

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

COMPARACIÓN DE DOS ESTRATEGIAS METODOLÓGICAS PARA LA
ESTRATIFICACIÓN SOCIOECONÓMICA DEL MARCO MUESTRAL DE VIVIENDAS DE
COSTA RICA

Trabajo final de investigación aplicada sometido a la consideración de la Comisión del
Programa de Estudios de Posgrado en Estadística para optar al grado y título de Maestría
Profesional en Estadística

ADRIANA GUZMÁN RIZO

Ciudad Universitaria Rodrigo Facio, Costa Rica

2020

AGRADECIMIENTOS

El presente trabajo lo dedico de manera especial a mi familia, la mayor bendición que Dios me ha dado: los amo infinitamente. Gracias por creer en mí, por el apoyo y la paciencia en todo momento. A Eve, por apoyarme y motivarme en este proceso. A mis compañeros y amigos de maestría, por las noches en vela que pasamos juntos para terminar algún trabajo y por la colaboración mutua que hicieron de esta maestría una aventura maravillosa. A los profesores, quienes me hicieron crecer como profesional. Por último, agradezco a todos los que me han ayudado incondicionalmente y me han impulsado a lograr esta meta.

“Este trabajo final de investigación fue aceptado por la Comisión del Programa de Estudios de Posgrado en Estadística de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Profesional en Estadística.”



M.Sc. Johnny Madrigal Pana
Profesor Guía



M.Sc. Giselle Arguello Venegas
Lectora



PhD. Guaner Rojas Rojas
Lector

Adriana Guzmán Rizo

Adriana Guzmán Rizo
Sustentante

Tabla de contenido

Resumen	vi
I. Introducción	1
1.1 Justificación	3
1.2 Planteamiento del problema	5
1.3 Objetivos	5
II. Marco teórico	6
2.1 Elementos estadísticos para una estratificación	6
2.2. Enfoque conceptual del nivel socioeconómico	8
2.3 Conceptos relacionados con el nivel socioeconómico	10
2.4 Marco de referencia.....	12
2.4.1 Estratificación socioeconómica en países de la región	12
2.4.2 Estudios relacionados con el nivel socioeconómico en Costa Rica.	21
III. Metodología	25
3.1 Descripción de las dos estrategias metodológicas	25
3.2 Estrategia metodológica A	26
3.2.1 Construcción de indicadores en estrategia metodológica A	27
3.3 Estrategia metodológica B	38
3.1.1 Construcción indicadores estrategia metodológica B	39
3.4 Variante estrategia B (estrategia mixta)	41
3.5 Datos	41
3.6 Delimitación geográfica	42
3.7 Métodos y técnicas estadísticas	43
3.8 Validación interna	45
3.9 Estudio de simulación para la validez externa	46
3.10 Software utilizado	47
IV. Resultados	48
4.1 Descriptivos variables generales	48

4.2. Resultados de la estrategia A	50
4.2.1 Análisis descriptivos Estrategia A.....	51
4.2.2 Modelo de regresión aplicado en estrategia A.....	54
4.2.3 Grupos K medias estrategia A.....	57
4.3 Resultados estrategia metodológica B.....	58
4.3.1 Descriptivos estrategia metodológica B.....	58
4.3.2 Conformación de grupos utilizado la técnica de K medias en la estrategia metodológica B	60
4.4 Resultados variante estrategia B	61
4.4.1 Grupos K medias para variante de estrategia metodológica B	62
4.5 Validación de los resultados	63
4.6 Resultado del estudio de simulación.....	66
VI. Recomendaciones	78
Bibliografía	79
Anexos	83

Resumen

El objetivo del presente trabajo es comparar dos estrategias metodológicas para estratificar el marco muestral de viviendas de Costa Rica, con el propósito de reducir la variabilidad de las estimaciones de las principales temáticas investigadas en las encuestas que realiza el INEC. Para esta institución, este estudio es relevante, ya que es un insumo que contribuye a identificar la mejor estratificación que favorezca los diseños muestrales que se implementarán con el nuevo marco muestral de viviendas que se obtendrá como producto del Censo 2021.

La estrategia metodológica A consiste en construir un modelo de regresión lineal múltiple para estimar el logaritmo del ingreso per cápita neto, a partir de siete variables independientes: índices de características de la vivienda, índice del patrimonio acumulado, miembros ocupados, número de miembros del hogar, años de escolaridad del jefe, índice del tipo de aseguramiento del jefe e índice de ocupación del jefe. Este modelo se construye con información de la ENAHO 2015 y se aplica al Censo 2011 con el objetivo de estimar el ingreso de los hogares y posteriormente obtener una estimación agregada del ingreso a nivel de las unidades primarias de muestreo (UPM) que conforman las unidades del marco muestral de viviendas. De esta manera, se procede a estratificar las UPM del marco por medio del análisis de conglomerados con el procedimiento de K-medias generando siete grupos, posteriormente se hace una recodificación de estos grupos en tres estratos socioeconómicos denominados bajo, medio y alto.

Por otro lado, la estrategia metodológica B consiste en crear indicadores independientes con los datos del Censo 2011 para elaborar una estratificación multivariada. Los indicadores construidos son: porcentaje de hogares con al menos una necesidad básica insatisfecha (NBI), tasa de desempleo abierto (TDA), tipo de seguro social del jefe y porcentaje de hogares con al menos un universitario presente en la vivienda. Estos indicadores se construyen por unidad primaria de muestreo y se estratifican usando el procedimiento K-medias. Por último, se genera una variante a la estrategia B al mezclar el indicador del ingreso estimado con las cuatro variables construidas en la estrategia B.

Como resultado se obtuvo que las estrategias presentan una fuerte validez interna y externa, y con el análisis se detecta un comportamiento similar en términos de precisión. A partir del resultado anterior, para estratificar socioeconómicamente las UPM del nuevo marco muestral que realizará el INEC a partir del Censo 2021, puede utilizarse cualquiera de las metodologías aplicadas; no obstante, por practicidad es preferible implementar la metodología B.

Lista de cuadros

Cuadro 1. Ingreso principal neto total promedio del hogar por categoría ocupacional según ocupación del empleo principal del jefe de hogar	47
Cuadro 2. Estadísticos descriptivos del índice de ocupación.....	48
Cuadro 3. Distribución porcentual de las características del hogar y la vivienda según zona de residencia	60
Cuadro 4. Distribución porcentual de los jefes de hogar según algunas características de interés	61
Cuadro 5. Distribución absoluta y relativa de la cantidad de pertenencias acumuladas por el hogar	63
Cuadro 6. Distribución porcentual de la agrupación de la variable tipo de seguro	63
Cuadro 7. Estadísticos descriptivos de la Estrategia A.....	64
Cuadro 8. Promedio de los indicadores según archivo del Censo y ENAHO 2015	65
Cuadro 9. Coeficientes de correlación de Pearson para las variables utilizadas en el modelo de regresión múltiple	66
Cuadro 10. Resultados de la regresión múltiple usando la estrategia metodológica A	68
Cuadro 11. Estadísticos de los k grupos, aplicando estrategia metodológica A.....	70
Cuadro 12. Estadísticos descriptivos usando Estrategia metodológica B	71
Cuadro 13. Estadísticos de los k grupos aplicando estrategia metodológica B.....	72
Cuadro 14. Variante a la estrategia metodológica B incorporando el logaritmo Ingreso per cápita del hogar.....	74
Cuadro 15. Porcentaje de viviendas que cumple con ciertas características nivel según estrategia aplicada	76
Cuadro 16. Porcentaje de hogares con jefes de hogar por nacionalidad según estrategia metodológica.....	76
Cuadro 17. Promedio de las variables de interés según estrategia metodológica aplicada	77
Cuadro 18. Distribución del número de UPM por estrato según estrategia y selección de muestra para la primera etapa	78
Cuadro 19. Resumen de los resultados obtenidos en la simulación de variables según estrategia metodológica aplicada	83

Lista de tablas

Tabla 1. Distribución de los hogares de México por Estratos Socioeconómicos.....	14
Tabla 2. Variables utilizadas para la construcción del nivel socioeconómico APEIM Perú	16
Tabla 3. Distribución de los hogares de Perú por estratos socioeconómicos.....	16
Tabla 4. Estratos Socioeconómicos de Chile 2018.....	18
Tabla 5. Variables utilizadas para la construcción del nivel socioeconómico INSE general y reducido	20
Tabla 6. Distribución de los hogares de Uruguay por estratos socioeconómicos.....	21
Tabla 7. Jerarquía de importancia de las variables utilizadas para definir nivel socioeconómico INEC Ecuador.....	22
Tabla 8. Distribución de los hogares de Ecuador por Estratos Socioeconómicos	23
Tabla 9. Variables consideradas para la construcción del modelo del ingreso <i>per cápita</i> del hogar.....	33
Tabla 10. Comparación de variables independientes modelo base y modelo propuesto ...	33
Tabla 11. Comparación de variables independientes modelo de Madrigal y modelo propuesto en esta investigación	34
Tabla 12. Comparación de variables independientes modelo base y modelo propuesto para el índice que contempla el equipamiento del hogar	35
Tabla 13. Comparación en la construcción de la categoría ocupacional en el modelo base y el propuesto en esta investigación.....	36
Tabla 14. Comparación de los aspectos favorables y desfavorables de las estrategias metodológicas	78

Lista de figuras

Figura 1. Estrategia de análisis para la construcción del nivel socioeconómico en Ecuador. Fuente: INEC Ecuador (2011)	30
Figura 2. Estructura metodológica empleada en la investigación	35

Lista de gráficos

Gráfico 1. Histograma del ingreso <i>per cápita</i> neto del hogar	59
Gráfico 2. Histograma del logaritmo del ingreso <i>per cápita</i> neto del hogar	59
Gráfico 3. Codo de Jambú: Inercia intra clase, según cantidad de conglomerados considerando el ingreso <i>per cápita</i> neto, estrategia A	63
Gráfico 4. Codo de Jambú: Inercia intra clase, según cantidad de conglomerados considerando los 4 indicadores de la estrategia B.....	66
Gráfico 5. Codo de Jambú: Inercia intra clase, según cantidad de conglomerados considerando el ingreso y los 4 indicadores de la estrategia VB	68
Gráfico 6. Distribución de densidad para las variables NBI, TDA y número de ocupados por zona	72
Gráfico 7. Distribución de densidad para las variables porcentaje de carros, internet y computadora portátil por zona.....	73
Gráfico 8. Comparación variabilidad del Coeficiente de variación según estrategia	76
Gráfico 9. Comparación variabilidad del Coeficiente de variación para totales según estrategia metodológica	77
Gráfico 10. Comparación variabilidad del coeficiente de variación para porcentajes de pertenencias según estrategia metodológica.....	78
Gráfico 11. Efecto del diseño de las principales variables de comparación según estrategia metodológica aplicada	80

Lista de abreviaturas

ENAH0 Encuesta Nacional de Hogares

ECE Encuesta Continua de Empleo

ENIGH Encuesta Nacional de Ingresos y Gastos de los Hogares

ENAMEH Encuesta Nacional de Microempresas de los Hogares

INEC Instituto Nacional de Estadística y Censos

INEGI Instituto Nacional de Estadística y Geografía

INEI Instituto Nacional de Estadística e Informática

IPM Pobreza Multidimensional

LP Línea de Pobreza

MMV-2011 Marco muestral de viviendas 2011

NBI Necesidades básicas insatisfechas

OCDE Organización para la Cooperación y el Desarrollo Económico

UPM Unidad Primaria de Muestreo



UNIVERSIDAD DE
COSTA RICA

SEP Sistema de
Estudios de Posgrado

Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.

Yo, Adriana Guzmán Rizo, con cédula de identidad 5-0383-0021, en mi condición de autor del TFG titulado Comparación de dos estrategias metodológicas para la estratificación socioeconómica del marco muestral de viviendas de Costa Rica

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. SI NO *

*En caso de la negativa favor indicar el tiempo de restricción: _____ año (s).

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

INFORMACIÓN DEL ESTUDIANTE:

Nombre Completo: Adriana Guzmán Rizo

Número de Carné: A92950 Número de cédula: 5-0383-0021

Correo Electrónico: adri5gemela2@hotmail.com

Fecha: 15-05-2020 Número de teléfono: 8607-70-53

Nombre del Director (a) de Tesis o Tutor (a): Johnny Madrigal Pana

Adriana Guzmán Rizo

FIRMA ESTUDIANTE

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

I. Introducción

Las encuestas por muestreo son ampliamente utilizadas en muchos países, incluyendo Costa Rica, dado los múltiples beneficios que generan en diversas áreas de investigación. Una de las ventajas de aplicarlas consiste en minimizar el tiempo y los costos para realizarlas, si se compara con un censo. Otro aspecto importante es que permiten proveer de información estadística a la sociedad sobre diversas áreas de interés, que facilitan la formulación, seguimiento y evaluación de políticas públicas, así como de acciones privadas y civiles orientadas hacia el desarrollo.

En el contexto costarricense, las encuestas se realizan desde hace muchos años y presentan diferentes enfoques. Uno de ellos es determinar el nivel de bienestar de la población, centrados en la conformación del ingreso de los hogares, su distribución, sus características y la población en situación de pobreza. Otros enfoques están relacionados con temas de empleo y sus condiciones, así como temas culturales, uso del tiempo, personas en situación de discapacidad, entre otros.

Para llevar a cabo encuestas por muestreo, es indispensable contar con un marco muestral de viviendas, del cual se selecciona la muestra con procedimientos y técnicas estadísticas ampliamente desarrolladas en la literatura académica. El marco muestral es un instrumento básico para estos análisis, debido a que está conformado por el total de los elementos que componen el universo de estudio.

El Instituto Nacional de Estadísticas y Censo (INEC) construye el marco muestral de viviendas a partir de los censos nacionales de población y vivienda que se desarrollan en el país aproximadamente cada diez años. Los censos también permiten elaborar nuevas proyecciones de población, las que son comúnmente utilizadas para ajustar las estimaciones obtenidas de las muestras y corregir la desactualización de los marcos de muestreo durante el periodo intercensal.

La estratificación de los marcos de muestreo se realiza principalmente con dos objetivos: obtener representatividad y precisión. En primer lugar, aumenta la representatividad de subpoblaciones de interés en las encuestas, es decir, permite que las características que se quieren observar en la población estén expresadas apropiadamente en la muestra y, de esta manera, hacer inferencias al resto de la población con un margen de error conocido.

En segundo lugar, la estratificación por lo general reduce la variabilidad de las estimaciones, debido a que existe una mayor homogeneidad dentro de cada estrato respecto a la totalidad de la población, lo que es beneficioso en términos de estimación.

El marco muestral de viviendas en 1984 se estratificó por primera vez con una metodología desarrollada por Johnny Madrigal Pana, quien construyó un indicador del ingreso familiar de los miembros que residen en una vivienda (Madrigal, 1986). De igual manera, en el año 2000 se encargó de realizar la estratificación del marco muestral utilizando una metodología más rigurosa, lo que generó avances para el INEC en esta temática (Madrigal, 2004). En el Censo del 2011, el personal designado por el INEC con el apoyo de Madrigal replicaron la metodología desarrollada en el 2000.

El presente estudio trata sobre el desarrollo de una estrategia metodológica alternativa con el propósito de estratificar el marco muestral de viviendas. Además, de generar una versión actualizada de la propuesta metodológica desarrollada por Madrigal (2004) y compararla con la nueva. Los insumos recabados podrán ser utilizados para estratificar el marco muestral de viviendas que será derivado del Censo del año 2021 y así propiciar un impacto positivo en la precisión de las futuras encuestas que realizará el INEC.

Este documento se conforma de seis secciones. En la primera, se elabora un capítulo introductorio que establece el contexto del problema por investigar. En la segunda sección, se presenta el marco teórico, donde se detallan los conceptos estadísticos que intervienen en una estratificación y el nivel socioeconómico. Posteriormente, en la tercera, se describen los procedimientos metodológicos necesarios para alcanzar los objetivos del estudio: fuente de datos, cobertura de la estratificación (delimitación geográfica), descripción de las estrategias metodológicas, la construcción de los indicadores, métodos y técnicas estadísticas, así como la validación de las estrategias.

En la cuarta sección, se analizan los resultados de ambas metodologías. Finalmente, se presenta un apartado de conclusiones y recomendaciones en donde se sintetizan los hallazgos obtenidos de la investigación y se plantean sugerencias al INEC para que sean consideradas en el próximo censo.

1.1 Justificación

Existen diversas disciplinas que utilizan las encuestas por muestreo, debido a los beneficios que producen en las investigaciones. No obstante, se dimensiona en menor medida la importancia de contar con un marco muestral de calidad con criterios de estratificación óptimos para generar diseños de muestras eficientes y precisos.

Países latinoamericanos como México, Chile, Perú, Uruguay y Ecuador han reconocido la importancia de elaborar un marco de muestreo de calidad, así como de una buena estratificación. De esta manera, se han identificado esfuerzos en tiempo y recursos para desarrollar este tipo de análisis. En México, las empresas de mercadeos consideran necesario elaborar una actualización de los estratos socioeconómicos al menos cada dos años (Asociación Mexicana de Inteligencia de Mercado y Opinión, 2018).

Según Gutiérrez (2013), una categorización socioeconómica o exclusivamente económica, definida apropiadamente, da como resultado agrupaciones suficientemente heterogéneas entre sí, que contribuyen al análisis porque generan resultados más precisos y útiles para orientar de manera adecuada estudios o acciones dirigidos a una población específica.

Los especialistas en ciencias sociales realizan investigaciones en torno al nivel socioeconómico, al producir conocimientos sobre los mecanismos de diferenciación social por medio del uso de criterios categóricos para la clasificación. Estos investigadores se especializan en establecer las relaciones económicas de la población en ámbitos principalmente ocupacionales, como ejes sobre los cuales se distribuyen los grupos sociales (Vega, Castro, Gutiérrez y Rodríguez, 1995).

Por otra parte, los investigadores en el área de mercadeo han construido metodologías aplicadas para la identificación de grupos de consumidores y análisis de demanda, con énfasis en el componente socioeconómico. Estas metodologías, por su simplicidad, se han instalado como medidas estándar para la comprensión de los grupos en la sociedad (García, 2002).

Tanto en las investigaciones sociales como en las de mercado, segmentar o caracterizar la población permite reconocer a la población objetivo (Egüez y Pérez, 2017). En el ámbito del mercadeo, cualquiera que sea el producto o servicio, ha sido creado para llegar a un segmento de la población en particular y, de esta manera, se logra determinar quién puede comprar o al menos interesarse en el producto o servicio que se ofrece.

A nivel social, la caracterización facilita la investigación, orientación, creación, ejecución de políticas sociales y la valoración de estas; sin embargo, el énfasis se sitúa en conocer un grupo específico de la población más que en crear una clasificación global. Clasificar solo una parte de la población no es provechoso para los fines estadísticos que se desean desarrollar en esta investigación.

En estadística, específicamente en el área de muestreo, la clasificación de los hogares o viviendas es de gran interés porque provee los insumos para estratificar los marcos muestrales, con el fin de generar mejoras en los diseños muestrales.

Actualmente, en las encuestas por muestreo que elabora el INEC, se estratifican las Unidades Primarias de Muestreo (UPM), y como consecuencia los hogares a lo interno de estas, a partir de un indicador del ingreso con base en un modelo de regresión elaborado con datos de la Encuesta Nacional de Hogares 2011 (ENAH0), que contempla información del jefe de hogar y de la vivienda. No obstante, se ha detectado la necesidad de generar estratos estadísticos que permitan reducir aún más la variabilidad de las estimaciones, debido a que, con el paso del tiempo, el INEC por tema de comparabilidad ha mantenido la misma metodología, sin contemplar otras alternativas que pueden influir de manera positiva en la reducción de la variancia.

Además, algunos consultores han manifestado que generar la estratificación del marco muestral a partir de una estimación de ingreso, no es la metodología más adecuada y sugieren buscar variables *proxys* al nivel socioeconómico en el archivo del censo para estratificar el marco (Medina, 2018). Por lo tanto, este estudio pretende responder, de manera científica, cuál es la metodología de estratificación más conveniente para generar las estimaciones de las encuestas por muestreo realizadas por el INEC, según un conjunto de variables de interés.

En las encuestas a hogares que realiza el INEC, las temáticas más importantes tienen componentes relacionados con el nivel socioeconómico. Por ejemplo, en la Encuesta Nacional de Hogares (ENAH0), los principales temas de investigación son la distribución de los ingresos y nivel de pobreza de la población; en la Encuesta Continua de Empleo (ECE) son las características del empleo y desempleo de la población; por su parte, en la Encuesta Nacional de Ingresos y Gastos son la distribución del ingreso y del gasto de los hogares. Todas estas temáticas, están sumamente correlacionadas con el nivel socioeconómico de la población.

Dado lo anterior, la elaboración de una estratificación con base en elementos socioeconómicos en el marco muestral de viviendas puede contribuir con la reducción de la variancia de las estimaciones, ya que la mayoría de las variables de diseño presentes tienen relación con el nivel socioeconómico.

De igual forma, es importante señalar que la utilidad fundamental de la estratificación tiene como fin el muestreo y no para obtener estimaciones, ni tampoco para el análisis económico de la población de estos estratos.

1.2 Planteamiento del problema

¿Cuál metodología de estratificación socioeconómica del marco muestral de viviendas de Costa Rica proporciona diseños de muestras más representativos y precisos?

1.3 Objetivos

1.3.1 Objetivo general

Determinar cuál de los procedimientos propuestos para estratificar el marco muestral de viviendas de Costa Rica reduce más la variabilidad de las estimaciones de las principales encuestas que realiza el INEC.

1.3.2 Objetivos específicos

- Elaborar una estrategia de estratificación socioeconómica de las Unidades Primarias de Muestreo que conforman el marco muestral de viviendas, por medio de la técnica de regresión lineal múltiple, que modele el ingreso *per cápita* de los hogares costarricenses a través de un conjunto de variables independientes obtenidas de la Encuesta Nacional a Hogares 2015 y estimada en el Censo de Población y Vivienda del 2011.
- Elaborar una estrategia de estratificación socioeconómica de las Unidades Primarias de Muestreo que conforman el marco muestral de viviendas, con variables *proxys* al ingreso *per cápita* de los hogares, a partir de información del Censo 2011.
- Evaluar la estrategia de estratificación que genera la mayor reducción de la variabilidad en las variables de estudio de las principales encuestas realizadas por el INEC mediante simulaciones del diseño muestral de la ENAHO para valorar la eficacia de las estratificaciones construidas con cada estrategia metodológica.

II. Marco teórico

2.1 Elementos estadísticos para una estratificación

Cuando se trabaja con encuestas probabilísticas, se debe poseer un marco muestral, que es el listado o forma de identificar a los individuos de la población o unidades de muestreo, para aplicar sobre ellos algún proceso de selección. Es decir, el marco muestral es el conjunto de materiales fuente (mapas, listas, directorios, etcétera) de los que se selecciona la muestra (United Nations, 1995; Kish, 1975). Por esta razón, se debe garantizar la inclusión de todos los elementos definidos en la población, sin duplicaciones ni omisiones, así como estar referidos en tiempo y lugar.

Existen al menos dos tipos de marcos muestrales: de listas y de áreas. El marco muestral de lista es un listado de todos y cada uno de los elementos de la población de interés en forma de base de datos (Kish, 1975). El marco muestral de áreas está conformado por áreas geográficas, cuyos límites deben ser fácilmente identificables en el campo. El tamaño de estas áreas geográficas, en términos de población o vivienda, depende de la fácil delimitación, el grado de homogeneidad de las principales características por investigar, de las cargas de trabajo y del período de vida útil del marco (INEC, 2014).

Los marcos muestrales de áreas incluyen información básica sobre las características sociodemográficas y económicas de la población asociada a esas áreas geográficas; además, se acompañan con insumos cartográficos para la correcta ubicación en el campo de los elementos que lo conforman. Generalmente, se construye a partir de la información que brindan los censos nacionales.

El marco muestral de viviendas del INEC corresponde al denominado marco de áreas, ya que sus unidades de muestreo son áreas geográficas. Específicamente, el INEC define estas áreas como Unidades Primarias de Muestreo (UPM), las que se construyen dividiendo cada distrito de la división territorial administrativa de Costa Rica en zonas más pequeñas.

Una construcción adecuada del marco muestral de viviendas es importante, ya que este será el instrumento básico para la ejecución de las encuestas por muestreo que se planean realizar en el país al menos durante los próximos diez años, tanto por el INEC, como por los institutos de investigación universitarios y la empresa privada.

En el caso del marco muestral de viviendas 2011 (MMV-2011) de Costa Rica, según el INEC (2017) las UPM se definen como:

Áreas geográficas en las que se divide cada distrito del país y que contienen en promedio 150 viviendas en las zonas urbanas y 100 viviendas en promedio en las zonas rurales. En total, el MMV-2011 lo conforman 10 461 UPM, clasificadas como urbanas un 64,5 % y como rurales 35,5 %, y 1 360 055 viviendas, 70,6 % en zonas urbanas y 29,4 % en zonas rurales. Además, del conjunto de UPM, el MMV-2011 se complementa con información demográfica y socioeconómica proveniente del censo nacional, y con mapas cartográficos que permiten la ubicación, ordenamiento y estratificación del marco según los requerimientos específicos de cada encuesta. (p.2).

Dada la información que recopila el censo y considerando las UPM como la unidad geográfica más pequeña que se deben caracterizar, la estratificación consiste en clasificar las UPM en subgrupos, de manera que las unidades dentro de cada grupo o estrato sean lo más homogéneas posibles, según una o varias características pertinentes a las estimaciones que se deseen realizar. La disminución de la variancia de las estimaciones se logra construyendo estratos lo más heterogéneos posibles entre sí y lo más homogéneos a lo interno (Kish, 1975).

Para estratificar, es fundamental que las unidades del marco muestral se agrupen de acuerdo con características homogéneas. Esta situación contribuye a reducir el número de selecciones de primera y segunda etapa, lo que ayuda a disminuir la variancia de los estimadores. La creación de estos estratos estadísticos también tiene la finalidad de mejorar los procedimientos de selección de las muestras probabilísticas.

Kish (1975) menciona que este tipo de estratificación tiene principalmente tres usos: la disminución de varianzas en las estimaciones de las muestras, la posibilidad de aplicar distintos métodos o procedimientos en cada subgrupo de la población o estrato y, por último, si se define los estratos también como dominios de estudios, beneficia el análisis comparativo de datos entre los grupos.

Es importante puntualizar que las unidades de observación y análisis que se involucran en el proceso de estratificación son las UPM. En este sentido, se hace un uso exhaustivo de la información asociada a las características de las viviendas y los hogares en materia socioeconómica para esas áreas. Los hogares se definen según Dahrendorf (1956) e INEC (2011) como la unidad social conformada por una persona o un grupo de personas que se asocian para compartir el alojamiento y la comida. Es decir, el hogar es el conjunto de

personas que residen habitualmente en la misma vivienda, que están unidas o no por lazos de parentesco y que cocinan en común para todos sus miembros. La UPM es la unidad de análisis que debe ser caracterizada a partir de la información de las viviendas y los hogares.

2.2. Enfoque conceptual del nivel socioeconómico

Desde la perspectiva social, la estratificación es vista mediante el concepto de estrato social. Según Dahrendorf (citado en Francés, 2009) un estrato es “una categoría de personas que ocupan una posición similar en una escala que jerarquiza ciertas categorías propias de una situación, como pueden ser el ingreso, el prestigio o el estilo de vida” (p.46).

Las teorías sobre estratificación social tienen dos enfoques: la teoría de Karl Marx, quien establece que las clases sociales están en función de quien posee y controla los medios de producción y la teoría de Max Weber, quien acepta que las clases se basan en condiciones económicas objetivas. Para este último autor, el control y propiedad de los medios de producción solo son uno de los factores económicos que influyen en la determinación de las clases, pero existen otros como los conocimientos técnicos y las cualificaciones que afectan a los tipos de trabajos que las personas pueden obtener (Dahrendorf, 1956, citado en Francés, 2009). Esta investigación se desarrollará sobre este segundo enfoque.

La perspectiva estadística se centra en la conformación de estratos estadísticos, los cuales tienen el objetivo de generar mejores diseños muestrales, pues las agrupaciones de la población en estratos homogéneos reducen la variancia de los estimadores y facilita el estudio y caracterización de cada subgrupo.

Si bien el nivel socioeconómico ha sido estudiado extensivamente por las ciencias sociales, muchos autores concuerdan en la complejidad para llegar a una definición única, sobre todo por su carácter multidimensional y su estrecha relación con el contexto de la realidad que pretende evaluar o medir. Aunque es complejo, existe cierto consenso que señala que el nivel socioeconómico permite el ordenamiento relativo de individuos, a partir de sus características económicas y sociales dentro de una estructura social jerárquica (Caro y Cortés, 2012).

Una manera en que los estudios se aproximan al concepto del nivel socioeconómico de los hogares es mediante tres dimensiones: el ingreso familiar, el nivel educativo de los padres y el prestigio ocupacional de los mismos, esquema que responde a la definición propuesta por Duncan, Featherman, y Duncan (1972).

Egüez y Pérez (2017) afirman que el ingreso siempre será uno de los factores más importantes para clasificar o estratificar a una sociedad, pues la desigualdad del ingreso establece relaciones entre categorías de individuos. Pero las estratificaciones siempre son variables, los modelos siempre dependerán de las necesidades de cada investigación y los modifican para acomodarlos a “sus” realidades. Por este motivo, estos autores definen la estructura social o la clase social como “un grupo donde los individuos comparten una característica común que lo vincula social o económicamente; dado que, la composición de ese status no depende específicamente solo del ingreso, sino de otras características más: culturales, económicos, demográficos y sociales” (Egüez y Pérez, 2017, p.2).

Dadas las limitaciones de recursos económicos para medir los ingresos de los hogares en algunos países y por la incertidumbre de obtener un dato de calidad es esta variable de carácter sensible para el entrevistado, desde hace muchos años, se han desarrollado métodos con variables proxys que permiten construir índices de nivel socioeconómico como medida de aproximación a la estimación del ingreso. Esta estrategia ha sido bien aceptada, al generar que los datos se recopilen más rápidamente, de manera más fácil y con mayor credibilidad (Egüez y Pérez, 2017).

