



**MÁSTER OFICIAL EN ECONOMÍA: INSTRUMENTOS DE
ANÁLISIS ECONÓMICOS.**

CURSO ACADÉMICO: 2018-2019

TRABAJO FIN DE MÁSTER

**ESTIMACIÓN DE ÍNDICE DE POBREZA MULTIDIMENSIONAL A
NIVEL PROVINCIAL PARA ECUADOR.**

**ESTIMATION OF MULTIDIMENSIONAL POVERTY INDEX TO
PROVINCIAL LEVEL FOR ECUADOR.**

AUTOR:

LUIS DANIEL VISCAINO CAICHE

TUTOR:

FRANCISCO JAVIER PARRA RODRIGUEZ

DANIEL DIAZ FUENTES

SEPTIEMBRE, 2019

TABLA DE CONTENIDO

1	RESUMEN	2
2	INTRODUCCIÓN	3
3	MEDIDAS DE POBREZA	5
3.1	ENFOQUE MULTIDIMENSIONAL PARA EVALUAR EL BIENESTAR.	5
3.2	MEDICIONES MULTIDIMENSIONALES DE POBREZA	5
3.2.1	Indicadores Monetarios.	7
3.2.2	Enfoque Multidimensional.	8
4	METODOLOGÍA	13
4.1	EL MATCHING ESTADÍSTICO: METODOLOGÍA PARA EL CÁLCULO DE LA TASA DE POBREZA EN ECUADOR.	13
4.1.1	Variables Comunes entre el Censo de Población y Vivienda 2010 con la Encuesta de Condiciones de Vida del año 2014 para Ecuador.	14
4.1.2	Tratamiento de los Datos.....	15
4.1.3	Selección de Variables y Aplicación de las Técnicas de Fusión.	19
4.1.4	Técnicas y Regularización del Mejor Modelo.	28
4.1.5	Cartografía.....	39
5	CONCLUSIONES	41
6	REFERENCIAS BIBLIOGRÁFICAS.....	42

1 RESUMEN

El presente estudio trata de analizar diversos factores que pueden influir de manera directa o indirectamente en una determinada persona, teniendo además en consideración las condiciones de vida que presenta, permitiendo así categorizarla como una persona de escasos recursos (pobre), a través de la aplicación de la respectiva metodología.

Dicha metodología implementada principalmente se aplica a 28970 viviendas lideradas por una persona la cual es asignada como Jefe(a) de Hogar, ubicadas en áreas urbanas y rurales dentro del conjunto de provincias que forman parte del territorio ecuatoriano según los datos recopilados de la respectiva Encuesta de Condiciones de Vida (ECV) para el año 2014 que fueron proporcionados por el Instituto Nacional de Estadística y Censo (INEC), realizando el correspondiente Matching Estadístico junto con los datos extraídos del Censo Nacional de Población y Vivienda para el año 2010.

El Matching Estadístico ha sido implementado para poder realizar la respectiva estimación del correspondiente Índice de Pobreza, en el cual se puede obtener resultados a nivel de regiones, provincias, cantones, pero para interés personal se realiza la correspondiente estimación para cada una de las provincias que forman parte del presente estudio, dicho método trata de estimar aquel índice recopilado de la ECV-2014 a partir de un conjunto de variables comunes seleccionadas que se presenta en ambas bases de datos.

Por lo que en base a los resultados se pudo observar que la provincia con un alto índice se presenta en la Región Sierra principalmente para Morona Santiago en contraste con la provincia de Galápagos que muestra un porcentaje bajo en relación a la pobreza.

The present study tries to analyze various factors that can directly or indirectly influence a given person, also taking into consideration the living conditions that it presents, thus allowing it to be categorized as a person of poor resources (poor), through the application of the respective methodology.

This methodology implemented mainly applies to 28970 homes led by a person who is assigned as Head of Home, located in urban and rural areas within the set of provinces that are part of the Ecuadorian territory according to the data collected from the respective Survey of Living Conditions (ECV) for the year 2014 that were provided by the National Institute of Statistics and Census (INEC), performing the corresponding Statistical Matching together with the data extracted from the National Population and Housing Census for the year 2010.

Statistical Matching has been implemented to be able to make the respective estimate of the corresponding Poverty Index, in which results can be obtained at the level of regions, provinces, cantons, but for personal interest the corresponding estimate is made for each of the provinces that As part of this study, this method tries to estimate that index compiled from the ECV-2014 from a set of selected common variables that is presented in both databases.

