza' *** descarga dashbord: https://www.mediafire.com/file/wfbhfjv9v00xw97/dashboard_ipm_gt.xlsx/file</p>

Fuente: elaboración propia con datos de la ENCOVI 2014, programa SPSS versión 25.