En lugar del ingreso, se indaga por información de variables correlacionadas a los ingresos que capturan los estándares de vida, como la tenencia y cantidad de bienes duraderos como la televisión, carro, motos, así como las características del estado de la infraestructura y vivienda, escolaridad del jefe, entre otros.

Como lo menciona Arias (2017) cualquier enfoque que se desarrolle representa un reto para el investigador, ya que, una vez establecida la metodología para la construcción de los estratos, debe determinar el número que quiere construir. Considerando algunas particularidades, si se crean muchos estratos, pueden producir grupos con pocas diferencias, lo que es poco útil para discriminar entre esos grupos. Por otro lado, si se crean pocos estratos, es posible cometer el error de incluir en un mismo grupo casos con características muy distintas contrarias a lo que se espera. Un balance entre el número de casos que conforman los grupos o estratos, niveles de variabilidad que estos presentan y el estadístico *eta* son algunas herramientas que se deben utilizar para obtener el número óptimo de conglomerados o estratos.

En diferentes estudios de América Latina, específicamente en el área del mercadeo, el número de estratos promedio generados oscilan entre cinco y siete, en los cuales se

concentra la mayor acumulación de la población en los grupos medios y bajos (INEGI, s.f; AMAI, 2018; APEIM,2018; MINEDU,2018; AIM, 2018 y INEC,2011).

2.3 Conceptos relacionados con el nivel socioeconómico

Existen términos estrechamente ligados al concepto de nivel socioeconómico. Litchfield (1999) y OCDE (2019) mencionan que estos conceptos son la pobreza y la desigualdad. Es importante señalar la diferencia entre desigualdad y pobreza, ya que suelen confundirse. La desigualdad se refiere a la variación de los estándares de vida de la población, independientemente de si dicha población está o no en pobreza (McKay, 2002). Es decir, puede existir desigualdad sin pobreza y pobreza sin desigualdad.

La desigualdad social implica una distribución de bienes y servicios inequitativa a lo largo del territorio nacional (Banerjee y Duflo, 2012), lo que provoca que la sociedad se concentre y se estratifique en grupos poblacionales con distintos niveles de acceso a oportunidades de desarrollo (Martínez, 2010). Las diferencias entre estos grupos poblacionales se deben principalmente a aspectos externos a las características de las familias, tales como la segregación urbana y la geografía, que ejercerían una influencia importante en la calidad de vida de sus habitantes (Therborn, 2015). El estudio de Madrigal (2004) afirma que el ingreso es uno de los factores más comunes para generar la desigualdad, ya que posibilita la adquisición de bienes y servicios que son indispensables para satisfacer las necesidades elementales de la vida.

Según Egüez y Pérez (2017), la desigualdad se relaciona generalmente con la pobreza. Esta relación tiene una dirección lineal, es decir, cuanto más desigual es la distribución del ingreso, mayor será el porcentaje de la población pobre. McKay (2002) manifiesta que la desigualdad es un concepto más amplio que el de pobreza, porque este último se refiere fundamentalmente a quienes apenas tienen ingresos suficientes para satisfacer sus necesidades básicas, mientras que la desigualdad puede referirse a situaciones no relacionadas con el ingreso; la desigualdad de género se da también en los países con ingresos altos.

Para la ONU la pobreza es “la condición caracterizada por una privación severa de necesidades humanas básicas, incluyendo alimentos, agua potable, instalaciones sanitarias, salud, vivienda, educación e información. La pobreza depende no solo de ingresos monetarios sino también del acceso a servicios” (UN, 1995,p.57).

En Costa Rica, existen dos formas de medir la pobreza, de modo unidimensional a través del línea de Pobreza (LP) y de manera multidimensional mediante el Índice de Pobreza Multidimensional (IPM) basada en la metodología Alkire–Foster (2008), la cual realiza un recuento de las privaciones que presentan los hogares, esto sin establecer restricciones en la determinación de indicadores, pues más bien promueve el análisis de las dimensiones e indicadores que visibilicen de una mejor forma la realidad y el contexto específicos de cada país.

Así mismo, se han realizado esfuerzos por determinar la pobreza por medio de otros métodos, como el de Necesidades Básicas Insatisfechas (NBI), calculado a partir de información proveniente de los censos de población y vivienda (INEC, 2015, p.8-9). Los objetivos de cada medición varían. El IPM, más que ser un índice que identifique quiénes tienen una condición de pobreza y quiénes no, tiene como fin convertirse en una herramienta de trabajo para que las instituciones coordinen esfuerzos y busquen de manera articulada atender las necesidades de la población costarricense en situación de pobreza (INEC, 2015).

Estas metodologías de medición tienen un objetivo específico que responde a una problemática, algunas cuestiones relacionadas con su medición se vinculan con el nivel socioeconómico. Las mediciones convergen en varios aspectos, pues todas están ligadas a una necesidad país; son insumos para toma de decisiones en las políticas públicas. Además, se relacionan con un término que puede definirse como nivel socioeconómico, sinónimo de la calidad de vida y qué tanto las personas satisfacen las necesidades elementales de la vida.

2.4 Marco de referencia

En este apartado, se detallan estudios realizados en América Latina y Costa Rica, sobre la construcción de estratos socioeconómicos, con el objetivo de tener un panorama más amplio acerca de las diferentes metodológicas de análisis que se desarrollan actualmente, así como, conocer variables determinantes en la medición de este constructo, que serán un parámetro de referencia para el desarrollo de esta investigación.

2.4.1 Estratificación socioeconómica en países de la región

a) México

Generalmente, las empresas relacionadas con estudios de mercadeo son las que han desarrollado en la región la mayor cantidad de metodologías referentes a la medición del nivel socioeconómico. Es el caso de México, donde se identifica a la Asociación Mexicana de Agencias de Inteligencia de Mercado y Opinión Pública (AMAI), quienes para el año 2018 crean la regla de medición del Nivel Socioeconómico (NSE) basada en el desarrollo de un modelo estadístico que permite clasificar a los hogares de una manera objetiva y cuantificable de acuerdo con su nivel socioeconómico (Asociación Mexicana de Inteligencia de Mercado y Opinión, 2018).

Esta segmentación clasifica a los hogares, es decir, a todos sus integrantes, de acuerdo a su bienestar económico de espacio, salud e higiene, comodidad y practicidad, conectividad, entretenimiento dentro del hogar, planeación y futuro (Asociación Mexicana de Inteligencia de Mercado y Opinión, 2018).

Las variables utilizadas para crear el algoritmo llamado Regla NSE 2018 que permite estimar el nivel de satisfacción de las necesidades más importantes del hogar son: escolaridad del jefe del hogar, número de dormitorios, número de baños completos, número de personas ocupadas de 14 años y más, número de autos y tenencia de internet. El resultado final de esta medición son siete calificaciones que se distribuyen de la siguiente manera.

Tabla 1. Distribución de los hogares de México por Estratos Socioeconómicos

Niveles	Puntos de corte	Porcentaje	Descripción
A/B	205+	6,0%	El 82 % de los hogares que pertenecen a este nivel tienen un jefe de familia con estudios profesionales. De dichos hogares, el 98% cuenta con internet fijo e invierten 13% de sus ingresos en educación, y apenas 25 % de sus ingresos totales en alimentos.
C+	166 - 204	10,9%	El nivel de estudios del jefe de familia también es profesional. El 89 % de dichos hogares cuentan con uno o más vehículos de transporte y el 91 % tienen acceso a internet fijo. El 31 % de sus ingresos se destina a alimentación y lo que se destina a calzado y vestido ronda el 5 %.
C	136 - 165	13,3%	El 13,3 % pertenecen a este nivel. Aquí el 81 % de los jefes de familia tienen estudios mayores a primaria y 73 % cuentan con conexión a internet. Tienen un vehículo en el hogar y el 35 % de su ingreso se destina a la alimentación mientras el 9 % se destina a la educación.
C-	112 - 135	14,3%	El 73 % están encabezados por un jefe de familia con estudios mayores a primaria. El 47 % de ellos tienen acceso a internet y el 38 % del gasto de dichos hogares se asigna para alimentos. El gasto en educación se reduce al 8 %.
D+	90 - 111	15,0%	El nivel de estudios del jefe de familia del 62 % de ellos es mayor a primaria y solamente el 19 % cuenta con internet. Destinan el 41 % de su gasto a alimentación y 7 % a la educación.
D	48 - 89	30,0%	En el 56 % de los hogares de este nivel el jefe de familia tiene estudios hasta primaria y apenas el 4 % tiene internet fijo. Poco menos de la mitad de su gasto, el 46 %, va destinado a los alimentos.
E	0 - 47	10,5%	El 95 % de los jefes de familia de dicho sector no tienen estudios mayores a primaria y la posesión de internet fijo en esos hogares es de apenas 0,1 %. Este es el nivel en el que mayor parte del gasto se asigna a alimentos con un 52 % y apenas dedican 5 % de su ingreso a la educación.

Fuente: Asociación Mexicana de Agencias de Inteligencia de Mercado y Opinión Pública (AMAI), (2017)

Como se muestra en la Tabla 1, los estratos A/B y C+ corresponden a estratos altos que albergan a los hogares con mayores recursos económicos; los estratos C, C- y D+ corresponden a estratos de clase media; mientras tanto, los estratos D y E son las agrupaciones con nivel socioeconómicos bajo, es decir, categorías conformadas por usuarios con menores recursos y poseen una menor calidad de vida.

En México, los estratos bajos representan el mayor porcentaje de la población, ya que 40,5% de los hogares se concentran en la categoría D y E. Identificar este tipo de comportamientos en la población es de gran utilidad; pues proporciona herramientas a las instituciones gubernamentales para orientar políticas públicas que provean mejores oportunidades a esos grupos. Existen otras metodologías para la medición socioeconómica en México, como la regla Rankia, pero la metodología elaborada por el AMAI es la más reconocida (AMAI, 2017).

Por otra parte, el Instituto Nacional de Estadística y Geografía (INEGI) creó, en el año 2015, el módulo de condiciones socioeconómicas, con el objetivo de contar con datos sobre ingresos, salud, educación, seguridad social, calidad y espacios de la vivienda, servicios básicos, alimentación y cohesión social, los cuales serán recolectados cada dos años paralelamente al proceso de levantamiento de información de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) (INEGI, 2015).

Esta información es útil para el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), pues este organismo tiene la atribución y obligación de establecer los lineamientos y criterios para la definición, identificación y medición de la pobreza. El CONDEVAL, al evaluar los programas sociales, contribuye a la mejora continua de las acciones dirigidas a combatir la pobreza. Este caso ilustra otro beneficio que puede generar la construcción de una estratificación de socioeconómica adecuada.

Otro aspecto que destaca en México es el sistema web creado por el INEGI, pues consiente de la creciente necesidad que tienen los tomadores de decisiones y los usuarios en general, de contar con información de calidad, oportuna y adecuada, genera una herramienta que permite construir estratos, según la investigación que se esté realizando, se podría definir como “Estratos a la carta” (INEGI, s.f.).

El sistema permite al usuario, entre otras facilidades, la libertad de seleccionar aquellas variables que muestran mayor afinidad con el tema de su interés y elegir uno o más procedimientos de estratificación; de este modo será posible disponer de dos o más estratificaciones. Una de las cualidades que esta herramienta posibilita es que el usuario pueda seleccionar la estratificación más conveniente para sus interés, así como la cantidad de grupos a generar, el nivel de la estratificación y los métodos, en ese sistema se pueden seleccionar tres opciones de estratificación el métodos k-medias, Mulvar y Mclust (INEGI, s.f.)

b) Perú

En el año 2018, la Asociación Peruana de Empresas de Investigación de Mercado (APEIM) desarrolló el nuevo nivel socioeconómico para la población peruana. Esta asociación agrupa a las empresas de investigación de mercados y de opinión pública en ese país (APEIM, 2018).

El nivel socioeconómico se calcula con base en la ENAHO, encuesta realizada por el Instituto Nacional de Estadística e Informática (INEI) de Perú, para elaborar los indicadores de pobreza. La estratificación no se define a partir de sus ingresos, sino en función a un grupo de variables clasificadas en cuatro dimensiones que se detallan en la siguiente tabla.

Tabla 2. Variables utilizadas para la construcción del nivel socioeconómico APEIM Perú

Variables	Bienes y servicios	Equipamiento del hogar	Servicios Públicos
Nivel educativo alcanzado por el jefe del hogar.	Auto o camioneta de uso particular (no taxi ni auto de empresa)	Computadora/Laptop	Teléfono fijo
Sistema de salud al que está afiliado el jefe de hogar.	Servicio doméstico pagado (que vaya por lo menos 1 vez por semana).	Refrigeradora/ Congeladora	Televisión por cable
Material predominante en los pisos de la vivienda.		Lavadora	Internet
Material predominante de las paredes exteriores de la vivienda. Conexión del baño.		Horno Microondas	

Fuente: Asociación Peruana de Empresas de Investigación de mercado APEIM Perú, (2018)

Como resultado final se obtienen cinco estratos, que se distribuyen de la siguiente manera.

Tabla 3. Distribución de los hogares de Perú por estratos socioeconómicos

Niveles	A	B	C	D	E
Porcentaje	4,7%	23,3%	41,3%	24,4%	6,4%

Fuente: Asociación Peruana de Empresas de Investigación de mercado APEIM Perú (2018)

En la Tabla 3, se identifica que la mayor distribución de la población peruana se concentra en los niveles medios y bajos. Se percibe también una distribución muy pequeña en el estrato alto, cerca de 5% de los hogares.

Por otro lado, en el estudio “Desafíos en la medición y el análisis del estatus socioeconómico de los estudiantes peruanos”, elaborado por la Oficina de Medición de la

Calidad de los Aprendizajes del Ministerio de Educación de Perú, se utilizan cuatro dimensiones para crear un índice de nivel socioeconómico. Para ello, toman en cuenta las variables de educación de los padres, los materiales de la vivienda, servicios básicos en el hogar, la dimensión de activos y otros servicios del hogar. Esta recopilación de variables y dimensiones obtenidas en estos estudios se tomarán de referencia para el análisis aquí planteado. Como se nota, todas las variables apuntan a condiciones relacionadas con el poder adquisitivo de los hogares (MINEDU, 2018).

c) Chile

La empresa de mercado llamada Asociación de Investigadores de Mercado (AIM), adoptó la clasificación socioeconómica de los hogares en Chile y en febrero del 2018 presentó la nueva metodología de segmentación y clasificación socioeconómica. Esta nueva clasificación se realiza a partir de un índice socioeconómico que resume la información de tres indicadores, el cual permite jerarquizar los hogares de Chile según su nivel de bienestar, y es la base para definir los Grupos Socioeconómicos (GSE) (AIM, 2018).

El índice que estima el grado de bienestar relativo de un hogar y sus miembros se calcula a partir de su tramo de ingreso (*per cápita* equivalente) de los niveles educacionales y ocupacionales de su principal sostén económico. La manera en que se construye este indicador es partiendo de un umbral que corresponde a la línea de pobreza definida por el INE de Chile y establece intervalos regulares del índice que delimitan a los demás grupos socioeconómicos.

La fórmula de cálculo:

$$\text{Índice socioeconómico} = A * B * C$$

A= tramo de ingreso *per cápita* equivalente (ajustado según economía de escala que corresponde al tamaño del hogar).

B= nivel educacional de principal sostén económico del hogar.

C= nivel ocupacional del principal sostén económico del hogar.

La metodología de clasificación se basa íntegramente en el análisis de estadísticas públicas que caracterizan en detalle los hogares chilenos y con muestras de gran tamaño. La fuente de datos utilizada para elaborarlo es la Encuesta de Presupuestos Familiares (EPF) que registra detalladamente los gastos de los hogares chilenos y la Encuesta

de Caracterización Socioeconómica Nacional (CASEN). A partir de la información, se obtienen siete niveles socioeconómicos que se detallan a continuación:

Tabla 4. Estratos Socioeconómicos de Chile 2018

Grupo	%	Ingreso total promedio del hogar	Educación	Ocupación	Sistema de salud	Tarjeta de crédito Bancaria	Vehículo Particular	Teléfono Móvil
E	13%	M\$324	La (54%) mayoría no va más allá de la enseñanza básica.	98% trabaja en oficios sin requisitos de educación formal, principalmente trabajadores no calificados.	95% público (FONASA).	5%	14%	84% tiene celular de prepago y 8% con contrato.
D	37%	M\$562	La mayoría (54%) llega a la enseñanza media pero no va más allá de ella	93% trabaja en oficios sin requisitos de educación formal principalmente operadores y vendedores	90% público (niveles A Y B de FONASA)	10%	22%	74% tiene celular con prepago y 17% con contrato.
C3	25%	M\$899	La mayoría 53% completa la enseñanza media y el 27% va más allá de ella	70% trabaja en oficios que no requiere educación formal y 30% accede a empleos que si lo requiere	76% público (niveles A Y B de FONASA)	24%	32%	58% tiene celular con prepago y 37% con contrato.
C2	12%	M\$1360	Mayormente profesionales técnicos (45%) y en menor medida universitarios (26%).	58% trabaja en oficios que requieren educación formal, 47% son técnicos y profesionales.	56% público (niveles B,C y D de FONASA) y 32% privado (ISAPRE).	41%	44%	59% tiene celular con contrato y 39% con prepago.
C1b	6%	M\$1986	Mayormente profesionales universitarios (71%).	81% trabaja en oficios que requiere educación formal principalmente directiva y profesional nivel alto (56%)	51% privado (ISAPRE) y 39% público (nivel D de FONASA)	60%	49%	72% tiene celular con contrato 26% con prepago
C1a	6%	M\$2739	Profesionales universitarios (95%) unos pocos con posgrado (13%).	92% con directivos y profesionales de alto nivel.	76% privado (ISAPRE) y 19% público (nivel D de FONASA)	77%	59%	84% tiene celular con contrato y 16% con prepago.
AB	1%	M\$6452	Profesionales universitarios (99%) buen aparte con posgrado (30%).	100% son directivos y profesionales de alto nivel.	89% privado (ISAPRE)	92%	78%	89% tiene celular con contrato y 10% con prepago.

Fuente: Asociación de Investigadores de Mercado (AIM,2018).

En la Tabla 4, se evidencia una concentración de la población chilena, para el año 2018, en las clases bajas, pues el 50% de la distribución total de la población pertenece a esa

categoría. De esta tabla resalta el comportamiento del tipo de seguro, ya que se nota que los grupos con un nivel económico alto presentan mayores porcentajes de servicios de salud privados, mientras que el caso de los grupos con mayores necesidades utiliza el servicio público.

d) Uruguay

En el 2018, se realiza una actualización del índice de nivel socioeconómico (INSE) construido por la Centro de Investigaciones Económicas (CINVE) y una versión reducida del mismo (INSE Reducido). La metodología del índice consiste en la estimación de un modelo de regresión lineal en el cual se introduce como variable dependiente una medida del poder adquisitivo de los hogares de Uruguay (Y) en función de un conjunto de variables independientes, a partir de los coeficientes β estimados por mínimos cuadrados ordinarios. Mediante un análisis de conglomerados, construyen una clasificación de los hogares en siete niveles, lo que implica establecer los valores del INSE que delimitan los siete intervalos (CEISMU, 2018).

La fuente de información utilizada es la última Encuesta Continua de Hogares (ECH) disponible, correspondiente al año 2017. La ECH ofrece información detallada sobre los ingresos de los hogares y diversas características socioeconómicas de los mismos (composición del hogar, nivel educativo, características de la vivienda y equipamiento).

Un aspecto relevante que se menciona en este estudio es el periodo útil del tipo de instrumentos como resultado del aumento del nivel de vida de la población, los cambios en la composición de los hogares y los nuevos patrones de consumo. Debido a que las variables consultadas tienden a perder capacidad discriminante a mediano y largo plazo. Por lo tanto, es fundamental cada cierto tiempo, someter a revisión del instrumento. En este caso, la actualización se realiza cada dos años.

La variable objetivo es el ingreso equivalente o ingreso por unidad de consumo. La medida del poder adquisitivo “toma en cuenta el tamaño y la composición del hogar, así como las economías de escala producto de la convivencia en una misma vivienda” (CEISMU, 2018, p.6).

El ingreso equivalente se obtiene dividiendo el ingreso total de un hogar por las unidades equivalentes de consumo de dicho hogar. Con miras de hacer esta definición operativa, es necesario disponer de las llamadas escalas de equivalencia, que consisten en un criterio para contar las unidades de consumo dentro de un hogar en función del número de

personas y de las características de estas. También se calcula el INSE Reducido, índice que pretende ser un instrumento para la clasificación de los hogares, pero con menos requerimientos de información que el INSE General.

La selección de las variables que conforman el índice reducido tiene como propósito utilizar variables simples de preguntar con el fin de obtener menor sesgo de las respuestas (dentro de las que conforman del INSE General). Otro aspecto que hace importante generar el índice del nivel socioeconómico reducido es mantener un razonable poder predictivo del ingreso de los hogares. Las variables utilizadas para la construcción del nivel socioeconómico en Uruguay se presentan en la Tabla 5.

Como resultado final, se obtienen siete estratos. A diferencia de los países citados anteriormente, en Uruguay los niveles medios presentan la mayor concentración de los hogares y representa 61% de la distribución total. Mientras que los niveles altos tan solo lo conforman el 15% de esta distribución.

Tabla 5. Distribución de los hogares de Uruguay por estratos socioeconómicos

Niveles	B-	B+	M-	M	M+	A	A+
Porcentaje	7,0%	16,0%	21,0%	22,0%	18,0%	11,0%	5,0%

Fuente: Cámara de Empresas de Investigación Social y de Mercado del Uruguay (CEISMU2018)

e) Ecuador

En Ecuador, existen diferentes estudios relacionados con el nivel socioeconómico elaborados por las empresas de mercadeo y las oficinas de estadística. Los resultados de estas investigaciones presentan entre cinco y siete estratos finales. El INEC de Ecuador desarrolla la Encuesta de Estratificación del Nivel Socioeconómico (NSE) que contribuye a homologar las herramientas de estratificación, así como construir una adecuada segmentación del mercado de consumo.

Para la construcción de la estratificación, se plantea un procedimiento detallado. En la primera etapa de esta estrategia, se exploran las variables a través de análisis descriptivos. Luego, mediante el escalamiento óptimo, se generan las variables de manera continua. En la tercera etapa, se crean puntajes por dimensión; mientras que en la cuarta se genera el índice y con el análisis de conglomerados se construyen los umbrales, los que se reajustan para estandarizar los resultados. Por último, se establece por medio de una regresión, el índice de nivel socioeconómico (INEC [Ecuador], 2011).

La estrategia de estratificación que plantea el INEC Ecuador se resume de la siguiente manera:

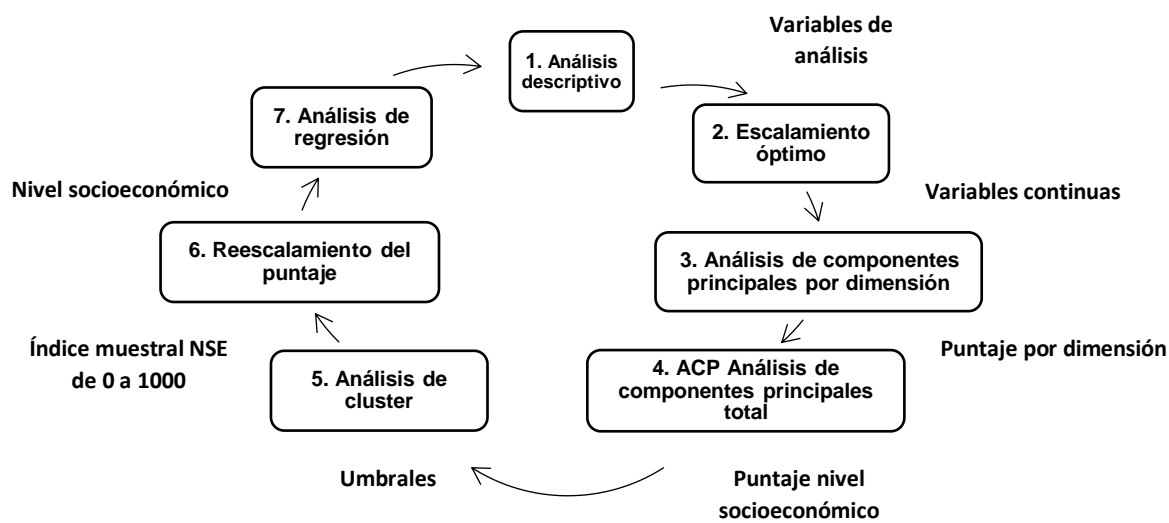


Figura 1. Estrategia de análisis para la construcción del nivel socioeconómico en Ecuador. Fuente: INEC Ecuador (2011).

Con esta estrategia metodológica, logran construir seis dimensiones que consisten en las características de la vivienda, el nivel de educación, actividades económicas del hogar, posesión de bienes, acceso a tecnología y hábitos de consumo. La jerarquía de importancia de las variables que utiliza el INEC Ecuador son las siguientes:

Tabla 6. Jerarquía de importancia de las variables utilizadas para definir nivel socioeconómico INEC Ecuador

Variables	Puntaje
Nivel de educación del jefe del hogar	171
Ocupación del jefe del hogar	76
Tipo de vivienda	59
Material predominante de las paredes exteriores de la vivienda	59
Afiliación al seguro privado/ seguro internacional/ seguro de vida	55
Material predominante del piso de la vivienda	48
Tiene este hogar servicio a Internet	45
Número de celulares activados	42
Tiene computadora portátil	39
Afiliación al seguro social (IESS, ISSFA o ISSPOL)	39
Tipo de servicio higiénico	38
Tiene computadora de escritorio	35
Número de televisores a color	34
Número de cuarto de baño con ducha de uso exclusivo del hogar	32
Tiene refrigeradora	30
Tiene cocina con horno	29
Registro a una página social	28
Uso de correo electrónico que no es de trabajo	27

Fuente: INEC Ecuador, (2011).

El resultado final de este análisis son cinco niveles socioeconómicos. La distribución y los umbrales se presentan en la siguiente tabla.

Tabla 7. Distribución de los hogares de Ecuador por Estratos Socioeconómicos

Niveles	Umbrales	Porcentaje
A	De 845,1 a 1000 puntos	1,9%
B	De 696,1 a 845 puntos	11,2%
C+	De 535,1 a 696 puntos	22,8%
C-	De 316,1 a 535 puntos	49,3%
D	De 0 a 316 puntos	14,9%

Fuente: INEC Ecuador, (2011).

De acuerdo con la tabla anterior, el estrato A corresponde al estrato alto, que alberga a los hogares con mayores recursos económicos. Los estratos B, C+ y C- corresponden a estratos de clase media y el estrato D es el estrato bajo, conformado por la población con menores recursos y que poseen una menor calidad de vida.

En la investigación de Egüez y Pérez (2017), también se estudia el nivel socioeconómico. Las variables analizadas por los autores se dividen en tres dimensiones: la vivienda (tipo de vivienda, número de piezas, número de piezas exclusivas para dormir, etc.), el hogar y jefe del hogar (años de estudio, nivel cursado, ocupación, nivel de ingresos y el hacinamiento, número de personas del hogar, etc.). Este estudio coincide con los anteriores, respecto a incluir un componente de la vivienda, así como un componente ocupacional y otro educativo.

2.4.2 Estudios relacionados con el nivel socioeconómico en Costa Rica.

En Costa Rica, a diferencia de los países latinoamericanos, las empresas de mercadeo no publican estudios oficiales sobre niveles socioeconómicos o estratos sociales, lo que podría explicarse por el alto costo que requiere el levantamiento de la investigación y el análisis. Otro aspecto que puede explicar este comportamiento es la ventaja competitiva que este tipo de investigación les brinda a estas empresas, por lo que, aunque las realicen, no optan por compartir la información.

Los enfoques de estratificación realizados en Costa Rica son variados. El estudio más reconocido en el ámbito estadístico es el realizado por Madrigal (2004). Este investigador nacional construye la estratificación de los hogares y segmentos del Censo del año 2000, a partir de la estimación de los ingresos generando un modelo de regresión con datos de la

Encuesta de Hogares de Propósitos Múltiples y usando variables predictoras como la educación, ocupación, pertenencias de la familia, hacinamiento, dependientes y ocupados. De esta manera, con el modelo final, estima el logaritmo del ingreso *per cápita* del hogar y lo aplica a los hogares del Censo 2000. El procedimiento K-Medias, define un total de seis estratos, que posteriormente reduce a tres (alto, medio y bajo), cambio que se hace por la comodidad de interpretar y uso.

El INEC utilizó para la estratificación del marco muestral del 2000 la metodología de Madrigal y en 2011 actualizó la estratificación con el mismo enfoque. Sin embargo, aunque existe una documentación oficial de los resultados obtenidos para el año 2011, esta es de uso interno.

Paralelo a esa investigación, García (2002) realizó un estudio que consistió en generar una segmentación social de los hogares. A diferencia del caso anterior, en esta investigación se utiliza la información demográfica contenida en el formulario a partir de indicadores que proveen datos sobre la composición por edad, nivel educativo del hogar y el ciclo de vida. Para ello se escogieron diecinueve variables cuya estructura factorial reportó buenas adecuaciones muestrales. Este análisis definió seis dimensiones a saber:

- a) Miembros con posibilidad de aportar al ingreso familia
- b) Desarrollo social
- c) Orientación a educación
- d) Fortaleza del jefe del hogar
- e) Tipología del hogar
- f) Movilidad del jefe

Con base en los indicadores, se corrió el procedimiento de Clúster con el algoritmo KMeans del SPSS versión 9.0 y se agruparon los hogares en diez segmentos. Posteriores análisis permiten reagrupar esos segmentos en un menor número de categorías según el interés del investigador (García, 2002, p.1).

Esta autora se centra en una segmentación de mercado basada en el Censo del 2000, la cual pretende introducir dentro del marco muestral como un indicador para estratificación socio-demográfica de los hogares.