Therefore, based on the results, it was observed that the province with a high index appears in the Sierra Region mainly for Morona Santiago in contrast to the province of Galapagos, which shows a low percentage in relation to poverty.

2 INTRODUCCIÓN

Partiendo del contexto de la crisis económica, la cual se entiende como aquella fase caracterizada por tener efectos negativos, tales como recesión, contracción o depresión económica causando de manera global daños sociales, patrimoniales e incluso políticos. Por lo que la crisis económica que atravesó el territorio ecuatoriano, ha venido siendo tema de debate desde 1998 cuando la convergencia de tres eventos negativos, condujo a un respectivo deterioro en la economía originando una fuerte inestabilidad social que fue motivo para profundización de problemas estructurales de inequidad social, exclusión y pobreza.

Tal que en Ecuador se ha establecido 12 objetivos y 93 metas del Plan Nacional para el Buen Vivir de los cuales 9 objetivos y 53 metas tienen relación con la lucha contra la pobreza

¹Por lo que la utilización de técnicas como el Matching Estadístico ha sido de gran aportación para varios estudios como estimación de la privación material y el gasto de los hogares para el estudio de Statistical Matching of European Union statistics on income and living conditions (EU-SILC) and the household budget survey (Serafino and Tonkin, 2013), donde realiza la respectiva técnica entre la Encuesta de Presupuestos Familiares (HBS) y la Encuesta Europea de Ingresos y Condiciones de Vida (EU-SILC), con el objetivo de estimar el gasto y la privación material a partir de un número limitado de variables comunes (región, grupo de edad, estado de tenencia del hogar, el tipo de vivienda, el tamaño del hogar, la tenencia o no de automóvil y el nivel de ingresos disponibles en el hogar).

La amplia literatura acerca del ²Índice de Pobreza Multidimensional (IPM) la cual trata de identificar diversas carencias a nivel de hogares y personas en diferentes ámbitos como: salud, educación y nivel de vida.

Por lo que el IPM se encarga de estudiar tanto la prevalencia del conjunto de carencias multidimensionales como su respectiva intensidad, dando como resultado el conocimiento de múltiples carencias que sufren los individuos al mismo tiempo.

Por otra parte, también es usado para tener una idea general de manera proporcional del número de personas que viven en condiciones de pobreza, de la misma manera permitiendo hacer diversas comparaciones entre niveles nacionales, regionales y mundiales de factores como Grupos Étnicos, Zonas Rurales o Urbanas, así como otras cualidades relevantes de los diferentes hogares y comunidades.

Dada esta previa definición. ³Para Ecuador, el Índice de Pobreza Multidimensional contiene 4 dimensiones y 12 indicadores que evalúan de manera simultánea el cumplimiento mínimo de los derechos de cada individuo que se encuentra dentro del territorio.

¹ (Parra Rodriguez, 2018)

² (Report, 2019)

³ (Castillo Añazco, 2016)

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

Es identificada una persona como pobre multidimensional cuando esta tiene privaciones en una tercera parte o más de los indicadores ponderados y como pobre extremo multidimensional cuando tiene al menos la mitad de privaciones en los indicadores ponderados.

El enfoque del presente estudio se basa en utilizar una muestra contenida de la Encuesta de Condiciones de Vida para el año 2014, obtenida del sitio web del Instituto Nacional de Estadística y Censo (INEC) para estimar un modelo explicativo, que estudie la relación mediante el nivel de significancia que aporta los diversos factores, que se encuentran establecidos por el IPM con respecto a la variable pobreza, la cual ya viene proporcionada en dicha Encuesta.

Por lo que se puede estimar una distribución conjunta de Y_i con un vector de covariables, denotado por X_{ij} ambas covariables deben estar implementadas tanto en la Encuesta de Condiciones de Vida (2014) como en el último Censo Nacional de Población y Vivienda (2010).

Como línea de estudio, se propone la estimación de tasa de pobreza de cada una de las provincias que comprenden el amplio territorio ecuatoriano, utilizando los datos de la Encuesta de Condiciones de Vida para el 2014 (ECV) y del Censo de Población y Vivienda de 2010, ambas estadísticas son operaciones oficiales del Instituto Nacional de Estadística y Censos de Ecuador (INEC).