Otros estudios relacionados con esta temática determinan las clases sociales a través de una estructura ocupacional, como es el caso de Vega, Castro, Gutiérrez y Rodríguez (1995). Entre los resultados más relevantes de su investigación obtiene cuatro clases

sociales (alta, media alta, media y baja). Esta agrupación coincide con la propuestas por Castro, Gutiérrez, Rodríguez y Barahona (2007), ya que estos investigadores establecen cuatro estratos sociales (alto, medio-alto, medio-bajo y bajo) que se agrupan en tres estratos, unificando los dos estratos medios.

El Ministerio de Vivienda y Asentamientos Humanos (MIVAH) creó en el 2012 una estratificación socioeconómica a partir de datos que brindan información específica para conocer las condiciones de vida de esta población, particularmente en materia de vivienda. Con esa información genera tres grupos de estratos socioeconómico medio-alto (ESM-A), medio-medio (ESM-M), y medio-bajo (ESM-B). Este estudio solo clasifica a la población media, por consiguiente, es de poca utilidad para este análisis. Lo que se puede rescatar es la implementación de preguntas relacionadas con la vivienda para la construcción de la clasificación (MIVAH, 2012).

Aunque se han identificado algunos estudios de estratificación de los hogares en el país, estos se realizaron hace una década; por lo tanto, existió un periodo en el cual no se publicaron investigaciones de esta temática. Es hasta el 2017 que Arias (2017) presenta su estudio: “Propuesta de indicador del ingreso de los hogares urbanos a partir de información del consumo: clasificación óptima de hogares urbanos por niveles de ingreso en Costa Rica”.

En este estudio, se aproxima el ingreso a través de variables menos sensibles y compara tres modelos que se definen por medio de una regresión múltiple. El modelo A lo define con base en el modelo de Madrigal (2004), donde se contemplan información del jefe del hogar y la vivienda. Los otros dos se construyeron con base en la Encuesta Nacional de Ingresos y Gastos de los Hogares del 2013 y en los gastos de consumo mensual de servicios de manera dicotómica (B) y continua (C). De los tres modelos, la autora concluye que el modelo B resulta mucho más fácil de aplicar y la forma de captar la información resulta más sencilla para el entrevistador y para el informante (Arias, 2017).

El segundo estudio es un seguimiento al anterior. Para Arias (2007), el objetivo es determinar el método más adecuado para construir niveles de ingreso de los hogares urbanos en Costa Rica, así como definir un número óptimo de categorías que permita segmentar la población en grupos que sean suficientemente heterogéneos entre ellos. De esta manera, aplica diferentes técnicas de estratificación tales como K-means, el método de Dalenius-Hodgens y el de Lavallée-Hidiroglou.

Los hallazgos más relevantes muestran que el ingreso se puede clasificar de manera adecuada en tres niveles o más, según los resultados del codo de Jambú. Además, al evaluar las técnicas de estratificación, deduce que el método de K-means y con Dalenius es muy similar, en todos los aspectos evaluados. Por último, concluye que la mejor predicción del ingreso se realiza al implementar el algoritmo de máquinas de soporte vectorial (Arias, 2017).

A partir de la revisión bibliográfica, se determina que el número de estratos construidos en el país con indicadores que tienen relación con el nivel socioeconómico es variable, entre tres y seis, aunque la mayoría se inclinan por establecer tres estratos; mientras que en los países de Latinoamérica este valor oscila entre cinco, seis y hasta siete grupos. Por otro lado, se detectan estudios más actuales en los países de la región implementados principalmente por oficinas de mercadeo. En Costa Rica no se acostumbra a publicar estudios de este tipo. Otro dato interesante es que se trabaja en tres dimensiones como mínimo: educación, ocupación y características del hogar.

III. Metodología

En este apartado, se muestran los principales aspectos metodológicos desarrollados con el fin de cumplir con los objetivos planteados. Inicialmente, se hace una descripción de las estrategias para construir los estratos socioeconómicos y, seguidamente, se desarrolla cada una de estas estrategias. Además, se presenta la fuente de datos requerida para el análisis y el planteamiento de las técnicas estadísticas. Por último, se detalla la validación de las estratificaciones elaboradas.

3.1 Descripción de las dos estrategias metodológicas

Esta investigación establece dos estrategias metodológicas para estratificar socioeconómicamente el marco muestral de viviendas de Costa Rica del 2011. A continuación, se presenta la estructura de análisis de cada metodología.



Figura 2. Estructura metodológica empleada en la investigación.

La estrategia metodológica A consiste en generar una estratificación mediante la estimación del indicador del ingreso, construido y validado con información de la ENAHO 2015. Una vez definido el modelo, se estimará en el archivo de datos del Censo 2011 y se procederá a estratificar el indicador final.

Por otro lado, la estrategia metodológica B consiste en crear indicadores con las variables de la base de datos del Censo 2011 para, posteriormente, hacer una estratificación multivariada. Adicionalmente, se realizará una variante a la metodología B, que consiste en hacer una combinación del indicador del ingreso generado en la estrategia A y los indicadores considerados en la metodología B, con el objetivo de complementar ambas estrategias.

En la Figura 2, se muestra cada una de las etapas que abarcan las estrategias A y B y la variante B. La etapa inicial consiste en identificar los indicadores o las variables requeridos para generar la estratificación. En todas las estrategias se realiza este paso, quizás uno de los más relevantes en esta investigación. Una diferencia importante entre estas metodologías es la fuente de datos que se utiliza para desarrollar los procedimientos. En la estrategia metodológica B solo se utiliza información del censo 2011; mientras que en la estrategia metodológica A y en la variante de la metodología B se utiliza tanto el archivo de datos del censo como el de la ENAHO 2015. Existen otros pasos que se comparten, como es el caso de la agregación de la información de las unidades (hogares) por UPM y la implementación del método de estratificación, lo que permitirá evaluar qué tan eficiente, en términos de variabilidad, es una estrategia en comparación con la otra.

3.2 Estrategia metodológica A

La estrategia metodológica A tiene como base la metodología desarrollada por Madrigal (2004), en el estudio *Estratificación de hogares y segmentos por niveles de ingreso en el Censo 2000* (Madrigal, 2004). Con esta estrategia, se busca construir un modelo de regresión lineal para estimar el ingreso *per cápita* neto de los hogares con el fin de reforzar y actualizar las variables independientes a partir de información de la ENAHO 2015.

Una vez obtenido el modelo, se replicará la regresión en el archivo del Censo de Población y Vivienda 2011 a nivel de hogares y después se agrega a nivel de UPM. Por último, ese indicador estimado del ingreso es el que se estratifica usando la técnica de k-medias. Pese a que existen varias técnicas de estratificación, en este estudio solo se considera el

procedimiento de k-medias dado que en los resultados obtenidos por Arias (2017) se identifica que genera estimaciones más precisas.

Madrigal (2004) incorpora como variable dependiente el logaritmo natural del ingreso *per cápita* y las variables independientes relacionadas con el jefe de hogar tales como nivel educativo, ocupación principal, categoría ocupacional y sector institucional. Además, incluye variables relacionadas con las pertenencias del hogar, aspectos económicos y demográficos. En ese entonces para la modelación se utilizaron seis indicadores que intentan describir el panorama socioeconómico del hogar mediante la siguiente ecuación:

$$\ln(\text{ingreso per cápita del hogar}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6$$

En donde:

X1= Nivel educativo del jefe de hogar

X2= Índice de pertenencias

X3= Dependencia económica

X4= Hacinamiento

X5= Miembros ocupados en el hogar

X6= Índice de ocupación del jefe de hogar

A continuación, se muestra una comparación de las variables utilizadas originalmente y las que se deciden utilizar en este estudio.

3.2.1 Construcción de indicadores en estrategia metodología A

En las investigaciones revisadas anteriormente (apartado 2.4), se menciona que el nivel socioeconómico está estrechamente ligado con el ingreso. Al no existir una medición directa de esta variable en el Censo Nacional de Población y Vivienda, se debe implementar alguna metodología para aproximarla. Algunas de las variables citadas por diversos autores relacionados con el nivel socioeconómico son: la educación, la ocupación y las pertenencias del hogar, entre otros.

La estrategia utilizada en este estudio es una actualización de la implementada por Madrigal (2004). Las variables independientes que se utilizan para estimar el ingreso *per cápita* del hogar son el índice de características de la vivienda, que incorpora variables como la tenencia de cielo raso y el estado de la vivienda, el índice del patrimonio acumulado del hogar conformado por el número de tenencias de interés en el hogar, el número de

miembros del hogar, el número de miembros ocupados, así como variables relacionadas con el jefe del hogar como el índice del tipo de aseguramiento, años de escolaridad y su ocupación.

Definición del jefe

Para desarrollar la estimación del ingreso, es común utilizar información laboral y personal del jefe del hogar. Algunas de las variables utilizadas por los investigadores son la condición de actividad, la ocupación de su empleo principal y su nivel educativo, entre otras. Sin embargo, el concepto del jefe de hogar varía según el estudio que se esté analizando y es por esta razón que se requiere plantear la definición que se utilizará en esta investigación.

En el caso de la ENAHO, el jefe no necesariamente es quien mantiene económicamente el hogar, sino la persona o personas consideradas como jefatura por los demás miembros del hogar. Generalmente, se entiende como la persona que lo lidera y toma decisiones. Por este motivo, es común encontrar casos en los que el jefe de la familia es el miembro de mayor edad (usualmente las personas mayores de la vivienda, que suelen ser los padres o abuelos), aunque estos se encuentren fuera de la fuerza de trabajo (Arias, 2017).

Para el INEC, esta es la persona de referencia, por ende, corresponde asignarle la línea uno del hogar y respecto a ella se establecen las relaciones de parentesco de los demás miembros del hogar. Para casos especiales, como los grupos no familiares, la jefatura es la persona que tiene la autoridad máxima, la que lleva la administración del hogar, la que tiene más tiempo de residir ahí o la de más edad. Estos criterios son válidos también si la persona informante de un hogar que constituye un grupo familiar no logra señalar cuál es el jefe del hogar.

Debido a que el interés en esta investigación es generar modelos que permitan obtener un indicador del ingreso *per cápita* del hogar, algunas de las variables utilizadas requieren la información de la persona que mantiene económicamente el hogar. Es por esta razón que se considera un orden de prioridad para captar la información, ya que únicamente el 69,9% de los hogares en la ENAHO 2015 registra jefes ocupados. Esta condición generaría una reducción del archivo de datos de un 29,1% de los hogares del país. Para evitar esta situación, se modificó la definición de jefe del hogar de la siguiente manera:

1. Se seleccionan los jefes de hogares que se encuentren ocupados.

2. Si la persona definida como jefe se encuentra desempleada, fuera de la fuerza de trabajo o inactivo, se selecciona a la persona del hogar que sea residente habitual de mayor edad y que estaba ocupado. Se excluye el servicio doméstico.
3. En hogares donde no existe ningún miembro ocupado, se toma el jefe de hogar desocupado.

Una vez aplicada esta definición para seleccionar los jefes de hogares, la cantidad de hogares cuyo jefe no tiene ocupación se reduce a 15,8%. Este porcentaje afecta las estimaciones, debido a que se construye un índice ocupacional por jefe de hogar, para introducirlo al modelo lineal. Por esta razón, fue necesario imputar el índice ocupacional en los jefes de hogar en la condición de desocupado. La imputación se realizó usando una regresión lineal donde se incluyeron como variables explicativas el sexo, años de escolaridad e índice de tenencia. Para dicho modelo se obtuvo un R cuadrado ajustado de 0,49. La ecuación de regresión resultante es la siguiente:

$$\begin{aligned} \mathbf{Imp\ Iocup} = & 0.250 + 0.362 * \mathbf{Sexo} + 0.217 * \mathbf{Años\ Escolaridad} + 0.297 \\ & * \mathbf{Índice\ patrimonio\ acumulado} + \mathbf{Edad} * 0.007 \end{aligned}$$

Imp Iocup = Modelo de imputación para el índice ocupacional.

Variables explicativas

Para iniciar el análisis, es necesario adaptar el archivo de datos de la ENAHO, ya que su estructura es a nivel de persona y se requiere a nivel de hogares en el cual se selecciona el jefe de hogar con base en los tres criterios mencionados anteriormente.

Variable dependiente el ingreso

Para la elección de las variables, se realizaron pruebas con el ingreso *per cápita* del hogar y el ingreso neto total del hogar; ambas variables funcionan como *proxy* del nivel socioeconómico. Sin embargo, para continuar con la línea base generada por Madrigal (2004), también se utiliza como variable dependiente el logaritmo natural del ingreso neto *per cápita* del hogar.

La variable ingreso requiere de un tratamiento especial. En primer lugar, se analiza el comportamiento identificando si existen valores perdidos, valores de cero o extremos. En

general no se detectaron valores faltantes, pero sí valores con ingresos iguales a cero (lo que genera un problema a la hora de calcular el logaritmo natural del ingreso, ya que el valor se indefine) y valores extremos en ambas colas. Para remediar esta situación, se calculó el percentil 2,5 y el 98,5 del ingreso *per cápita* y las observaciones fuera de este rango fueron imputadas a partir de un modelo de regresión lineal que considera variables de escolaridad, índice de patrimonio acumulado, índice de aseguramiento, tamaño del hogar y cantidad de ocupados, debido a que los valores extremos pueden ocasionar problemas en el ajuste de modelo.

Variables independientes del modelo base y el actualizado

Una evolución y actualización de los indicadores del modelo base siempre será una buena práctica debido a que todo cambia. Lo que en un momento era un buen predictor, puede dejar de serlo y otras variables, que antes no se consideraban, pueden tener un aporte fundamental en la estimación del ingreso *per cápita* neto del hogar.

Las empresas de mercado suelen desarrollar actualizaciones cada dos o tres años, pues consideran que ese es el periodo de vida útil que se recomienda para generar una estratificación socioeconómica. Sin embargo, en este caso es complicado establecer este tiempo, puesto que el marco se construye a partir del censo, por lo que esa actualización, como mínimo, se puede hacer cada diez años.

A continuación, se presentan las variables individuales utilizadas en el estudio y sus características. Estas variables, posteriormente, serán transformadas a índices para que funcionen como variables independientes en el modelo de regresión. En total se utilizaron 15 preguntas de la ENAHO 2015, tal y como se muestra en la siguiente tabla. Estas fueron escogidas después de una revisión exhaustiva de la bibliografía relacionada con esta temática y determinar que teóricamente pueden explicar el ingreso o bienestar económico de un hogar. Estas variables se eligen por el criterio de validez lógica o facial (Babbie, 1989).

Tabla 8. Variables consideradas para la construcción del modelo del *ingreso per cápita* del hogar

Grupo	Variables
Relacionadas con el jefe del hogar	1. Años de escolaridad 2. Posición en el empleo principal 3. Ocupación en el empleo principal 4. Sector institucional 5. Tipo de seguro social
Relacionadas con la característica de la vivienda	6. Estado físico de la vivienda 7. Tenencia de cielo raso
Relacionadas con el patrimonio acumulado del hogar	8. Sistema de agua caliente 9. Computadora portátil 10. Automóvil 11. Teléfono residencial 12. Internet 13. Pantalla LCD o LED
Económica	14. Condición de actividad
Hogar	15. Número de miembros del hogar

Las 15 variables seleccionadas fueron resumidas posteriormente en 7 covariables. Al comparar los indicadores de esta investigación, en relación con los utilizados por Madrigal (2004), se identifican algunas diferencias. En la Tabla 10, se muestran las variables utilizadas en las dos propuestas y se evidencia la incorporación de dos nuevos indicadores y el descarte de otros dos (la condición de hacinamiento y la proporción de dependencia económica). Estos indicadores se descartan dado a que se pudo corroborar que su exclusión no afecta de manera importante el modelo final obtenido.

Tabla 9. Comparación de variables independientes modelo base y modelo propuesto

Propuesta base Madrigal 2004	Nueva propuesta
Nivel educativo del jefe de hogar	Años de escolaridad del jefe de hogar
Índice de pertenencias	índice del patrimonio acumulado del hogar
Dependencia económica	índice del tipo de seguro
Hacinamiento	índices de características de la vivienda
Miembros ocupados en el hogar	Miembros ocupados en el hogar
Índice de ocupación del jefe de hogar	Índice de ocupación del jefe de hogar
	Número de miembros del hogar

A continuación, se presenta la construcción de cada una de las variables propuestas en esta actualización del modelo de regresión.

A. Nivel de escolaridad del jefe del hogar

La variable escolaridad es común utilizarla para estimar el nivel socioeconómico de los hogares. El comportamiento esperado de este indicador es que haya una relación positiva entre la cantidad de años de escolaridad y el poder adquisitivo de las familias y, por ende, con un mayor nivel socioeconómico. Esto se explica porque el aumento de los años de educación permitiría aumentar las probabilidades de obtener mejores empleos y con mejores condiciones, de acuerdo con la preparación y especialización. Existen muchas formas de incorporar esta variable en el modelo; sin embargo, para este caso se incluirán los años de escolaridad del jefe de hogar. Se asume que, a mayor número de años de educación del jefe, es más probable que el hogar posea un mayor nivel socioeconómico y viceversa.

En el estudio de Madrigal (2004), se utiliza el nivel de escolaridad en siete categorías, dado que no existían diferencias significativas entre el uso de la variable discreta o categórica (p.35). Por otro lado, en la investigación realizada por Arias (2017), se utiliza la variable años de escolaridad. Por practicidad, en este estudio se utiliza los años de escolaridad del jefe como variable independiente.

En la Tabla 11, se presenta una comparación entre la variable utilizada por Madrigal y la que se utilizará en esta investigación para generar la estimación del ingreso.

Tabla 10. Comparación de variables independientes modelo de Madrigal y modelo propuesto en esta investigación.

Propuesta base Madrigal 2004		Nueva propuesta
Nivel Educativo	Escala	
Sin escolaridad	0	
Primaria incompleta	1	
Primaria completa	2	
Secundaria Incompleta	3	Años de escolaridad Variable continua
Secundaria completa	4	
Universidad incompleta	5	
Universidad completa	6	

Los valores faltantes se imputaron asignando la mediana de los cinco casos más cercanos (3% de los casos). Para imputar el valor más acertado se ordenó previamente el archivo de datos por UPM, ingreso y edad.

B. Índice del patrimonio acumulado del hogar

En la versión base, al índice que contempla la tenencia de equipamiento del hogar se le asignó el nombre de índice de pertenencia. Para ese entonces, la construcción se realizó a partir de nueve variables. En esta nueva versión, se denominará índice de patrimonio acumulado del hogar. Para elaborarlo, se toman únicamente seis variables referentes al equipamiento del hogar.

Este cambio se da debido a que existen variables que ya no discriminan el ingreso de las familias costarricenses. Al igual que Madrigal (2004), para elegir las seis variables, se utilizan los betas estandarizados de una regresión lineal múltiple con el propósito de identificar cuál de los artefactos incluidos en la ENAHO 2015 y presentes en el Censo 2011 es el más importante en términos de significancia y del peso de sus betas con la intención de explicar el logaritmo natural del ingreso *per cápita* neto. Para construir este indicador, se tomó la decisión de incluir aquellas pertenencias que tienen una mayor importancia en la ecuación.

Las variables incluidas en la regresión son dicotómicas, de manera que se le asignará el valor de uno si hay presencia de ese equipamiento en el hogar y cero si no lo hay. El razonamiento del comportamiento de esta variable sugiere que un número mayor de artefactos o pertenencias en el hogar es producto de un ingreso mayor. Por ello, cuando la presencia es nula, se espera que el ingreso sea menor que si se tuviera el total de artefactos. La siguiente tabla muestra la composición del índice de pertenencias del modelo base y el índice propuesto en este estudio, el cual reduce la cantidad de variables que lo conforman de nueve pertenencias a seis.

Tabla 11. Comparación modelo base y modelo propuesto para el índice de pertenencias

Equipamiento del Hogar	Modelo base	Propuesta¹
Lavadora	x	
Microondas	x	
Refrigeradora	x	
Ducha para agua caliente	x	
Tanque para agua caliente	x	x
Televisor a color	x	
Televisor plasma. LCD o LED		x
Teléfono habitacional	x	x
Televisión por cable o satelital		
Computadora portátil	x	x
Automóvil		x
Sistema de agua caliente para toda la casa		x
Total	9	6

El alfa de Cronbach de este índice es de 0,715 el cual se considera bastante bueno, con base en las variables escogidas, se creó el índice del patrimonio acumulado del hogar por medio de una suma y se considera que a mayor valor del índice se espera que el hogar tenga un mayor nivel socioeconómico (Ver Cuadro 5). La fórmula de cálculo es la siguiente:

$$\text{Índice de patrimonio acumulado del hogar} = \sum_{i=0}^6 x_i$$

C. Índice de ocupación

Para elaborar este índice, se utiliza el ingreso principal neto total promedio del jefe del hogar por ocupación en el empleo principal y la categoría ocupacional. Esta variable se construye mediante la combinación de la posición en el empleo principal y el sector institucional. Este índice es similar al desarrollado en la propuesta de referencia; sin embargo, en ese caso Madrigal utilizó el ingreso mensual neto. Esta diferencia se da porque en la ENAHO 2015 se eliminó la variable ingreso mensual neto; así que se procede a buscar una variable equivalente. Además, la información de la categoría ocupacional se recopiló de forma distinta a la encuesta del 2015 y por tanto las clasificaciones realizadas en este trabajo difieren un poco.

Tabla 12. Comparación en la construcción de la categoría ocupacional en el modelo base y el propuesto en esta investigación

Categoría Ocupacional original	Categoría Ocupacional propuesta
1. Empleado del Estado	1. Empleado institución pública
2. Empleado institución autónoma	2. Empleado empresa privada
3. Empleado empresa privada	3. Cuenta propia
4. Cuenta propia	4. Servicio doméstico
5. Patrono socio activo	5. Patrono
6. Servidor doméstico	

Al realizar el cruce entre la variable categoría ocupacional y ocupación en el empleo principal, se genera el Cuadro 1. En ese cuadro, se observa cómo los jefes con condición de asalariado del hogar privado, es decir, las trabajadoras domésticas, son las que reciben menos remuneración (promedio de 148 834 colones). Por otro lado, los patronos son los que reciben en promedio mayor ingreso (956 456 colones).

Al analizar los ingresos según la ocupación del empleo principal, también se encuentran resultados esperados: la población que se dedica a realizar ocupaciones elementales tiene un salario de 209 753 colones, mientras que la remuneración de los directores y gerentes

es casi nueve veces mayor. Por otro lado, los profesionales científicos e intelectuales presentan un salario mensual de 1 106 353 colones.

Para conformar el índice de ocupación, se toman los datos del Cuadro 1 con el fin de establecer categorías de acuerdo con los ingresos de los hogares. Para eso, se realizan grupos de individuos que tienen ingresos similares en ocupación del empleo principal y categorías ocupacionales usando la técnica de K medias. Cuando se generan los grupos, se realiza un ordenamiento de manera ascendente de acuerdo con el ingreso centroide de cada grupo, con el fin de que, al aumentar la categoría del índice, incrementemente también el ingreso promedio.

La construcción de los grupos se hizo con tamaños de 5, 9,10 y 12. La selección final fue de 12 grupos y se realizó considerando las diferencias en el ingreso entre los grupos (ya que se buscaban las mayores diferencias posibles) y también se consideró el valor de la eta cuadrado (0,45) (Ver Anexo 1).

Cuadro 1. Ingreso principal neto total promedio del hogar por categoría ocupacional según ocupación del empleo principal del jefe de hogar

Ocupación del empleo principal	Categoría ocupacional					Total
	Empleado público	Empleado privado	Asalariado del hogar privado	Cuenta propia	Patrono	
Directores y gerentes	1 618 128	1 846 254		2 758 028	2 339 187	1 882 375
Profesionales científicos e intelectuales	1 182 993	991 874		802 890	1 468 286	1 106 353
Técnicos y profesionales de nivel medio	792 360	625 591		549 089	906 582	673 399
Personal de apoyo administrativo	575 807	395 024		417 399	600 619	441 865
Trabajadores de los servicios y vendedores de comercios y mercados	515 856	312 916	123 958	241 980	799 920	338 504
Agricultores y trabajadores calificados agropecuarios, forestales y pesqueros		319 404		292 361	563 911	338 871
Oficiales, operarios y artesanos de artes mecánicas y de otros oficios	589 319	340 840	26 833	202 498	580 700	310 683
Operadores de instalaciones y máquinas y ensambladores	526 833	369 419	492 413	290 258	802 486	367 479
Ocupaciones elementales	350 740	232 735	152 963	115 922	252 224	209 753
Total	852 734	419 290	148 834	287 605	956 456	467 242

En el Cuadro 2, se presenta el resultado del índice ocupacional. En general, se puede identificar un comportamiento esperable, pues el primer grupo se conforma por las

ocupaciones de operadores de instalaciones, máquinas y ensambladores, así como las ocupaciones elementales y trabajadores de los servicios, vendedores de comercios y mercados. La categoría ocupacional predominante en este grupo es el de las trabajadoras domésticas y su promedio de ingreso es 137 604 colones. El servicio doméstico es la única categoría ocupacional que solamente se observa en los primeros 4 grupos.

A partir del séptimo grupo, se identifican las ocupaciones de mayor nivel, como lo es el caso de profesionales científicos e intelectuales en el empleo público. Este grupo, en promedio, presenta un ingreso de 1 182 933 colones. Además, del grupo nueve en adelante, predominan los directores y gerentes, quienes presentan salarios promedios que varían desde 1 618 128 a 2 758 028 colones mensuales.

Cuadro 2. Estadísticos descriptivos del índice de ocupación

Grupo	No ponderado	Ponderado	Ingreso	
			promedio	Diferencia ¹
1	859	109 948	137 604	91 320
2	2 513	313 780	228 924	108 428
3	3 207	417 203	337 352	245 852
4	1 236	163 599	583 204	214 221
5	498	66 346	797 426	180 256
6	348	48 997	977 682	205 311
7	535	71 671	1 182 993	285 293
8	82	11 405	1 468 286	149 841
9	53	7 541	1 618 128	228 126
10	97	14 324	1 846 254	492 933
11	33	4 868	2 339 187	418 841
12	2	327	2 758 028	

^{1/} Corresponden a las diferencias entre los grupos $x+1$ y x

D. Índice del tipo de seguro

En Chile, el índice que estima el grado de bienestar relativo de un hogar y sus miembros desarrollado con información de las Encuestas sobre Caracterización Socioeconómica Nacional (CASEN) y la Encuesta de Presupuestos Familiares (EPF) utiliza variables relacionadas con el sistema de salud para perfilar a los grupos socioeconómicos de ese país. En esa Tabla 4, se muestra cómo los hogares que se encuentran en condiciones más

favorables optan por seguro privado, mientras que los hogares que forman un nivel socioeconómico bajo, tienden a asistir al seguro público o hasta incluso no disponen de él.

Al percibir una validez lógica o facial, se buscó en la ENAHO 2015 y en el Censo 2011 una variable relacionada a esta temática y se identificó la pregunta tipo de seguro social del jefe. Aunque esta variable estaba presente en ambos archivos, existía una diferencia en las categorías de medición, por lo tanto, el primer paso fue homologarlas. Posteriormente, se hizo una recodificación de las categorías para generar seis grupos; la agrupación se realizó utilizando dos criterios: se calcula el promedio del ingreso *per cápita* y se hace un ordenamiento según esta variable, con el objetivo de identificar cuáles categorías tienen un comportamiento similar y mediante criterio de experto, se definen los seis grupos.

Según esta lógica, dentro del primer grupo, denominado grupo cero, se encuentran los jefes que no dan una contribución económica al Estado por este servicio, es decir, los asegurados por el estado, los que no tienen seguro social de CCSS y los asegurados por el régimen no contributivo, tanto de monto básico como los de gracia o guerra. En la categoría de agrupación uno, se encuentran los jefes que contribuyen de manera indirecta, es decir, son asegurados por algún familiar.

En la agrupación dos, se encuentran aquellos jefes que tienen un tipo de seguro voluntario, cuenta propia o por convenio. Este grupo presenta un ingreso *per cápita* promedio de 365 980 colones. Los grupos 3 y 4 tienen un promedio *per cápita* similar. Por último, la categoría cinco presenta el mayor ingreso *per cápita* promedio de 1 135 281 colones (Ver Anexo 3).

Este índice de tipo de seguro social presentaba 12 casos faltantes, por tanto, se imputaron los casos utilizando el vecino más cercano, después de un ordenamiento por ingreso e índices de patrimonio acumulado.

E. Índice de características de la vivienda

Este índice se construye a partir de la suma de dos variables. Una de ellas es el estado general de la vivienda, que se obtiene combinando la respuesta sobre el estado de las paredes exteriores, techo y piso (bueno, regular y malo). La otra es la variable dicotómica que consulta si la vivienda tiene cielo raso. Es decir, el valor máximo de este indicador es cuatro y el valor mínimo es cero. Se espera que las viviendas en mejor condición y que presenten cielo raso se encuentren habitadas por personas de mayores ingresos.

F. Número de ocupados

El número de ocupados fue de los últimos índices construidos. Para calcularlo, solamente se contó a los residentes habituales del hogar que cumplieran con la definición de ocupado. Es de esperar que, a mayor número de ocupados, mayor sea la condición económica del hogar.

G. Número de miembros del hogar

El número de miembros del hogar se construye con la suma de los residentes habituales del hogar; es un cálculo simple, pero con un gran peso en la predicción del ingreso. El comportamiento de esta variable indica que a mayor número de miembros en el hogar existe una tendencia negativa en el ingreso del mismo, el bienestar o poder adquisitivo de las familias. Esto puede explicarse dado que a mayor número de miembros los ingresos deben ser distribuidos en mayor cantidad de personas, lo que limita en algunos casos las oportunidades de los miembros del hogar.

3.3 Estrategia metodológica B

La estrategia metodología B consiste en identificar variables de utilidad para llevar a cabo la estratificación del marco. Generalmente, son variables relacionadas con la pobreza y el empleo, entre otras, pues esas variables son las de diseño de las principales encuestas a hogares que produce el INEC. Una vez elegidas las variables se construyen los indicadores que serán el insumo principal para construir la estratificación socioeconómica.

El propósito en esta estrategia es generar indicadores a partir de variables binarias para cada una de las viviendas y hogares que conforman la UPM. Si la vivienda posee la característica de interés, el indicador asumirá el valor uno y cero en caso contrario.

Posteriormente, en cada una de las UPM se deberán sumar los valores reportados en las viviendas u hogares que cumplen con la condición. Ese valor se deberá dividir entre el total de viviendas de la UPM, de manera que se obtenga el porcentaje de viviendas con la característica de interés en cada unidad del marco. Además, existe la posibilidad de conformar algún indicador de carácter numérico.

De esta manera, construidos los indicadores y agregados por UPM, se procederá a construir una estratificación multivariada. La lógica de esta estrategia es generar grupos o estratos a partir de algunas variables de interés.

Una vez validados cada uno de los indicadores construidos para la estratificación, se realiza una de las etapas más complejas para cualquier investigación de este tipo, que consiste en elaborar el número de grupos o estratos. Para hacerlo, se utilizarán diferentes estrategias con el fin de determinar el número óptimo de estratos. Dado los estudios anteriores, es preferible conformar un número de estratos impar, como se explicará más adelante.

3.1.1 Construcción indicadores estrategia metodológica B

Como se evidencia teóricamente, el concepto del nivel socioeconómico no es fácil de definir, por lo que es más difícil contar con una definición que satisfaga todas las áreas del conocimiento. Por consiguiente, en esta estrategia se desarrolla una aproximación a este concepto mediante diferentes variables. Para elaborar la estrategia metodológica B y así lograr un mayor aprovechamiento de la información censal, se trabajará con las siguientes variables:

Porcentaje de hogares en la UPM con al menos una necesidad básica insatisfecha (NBI): el NBI es una medida multidimensional que tiene como objetivo identificar cuáles hogares son pobres en función de un conjunto de necesidades socialmente consideradas básicas, de manera que se clasifican como hogares pobres aquellos que no logran satisfacer los umbrales mínimos determinados para cada necesidad; por consiguiente, se busca la identificación y caracterización de los hogares que no satisfacen aquellas necesidades humanas que dependen de condiciones económicas.

Para analizar las NBI, se construye una variable dicotómica por hogar, donde el cero indica que el hogar no tiene carencias y el uno indica que el hogar tiene al menos una carencia. Por lo tanto, al dividir la suma de hogares que cumple con la característica de interés, entre el total de hogares por UPM multiplicado por 100, se obtiene el porcentaje de hogares con al menos una necesidad básica insatisfecha por UPM. La manera de interpretar este indicador es que a mayor porcentaje de hogares con NBI, el nivel socioeconómico de la UPM se clasificará como menor.

Las dimensiones utilizadas por el INEC para definir el NBI son cuatro y se definen de la siguiente manera Méndez y Trejos (2004):

- Acceso a albergue digno: considera la calidad de la vivienda, el hacinamiento por cantidad de dormitorios y el acceso al alumbrado eléctrico.

- Acceso a vida saludable: considera el abastecimiento de agua de buena calidad para satisfacer las necesidades de alimentación e higiene, y la eliminación de excretas.
- Acceso al conocimiento: para cuantificar esta carencia, se consideraron dos componentes: asistencia escolar y rezago escolar, ambos para la población de 7 a 17 años. Todos aquellos hogares con al menos un miembro entre 7 y 17 años que no asistiera a la escuela o al colegio se cuantificaban con carencia. Sucedió lo mismo si un miembro o más asistía a dichas instituciones educativas, pero presentaba un rezago de más de dos años.
- Acceso a otros bienes y servicios: se refiere a la capacidad de consumo que refleja la disponibilidad potencial de recursos del hogar. Para esta dimensión, no se definieron componentes, pero las variables utilizadas para medir dicha carencia fueron la edad y educación del jefe del hogar, la presencia y educación de los perceptores regulares del hogar y el número de dependientes de estos.

Tasa de desempleo por UPM: este indicador se construye a partir del número de personas desocupadas por UPM, dividido entre el total de población que pertenece a la fuerza de trabajo en la UPM. La fórmula de cálculo de es la siguiente:

$$\text{Tasa de desempleo abierto TDA} = \frac{\text{Población desocupada}}{\text{Fuerza de trabajo}} * 100$$

Se espera que a mayor tasa de desempleo en la UPM esta tenga una condición más desfavorable, por ende, se debe clasificar en un nivel socioeconómico bajo.

Tipo de seguro social (Sin seguro o asegurado por el estado): en la estrategia A, se identificó que el tipo de seguro social del jefe es importante para determinar el nivel socioeconómico del hogar y que tanto los jefes asegurados por el estado como los que no están asegurados representaban a la población más vulnerable en términos de pobreza; por lo que en esta nueva estrategia se construye un indicador dicotómico, donde el jefe que se encuentre asegurado por el estado o no presente seguro social, se le asigna un uno y aquellos donde existe un tipo de seguro social diferente a estos, se le asigna un cero.

Posteriormente, se agregan los hogares por UPM que cumplen con la característica, se dividen entre el número de hogares de la UPM y se multiplica por 100, con el objetivo de

obtener el porcentaje de hogares con una condición desfavorable en la UPM. Se asume entonces que, a mayor porcentaje, menor será el nivel socioeconómico.

Al menos un universitario en el hogar: como se ha mencionado, la educación es una variable que tiene una estrecha relación con el nivel socioeconómico y una manera de incorporarla en esta estrategia es mediante la construcción de una variable dicotómica. Se asigna un uno cuando existe al menos un universitario en el hogar y cero cuando no sea así. De igual manera, se calcula un porcentaje en la UPM y a mayor porcentaje, más alto será el nivel socioeconómico de la UPM.

3.4 Variante estrategia B (estrategia mixta)

A raíz de las dos estrategias desarrolladas anteriormente (apartado 3.2 y 3.3), se elabora una variante a la metodología B, la que consiste en hacer una estratificación multivariada, en la que se combinan la estimación del ingreso generado en la estrategia A y los indicadores considerados en la estrategia B, con el objetivo de complementar ambas metodologías. Esta variante trabaja con cinco indicadores.

3.5 Datos

Para elaborar la estrategia metodológica A, se utiliza la base del Censo 2011 y para la construcción del modelo que estima el ingreso se utilizan variables de la ENAHO 2015. Este archivo de datos se divide en dos partes: la primera donde se construye el modelo contempla el 80% del total de los hogares y en el restante 20% se utiliza para validar el modelo generado.

La razón por la que se utiliza la encuesta del 2015 se debe a que para ese año se da el cambio completo de los segmentos (Metodología aplicada en el censo del 2000) por UPM (Metodología aplicada en el censo del 2011). Además, como todas las encuestas de hogares, contiene un conjunto de variables similares a las que indaga el censo tanto en el tema de población como en el de vivienda; por lo que su aprovechamiento es importante para desarrollar el presente trabajo. Para generar algún indicador que aproxime el ingreso, es vital contar con variables iguales en ambos archivos de datos, o por lo menos variables *proxys* que permitan reconstruir los indicadores en el archivo de datos del Censo, para poder estimarlo y así estratificarlo.

Una de las ventajas presentes en la ENAHO, en relación con el censo, es que se emplean preguntas más detalladas en ciertas áreas y también cuenta con el ingreso que, aunque representa una pregunta sensible, es variable *proxy* del nivel socioeconómico.

La estrategia general para realizar la estrategia A se ejecutó en cuatro pasos, los que se resumen a continuación:

1. Se usó el archivo de datos de la ENAHO-2015 para construir una ecuación que estimara el ingreso *per cápita* de los hogares (80%).
2. Con la ecuación, se estima el ingreso *per cápita* de los hogares del Censo 2011.
3. Se genera un indicador del ingreso por UPM.
4. Se estratificaron las UPM del MMV-2011.

Para trabajar la estrategia metodológica B y su variante, se utiliza la base del Censo 2011. Con el objetivo de realizar la estratificación, se debe construir un archivo en el ámbito del UPM al aplicar cada estrategia, en el que se hace un agregado con los hogares de las 6678 UPM que corresponden al marco muestral de la zona urbana del país. Posteriormente, cada UPM será clasificada en cada uno de los estratos socioeconómicos del marco muestral de viviendas.

3.6 Delimitación geográfica

En este trabajo, se utilizan los datos de los hogares urbanos y rurales para elaborar el modelo regresión que estima el ingreso *per cápita* de los hogares (ingresos e información del hogar y del jefe de hogar, entre otras características de interés). Sin embargo, solo se va a estratificar la zona urbana. Las investigaciones elaboradas en el país y en Latinoamérica limitan el análisis solo a la zona urbana. Esta decisión se justifica por la dinámica de la economía, pues existe un comportamiento distinto en la zona urbana y en la rural. De esta manera, se evidencia que la forma de captar los ingresos debe ser analizada independientemente en cada zona (Da, Xiao, y Zhuo, 2009).

Valdés, Foster, Pérez, y Rivera (2008) y Arias (2017) indican que excluir la población rural del análisis es una práctica común, porque el sector rural involucra estipendios no monetarios en mayor proporción (más difícil de medir) que en la zona urbana.

3.7 Métodos y técnicas estadísticas

Para crear los índices, el tratamiento de los archivos y el análisis de la información que contribuye en la generación de estratos socioeconómicos, se utilizan diferentes técnicas estadísticas. En este apartado, se explicará brevemente sobre los modelos predictivos, el análisis de conglomerados y los estadísticos necesarios para identificar el número óptimo de conglomerados.

a) Modelos predictivos

Existen diferentes técnicas para realizar análisis predictivo, entre los que se encuentra la regresión múltiple. Esta técnica es prácticamente igual que el modelo lineal simple; no obstante, la variante principal surge al considerar más de una variable explicativa (Cuadras, 2010). Este concepto consiste en obtener una variable respuesta Y sobre m variables explicativas, X_1, \dots, X_m , lo que matemáticamente es expresado de la siguiente manera:

$$y_i = \beta_0 + x_{i1}\beta_1 + \dots + x_{im}\beta_m + e_i, \quad i = 1, \dots, n,$$

La expresión y_i es la i -ésima observación de Y , lo que en este caso será el ingreso *per cápita* neto de cada uno de los hogares que responde a la ENAHO. De igual manera, las $X_{i1} \dots X_{im}$, representan las i -ésimas observaciones de las covariables o variables explicativas y e_i es el vector que contiene los términos de error aleatorio que tienen media igual a cero ($E(e_i)=0$) y varianza constante $\sigma^2(e_i)=\sigma^2$ (Da, Xiao y Zhuo, 2009).

La expresión matricial de este modelo es el siguiente:

$$Y_{n \times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad X_{n \times p} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1,p-1} \\ 1 & X_{21} & \dots & X_{2,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \dots & X_{n,p-1} \end{bmatrix} \quad \beta_{p \times 1} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \quad \varepsilon_{n \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Los parámetros se estiman a partir del método de mínimos cuadrados ponderados para β_0 y β_1 la expresión matemática es la siguiente.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Los coeficientes β se interpretan como el peso que tiene el incremento en una unidad de la variable explicativa, en la variable respuesta Y . Para garantizar un buen poder predictivo en el uso de esta técnica, aparte de requerir de un coeficiente de determinación (R^2) alto, es necesario el cumplimiento de algunos supuestos como lo son la linealidad de la variable dependiente, la homocedasticidad, normalidad en los errores y la independencia de las variables explicativas (Rojo, 2007).

Para estimar el ingreso, en la estrategia metodológica A se utilizará este método, a partir de variables identificadas en la ENAHO 2015. Además, esta técnica se utilizará para la imputación de algunos casos faltantes.

b) **Análisis de conglomerados**

El análisis de agrupamiento es un conjunto de técnicas para formar grupos o “clúster” de individuos u objetos que sean semejantes según algún criterio. Hernández (2013) menciona que estos grupos se conforman “con base a ciertas características o variables, el punto de partida para la creación de grupos es un conjunto de individuos u objetos, llamados también casos, en cada uno de los cuales se han medido u observado las p variables” (p.227).

Por otro lado, Fernández (2011) lo define como “el procedimiento estadístico multivariante que comienza con un conjunto de datos conteniendo información sobre una muestra de entidades e intenta reorganizarlas en grupos relativamente homogéneos a los que llamaremos clúster o conglomerados” (p.29).

Existen diversos métodos para conformar las agrupaciones. Para este análisis, se usará el procedimiento K-medias, pues se utiliza cuando la cantidad de datos es grande. La técnica K-medias inicia después de haberse especificado el número deseado de grupos por construir; los primeros K casos en el conjunto de datos se toman como los K grupos iniciales. Es decir, cada uno de estos K casos se constituyen en su propio grupo, cada uno

tiene como centro o centroide el vector de las p observaciones correspondientes a las p variables del caso que se define como grupo. Luego, se procede a asignar cada uno de los restantes $n-K$ casos, uno a la vez, al grupo cuyo centro está más cerca de él (Hernández, 2013).

c) Número de conglomerados

Para determinar el número de estratos en las diferentes estrategias implementadas, se utilizaron algunos estadísticos como el coeficiente Eta Cuadrado, el cual se calcula de la siguiente manera:

$$Eta = \sqrt{\frac{\sum n_j (m_j - G)^2}{(\sum n_j (m_j - G)^2) + (\sum (n_j - 1) \times s_i^2)}} = \sqrt{\frac{SCEntre}{SCEntre + SCResidual}}$$

Además, se desarrollaron dos métodos adicionales para la detección del número de conglomerados: el coeficiente de variación por grupo y el gráfico de sedimentación o codo de Jambú.

Por otro lado, existía un gran interés en obtener un número impar de grupos dadas las diferentes experiencias detalladas en el marco de referencia, ya que un número impar de grupos facilita la interpretación de los resultados (INEGI, s, f; AMAI, 2018; APEIM,2018; MINEDU,2018; AIM, 2018 y INEC,2011).

3.8 Validación interna

Se realiza una validación interna de la estimación del ingreso por medio de correlaciones. Esta validación permite determinar si la estimación está midiendo lo estipulado. Para la estrategia A, se dividió el archivo de datos de la ENAHO 2015 en dos, se considera el 80% de los hogares para el cálculo del modelo y el restante 20% como prueba. En este último archivo, se calculó una correlación de Pearson entre el logaritmo natural del ingreso real

contra el logaritmo del ingreso estimado por el modelo y dio como resultado una correlación de 0,812. Así que existe una relación fuerte y positiva entre estas variables.

Por otro lado, para validar la estratificación resultante de cada estrategia, los grupos socioeconómicos conformados se cruzaron con variables que tuvieran relación con el nivel socioeconómico tales como, vivienda en precario, la nacionalidad del jefe de hogar, vivienda hacinada y tubería dentro de la vivienda, entre otros.

3.9 Estudio de simulación para la validez externa

La validez externa de las estratificaciones construidas con las diferentes metodologías se realiza mediante un estudio de simulación que consiste en adoptar el diseño de muestra de algún estudio importante para el INEC, como la ENAHO. Esta simulación es una alternativa de validación externa ante la carencia del tiempo disponible para hacer una prueba de campo.

El procedimiento consiste en seleccionar del marco muestral (que fue construido con base en el censo) muestras simuladas por el uso del diseño muestral de la ENAHO 2015 con cada una de las propuestas de estratificación y la comparación entre las variabilidades obtenidas de las diferentes variables de diseño.

La ejecución la simulación consiste en la selección iterada de la muestra de 1120 UPM, según la asignación propuesta en el diseño ENAHO 2015 y, a partir de esas muestras, se estiman las variables de interés: NBI, TDA, porcentaje de carros, internet y computadoras portátiles, así como el total de personas, hombres, mujeres y número de ocupados.

Este procedimiento se itera 1005 veces, debido a que a partir de ese número se estabilizó la solución, por tanto, se tendrá 1005 estimaciones por cada una de las variables de interés y propuesta metodológica, para facilitar el análisis cada estimador será graficado con el objetivo de conocer su distribución. Así mismo, se incluirá en el gráfico los percentiles 2,5 y 98,5, el valor poblacional y las estimaciones promedio de cada variable utilizando cada una de las estrategias.

Para reforzar el análisis de cada una de las muestras, se extraerá el coeficiente de variación y el efecto del diseño para las variables de interés. El comportamiento de estos estadísticos también será evaluado en esta sección.

El diseño de muestra de la ENAHO será utilizado para este análisis. Según el INEC, este diseño es probabilístico, de conglomerados, estratificado y bietápico. Se denomina probabilístico debido a que se conoce la probabilidad de inclusión de cada elemento en la muestra de forma precisa; de conglomerados porque las unidades de muestreo son áreas geográficas o UPM; es estratificado, ya que para la distribución y selección de las viviendas se definieron estratos, y bietápico porque en una primera etapa se selecciona las UPM, y en la segunda etapa se selecciona viviendas dentro de las UPM (INEC, 2019).

3.10 Software utilizado

Para el desarrollo de esta investigación, se usó el paquete estadístico SPSS, versión 22, particularmente para la construcción de los indicadores. Se utilizó la plataforma del lenguaje de programación de R, versión 3.6 con el fin de determinar el número de estratos, construcción de los modelos, desarrollo de validaciones y escritura de códigos con el propósito de implementar el estudio de simulación.

IV. Resultados

En este apartado, se presentan los principales resultados de las estrategias metodológicas desarrolladas en esta investigación.

4.1 Descriptivos variables generales

Las distribuciones de las variables individuales requeridas para el análisis se muestran en el Cuadro 3. Las pertenencias del hogar presentan un amplio rango de variación. La pantalla LED o LCD es el artefacto más común en el hogar (59,7%), seguido por el carro (40,4%) y la computadora portátil (37%). El tanque para agua caliente es el menos común (5,6%). Sin excepción, todos los porcentajes son más altos en los hogares urbanos que en los rurales, principalmente el teléfono, computadora portátil, carro e internet.

Cuadro 3. Distribución porcentual de las características del hogar y la vivienda según zona de residencia

Variables	Urbano	Rural	Total
Casos en la muestra (n)	6269	2752	9021
Patrimonio del hogar			
Internet	37,1	11,3	29,3
Teléfono residencial	46,4	28,0	40,8
Computadora portátil	42,8	23,7	37,0
Carro	43,6	33,1	40,4
Sistema de agua caliente	7,4	1,6	5,6
Pantallas LED o LCD	65,0	47,5	59,7
Características de la vivienda			
Cielo Raso	76,4	47,4	67,5
Estado de la vivienda			
Malo	8,8	11,4	9,6
Regular	27,3	43,8	32,3
Bueno	64,0	44,9	58,1
Tipo de aseguramiento			
Asalariado	49,3	35,3	45,0
Cuenta propia, voluntario o convenio	14,9	22,6	17,2
Régimen no contributivo (recibe pensión)	2,0	5,2	3,0
Pensionado(a) de CCSS, Magisterio, Hacienda u otro	9,9	5,7	8,1
Asegurado(a) familiar	8,1	7,2	7,8
Asegurado(a) por el Estado	2,1	6,4	3,4
Otras formas	0,5	0,4	0,5
No tiene seguro social de la CCSS	13,3	17,1	14,4
Ignorados	0,1	0,1	0,1
Promedio número de ocupados del hogar	1,5	1,3	1,4
Número de miembros del hogar	3,3	3,3	3,3

1/ Solamente se encontraron cinco valores faltantes en la respuesta de tipo de seguro.

Fuente: Elaboración propia, a partir de INEC-Archivo de datos de Encuesta Nacional de Hogares (2015) (80 % de la muestra)

Es interesante el comportamiento de la variable de cielo raso. Aunque podría parecer que esta característica es común en las viviendas costarricenses, únicamente 67,5% cuentan con este y son las viviendas de la zona urbana las que la presentan con mayor frecuencia. Cuando se evalúa el estado de la vivienda, el comportamiento es similar y, de igual manera, existe una mayor concentración de viviendas en buen estado en la zona urbana.

El tipo de aseguramiento de los jefes del hogar en su mayoría se da por la condición de asalariados (45,0%), seguido por el tipo de seguro cuenta propia, voluntario o convenio (17,2%). Otra categoría con un peso importante son los jefes que no poseen seguro, este grupo representa 14,4% del total de jefes de los hogares costarricenses.

El promedio de número de ocupados es 1,4, cifra que es muy similar entre la zona urbana y rural (1,5 y 1,3, respectivamente). Por último, el promedio de miembros del hogar en cada zona es similar, el promedio nacional es de aproximadamente 3,3 miembros en el hogar.

Por otro lado, en el Cuadro 4 se muestra la información para el jefe del hogar y se observa un mayor porcentaje de jefes con ocupaciones elementales en la zona rural, debido a que 40,3% de la población rural se concentra en esta categoría. Mientras tanto, para la zona urbana, este porcentaje es 19,6%. Esta es una de las categorías que mayor peso tiene entre las ocupaciones.

Las categorías con una condición mejor en la ocupación del empleo principal se concentran, en su mayoría, en la zona urbana. Los directores y gerentes, los profesionales científicos e intelectuales, los técnicos, profesionales de nivel medio y el personal de apoyo administrativo presentan un mayor peso en la distribución total de jefes en la zona urbana si se comparan con las distribuciones de los jefes de la zona rural.

Además, al analizar la categoría ocupacional, 51,2% de los jefes son asalariados privados, se evidencia entonces que, en la zona urbana, hay una mayor concentración de asalariados públicos. Es interesante el comportamiento de los jefes que se encuentran en la categoría de cuenta propia en la zona rural, ya que representan la segunda agrupación más frecuente en esta zona (26,3% de la población rural). Por último, en promedio, los años de escolaridad de los jefes son 8,5 años, cifra que es 6,5 años en la zona rural y 9,3 en la urbana, evidenciando una diferencia importante entre las zonas.

Cuadro 4. Distribución porcentual de los jefes de hogar según algunas características de interés

Variables	Urbano	Rural	Total
Casos en la muestra (n)	5315	2266	7581
Total	100,0	100,0	100,0
Ocupación en el empleo principal ^{1/}			
Directores y gerentes	2,4	0,8	1,9
Profesionales científicos e intelectuales	13,6	4,0	10,7
Técnicos y profesionales de nivel medio	11,5	4,8	9,5
Personal de apoyo administrativo	7,5	2,5	6,0
Trabajadores de los servicios y vendedores de comercios y mercados	20,7	17,2	19,8
Agricultores y trabajadores calificados agropecuarios. forestales y pesqueros	2,1	13,1	5,4
Oficiales. operarios y artesanos de artes mecánicas y de otros oficios	12,9	10,5	12,2
Operadores de instalaciones y máquinas y ensambladores	9,6	6,9	8,7
Ocupaciones elementales	19,6	40,3	25,7
Categoría ocupacional ^{1/ 2/}			
Asalariado institución pública	16,8	10,1	14,8
Asalariado institución privada	51,0	51,6	51,2
Auxiliar no remunerado	0,3	0,5	0,4
Asalariado del hogar privado	6,5	6,5	6,5
Empleadora	5,2	5,0	5,2
Cuenta propia	20,1	26,3	22,0
Promedio años de escolaridad ^{3/}	9,3	6,5	8,5

1/ Incluye jefes de hogar ocupados

2/ Variable reconstruida con base en el sector institucional

3/ Incluye al total de jefes de la muestra

Fuente: Elaboración propia a partir de INEC-Archivo de datos de Encuesta Nacional de Hogares (2015) (80 % de la muestra que cumplen con la condición de ocupados)

4.2. Resultados de la estrategia A

La estrategia metodológica A consiste en generar un modelo de regresión que estime el ingreso *per cápita* neto del hogar a partir de siete variables *proxys* al nivel socioeconómico en el archivo de datos de la Encuesta Nacional de Hogares. Con esta técnica, se obtiene la ecuación de regresión que se utilizará para estimar la variable ingreso en el archivo del Censo 2011. Una vez constituida, se agrega por UPM y se procede a estratificarla mediante el método estadístico de K medias. Por último, se genera una validación interna y externa de la conformación de estos grupos.

4.2.1 Análisis descriptivos Estrategia A

En este subapartado, se presentan las estadísticas descriptivas de los indicadores contruidos para la aplicación de la estrategia metodología A, mismas que se incorporan a la regresión lineal múltiple para estimar el ingreso neto de los hogares costarricenses.

a) Índice de patrimonio acumulado del hogar

En el Cuadro 5, se muestra la distribución absoluta y relativa de la cantidad del patrimonio acumulado presente en los hogares costarricenses; se identifica un grupo importante de hogares que únicamente tienen dos o menos de los artefactos seleccionados para este indicador (63%). Además, solo el 2,3% de los hogares cuenta con la totalidad de artefactos seleccionados.

Cuadro 5. Distribución absoluta y relativa de la cantidad de pertenencias acumuladas por el hogar

Pertenencia	N	Muestra	
		no ponderada	Porcentaje
0	229 283	1 768	19,6
1	288 944	2 228	24,7
2	215 247	1 660	18,4
3	161 434	1 246	13,8
4	133 359	1 028	11,4
5	114 642	884	9,8
6	26 906	207	2,3
Total	1 169 815	9 021	100,0

Fuente: Elaboración propia a partir de INEC-Archivo de datos de Encuesta Nacional de Hogares (2015) (80 % de la muestra)

b) Índice del tipo de aseguramiento del jefe

Este índice se conforma por seis categorías. En el grupo cero, se encuentran los jefes que no dan una contribución económica al estado y representa el 20,8% de los jefes de hogar. En la categoría de agrupación uno, se encuentran los jefes que contribuyen de manera indirecta, es decir, es asegurado por algún familiar, quienes representan el 7,8%.

En la agrupación dos, se encuentran aquellos jefes que tienen un tipo de seguro voluntario, cuenta propia o por convenio. Este grupo equivale al 17,2% de los jefes y representa el tercer grupo con mayor frecuencia a nivel país. El grupo tres y cuatro lo integran el 8,1% y 45% respectivamente. Por último, la categoría cinco está integrada únicamente por 0,6 % de los jefes de hogares costarricenses (Ver Anexo 3).

Cuadro 6. Distribución porcentual de la agrupación de la variable tipo de seguro

Tipo de seguro	Agrupación	Porcentaje
Asegurado(a) por el estado		
No tiene seguro	0	20,8
Régimen no contributivo (reciben pensión)		
Asegurado(a) familiar	1	7,8
Cuenta propia. voluntario o convenio	2	17,2
Pensionado de la CCSS. Magisterio de Hacienda u otro	3	8,1
Asalariado	4	45,0
Otras formas (privado)	5	0,6

c) Otros índices utilizados en estrategia A

En el Cuadro 7, se presenta un resumen de cada una de las variables que se incluirán en el modelo de regresión, así como del logaritmo del ingreso, cuyo promedio es de 12,2, es decir, casi doscientos mil colones por persona.

El índice de ocupación varía en un rango de 0 a 12. Cuando el valor obtenido por un jefe de hogar es cercano a cero, se asocia con una condición económica baja. Por el contrario, si el valor obtenido en el índice por el jefe de hogar es 12, este tendrá una condición económica alta. El promedio aproximado de este índice es de 3,2 a nivel nacional, lo que indica una mayor concentración de jefes en ocupaciones de las primeras cuatro categorías, relacionados con niveles económicos bajos.

Por otro lado, el índice del tipo de aseguramiento del jefe de hogar presenta un comportamiento interesante, pues en promedio se obtiene un valor de 2,5, lo que establece que los jefes se concentran en el nivel medio. También es de interés el comportamiento del índice sobre las características de la vivienda, ya que registra un promedio alto, de 2,2 en relación con su valor máximo de 3.

Finalmente, los años de escolaridad del jefe varían de 0 a 20 años y el promedio es alrededor de 8 años, lo que equivale a segundo año de secundaria. Un dato relevante es la identificación de 1,4 persona ocupada por hogar. Además, la cantidad promedio de miembros por hogar es de 3,3 personas. Por último, el índice de patrimonio acumulado del hogar varía entre 0 y 6, y el promedio obtenido es de 2,1 pertenencias.

Cuadro 7. Estadísticos descriptivos del Estrategia A

Variabes	Mínimo	Máximo	Media	Desviación Estándar
Logaritmo natural ingreso <i>per cápita</i> del hogar	8,8	14,7	12,2	0,7
Índice de ocupación del jefe	1	12	3,2	1,7
Índice de patrimonio acumulado del hogar	0	6	2,1	1,7
Índice del tipo de aseguramiento del jefe	0	5	2,5	1,6
Años de escolaridad del jefe	0	20	8,4	4,5
Cantidad de miembros ocupados en el hogar	0	7	1,4	1
Índice de las características de la vivienda	0	3	2,2	0,9
Número de miembros del hogar	1	16	3,3	1,6

Fuente: Elaboración propia a partir de INEC-Archivo de datos de Encuesta Nacional de Hogares (2015) (80 % de la muestra).

d) Comparación indicadores Censo y ENAHO

Se realizó un análisis detallado entre la distribución de los índices utilizados en la estrategia metodológica A, tanto para el censo como para la ENAHO (Ver cuadro 8). Se puede observar un comportamiento similar entre ambas distribuciones, lo que es importante para garantizar un apropiado funcionamiento de la regresión. Únicamente en el índice de la característica de la vivienda, se identifica una diferencia entre los diferentes archivos de datos.

Cuadro 7. Promedio de los indicadores según archivo del Censo y ENAHO 2015

Variabes	Censo	ENAHO
N	1236981	1169815
n	-	9021
Índice Ocupacional	3,3	3,2
Índice de patrimonio acumulado del hogar	1,9	2,1
Años de escolaridad	8,3	8,4
Índice tipo de aseguramiento	2,4	2,5
Cantidad de miembros ocupados	1,3	1,4
Índice de la característica de la vivienda	1,3	2,2
Número de miembros del hogar	3,5	3,3

Fuente: Elaboración propia, a partir de INEC-Archivo de datos de Encuesta Nacional de Hogares (2015) (80 % de la muestra) y Censo 2011 INEC.

4.2.2 Modelo de regresión aplicado en estrategia A

Como es común, la distribución del ingreso *per cápita* neto del hogar presenta un comportamiento log normal y para asegurar que se cumpla el supuesto de normalidad de los residuos en el modelo, se aplica el logaritmo natural.