El desarrollo del presente estudio está estructurado en los siguientes apartados, en el tercer apartado se exponen las medidas de pobreza, en el cuarto apartado, se detalla la metodología implementada, se describe el procedimiento utilizado para estimar el índice de pobreza en las respectivas regiones del Ecuador, junto con la selección de covariables comunes entre la ECV y el Censo de Población, posteriormente se analizan los resultados obtenidos en la estimación del índice de incidencia de la pobreza y la brecha de pobreza sobre un umbral de necesidades consideradas básicas o insatisfechas, y finalmente en el quinto apartado, se exponen las conclusiones obtenidas del respectivo modelo.

3 MEDIDAS DE POBREZA

3.1 ENFOQUE MULTIDIMENSIONAL PARA EVALUAR EL BIENESTAR.

El típico enfoque en base al cual se construye la definición respectiva acerca del bienestar se origina a partir de la Economía Neoclásica, donde el término “bienestar” es referido a la satisfacción de necesidades preestablecidas como ilimitadas en un mundo de recursos limitados.

Para el territorio ecuatoriano, la visión de bienestar corresponde al concepto del ⁴“Buen Vivir” que teóricamente se compara con el enfoque clásico de acumulación y de consumo y es compatible con el enfoque de capacidades o de derechos.

⁵Son entonces los derechos del Buen Vivir. Plenamente identificados en el segundo capítulo de la Constitución, los elementos que impulsan a generar una nueva métrica de bienestar más acorde al Mandato Constitucional.

3.2 MEDICIONES MULTIDIMENSIONALES DE POBREZA

Con el pasar de los tiempos se han venido introduciendo mediciones multidimensionales en relación a la pobreza.

Los primeros estudios acerca del tema se originaron con el Método de la pobreza por Necesidades Básicas Insatisfechas (NBI), el cual se basa en poder identificar carencias críticas que posee una determinada población, utilizando un conjunto de indicadores que se destacan principalmente en cuatro áreas de necesidades básicas que presenta cada individuo, dicho concepto fue introducido en 1981 por Paul Streeten e implementada por la Comisión Económica para América Latina y el Caribe (CEPAL).

Posteriormente, Ruben Kaztman en 1989 añade el tema de la pobreza monetaria como complemento a las Necesidades Básicas Insatisfechas para difundir una nueva tipología, tomando en consideración el contingente estado transitorio o crónico de la pobreza.

Desde 1990, el Reporte de Desarrollo Humano incluye el Índice de Desarrollo Humano (IDH) ideado por Mahbub, en función del enfoque de capacidades de Amartya-Sen que agrupa tres dimensiones de bienestar (salud, educación y nivel de vida).

Posteriormente, en el año 2010, con motivo del 20º aniversario de la publicación del Informe de Desarrollo Humano, emerge un nuevo indicador, elaborado por el PROGRAMA DE LAS NACIONES UNIDAS PARA EL DESARROLLO con la colaboración del Oxford Poverty and Human Development Initiative ⁶(OPHI). Se trata del IPM, método desarrollado por Alkire y Foster (AF) en el año 2007, que se ha

⁴ Buen Vivir: Concepto que implica vivir en dignidad, con las necesidades básicas satisfechas, en el goce pleno de los derechos, en armonía con uno mismo, con el resto de la comunidad y con las distintas culturas y con la naturaleza (León, 2015).

⁵ (SENPLADES, 2017)

⁶ OPHI: OPHI cuenta con una red global de trabajo en la región que la integran países como: Brasil, Chile, Colombia, Costa Rica, Cuba, República Dominicana, El Salvador, Honduras, México, Paraguay, Perú, Uruguay y Ecuador.

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

convertido en la práctica estadística más utilizada para evaluar la pobreza multidimensional a nivel mundial.

En la actualidad el Enfoque Multidimensional de la Pobreza forma parte de los 17 Objetivos de la ⁷Agenda 2030, que en su primer objetivo establece “Poner fin a la pobreza en todas sus formas y dimensiones al 2030”.

Desde hace casi cuatro décadas, la ⁸Comisión Económica para América Latina y el Caribe (CEPAL) ha venido realizando constante mediciones al tema de pobreza, basándose de los ingresos que se perciben en los hogares, dicha información ha sido recopilada de la encuesta de hogares de los países que forman parte de América Latina para la estimación de la magnitud de pobreza en la región y su evolución en los distintos países por medio de una metodología común.