En los gráficos 1 y 2, se observa que la distribución de la variable original del ingreso y el resultado de la transformación.

Gráfico 1. Histograma del ingreso *per cápita* neto del hogar

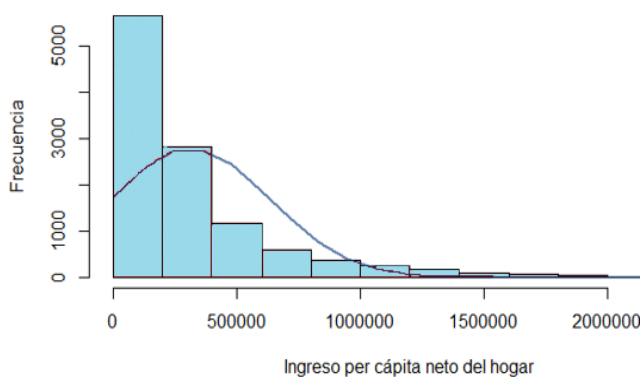
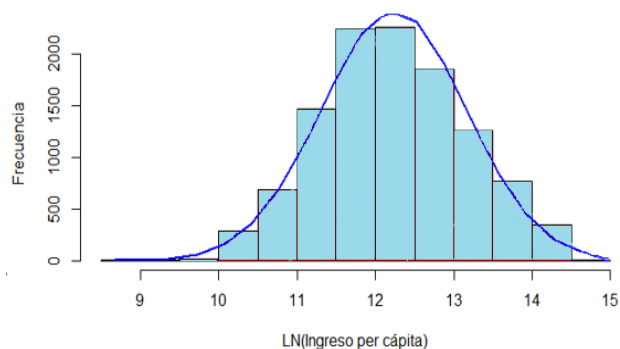


Gráfico 2. Histograma del logaritmo del ingreso *per cápita* neto del hogar



Para generar el modelo de regresión, se utilizan siete indicadores que asumen el papel de variables independientes. En el Cuadro 9, se presenta el coeficiente de correlación de

Pearson con el fin de determinar la relación que existe entre las variables usadas en el modelo de regresión.

Cuadro 8. Coeficientes de correlación de Pearson para las variables utilizadas en el modelo de regresión múltiple

VARIABLES	Ln(ipcn)	Índice de ocupación del jefe	Índice de patrimonio acumulado del hogar	Años de escolaridad del jefe	Índice del tipo de seguro del jefe	Cantidad de miembros ocupados en el hogar	Índice de las características de la vivienda	Número de miembros del hogar
Ln(ipcn) ¹	1,00	0,59	0,60	0,57	0,42	0,31	0,33	-0,24
Índice de ocupación del jefe		1,00	0,57	0,68	0,39	0,12	0,26	-0,01
Índice de patrimonio acumulado del hogar			1,00	0,57	0,31	0,24	0,37	0,07
Años de escolaridad del jefe				1,00	0,35	0,11	0,31	-0,03
Índice del tipo de seguro del jefe					1,00	0,16	0,17	0,06
Cantidad de miembros ocupados en el hogar						1,00	0,08	0,48
Índice de las características de la vivienda							1,00	-0,05
Número de miembros del hogar								1,00

¹Ip cn = Ingreso *per cápita* neto de los hogares

Fuente: Elaboración propia a partir de INEC-Archivo de datos de Encuesta Nacional de Hogares (2015) (80 % de la muestra).

La relación entre el logaritmo del ingreso y la mayoría de las variables independientes son altas y positivas. Las variables que tienen una mayor relación lineal con el logaritmo del ingreso son el índice de patrimonio acumulado ($\rho=0,60$), el índice de años de escolaridad ($\rho=0,57$) y el índice de ocupación ($\rho=0,59$).

La variable que determina el número de miembros del hogar presenta un valor de $\rho=-0,24$, señala entonces que posee una relación negativa con el logaritmo del ingreso. Por otro lado, existen relaciones entre variables independientes que tienen correlaciones altas, como el caso del índice de ocupación y el índice de escolaridad ($\rho=0,68$). Sobre este tema, se corroboró que los supuestos de multicolinealidad para la regresión se cumplen sin problema, ya que existen Factor de Inflación de Varianza (VIF) menores a 3 y una tolerancia con valores entre 1 y 0,1 (Ver Anexo 4).

Existen algunas condiciones necesarias para desarrollar los modelos de regresión lineal múltiple. En el Anexo 4, se muestra el detalle de todas las pruebas gráficas y formales que garantizan el cumplimiento de los supuestos del modelo.

Para determinar si se cumple con el supuesto de variabilidad constante, se aplicó la prueba de Breusch-Pagan. Inicialmente, este supuesto no se cumplía, por lo que fue necesario

realizar la corrección por mínimos cuadrados ponderados y de esta manera, el resultado se corrigió de forma positiva para continuar con el análisis. Por otro lado, para identificar si los valores de cada observación son independientes de los otros, se realiza la prueba de hipótesis de Durbin-Watson y se identificó que el estadístico es mayor a 1,5. Por consiguiente, se concluye que este supuesto se cumple (Ver Anexo 4).

En el Cuadro 10, se pueden observar los resultados obtenidos en el modelo de regresión múltiple. El coeficiente de determinación es de 0,6861, es decir, las variables independientes de este modelo explican el 68,61% de la variabilidad observada en el logaritmo del ingreso. La probabilidad asociada del modelo es significativa ($p < 0.01$), por lo que se puede aceptar que el modelo no es por azar y al menos uno de los coeficientes parciales de regresión es distinto de cero. Por otro lado, se aprecia que todos los coeficientes son estadísticamente diferentes de cero, de manera que todas las variables predictoras son importantes para explicar el logaritmo del ingreso.

Cuadro 9. Resultados de la regresión múltiple usando la estrategia metodológica A

Variables	Coeficientes no estandarizados		Beta estandarizados	t	Significancia	Intervalo de confianza de 95% para B	
	B	Error Estándar				Límite inferior	Límite superior
Constante	11,31	0,02		517,54	0,00	11,27	11,35
Índice de ocupación del jefe	0,11	0,00	0,21	23,60	0,00	0,10	0,12
Índice de patrimonio acumulado del hogar	0,14	0,00	0,26	31,00	0,00	0,13	0,15
Años de escolaridad del jefe	0,03	0,00	0,14	15,40	0,00	0,03	0,03
Índice del tipo de aseguramiento del jefe	0,09	0,00	0,16	24,00	0,00	0,09	0,10
Cantidad de miembros ocupados	0,37	0,01	0,39	53,19	0,00	0,36	0,39
Índice de las características de la vivienda	0,06	0,01	0,05	7,83	0,00	0,04	0,07
Cantidad de miembros del hogar	-0,26	0,00	-0,45	-63,20	0,00	-0,27	-0,25

R^2 Múltiple = 0.6861 y R^2 Ajustado= 0.6859, p-value: < 2.2e-16

Fuente: Elaboración propia a partir de INEC-Archivo de datos de Encuesta Nacional de Hogares (2015) (80% de la muestra).

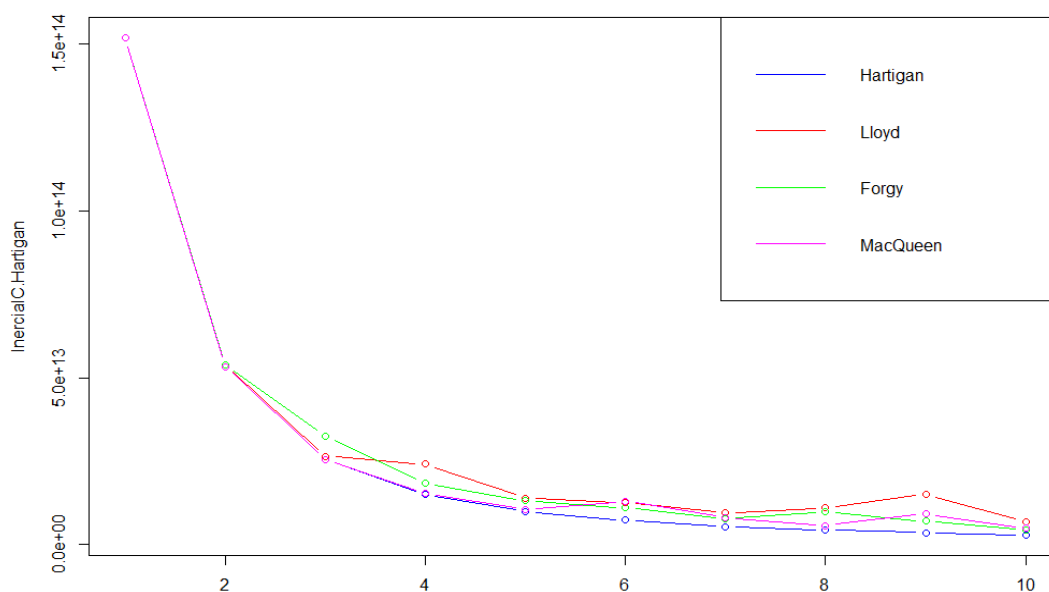
Otro aspecto importante de resaltar es la obtención de los parámetros betas estandarizados. En general, se puede decir que la variable cantidad de miembros ocupados, el índice de pertenencia del hogar y el índice de ocupación son las variables más importantes para explicar el ingreso *per cápita* neto del hogar.

4.2.3 Grupos K medias estrategia A

Una vez elaborada y verificada la ecuación del modelo de regresión para estimar el logaritmo del ingreso *per cápita* neto, se aplica a la base del censo 2011 y, con la estimación generada, se construye una base agregada por UPM, donde se obtiene el promedio de ingreso *per cápita* neto de cada UPM para iniciar con los procedimientos en los cuales se conformarán los grupos.

Para determinar el número de conglomerados por construir se utiliza el codo de Jambú, que muestra el comportamiento de la inercia intraclase según la cantidad de categorías definidas, la cual se espera que sea mínima. De esta manera, en el Gráfico 3, se puede observar cómo a partir de cinco grupos, los valores de la inercia intraclase son prácticamente iguales. Es decir, la inercia varía muy poco, lo que indica que es adecuado generar tres o más niveles de ingreso.

Gráfico 3. Codo de Jambú: inercia intra clase, según cantidad de conglomerados considerando el ingreso *per cápita* neto, estrategia A



Fuente: Elaboración propia con información del Censo 2011

En el Cuadro 11, se presentan los promedios de los grupos construidos. Se optó por seleccionar siete grupos debido a que el número de casos que cada uno de estos presentaba es importante. Además, el estadístico *eta* obtenido en esa clasificación es el más alto. Posteriormente, se decide agrupar los siete niveles en tres grupos finales,

denominados bajo, medio y alto. Se recodifican entonces los siete niveles originales de la siguiente manera: el bajo se forma con el grupo uno, el nivel medio es definido por la agrupación del grupo dos y tres (representando 52,5% del total de las UPM urbanas) y el nivel alto está compuesto por los grupos cuatro, cinco, seis y siete. En total, este grupo representa el 20,9% de las UPM de la zona urbana. El estadístico eta al agrupar estas categorías pasa de 0,98 a 0,88, lo que es esperable; sin embargo, sigue siendo adecuado para el análisis.

Cuadro 10. Estadísticos de los k grupos, aplicando estrategia metodológica A

Grupos	Ingreso	Desviación típica	Frecuencia	%	Eta ^{1/}	Eta cuadrado
Original						
1	161120	24976	1774	26,6		
2	238045	21839	1989	29,8		
3	316255	24360	1517	22,7		
4	410890	29686	818	12,2	0,98	0,96
5	524544	35069	348	5,2		
6	649156	38851	199	3,0		
7	789381	55300	33	0,5		
Recodificado						
Bajo	161120	24976	1774	26,6		
Medio	271885	45046	3506	52,5	0,99	0,88
Alto	482032	103579	1398	20,9		

1/ El eta se calcula entre la variable proxy del ingreso, es decir, el ingreso y los niveles contruidos.

Fuente: Elaboración propia con información del Censo 2011

4.3 Resultados estrategia metodológica B

Para desarrollar la estrategia B, se construyeron cuatro indicadores con respecto al UPM. En este apartado, se detallan los estadísticos descriptivos de estos indicadores y, además, se presenta el análisis de conglomerados realizado con esta estrategia.

4.3.1 Descriptivos estrategia metodológica B

El comportamiento del porcentaje de hogares con jefes sin seguro o asegurados por el estado por UPM presenta un promedio diferenciado por zona urbana y rural, pues los promedios son de 16,2% y 26,4% respectivamente, es decir, una diferencia de aproximadamente 10 puntos porcentuales. Cuando este valor se aproxima a 100%, indica que la UPM se relaciona con un nivel socioeconómico bajo.

Un comportamiento similar ocurre con la variable porcentaje de hogares en la UPM con al menos una necesidad básica insatisfecha y, de igual manera, se espera que, a mayor porcentaje obtenido, la UPM se relacione con un nivel socioeconómico bajo. Para ambas zonas, el rango varía entre 0% y 100 %, indica entonces que existen UPM donde la totalidad de las viviendas presenta al menos una necesidad básica insatisfecha y otras donde ningún hogar presenta esta característica. El promedio para la zona urbana es de 19,7%; mientras que para la zona rural es de 38,3%.

Cuadro 11. Estadísticos descriptivos usando Estrategia metodológica B

Zona ¹	VARIABLES	N ²	Mínimo	Máximo	Media	Desviación estándar
Urbano	% UPM de hogares sin seguro o asegurado por el estado	6678	0,0	75,8	16,2	8,7
	% de hogares en la UPM con al menos un NBI	6678	0,0	100,0	19,7	14,6
	% al menos un hogar con un universitario	6678	0,0	100,0	42,2	22,1
	TDA	6678	0,0	36,4	3,5	2,6
Rural	% UPM de hogares sin seguro o asegurado por el estado	3703	0,0	96,0	26,4	14,5
	% de hogares en la UPM con al menos un NBI	3703	0,0	100,0	38,3	18,6
	% al menos un hogar con un universitario	3703	0,0	100,0	18,8	12,1
	TDA	3703	0,0	38,4	3,2	3,3
Total	% UPM de hogares sin seguro o asegurado por el estado	10381	0,0	96,0	19,9	12,1
	% de hogares en la UPM con al menos un NBI	10381	0,0	100,0	26,4	18,4
	% al menos un hogar con un universitario	10381	0,0	100,0	33,8	22,2
	TDA	10381	0,0	38,4	3,4	2,9

1/ La *eta* se calcula entre la variable proxy del ingreso, es decir, el ingreso y los niveles construidos.

2/ La estratificación solo se realiza en la UPM de la zona urbana (6678).

Fuente: Elaboración propia con información del censo 2011.

El porcentaje de hogares que presenta al menos un miembro con estudios universitarios muestra que, si el valor obtenido es alto, la UPM se relaciona con un nivel socioeconómico alto. Esta relación tiene origen en las oportunidades que un alto nivel educativo alto proporciona a las personas. En la zona urbana, el promedio obtenido es 42,2%, diferencia importante se da con la zona rural, ya que el promedio es apenas 18,8%.

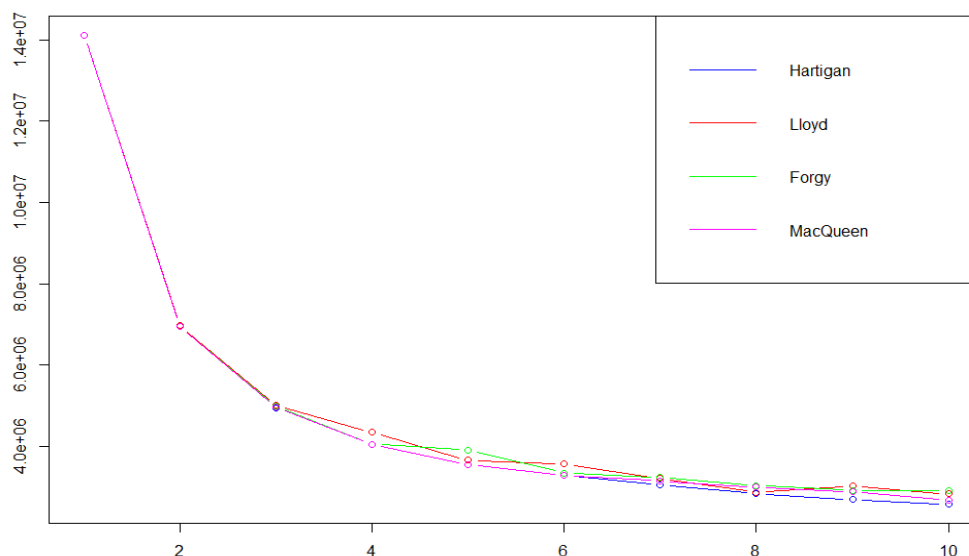
Por último, se calcula la tasa de desempleo abierto por UPM, variable que presenta un rango que oscila entre 0 y 38,4. El promedio para la zona urbana es de 3,5 y para la zona rural, de 3,2. La relación entre el nivel socioeconómico y este indicador es inversa, es decir, a mayor valor obtenido en el indicador, menor nivel socioeconómico presenta la UPM.

Cabe aclarar que cuando se construye una estratificación multivariada, el comportamiento de cada grupo debe realizarse con la comparación de los cuatro indicadores, por lo tanto, la interpretación debe darse de manera global.

4.3.2 Conformación de grupos utilizado la técnica de K medias en la estrategia metodológica B

Al analizar el codo de Jambú, que muestra el comportamiento de la inercia intraclase según la cantidad de categorías definidas, se observa una estabilidad en la inercia a partir de tres grupos (Gráfico 4). Por lo tanto, es adecuado generar tres o más niveles de ingreso.

Gráfico 4. Codo de Jambú: nercia intra clase, según cantidad de conglomerados considerando los 4 indicadores de la estrategia metodológica B



Fuente: Elaboración propia con información del Censo 2011

A partir del análisis gráfico de Jambú, se construyen siete estratos y se reagrupan en tres niveles: alto, medio y bajo. Debido a que se requiere grupos compuestos por una cantidad suficiente de observaciones (UPM), pues estos se utilizan para seleccionar muestras que permitan solventar las temáticas en el periodo intercensal; por tanto, estratos con pocos casos (pequeños) son inconvenientes puesto que generaría un rápido agotamiento de la muestra.

En el Cuadro 13, se muestra el comportamiento de la agrupación original y la recodificada al hacer este ajuste, el estadístico eta pasa de 0,89 a 0,84 y aunque hay un cambio de

magnitud, el valor obtenido con la agrupación sigue siendo aceptable. La distribución final revela 1706 UPM en el nivel bajo (25,5% del total de las UPM urbanas), el nivel medio contiene 3758 (56,3 total de UPM urbanas) y el nivel alto es integrado por UPM (18,2% de las UPM urbanas).

Los cuatro indicadores presentan el comportamiento esperado en cada uno de los grupos generados. En el nivel bajo, se observa un promedio mayor para el porcentaje de jefes del hogar sin seguro o asegurados por el estado. Además, conforme se aumenta de nivel socioeconómico, este promedio tiende a disminuir. Caso similar sucede con el porcentaje de hogares con NBI y la TDA, pues a mayor nivel existe una reducción en el promedio. Una situación inversa sucede con el indicador de la educación universitaria, ya que al aumentar de nivel socioeconómico se ve un incremento en el promedio.

Cuadro 12. Estadísticos de los k grupos aplicando estrategia metodológica B

Grupos	% sin seguro o asegurado por el estado	% NBI	% Al menos una miembro con educación Universitaria	TDA	Frecuencia	%	Eta ^{1/}	Eta cuadrado ^{1/}
Original								
1	39,1	68,5	10,2	6,3	128	1,9		
2	24,5	55,9	13,4	4,8	266	4,0		
3	24,2	30,8	18,5	4,4	1312	19,6		
4	16,8	19,2	33,0	3,5	1961	29,4	0,89	0,79
5	11,8	12,5	52,2	3,0	1797	26,9		
6	10,7	7,3	76,4	2,9	306	4,6		
7	8,7	5,9	77,6	2,3	908	13,6		
Recodificado								
Bajo	25,4	37,5	17,1	4,6	1706	25,5		
Medio	14,4	16,0	42,2	3,3	3758	56,3	0,84	0,71
Alto	9,2	6,3	77,3	2,4	1214	18,2		

^{1/} La eta se calcula entre la variable proxy del ingreso, es decir, el ingreso y los niveles construidos.

Nota: La estratificación solo se realiza en la UPM de la zona urbana (6678).

Fuente: Elaboración propia con información del censo 2011.

4.4 Resultados variante estrategia B

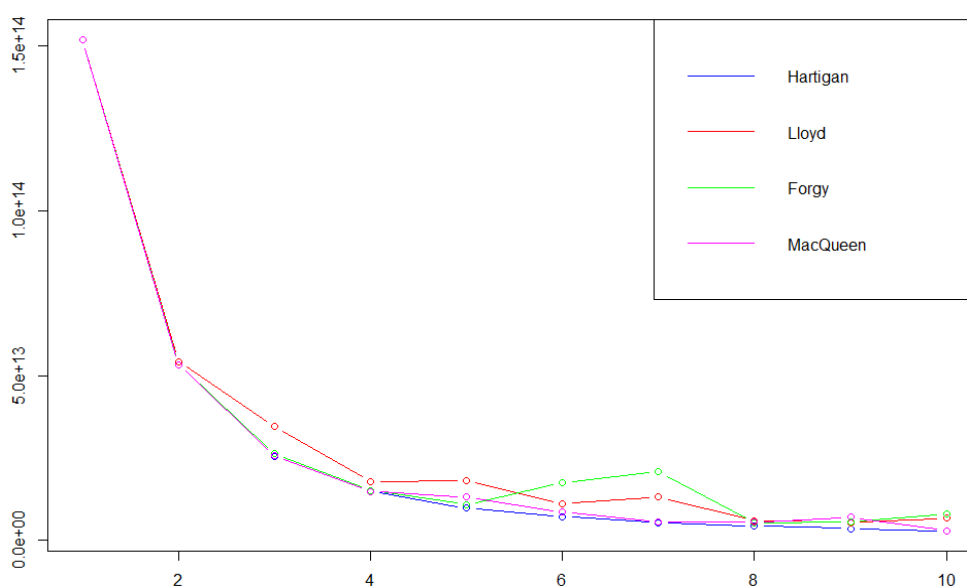
Esta variante se realiza con el objetivo de complementar los hallazgos obtenidos en las estrategias A y B, de manera que se utilizan los cuatro indicadores construidos en la estrategia B y el indicador obtenido con la regresión múltiple. En este apartado, se presentan la conformación de los grupos.

4.4.1 Grupos K medias para variante de estrategia metodológica B

De igual manera que en los casos anteriores, para la selección del número de conglomerados se parte de la interpretación del gráfico del codo de Jambú y se observa que a partir del grupo tres la inercia se estabiliza. Por lo tanto, se pueden seleccionar tres o más grupos.

Gráfico 5. Codo de Jambú: inercia intra clase, según cantidad de conglomerados considerando el logaritmo del ingreso *per cápita* de los hogares y los 4 indicadores de la estrategia metodológica

VB



Fuente: Elaboración propia con información del Censo 2011

Según el comportamiento del codo de Jambú, se opta por seleccionar siete grupos y posteriormente se recodifican en tres niveles. La distribución de las UPM por grupo es de 26,5%, 53,3% y 21,2% para los niveles bajo, medio y alto, respectivamente. Estas distribuciones son apropiadas para fines de muestreo. Para obtener este número final, se analizaron varios escenarios (Ver Anexo 7).

Los indicadores presentan el comportamiento deseable, pues el ingreso promedio tiende a ser mayor en el nivel alto, así como el porcentaje de hogares con al menos un universitario en el hogar. Por otro lado, los indicadores de TDA, porcentaje de hogares con NBI y porcentaje de jefes sin seguro o con seguro del estado, tienden a ser menor en el nivel alto.

Cuadro 13. Variante a la estrategia metodológica B incorporando el logaritmo Ingreso *per cápita* del hogar

Grupos	Ingreso	% jefes sin seguro o asegurado por el estado	% NBI	% Al menos una educación superior	TDA	Frecuencia	%	Eta ^{1/}	Eta cuadrado ^{1/}
Original									
1	161032	24,7	35,4	17,8	4,8	1770	26,5		
2	237607	16,6	19,8	34,5	3,4	1975	29,6		
3	315256	12,1	13,0	51,0	3,0	1518	22,7		
4	408016	9,5	8,7	67,0	2,6	809	12,1	0,98	0,96
5	518070	9,2	5,8	79,0	2,5	360	5,4		
6	641470	9,6	4,6	87,9	2,0	203	3,0		
7	773669	9,5	2,4	92,9	1,4	43	0,6		
Recodificado									
Bajo	161032	24,7	35,4	17,8	4,8	1770	26,5		
Medio	271352	14,7	16,8	41,6	3,2	3493	52,3	0,88	0,78
Alto	480620	9,5	7,2	73,9	2,4	1415	21,2		

1/ El *eta* se calcula entre la variable proxy del ingreso, es decir, el ingreso y los niveles construidos.

Fuente: Elaboración propia con información del Censo 2011

4.5 Validación de los resultados

Con el objetivo de probar la consistencia de las clasificaciones generadas con las diferentes estrategias metodológicas, se realiza una validación con indicadores básicos (Vivienda hacinada, agua por tubería dentro de la vivienda, servicio sanitario, vivienda en precario, promedio de aposentos de la vivienda y nacionalidad del jefe) presentes en el mismo archivo de datos (validación interna).

En el Cuadro 15, se muestra la información de cada uno de los indicadores, según la estrategia implementada. Para este análisis solo se utiliza la población o viviendas de la zona urbana, ya que es en esta donde se aplicó la estratificación.

Específicamente, se observa un comportamiento congruente entre las variables y los diferentes niveles de socioeconómicos obtenidos por estrategia. Del total de viviendas en condición de hacinamiento, casi el 55% se encuentran en el nivel bajo, mientras que una pequeña parte, alrededor del 5%, se concentra en el nivel alto. De igual manera, al analizar las viviendas que no cuentan con tuberías dentro de la vivienda, casi el 70 % de estas se identifican en el nivel más bajo.

Otra variable importante para validar la conformación de estos niveles socioeconómicos es la que indaga por la tenencia del servicio sanitario en el hogar, pues alrededor del 70% de las viviendas que no tienen se concentra en el nivel bajo. Por otro lado, las viviendas en

condición de precario se relacionan con los niveles socioeconómicos bajos y efectivamente, del total de viviendas con esta característica, casi el 90% pertenecen al nivel bajo. Este comportamiento se repite en todas las estrategias aplicadas. Con esta información, se puede comprobar con los estratos construidos con las diferentes estrategias cumplen con validez interna.

Cuadro 14. Porcentaje de viviendas que cumple con ciertas características nivel según estrategia aplicada

Variables	Nivel	Estrategia metodológica		
		A	B	VB
Total	Total	100	100	100
	Bajo	54,4	53,9	54,2
	Medio	40,7	42,9	40,6
Vivienda hacinada	Alto	5,0	3,2	5,2
	Bajo	67,3	71,0	67,3
	Medio	28,9	26,7	28,9
El agua no llega por tubería dentro de la vivienda	Alto	3,8	2,3	3,8
	Bajo	69,2	71,6	69,2
	Medio	28,5	27,3	28,5
No tiene servicio sanitario	Alto	2,3	1,1	2,3
	Bajo	89,7	91,2	89,7
	Medio	9,6	8,4	9,6
Vivienda en precario	Alto	0,8	0,4	0,8

Fuente: Elaboración propia con información del Censo 2011

Un dato interesante presente en el Cuadro 15 es que existe una gran similitud entre los resultados obtenidos con la estrategia A y la variante B, en términos de los porcentajes para cada indicador. Esto indica que el ingreso pesa más que las carencias en NBI para determinar las agrupaciones o estratos.

Adicionalmente, en el Cuadro 16, se presenta el porcentaje según la nacionalidad del jefe de hogar. Los jefes con nacionalidad nicaragüense en su mayoría se encuentran en el nivel bajo, mientras que los colombianos y estadounidenses se encuentran agrupados en el estrato alto.

Cuadro 15. Porcentaje de hogares con jefes de hogar por nacionalidad según estrategia metodológica

Estrategia metodológica	Niveles	Países				
		Costa Rica	Estados Unidos	Colombia	Nicaragua	Otro
	Total	100,0	100,0	100,0	100,0	100,0
A	Bajo	25,6	4,7	6,1	46,7	11,0
	Medio	55,2	35,9	42,3	44,4	39,1
	Alto	19,2	59,4	51,7	8,9	49,9
B	Bajo	24,3	7,8	7,2	47,2	11,5
	Medio	59,3	40,1	46,2	46,3	43,8
	Alto	16,4	52,1	46,6	6,5	44,7
VB	Bajo	25,5	4,6	6,0	46,6	11,0
	Medio	55,0	35,6	41,9	44,3	38,9
	Alto	19,5	59,8	52,0	9,0	50,2

Fuente: Elaboración propia con información del Censo 2011

Además, se calculó el promedio de aposentos y se identificó casi seis aposentos en la vivienda para los niveles promedio altos, mientras que en los niveles bajos es de 4,3 y el nivel medio es prácticamente cinco. Este comportamiento también es congruente con el resultado que se espera.

Cuadro 16. Promedio de las variables de interés según estrategia metodológica aplicada

Variable	Agrupación	Estrategia A	Estrategia B	Estrategia VB
Promedio de aposentos en la vivienda	Bajo	4,3	4,3	4,3
	Medio	4,9	5,0	4,9
	Alto	5,7	5,8	5,7

Fuente: Elaboración propia con información del Censo 2011

4.6 Resultado del estudio de simulación

La validez externa de las estratificaciones construidas con las diferentes metodologías se realiza mediante una simulación que consiste en aplicar las estrategias de estratificación al MMV-2011, seleccionar muestras con el mismo diseño de la ENAHO 2015 y determinar cuál estratificación funciona mejor en términos de precisión.

Para comparar las estrategias, se seleccionan 1005 muestras simuladas a partir del marco muestral de viviendas 2011 por medio del diseño de muestra de la ENAHO 2015, es decir, un diseño bietápico de conglomerados y estratificado. En este caso, los conglomerados están definidos por las UPM y los estratos se construyeron de la siguiente manera:

Cuadro 17. Distribución del número de UPM por estrato según estrategia y selección de muestra para la primera etapa

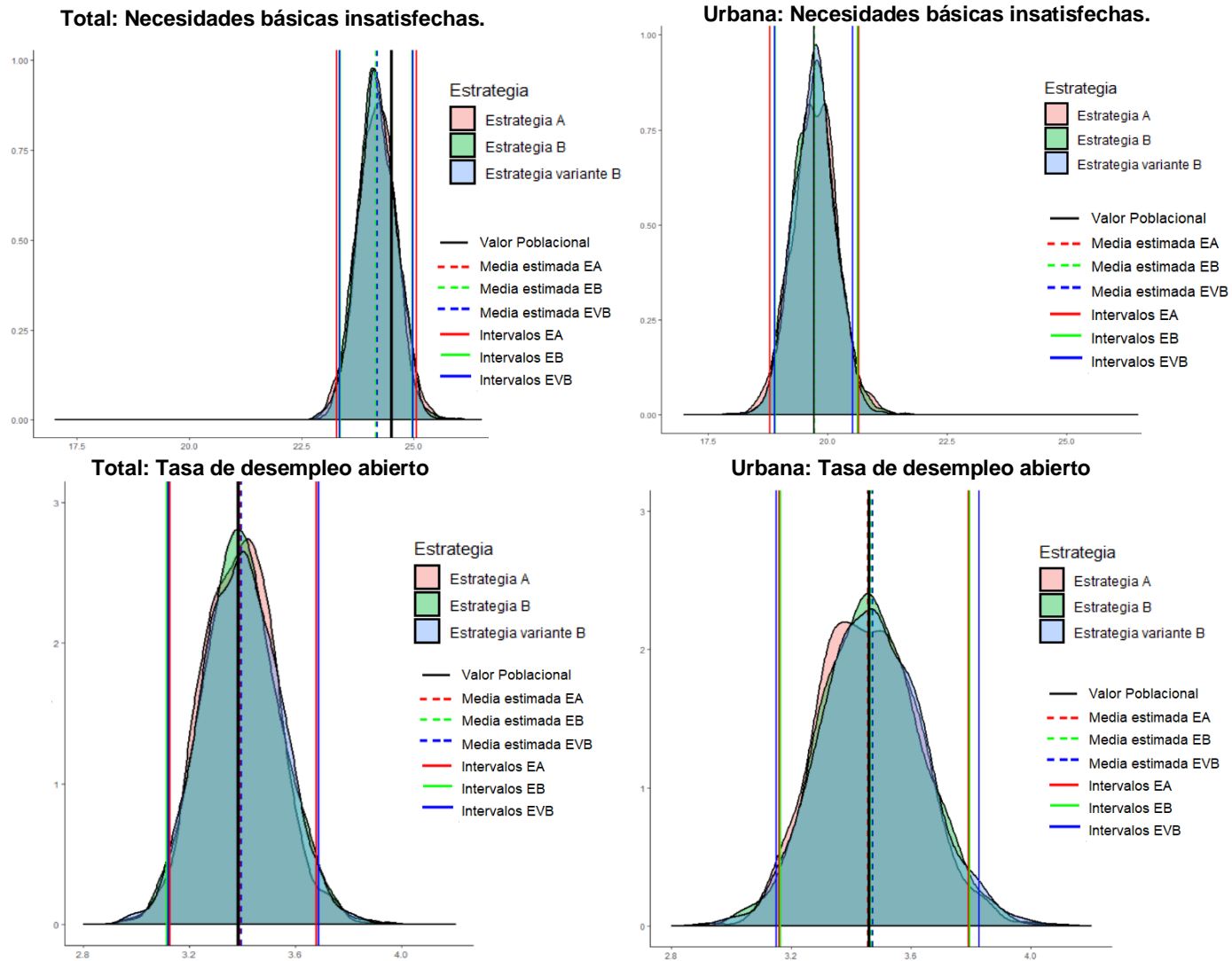
Grupos	UPM estrategia A	UPM estrategia B	UPM variante estrategia B	UPM a seleccionar según diseño
Central bajo	886	840	883	159
Central Medio	2599	2829	2587	313
Central Alto	1317	1133	1332	96
Central Rural	1109	1109	1109	96
Chorotega Urbano	445	445	445	52
Chorotega Rural	515	515	515	44
Pacífico Central Urbano	400	400	400	48
Pacífico Central Rural	320	320	320	28
Brunca Urbano	304	304	304	40
Brunca Rural	634	634	634	56
Huetar Atlántico Urbano	485	485	485	56
Huetar Atlántico Rural	506	506	506	48
Huetar Norte Urbano	242	242	242	28
Huetar Norte Rural	619	619	619	56

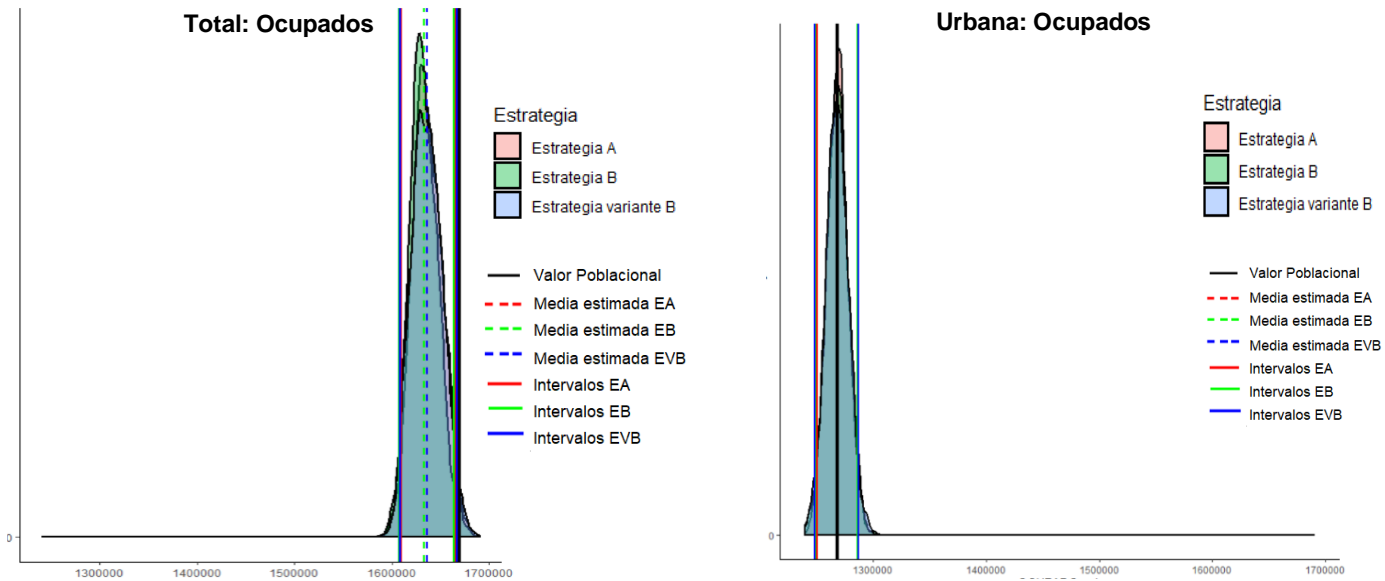
Fuente: Elaboración propia con información del Censo 2011

De cada una de las 1005 muestras, se estiman los principales indicadores de las encuestas a hogares como es el caso de NBI, TDA, porcentaje de computadoras, internet y carros, así como el total de personas, número de ocupados, entre otros; por tanto, se obtiene una distribución con 1005 observaciones por indicador y por estrategia. Además, de cada una de las muestras se extrae el coeficiente de variación y efecto del diseño.

A continuación, se presentan las distribuciones de los indicadores estimados, para cada una de las estrategias, sus respectivos intervalos de confianza y el valor poblacional, proveniente de la base del Censo. Con el propósito de identificar el efecto de las diferentes estrategias, el análisis se concentrará en la zona urbana, ya que es en esta zona donde se aplica la estratificación.

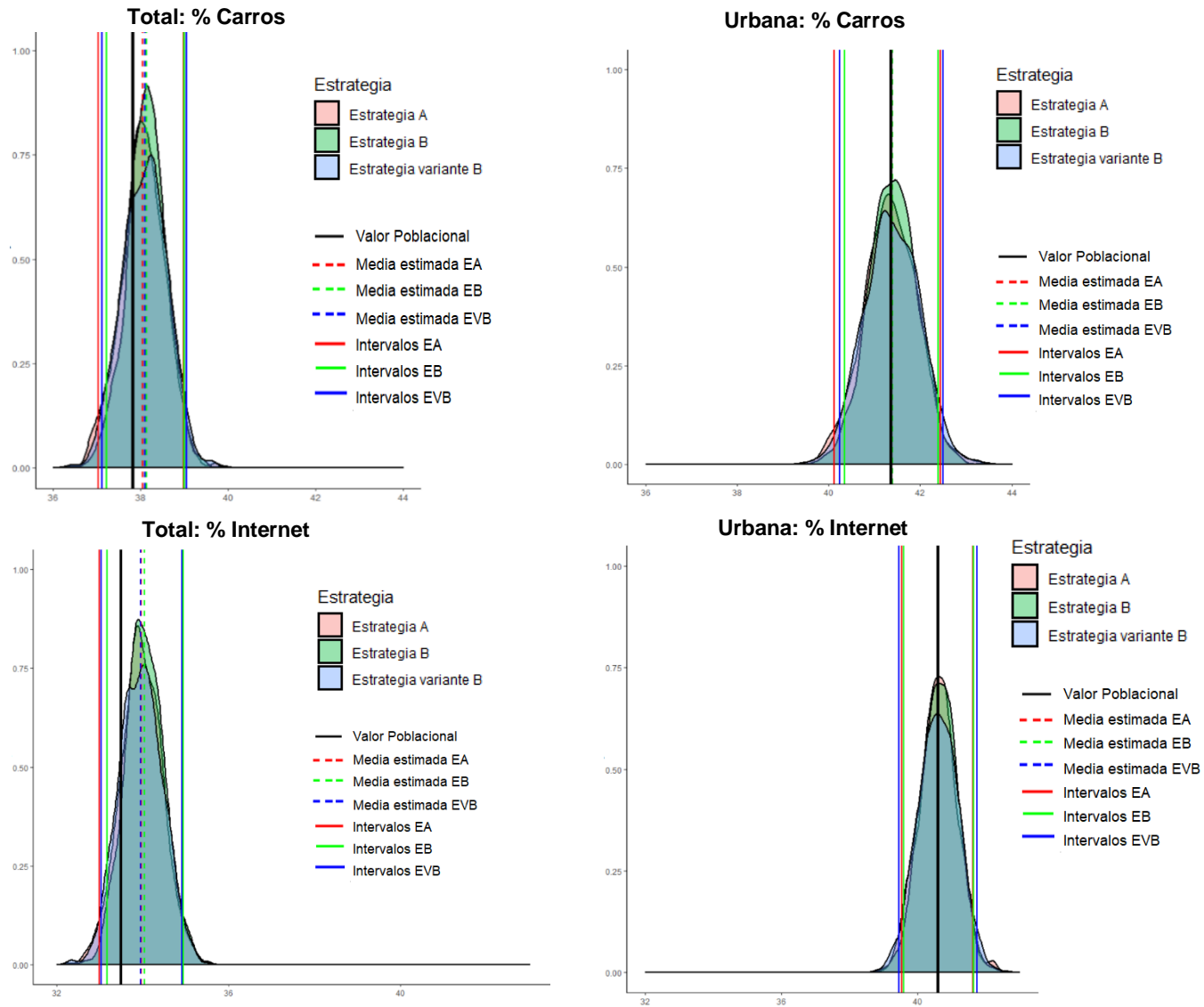
Gráfico 6. Distribución de densidad para las variables NBI, TDA y número de ocupados para el total país y zona urbana

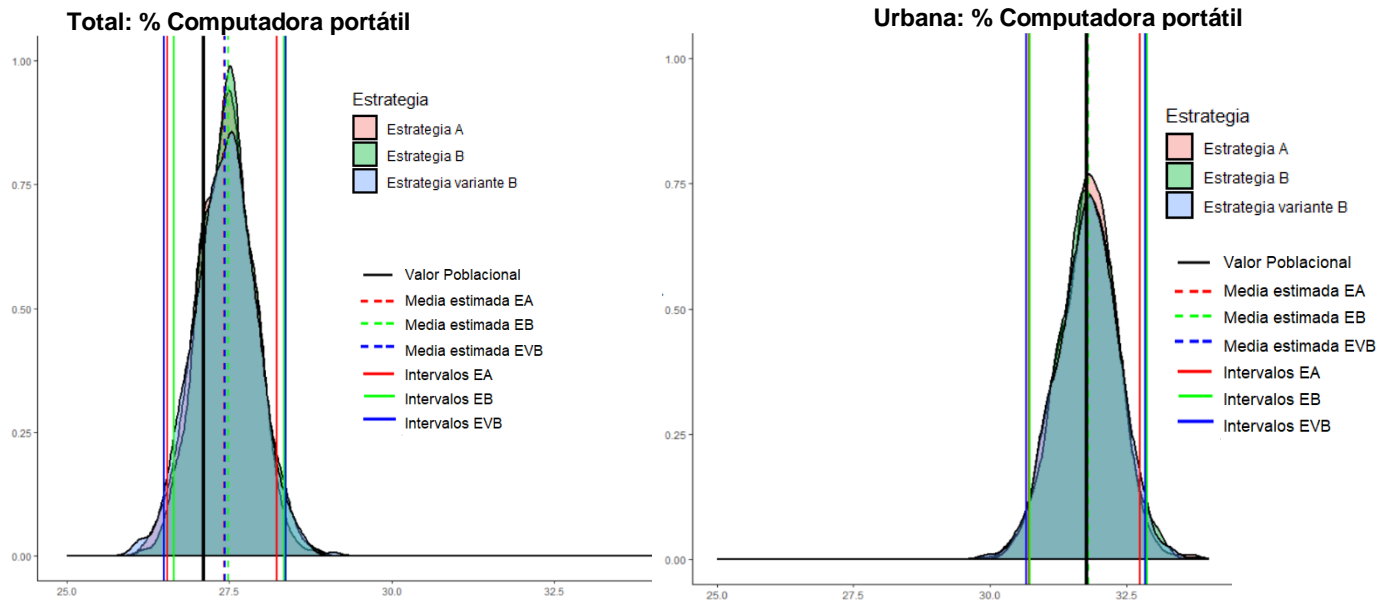




Las estimaciones utilizadas en la comparación de las tres estrategias generan un resultado positivo en términos de precisión, debido a que gráficamente se nota un valor similar entre el valor promedio estimado de las diferentes características y el valor poblacional proveniente del Censo 2011. Se puede concluir que, a partir de estas variables, el uso de cualquiera de las metodologías desarrolladas en esta investigación será adecuada para estratificar el marco.

Gráfico 7. Distribución de densidad para las variables porcentaje de carros, internet y computadora portátil por zona





Al comparar las características estimadas para cada una de las estrategias, las diferencias son mínimas (Cuadro 19). Esto se debe a los tamaños de muestra utilizados (de 13440 viviendas) y aunque se detectan diferencias leves entre las estimaciones, estas diferencias no son estadísticamente significativas. Debido a esto, es difícil concluir a partir de términos estadísticos que alguna de las estrategias aquí planteadas es mejor que otra, aspecto que lo muestra tanto la validez interna como la presente simulación. Sin embargo, existen otros factores que podrán determinar cuál es la más conveniente a partir de un balance de los aspectos favorables y desfavorables de cada estrategia, comparación que se presentará más adelante.

Cuadro 18. Resumen de los resultados obtenidos en la simulación de variables según estrategia metodológica aplicada

Variables	Estrategia A ^{1/}					Estrategia B ^{1/}					Estrategia variante B ^{1/}				
	EST	DE	P 2.5	P 97.5	CV	EST	DV	P 2.5	P 97.5	CV	EST	DV	P 2.5	P 97.5	CV
TDA	3,39	0,14	3,13	3,68	4,07	3,40	0,14	3,11	3,69	4,27	3,40	0,15	3,12	3,69	4,30
Urbano	3,46	0,16	3,16	3,79	4,72	3,47	0,17	3,16	3,80	4,81	3,47	0,17	3,15	3,83	4,90
Rural	3,16	0,29	2,63	3,78	9,19	3,15	0,29	2,57	3,74	9,06	3,13	0,29	2,58	3,74	9,39
NBI	24,19	0,44	23,28	25,06	1,83	24,16	0,42	23,36	24,97	1,72	24,18	0,42	23,33	24,98	1,75
Urbano	19,74	0,47	18,79	20,67	2,40	19,74	0,44	18,89	20,63	2,24	19,72	0,43	18,90	20,3	2,17
Rural	37,15	0,97	35,30	39,14	2,62	37,14	0,94	35,24	38,95	2,54	37,18	0,95	35,29	39,09	2,54
% Carro	38,05	0,49	37,02	38,99	1,30	38,13	0,45	37,20	38,99	1,18	38,09	0,51	37,11	39,05	1,34
Urbano	41,34	0,58	40,12	42,46	1,41	41,40	0,53	40,33	42,39	1,28	41,38	0,60	40,23	42,53	1,44
Rural	28,48	0,86	26,90	30,21	3,02	28,55	0,83	26,94	30,20	2,91	28,52	0,83	26,88	30,08	2,90
% Internet	33,96	0,48	32,99	34,94	1,41	34,04	0,45	33,17	34,95	1,31	33,96	0,50	33,03	34,93	1,46
Urbano	40,60	0,55	39,53	41,64	1,35	40,63	0,53	39,60	41,65	1,30	40,60	0,58	39,45	41,76	1,43
Rural	14,67	0,72	13,30	16,05	4,92	14,68	0,70	13,32	16,03	4,75	14,60	0,68	13,27	15,99	4,68
% Portátil	27,43	0,44	26,53	28,23	1,59	27,48	0,43	26,64	28,35	1,57	27,43	0,48	26,50	28,37	1,73
Urbano	31,78	0,54	30,70	32,76	1,69	31,80	0,53	30,72	32,89	1,68	31,79	0,57	30,65	32,87	1,78
Rural	14,77	0,65	13,52	16,07	4,39	14,79	0,64	13,47	16,04	4,35	14,74	0,65	13,47	16,06	4,41
Total Personas	4191682	38882	4125015	4264711	0,93	4181502	34715	4126614	4258389	0,83	4190978	36491	4132207	4262333	0,87
Urbano	3107320	17300	3071797	3139199	0,56	3107311	17166	3073248	3140408	0,55	3107189	15665	3074581	3137388	0,50
Rural	1084362	34136	1036097	1146271	3,15	1074192	30972	1033070	1138394	2,88	1083789	33914	1035516	1144937	3,13
Hombres	2046044	21033	2009449	2085471	1,03	2040832	18728	2008726	2083887	0,92	2045400	19793	2012049	2085041	0,97
Urbano	1494853	10918	1474102	1517431	0,73	1495236	11031	1473223	1517159	0,74	1494946	10010	1474373	1514088	0,67
Rural	551191	17575	524687	583922	3,19	545596	16009	523364	579400	2,93	550455	17552	525224	584293	3,19
Mujeres	2145638	20718	2108662	2186161	0,97	2140670	19309	2108632	2181022	0,90	2145578	19828	2111517	2184581	0,92
Urbano	1612467	10923	1590237	1633768	0,68	1612075	11177	1590869	1633656	0,69	1612243	10842	1590910	1633231	0,67
Rural	533171	17382	508192	565008	3,26	528596	15836	505833	562071	3,00	533335	17193	507669	564828	3,22
Ocupados	1636113	14821	1608713	1665337	0,91	1632694	14256	1607146	1663958	0,87	1635733	15325	1608509	1666813	0,94
Urbano	1268090	9090	1250118	1286745	0,72	1268075	9572	1248994	1286617	0,75	1267966	9686	1247793	1287233	0,76
Rural	368023	11833	350298	391103	3,22	364619	11047	348677	388684	3,03	367767	11861	349856	389967	3,23

Nota:

1/ EST (Estimación), DE (Desviación estándar de la media), P (Percentil) y CV (Coeficiente de variación de la media)

2/ n= 1005 muestras independiente para cada una de las estrategias de 13440 viviendas cada una. Replica diseño ENAHO.

También se buscó conocer la magnitud del coeficiente de variación obtenido en cada una de las estimaciones por estrategia. Se pudo notar un comportamiento similar al anterior, pues las diferencias no son estadísticamente significativas, por lo que esa información tampoco resulta fundamental para concluir sobre cuál de estas estrategias es mejor. Además, se nota una diferencia importante entre la zona urbana y la rural (Ver Anexo 10). Este efecto puede darse porque en la zona rural no se realiza ninguna estratificación socioeconómica, mientras que en la zona urbana sí.

Gráfico 8. Comparación variabilidad del Coeficiente de variación según estrategia

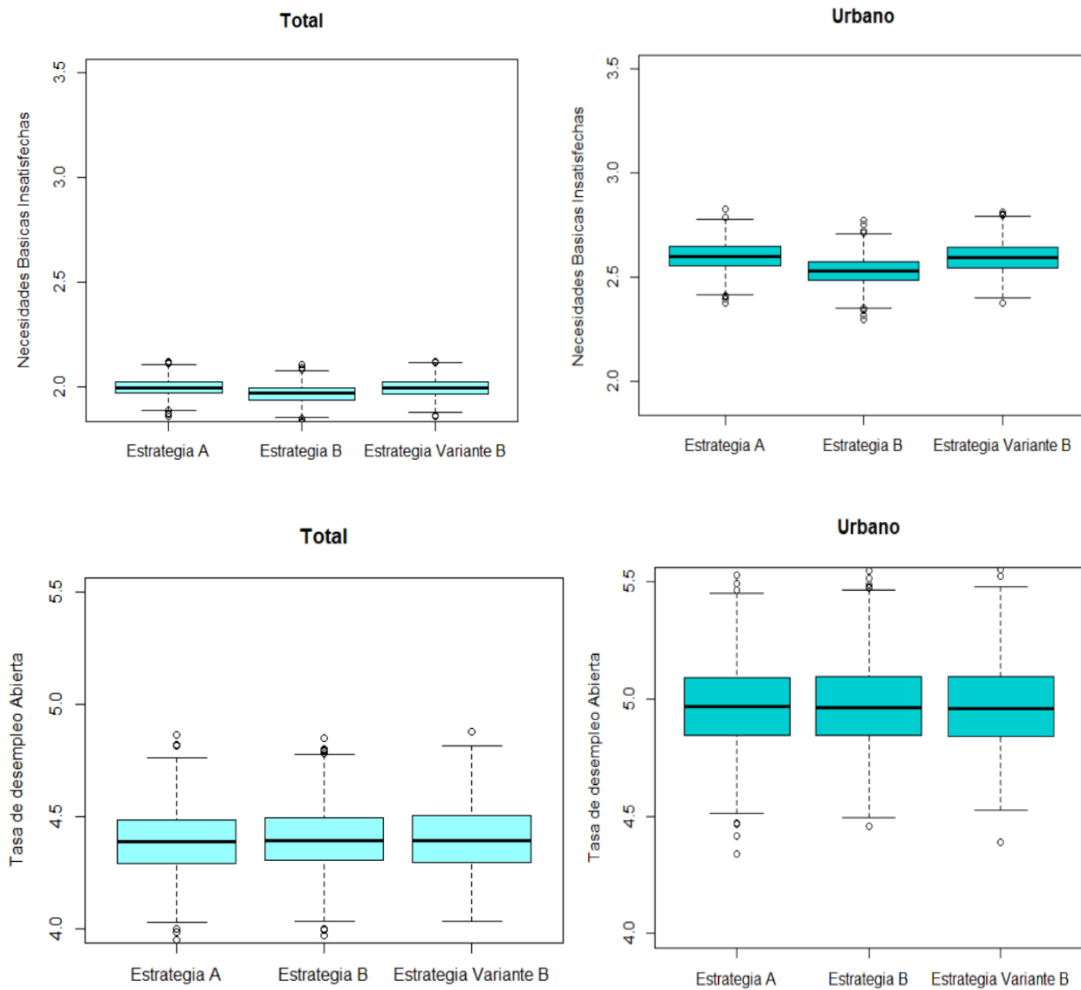
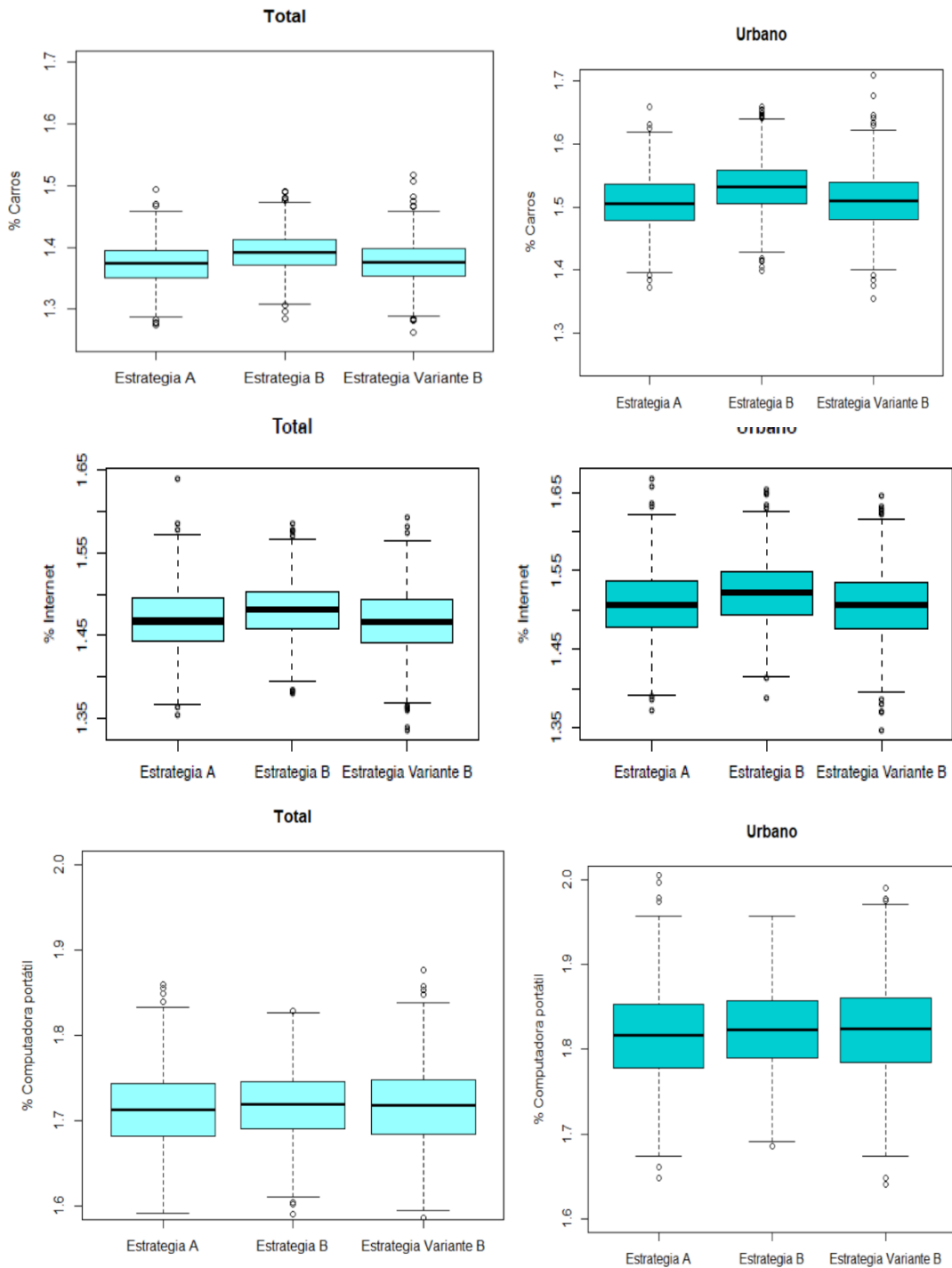
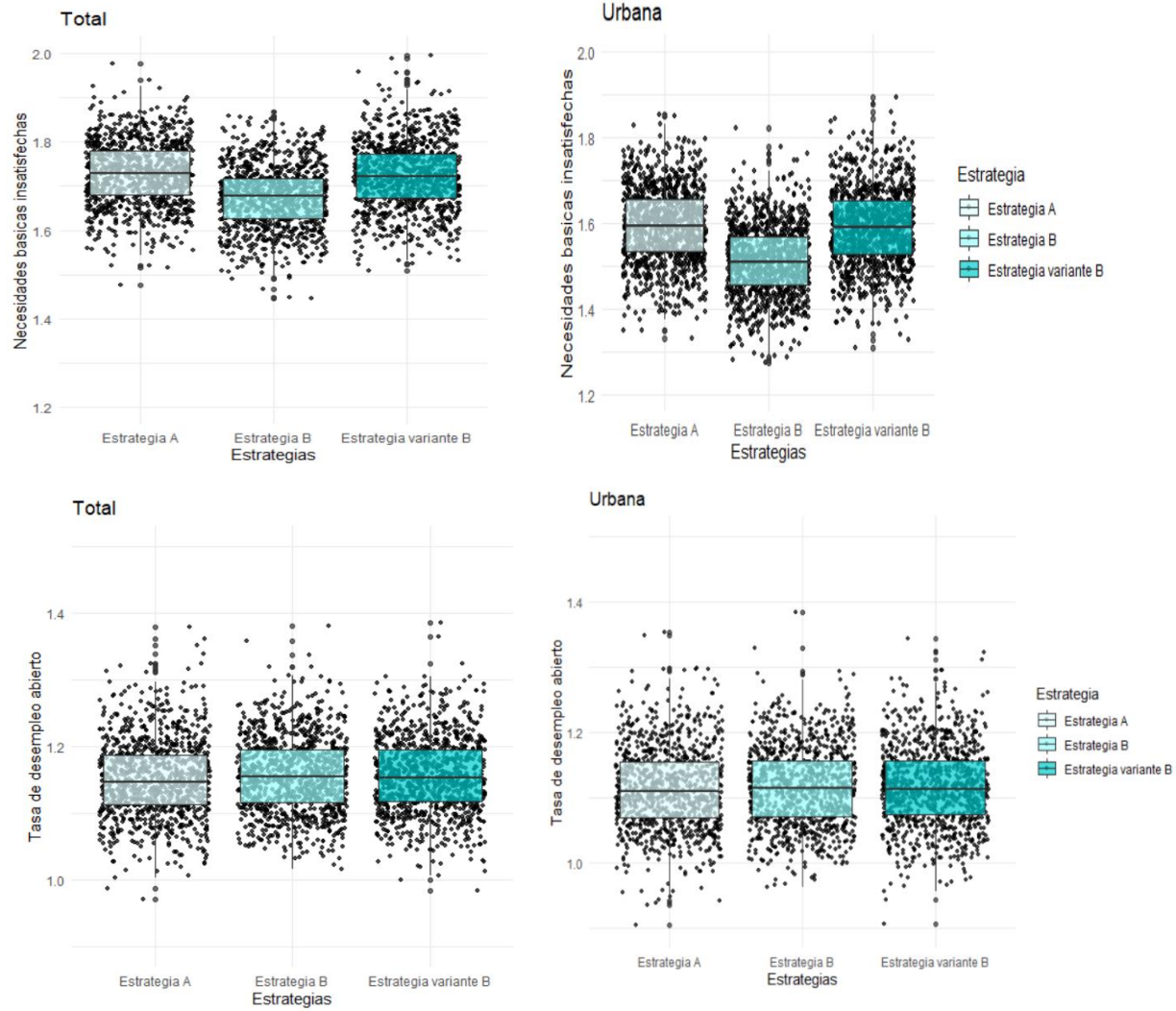


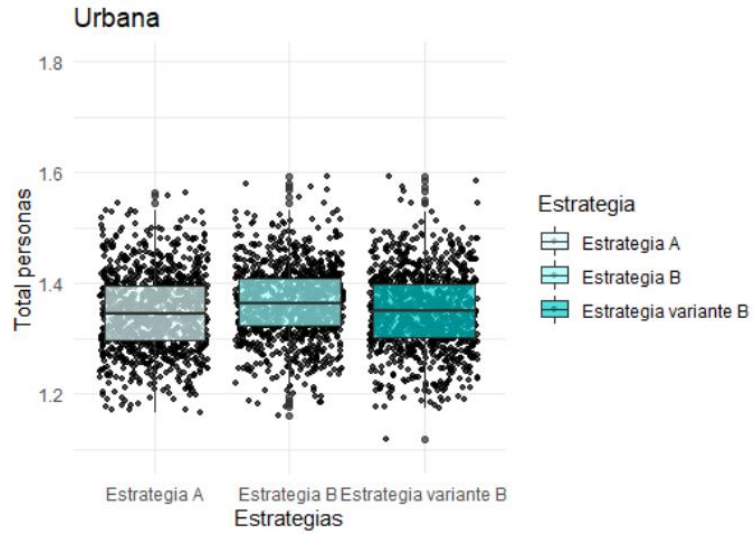
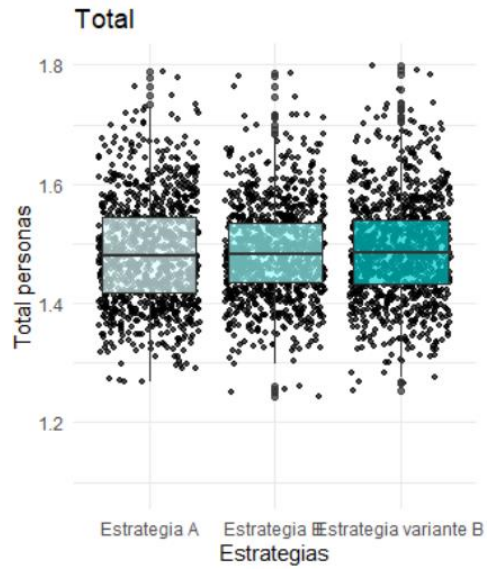
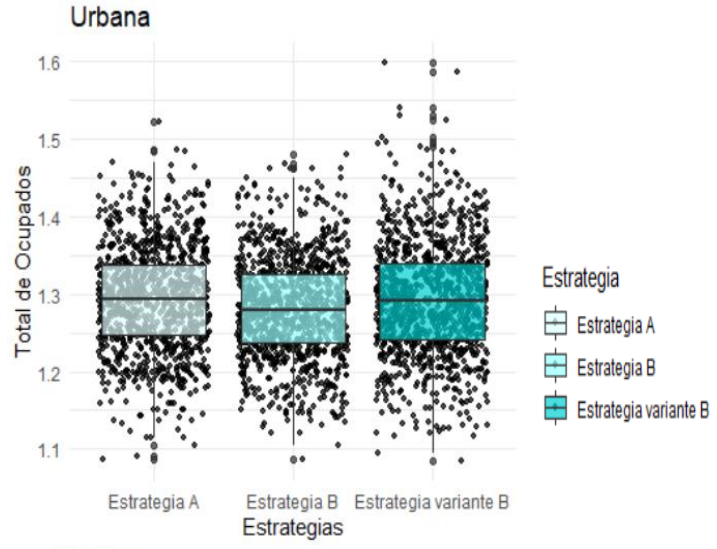
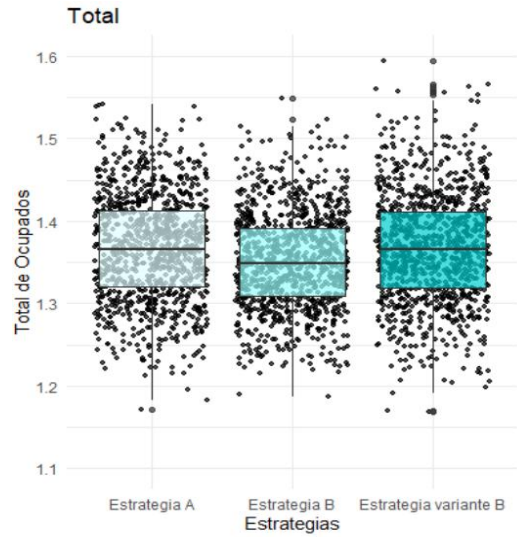
Gráfico 9. Comparación variabilidad del coeficiente de variación para porcentajes de pertenencias según estrategia metodológica



Seguidamente, al estudiar el comportamiento del efecto del diseño para cada estimación, en el caso de la TDA, el NBI, total de ocupados y número de personas, se puede decir que los patrones no son concluyentes.

Gráfico 100. Efecto del diseño de las principales variables de comparación según estrategia metodológica aplicada





Para finalizar con el análisis, se hace una tabla comparativa entre las diferentes estrategias con los aspectos favorables y desfavorables de cada una, con el objetivo de encontrar otro criterio (no estadístico).

Tabla 13. Comparación de los aspectos favorables y desfavorables de las estrategias metodológicas

Estrategias	Aspectos favorables	Aspectos desfavorables
Estrategia A	Obtención de un valor estimado del ingreso en la base del para estratificar. Presenta un buen resultado evidenciado en la validez interna y externa.	Procedimiento más elaborado, requiere mayor tiempo para su ejecución. Es necesario el cumplimiento de todos los supuestos del modelo de regresión para poder confiar en este criterio. Se requiere de una definición de jefes para cumplir con el indicador de ocupación, además de realizar criterios de imputación para los jefes no ocupados.
Estrategia B	No presenta un indicador de ingreso en el análisis. Presenta un buen resultado evidenciado en la validez interna y externa.	Es más sencilla en términos de procedimientos.
Estrategia variante B	Presenta un buen resultado evidenciado en la validez interna y externa.	Requiere mayor tiempo en ejecución y análisis, ya que es un complemento de las dos estrategias anteriores y no existen diferencias en los resultados obtenidos a nivel de precisión.

Considerando los puntos anteriores, se concluye que las estrategias no presentan diferencias importantes; sin embargo, al analizar otros aspectos como el tiempo y procedimiento de la elaboración que requiere la construcción de los estratos, se concluye que la mejor estrategia es la B, dada la facilidad metodológica.

V. Conclusiones

- El indicador estimado a partir del ingreso *per cápita* neto de los hogares y UPM del Censo 2011 es útil para estratificar por niveles de ingreso. Asimismo, el modelo de regresión obtenido es considerado aceptable para los propósitos del estudio, tomando en cuenta el cumplimiento de los supuestos, la significancia del modelo y el coeficiente de determinación.
- La nueva metodología de estratificación que propone el estudio (mediante la generación de cuatro indicadores) permite ampliar las posibilidades de estratificación, muestra buenos resultados a nivel de precisión y sobre todo es una metodología sencilla de implementar; por lo cual, se considera como un hallazgo relevante para el estudio.
- No se observan diferencias entre las estimaciones en las distintas estrategias metodológicas y se da un comportamiento apropiado en la validez interna y externa. Esto posibilita al INEC la ejecución en el nuevo marco proveniente del Censo 2021 y permite considerar cualquiera de estas metodologías; sin embargo, por practicidad es mejor implementar la estrategia B.
- Optar por la estrategia B es más práctico, pero evita que el censo posea una estimación del ingreso del hogar. El censo no pregunta por ingresos de los miembros del hogar por razones ya conocidas; aun así, existen algunos análisis que podrían desarrollarse a partir de una estimación del mismo. Poseer una estimación del ingreso posibilita otros análisis que podrían ser de interés, lo que no necesariamente se lograría con la estrategia B. La estimación del ingreso en el censo cumple varias funciones, entre ellas, estratificar el marco muestral.
- El desarrollo de una simulación como alternativa a un método de validez externa reduce tiempo y costos, si se comparara con una alternativa en campo. Con la simulación planteada, no se logran identificar diferencias entre las estrategias; sin embargo, se puede identificar un comportamiento adecuado de los grupos o niveles socioeconómicos construidos, ya que los promedios de las estimaciones son cercanos al valor poblacional utilizando las diferentes metodologías.

VI. Recomendaciones

- Para futuras investigaciones, es conveniente incorporar dentro del análisis la zona rural y generar una estratificación socioeconómica con el fin de mejorar las estimaciones en términos de variabilidad, ya que se observan mayores coeficientes de variación, así como rangos de intervalos de confianza mayores en las estimaciones de la zona rural, si lo comparamos con las obtenidas en la zona urbana.
- Es importante que el INEC haga una evaluación de las metodologías implementadas para la elaboración de estratificaciones de los marcos muestrales, puesto que tienen una repercusión importante en las estimaciones finales.
- Para el nuevo marco muestral que realiza el INEC a partir del Censo 2021, se recomienda utilizar alguna de estas metodologías con el objetivo de realizar la estratificación socioeconómica de las UPM. Por practicidad, es preferible que apliquen la metodología B.
- A las empresas de estudios de mercado, se les recomienda utilizar la estrategia B para sus análisis y estrategias de mercados, dado que es una metodología novedosa, fácil de implementar y, tal y como se identifica en esta investigación, genera resultados con altos niveles de precisión.
- Con el objetivo de evaluar los diseños muestrales de las encuestas a hogares que realiza el INEC, se recomienda utilizar la metodología de simulaciones como la que se implementa en esta investigación, porque representa una manera novedosa, diferente y muy útil para este tipo de análisis.

Bibliografía

- Arias, A. (2017). Propuesta de indicador del ingreso de los hogares urbanos a partir de información del consumo: Clasificación óptima de hogares urbanos por niveles de ingreso en Costa Rica (Tesis de maestría). Universidad de Costa Rica, San José, Costa Rica.
- Asociación de Investigadores de Mercado. (2018). Nueva metodología de segmentación y clasificación socioeconómica. Recuperado de https://www.pauta.cl/pauta/site/docs/20180622/20180622170601/aim_nuevo_gse_2018.pdf
- Asociación Mexicana de Inteligencia de Mercado y Opinión. (2018). ¿Cómo mide AMAI los Niveles Socioeconómicos? Recuperado de NSE Índices socioeconómicos AMAI website: <https://nse.amai.org/uncategorized/como-mide-amai-los-niveles-socioeconomicos/>
- Babbie, E. (1989). *The Practice of Social Research* (5a ed.). Belmont, California: Wadsworth Publishing Company.
- Banerjee, A., & Duflo, E. (2012). *Repensar la pobreza: Un giro radical en la lucha contra la desigualdad global*. TAURUS.
- Cámara de Empresas de Investigación Social y de Mercado del Uruguay. (2018). Índice de Nivel Socioeconómico (INSE) 2018. Recuperado de <https://www.ceismu.org/site/indice-de-nivel-socioeconomico-inse-2018/>
- Caro, D., & Cortés, D. (2012). Measuring family socioeconomic status: An illustration using data from PIRLS. Recuperado de http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_05_Chapter_1.pdf
- Castro, C., Gutiérrez, A., Rodríguez, C., & Barahona, M. (2007). *Transformaciones en la estructura social en Costa Rica: Estratos socio-ocupacionales, educación y trabajo*. (1a ed.). San José, Costa Rica: Editorial Universidad de Costa Rica.
- Cuadras, C. (2010). *Nuevos métodos de análisis multivariante* (CMC Editions).
- Da, W., Xiao, H., & Zhuo, M. (2009). Multiple regression analysis of the net income and consumption expenditure of Chinese rural households during 2007. *Asian Agricultural Research*, 1, 22–25. <https://doi.org/10.22004/ag.econ.53439>

- Dahrendorf, D. (1956). *Class and Class Conflict in Industrial Society*. Recuperado de <https://cominsitu.files.wordpress.com/2019/01/ralf-dahrendorf-class-and-class-conflict-in-industrial-society-1.pdf>
- Duncan, O., Featherman, D., & Duncan, B. (1972). *Socioeconomic background and achievement*. New York: Seminar Press.
- Egüez, V., & Pérez, M. (2017). *Caracterización multivariada del nivel socioeconómico para el área urbana del Ecuador, encuesta de condiciones de VIDA-ECV, Ronda 2014* (Tesis de licenciatura, Universidad Central del Ecuador). Recuperado de <http://www.dspace.uce.edu.ec/handle/25000/13979>
- Fernández, S. de la F. (2011). *Análisis de conglomerados*. Madrid: Universidad Autónoma de Madrid. Recuperado de <http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/CONGLOMERADOS/conglomerados.pdf>
- Francés, F. (2009). Elementos para el estudio de la estratificación social en las sociedades avanzadas: Estrategias operativas. *OBETS. Revista de Ciencias Sociales*, 3, 43–57. <https://doi.org/10.14198/OBETS2009.3.05>
- García, C. (2002). *Segmentación social de los hogares a la luz del CENSO 2000: Aplicación a los hogares del Marco Muestral de Viviendas*. Recuperado de http://www.inec.cr/sites/default/files/documentos/inec_institucional/publicaciones/07_segmentacion_social_de_los_hogares_a_la_luz_del_censo_2000.pdf
- Gutiérrez, J. P. (2013). Clasificación socioeconómica de los hogares en la ENSANUT 2012. *Salud Pública de México*, 55, S341–S346.
- Hernández, O. (2013). *Temas de análisis estadístico multivariante* (2. ed). Editorial Universidad de Costa Rica.
- INEGI. (2015). *Condiciones Socioeconómicas 2015*. Recuperado de <https://www.inegi.org.mx/programas/mcs/2015/>
- Instituto Nacional de Estadística y Censos (2014) *Nueva muestra a partir del Marco Muestral de Viviendas 2011 y Proyecciones de Población 2013*.
- Instituto Nacional de Estadística y Censos. (2011). *X Censo Nacional de Población y VI de Vivienda 2011*. Recuperado de http://www.inec.go.cr/sites/default/files/documentos/inec_institucional/estadisticas/resultados/reviviendcenso2011-06.pdf.pdf

- Instituto Nacional de Estadística y Censos. (2015). Índice de Pobreza Multidimensional. Recuperado de http://inec.cr/sites/default/files/documentos/pobreza_y_presupuesto_de_hogares/pobreza/metodologias/mepobrezaenaho2015-01.pdf
- Instituto Nacional de Estadística y Censos. (2017). Encuesta Nacional de Hogares Julio 2017. Recuperado de <http://sistemas.inec.cr/pad4/index.php/catalog/181>
- Instituto Nacional de Estadística y Censos. (2019). Encuesta Nacional de Hogares Julio 2019. Recuperado de <http://inec.cr/sites/default/files/documentos-biblioteca-virtual/reenaho2019.pdf>
- Instituto Nacional de Estadística y Censos [Ecuador]. (2011). Encuesta de Estratificación del Nivel Socioeconómico NSE 2011. Recuperado de <https://www.ecuadorencifras.gob.ec/encuesta-de-estratificacion-del-nivel-socioeconomico/>
- Instituto Nacional de Estadística y Geografía. (s/f). Estratificador INEGI Manual de usuario. Recuperado de https://documentop.com/estratificador-inegi_59f4c6f31723ddfd258676af.html
- Kish, L. (1975). Muestreo de encuestas. México: Trillas.
- Litchfield, J. (1999). Inequality: Methods and Tools. Recuperado de <https://siteresources.worldbank.org/INTPGI/Resources/Inequality/litchfie.pdf>
- Madrigal, J (1986). Metodología y Construcción de un Indicador del Ingreso Familiar: Aplicación a los Censos Nacionales de Vivienda y Población 1984. Asociación Demográfica Costarricense. San José, Costa Rica, 1986.
- Madrigal, J. (2004). Estratificación de hogares y segmentos por niveles de ingreso en el censo 2000. Recuperado de http://www.inec.cr/sites/default/files/documentos/inec_institucional/publicaciones/06_estratificacion_de_hogares_y_segmentos_por_niveles_de_ingreso_en_el_censo_2000.pdf
- Martínez, D. (2010). El aporte de Adolfo Figueroa al análisis de la relación entre equidad, productividad y competitividad. En F. Jiménez (Ed.), Teoría Económica y Desarrollo Social. Exclusión, desigualdad y democracia. Homenaje a Adolfo

- Figuroa (pp. 23–47). Recuperado de <https://ideas.repec.org/h/pcp/pucchp/ide-2010-01-02.html>
- McKay, A. (2002). Defining and Measuring Inequality. Recuperado de <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/3804.pdf>
- Medina, F (2018) Resumen de la consultoría Fernando Medina para la elaboración de una propuesta de SIEH en el INEC 2018.
- Méndez, F., & Trejos, J. D. (2004). Costa Rica: Un mapa de carencias críticas para el año 2000. Recuperado de <http://biblioteca.ccp.ucr.ac.cr/bitstream/handle/123456789/1219/Costa%20Rica%20Un%20mapa%20de%20carencias%20cr%C3%ADticas%20para%20el%20a%C3%B1o%202000.pdf?sequence=1&isAllowed=y>
- Ministerio de Vivienda y Asentamientos Humanos. (2012). Estrato Socio Económico Medio. Recuperado de https://www.mivah.go.cr/Documentos/investigaciones_diagnosticos/estratos_medios/ESM_2012_Tendencias_y_Lineamientos.pdf
- Organización para la Cooperación y el Desarrollo Económicos. (2019). Income inequality. Recuperado de OECD-iLibrary website: <https://data.oecd.org/inequality/income-inequality.htm>
- Rojo, J. M. (2007). Regresión lineal múltiple. Instituto de Economía y Geografía.
- Sirin, S. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, 75, 417–453. <https://doi.org/10.3102/00346543075003417>
- United Nations. (1995). Copenhagen Declaration on Social Development. Recuperado de https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_CONF.166_9_Declaration.pdf
- Valdés, A., Foster, W., Pérez, R., & Rivera, R. (2008). Evolución del ingreso agrícola real en América Latina, 1990-2005: Evidencia en base a cuentas nacionales y encuestas de hogares. *Revista Española de Estudios Agrosociales y Pesqueros*, 218, 71–98.
- Vega, M., Castro, C., Gutiérrez, A. L., & Rodríguez, C. R. (1995). Cambios en la estructura de clases costarricense. 1987-1994. Instituto de Investigaciones Sociales, Universidad de Costa Rica, San José.

Anexos

Anexo 1. Análisis de conglomerados de K medias en índice de ocupación

Opción 1: 5 grupos

Cuadro 20. Estadísticos descriptivos del índice de ocupación

Grupo	No ponderado	Ponderado	Ingreso promedio	Diferencia ¹
1	6 579	840 931	270 777	374 236
2	1 734	229 945	645 013	454 613
3	883	120 668	1 099 626	565 352
4	232	33 270	1 664 978	700 572
5	35	5 195	2 365 551	

1/ Corresponden a las diferencias entre los grupos x+1 y x. *Eta* cuadrado 0.429

Fuente: Elaboración propia a partir del ENAHO, 2015.

Opción 2: 9 grupos

Cuadro 21. Estadísticos descriptivos del índice de ocupación

Grupo	No ponderado	Ponderado	Ingreso promedio	Diferencia ¹
1	1.333	171 185	160 818	138 065
2	5.246	669 746	298 883	284 321
3	1.236	163 599	583 204	290 793
4	846	115 343	873 997	308 996
5	535	71 671	1 182 993	344 934
6	135	18 946	1 527 927	318 327
7	97	14 324	1 846 254	492 933
8	33	4 868	2 339 187	418 841
9	2	327	2 758 028	

1/ Corresponden a las diferencias entre los grupos x+1 y x. *Eta* cuadrado 0.444

Fuente: Elaboración propia a partir del ENAHO, 2015.

Opción 3: 10 grupos

Cuadro 22. Estadísticos descriptivos del índice de ocupación

Grupo	No ponderado	Ponderado	Ingreso promedio	Diferencia ¹
1	1.333	171 185	160 818	138 065
2	5.246	669 746	298 883	284 321
3	1.236	163 599	583 204	214 221
4	498	66 346	797 426	180 256
5	348	48 997	977 682	205 311
6	535	71 671	1 182 993	344 934
7	135	18 946	1 527 927	318 327
8	97	14 324	1 846 254	492 933
9	33	4 868	2 339 187	418 841
10	2	327	2 758 028	

1/ Corresponden a las diferencias entre los grupos x+1 y x. *Eta* cuadrado 0,447

Fuente: Elaboración propia a partir del ENAHO, 2015.

Anexo 2. Estadísticos descriptivos del índice de ocupaciones

Tabla 15. Estadísticos descriptivos del índice de ocupación

Grupo	Ocupación	Categoría Ocupacional	n muestral	N poblacional	Promedio
1	Trabajadores de los servicios y vendedores de comercios y mercados	Asalariado del hogar privado	109	14 141	123 958
	Oficiales. operarios y artesanos de artes mecánicas y de otros oficios	Asalariado del hogar privado	2	240	26 833
	Ocupaciones elementales	Asalariado del hogar privado	479	62 082	152 963
		Cuenta propia	267	33 229	115 922
2	Trabajadores de los servicios y vendedores de comercios y mercados	Cuenta propia	524	67 684	241 980
	Oficiales. operarios y artesanos de artes mecánicas y de otros oficios	Cuenta propia	474	61237	202 498
	Ocupaciones elementales	Empleado Privado	1501	183 062	232 735
		Empleadora	12	1 544	252 224
	Personal de apoyo administrativo	Empleado Privado	416	57 682	395 024
		Cuenta propia	6	754	417 399
3	Trabajadores de los servicios y vendedores de comercios y mercados	Empleado Privado	880	114 398	312 916
	Agricultores y trabajadores calificados agropecuarios. forestales y pesqueros	Empleado Privado	77	8 974	319 404
		Cuenta propia	360	42 328	292 361
	Oficiales. operarios y artesanos de artes mecánicas y de otros oficios	Empleado Privado	547	72 789	340 840
	Operadores de instalaciones y máquinas y ensambladores	Empleado Privado	544	71 685	369 419
		Cuenta propia	220	28 591	290 258
	Ocupaciones elementales	Empleado Público	157	20 002	350 740
4	Técnicos y profesionales de nivel medio	Empleado Privado	498	68 226	625 591
		Cuenta propia	126	16 923	549 089
	Personal de apoyo administrativo	Empleado Público	146	19 099	575 807
		Empleadora	12	1 862	600 619
	Trabajadores de los servicios y vendedores de comercios y mercados	Empleado Público	195	24 505	515 856
	Agricultores y trabajadores calificados agropecuarios. forestales y pesqueros	Empleadora	79	9 812	563 911
	Oficiales. operarios y artesanos de artes	Empleado Público	61	7 698	589 319
		Empleadora	70	8 757	580 700

	mecánicas y de otros oficios				
	Operadores de instalaciones y máquinas y ensambladores	Empleado Público	47	6 354	526 833
		Asalariado del hogar privado	2	363	492 413
	Profesionales científicos e intelectuales	Cuenta propia	120	16 295	802 890
	Técnicos y profesionales de nivel medio	Empleado Público	220	29122	792 360
5	Trabajadores de los servicios y vendedores de comercios y mercados	Empleadora	140	18492	799 920
	Operadores de instalaciones y máquinas y ensambladores	Empleadora	18	2437	802 486
6	Profesionales científicos e intelectuales	Empleado Privado	291	40 844	991 874
	Técnicos y profesionales de nivel medio	Empleadora	57	8 153	906 582
7	Profesionales científicos e intelectuales	Empleado Público	535	71 671	1 182 933
8	Profesionales científicos e intelectuales	Empleadora	82	11 405	1 468 286
9	Directores y gerentes	Empleado Público	53	7 541	1 618 127
10	Directores y gerentes	Empleado Privado	97	14 324	1 846 253
11	Directores y gerentes	Empleadora	33	4 868	2 339 186
12	Directores y gerentes	Cuenta propia	2	327	2 758 027

Fuente: Elaboración propia a partir del ENAHO, 2015.

Anexo 3. Estadísticos descriptivos del índice de tipo de seguro

Cuadro 23. Estadísticos descriptivos del índice de tipo de seguro social

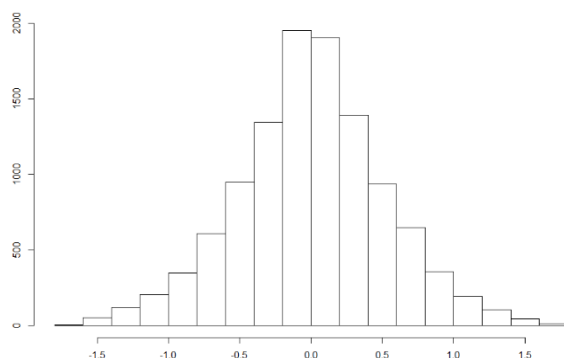
Tipo de seguro social	n muestral	N poblacional	Promedio Ingreso <i>per cápita</i> del hogar neto
No contribuyentes	2 345	294 051	172 624
Asegurados familiares	883	113 328	221 040
Cuenta propia. voluntario o convenio	1 954	249 164	365 980
Pensionado de la CCSS. Magisterio u otro	966	129 638	420 165
Asalariado	5 080	669 452	437 100
Otras formas (privado)	49	6 502	1 135.281
Total	11 277	1 462 135	356 648

Fuente: Elaboración propia a partir del Censo, 2011 y ENAHO, 2015.

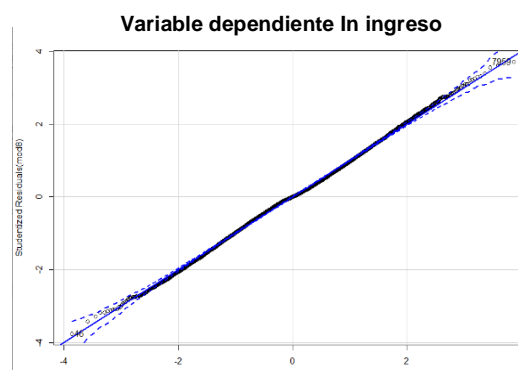
Anexo 4. Cumplimiento de los supuestos del modelo de regresión estrategia A.

a. Normalidad de los residuos

Histograma de los residuos del modelo



Normal PP Plot of Regression



Fuente: Elaboración propia a partir de ENAHO, 2015.

b. Análisis de varianza

En el análisis de varianza se observa que los resultados del modelo se ajustan a lo que se busca en la hipótesis de la regresión.

H₀ = pendientes es igual a cero H₁ = pendientes es diferente a cero

Cuadro 24. Análisis de varianza del modelo de regresión múltiple

Response: LN ingreso	Df	Suma cuadro de error	Cuadrado medio de error	F value	Pr(>F)
Índice Ocupacional	1	16419.4	16419.4	9917.36	2.2e-16
Índice de patrimonio Acumulado	1	4273.0	4273.0	2580.88	2.2e-16
Años de escolaridad	1	975.3	975.3	589.05	2.2e-16
Índice tipo de seguro	1	1102.2	1102.2	665.73	2.2e-16
Cantidad de miembros ocupados	1	1228.7	1228.7	742.14	2.2e-16
Índice de la características de la Vivienda	1	267.5	267.5	161.58	2.2e-16
Número de miembros del hogar	1	8063.6	8063.6	4870.43	2.2e-16
Residuos	8933	14789.7	1.7		

Fuente: Elaboración propia a partir del ENAHO, 2015.

c. Homocedasticidad

Cuadro 25. Test varianza no constante

Análisis de varianza no constante	
Variance formula: ~ fitted.values	
Chisquare = 1,962638. Df = 1. p = 0,16123	
Fuente: Elaboración propia a partir de ENAHO, 2015	

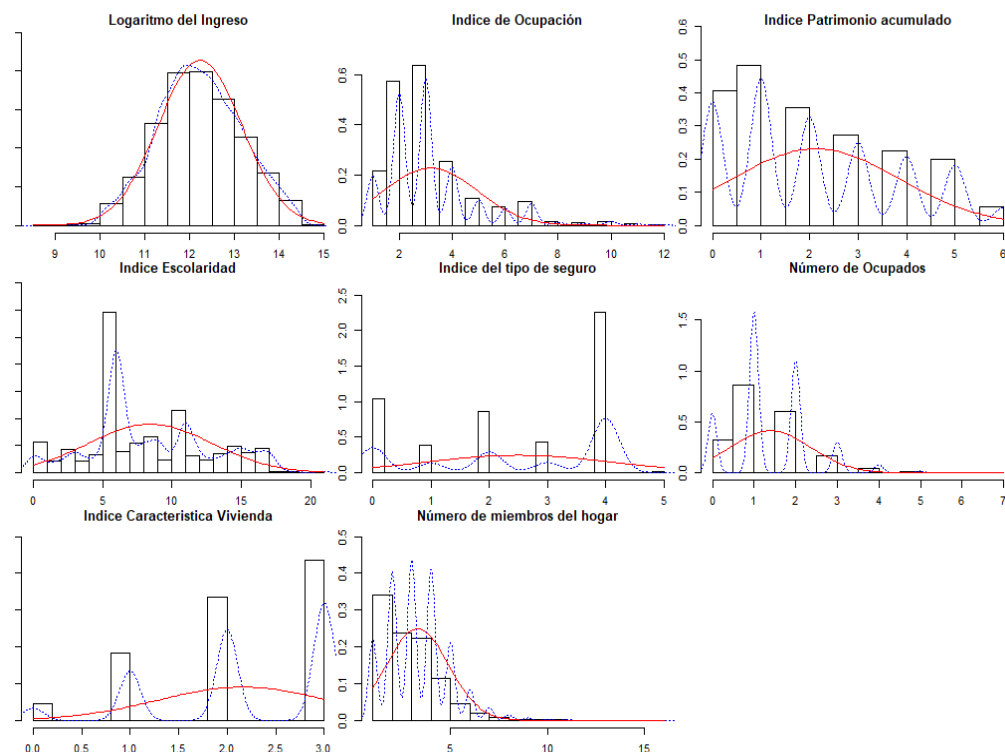
d. No autocorrelación (Independencia)

Cuadro 26. Prueba formal de Independencia

No autocorrelación	
Autocorrelación	0,1729
Estadístico Durbin-Watson	1,6541
Fuente: Elaboración propia a partir de ENAHO, 2015	

e. Comportamiento de variables del modelo

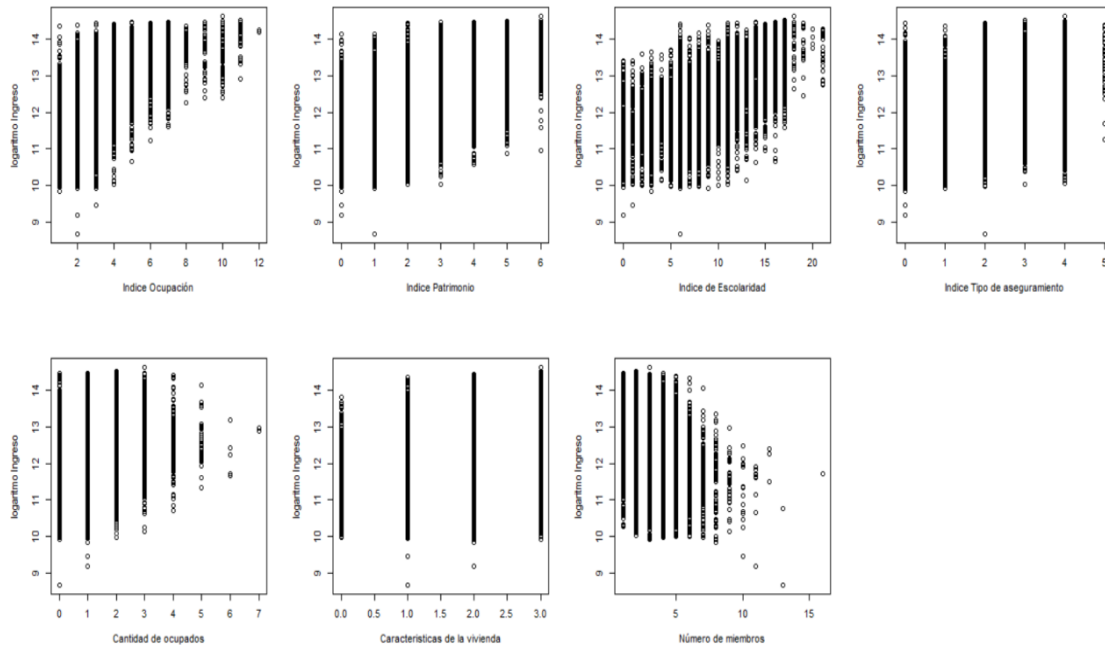
Gráfico 11. Distribución de las variables estrategias A



Fuente: Elaboración propia a partir de ENAHO, 2015.

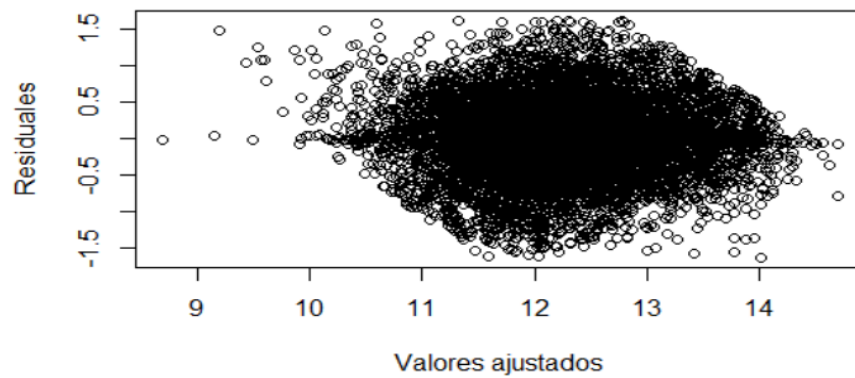
f. Comportamiento de la variable dependiente y el logaritmo del ingreso

Gráfico 12. Relación variable ingreso y covariables



Fuente: Elaboración propia a partir de ENAHO, 2015.

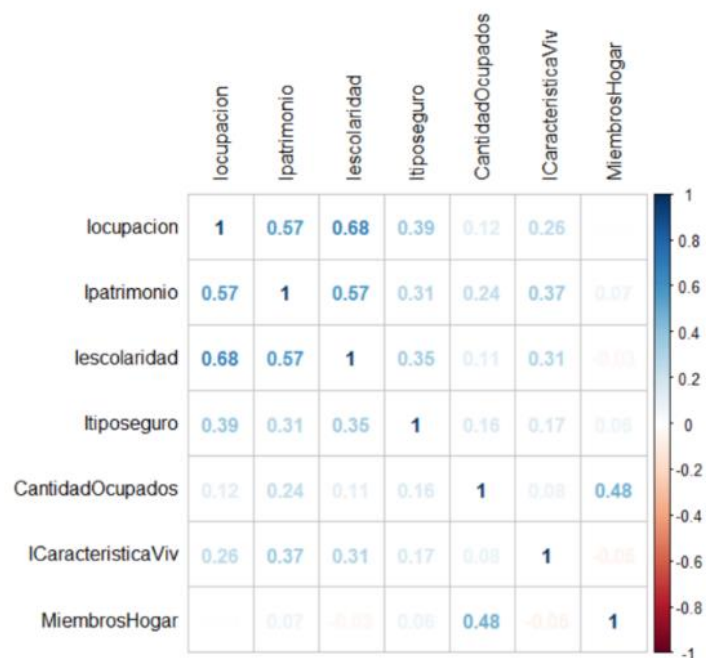
Gráfico 14. Relación variable ingreso y covariables



Fuente: Elaboración propia a partir de ENAHO, 2015

g. Colinealidad y Multicolinealidad

Gráfico 13. Correlación entre variables independientes



Fuente: Elaboración propia a partir de ENAHO, 2015

Cuadro 27. Análisis de tolerancia y VIF del modelo de regresión múltiple

Variables	Tolerancia	Vif
Ingreso <i>per cápita</i> neto		
Índice ocupacional	0,47	2,13
Índice de patrimonio Acumulado	0,57	1,76
Índice tipo de seguro	0,48	2,10
Años de escolaridad	0,85	1,17
Cantidad de miembros ocupados	0,71	1,41
Índice de la vivienda	0,85	1,17
Número de miembros del hogar	0,74	1,35

Fuente: Elaboración propia a partir de ENAHO, 2015

Anexo 5. Escenarios para conformación de niveles socioeconómicos, estrategia A

Cuadro 28. Estrategia A utilizando K medias en el ingreso

Grupos	Ingreso	Frecuencia	%	Eta 1/	Eta cuadrado	cv por grupo
1	198443,8	3620	54,2			21,79
2	340724,2	2404	36,1	0,910	0,829	14,72
3	568568,0	654	9,8			15,71
1	184981,8	2931	43,9			19,70
2	296446,1	2371	35,5	0,948	0,898	11,56
3	430497,2	1016	15,2			10,53
4	630536,9	360	5,4			11,58
1	176104,8	2491	37,3			18,29
2	276089,8	2397	35,9			11,16
3	387288,6	1162	17,4	0,962	0,925	9,21
4	526511,7	441	6,6			8,42
5	686018,0	187	2,8			8,93
1	167566,8	2084	31,2			16,74
2	254897,0	2225	33,3			10,15
3	346718,5	1382	20,7	0,972	0,945	8,62
4	457049,8	615	9,2			7,71
5	591730,4	268	4,0			6,33
6	724646,6	104	1,5			7,76

1/ El eta se calcula entre la variable proxy del ingreso, es decir, el logaritmo del ingreso y los niveles construidos.

Fuente: Elaboración propia a partir de Censo, 2011.

Anexo 6. Escenario conformación de niveles socioeconómicos, estrategia B

Cuadro 29. Estrategia B utilizando K medias en el ingreso

Grupos	% sin seguro o por el estado	% NBI	% Al menos una miembro con educación Universitaria	TDA	Frecuencia	%	Eta 1/	Eta cuadrado 1/
1	25,7	38,6	16,8	4,6	1622	24,3	0,88	0,78
2	15,6	17,6	37,1	3,4	2992	44,8		
3	9,8	8,2	69,2	2,6	2064	30,9		
1	27,1	42,8	15,2	4,9	1199	17,9		
2	18,0	21,0	29,7	3,7	2291	34,3	0,815	0,665
3	12,0	12,7	51,4	3,0	2026	30,3		
4	9,1	6,1	77,8	2,4	1162	17,4		
1	30,3	62,9	11,6	5,4	328	4,9		
2	24,2	31,8	18,3	4,4	1375	20,6		
3	16,7	19,1	33,2	3,5	2044	30,6	0,881	0,776
4	11,6	12,2	53,2	3,0	1830	27,4		
5	9,1	5,9	78,5	2,4	1101	16,5		
1	28,7	70,4	12,2	5,1	213	3,2		
2	29,0	39,0	14,6	5,2	674	10,1		
3	20,1	25,7	23,2	3,9	1470	22,1	0,903	0,815
4	15,3	16,6	38,7	3,3	1890	28,3		
5	10,7	11,1	57,8	2,8	1543	23,1		
6	9,0	5,3	81,1	2,4	888	13,3		

1/ El eta se calcula entre la variable proxy del ingreso, es decir, el l ingreso y los niveles construidos.

Fuente: Elaboración propia a partir de Censo, 2011.

Anexo 7. Escenarios conformación de niveles socioeconómicos, estrategia VB

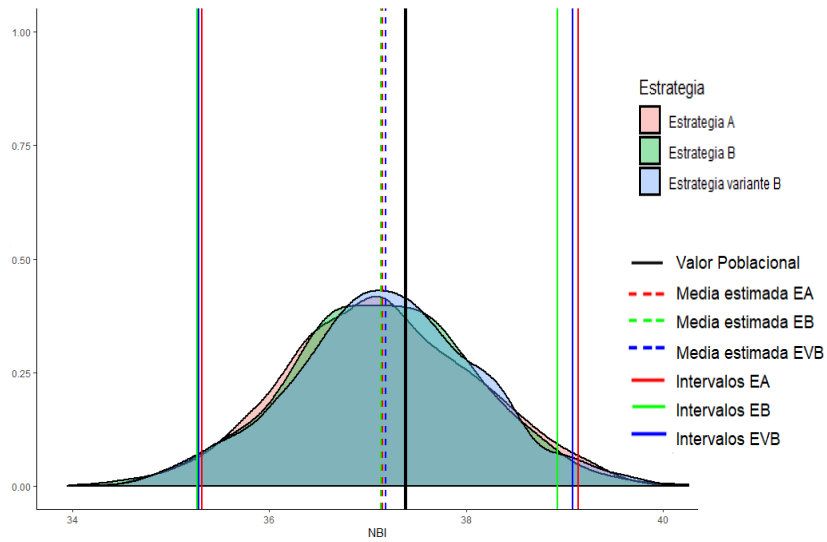
Cuadro 30. Variante a la estrategia B incorporando el ingreso

Grupos	Ingreso	% sin seguro o por el estado	% NBI	% Al menos una educación superior	TDA	Frecuencia	%	eta1	cv
1	198443,8	20,7	27,7	25,9	4,1	3616	54,2		21,79
2	340724,2	11,4	11,9	55,4	2,9	2408	36,1	0,829	14,72
3	568568,0	9,3	5,3	82,3	2,3	654	9,8		15,71
1	184981,8	22,0	30,1	22,9	4,3	2932	43,9		21,79
2	296446,1	13,1	14,4	47,1	3,1	2371	35,5	0,898	14,72
3	430497,2	9,5	8,1	69,7	2,5	1015	15,2		15,71
4	630536,9	9,4	4,4	86,7	2,1	360	5,4		11,33
1	176104,8	22,9	31,9	21,0	4,4	2491	37,3		21,79
2	276089,8	14,1	16,0	43,0	3,2	2397	35,9		14,72
3	387288,6	10,1	9,5	63,5	2,6	1162	17,4	0,925	15,71
4	526511,7	9,3	5,7	79,6	2,4	441	6,6		11,33
5	686018,0	9,5	4,2	90,0	1,8	187	2,8		8,29
1	167566,8	23,9	33,8	19,1	4,6	2084	31,2		21,79
2	254897,0	15,5	17,9	38,3	3,3	2229	33,3		14,72
3	346718,5	11,0	11,5	57,0	2,8	1383	20,7	0,945	15,71
4	457049,8	9,3	7,2	73,1	2,5	614	9,2		11,33
5	591730,4	9,4	4,7	84,4	2,3	267	4,0		8,29
6	724646,6	9,2	3,7	92,2	1,7	101	1,5		5,25

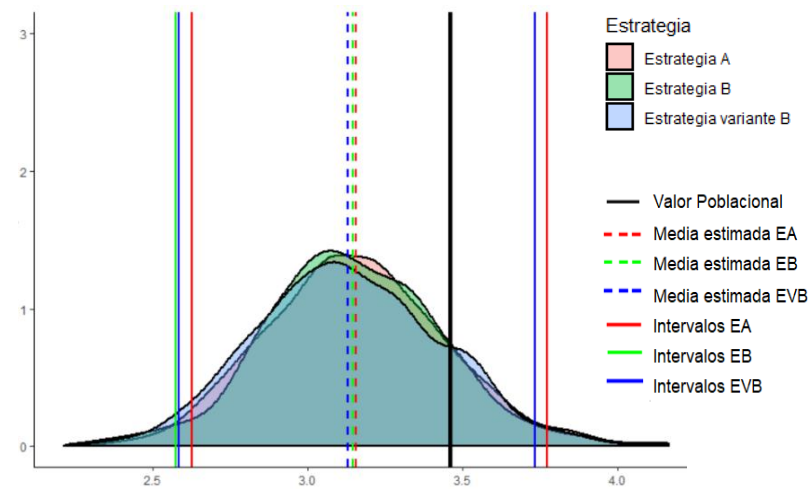
1/ El eta se calcula entre la variable proxy del ingreso, es decir, el logaritmo del ingreso y los niveles construidos. Fuente: Elaboración propia a partir de Censo, 2011.

Anexo 8. Gráficos estimadores por zona rural

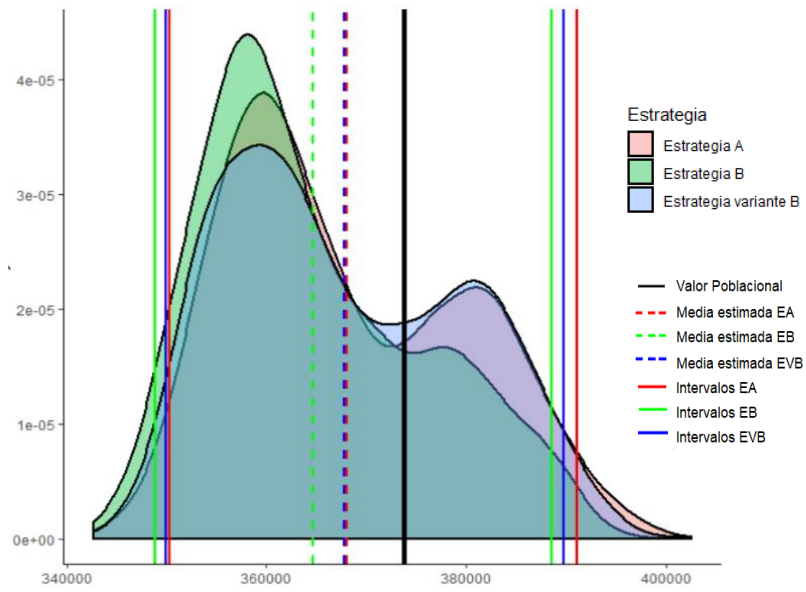
Rural: Necesidades básicas insatisfechas.



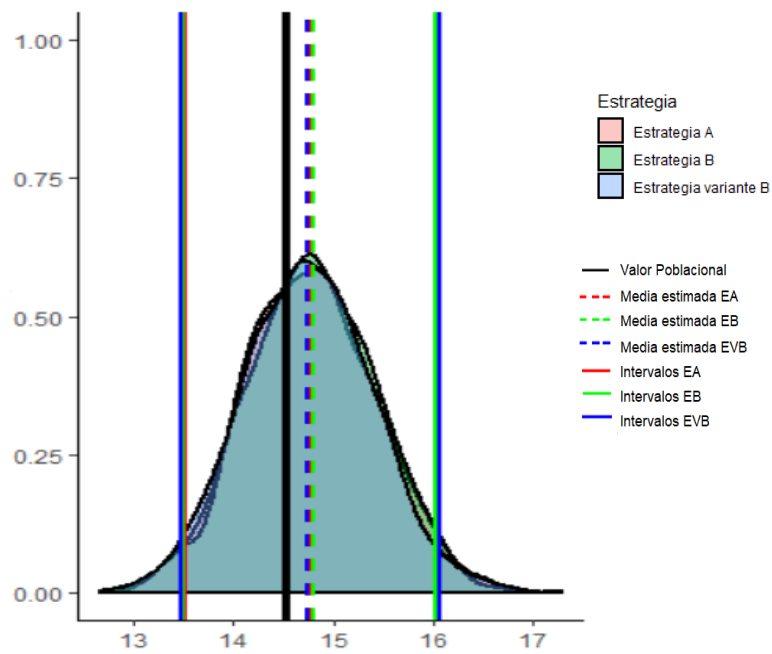
Rural: Tasa de desempleo abierto

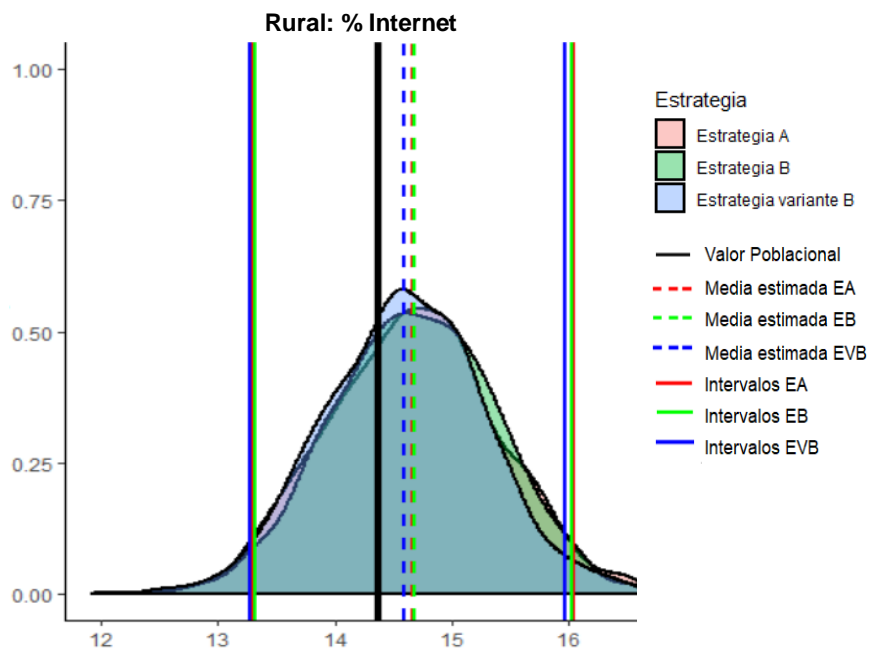
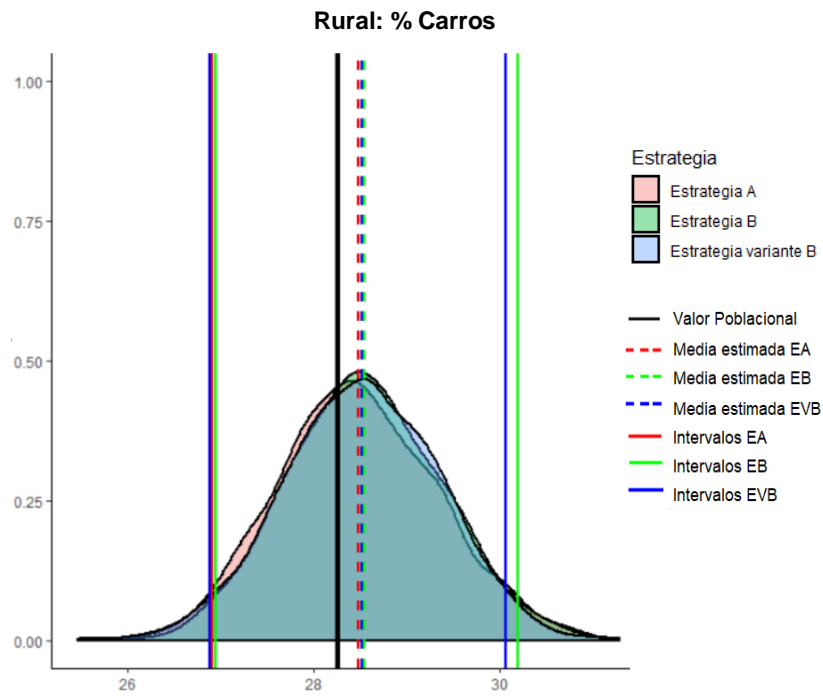


Rural: Ocupados

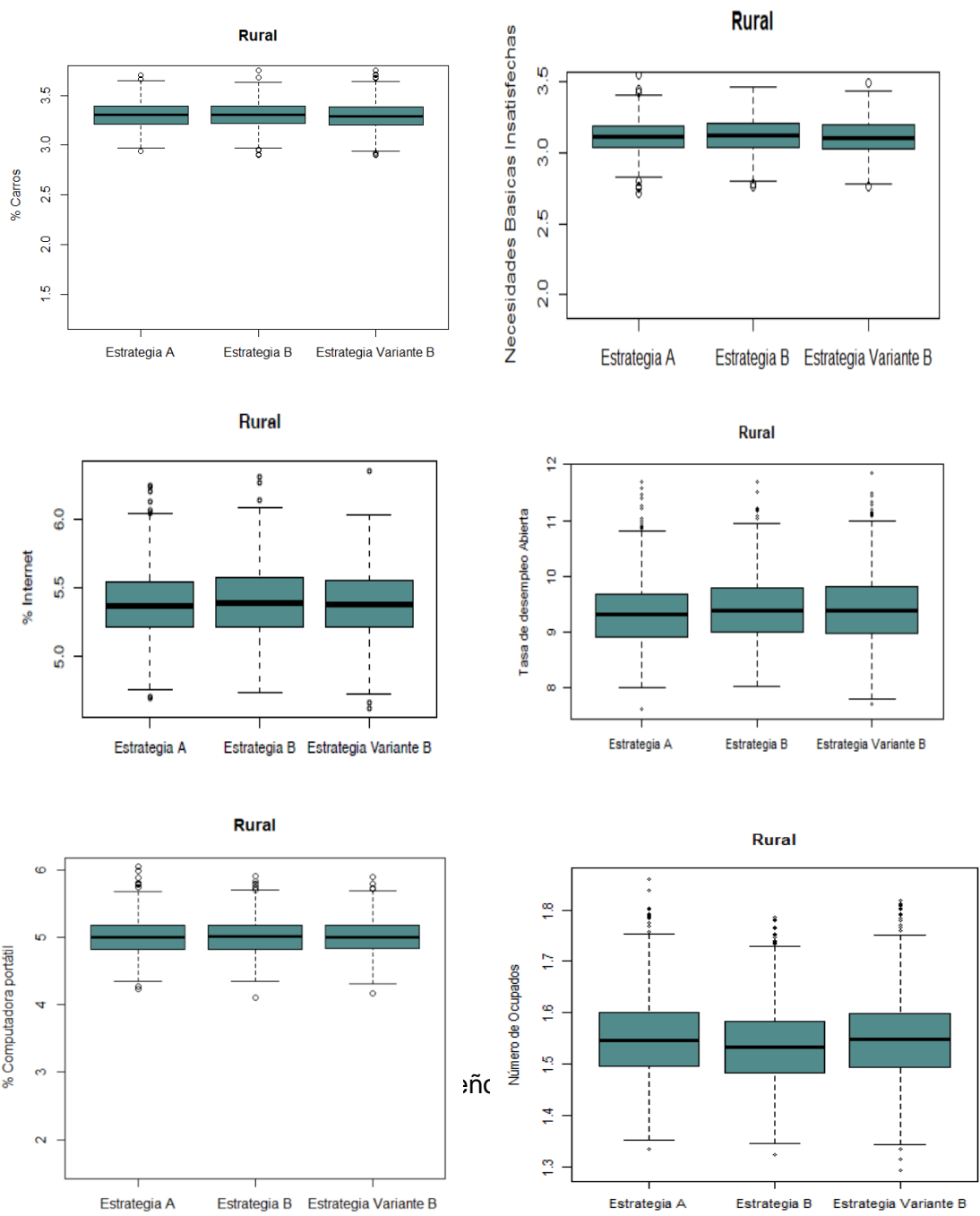


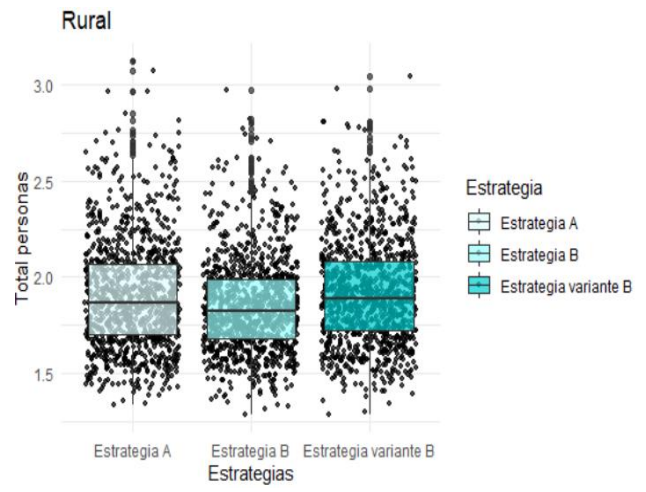
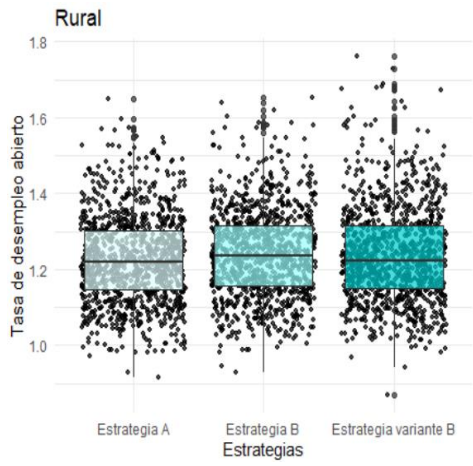
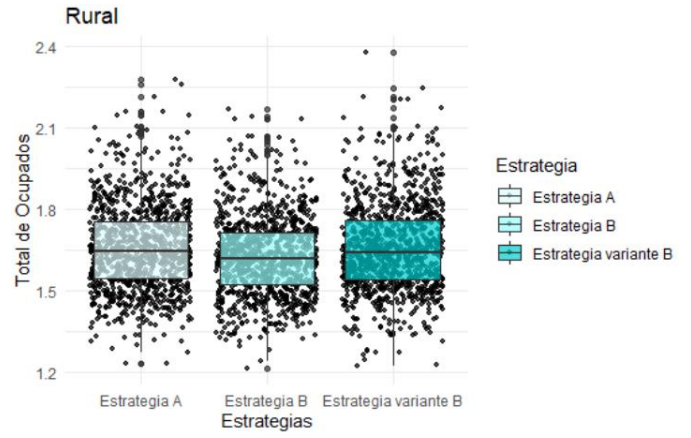
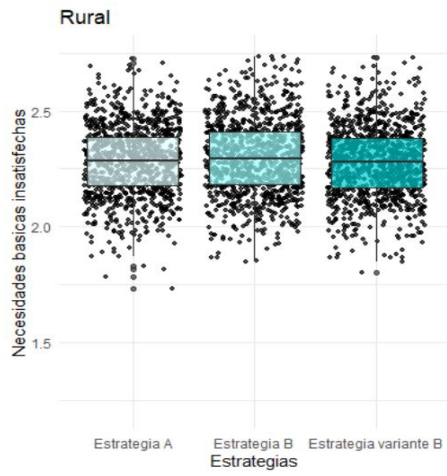
Rural: % Computadora portátil





Anexo 9. Gráficos del coeficiente de variación de los estimadores según zona rural





Anexo 11. Gráficos efecto del diseño para las variables porcentaje de carros, internet y portátil en Costa Rica según zona