Doce años después se mostraron los resultados de un segundo estudio, cuyo enfoque principal fue actualizar los valores de las líneas de pobreza a partir de las encuestas de ingresos y gastos de la década de 1980 (CEPAL, 1991). La CEPAL ha creado regularmente sus estimaciones de la pobreza regional sobre la base de las líneas determinadas en este estudio y de otras que se calcularon posteriormente para los países no incluidos en él.

⁹La noción de pobreza remite a una situación en que las personas no alcanzan un estándar, norma social o estado de situación deseable. Sen (1983) identificó un “núcleo irreductible” de la pobreza, que se refiere a la incapacidad de las personas de satisfacer sus necesidades básicas de supervivencia, como el acceso a una alimentación mínima o a un lugar donde se pueda guardar de la intemperie, que son necesidades cuya insatisfacción es signo de pobreza en cualquier sociedad contemporánea.

De esta manera, la definición de la pobreza puede incluir conceptos tan diversos como necesidad, estándar de vida, insuficiencia de recursos, carencia de seguridad básica, falta de titularidades, privación múltiple, exclusión, desigualdad, clase, dependencia y padecimiento inaceptable (Spicker, 1999).

¹⁰Relativamente los indicadores de pobreza en particular se clasifican en dos tipos, monetarios que se fundamenta en la distribución de los ingresos o los gastos, o de carencia o de necesidades básicas insatisfechas. Si bien como se apunta en Panorama Social de América Latina 2013, elaborado por CEPAL, ambas medidas se complementan:

“Es necesario evaluar también qué sucede cuando se integra la pobreza por ingresos como una dimensión más en la medición por carencias, en el entendido de que ambas son medidas incompletas del bienestar y su combinación permite reducir los errores de exclusión que surgen de utilizar solo una de estas medidas. Los resultados muestran que la complementariedad entre las medidas de pobreza multidimensional y monetaria es mayor en los países donde la pobreza medida por medio de carencias no monetarias es menos prevalente y, por tanto, es donde tiene mayor potencial para reducir vacíos de información y disminuir los errores de exclusión”, (CEPAL, 2013, p.21)

⁷ Agenda 2030: Se refiere a un plan de acción cuyo objetivo principal es velar por la protección de las personas, el planeta y la prosperidad

⁸ (CEPAL, 2013)

⁹ (CEPAL, 2018)

¹⁰ (Parra, 2015)

3.2.1 Indicadores Monetarios.

¹¹Los índices de pobreza monetarios que han sido utilizados en su mayoría, están asociados a la familia de índices paramétricos propuestos por Foster, Greer y Thorbecke (1984), FGT, se obtienen a partir de la siguiente expresión:

$$(1) P\alpha = \frac{1}{N} \sum_{i=1}^q \left[\frac{z - y_i}{z} \right]^\alpha, \quad \alpha = 0,1,2$$

Donde:

i- Individuo (viviendas, hogares, personas)

y_i- Ingreso del hogar por adulto equivalente (o gasto-consumo)

z.- Línea de pobreza. Ingreso del hogar inferior al valor de un monto mínimo (umbral) de una cesta de la compra.

(z – y_i)/z.- distancia relativa de y_i a z

n.- tamaño de la población (hogares, viviendas o personas)

q.- tamaño de la población pobre

α.- parámetro que hace sensible las medidas de pobreza a la distribución de ingresos.

FGT0; α = 0, obtiene el indicador incidencia de la pobreza (extensión o prevalencia), o proporción de población cuyo ingreso es inferior a la línea de pobreza.

FGT1; α = 1, obtiene la brecha de pobreza (o indigencia), intensidad de la pobreza. Es decir el promedio del déficit relativo de la pobreza de la población (las distancias de los no pobres son cero). Mide la brecha (en proporción) entre z y el ingreso promedio de los hogares.

FGT2; α = 2, se obtiene la severidad de la pobreza; lo que supone dar un peso mayor a las distancias relativas de los más pobres.

En cuanto a la línea de pobreza se calcula en base a un criterio estadístico, el cual representa el 60% de la mediana correspondiente a la distribución de los ingresos por unidad de consumo fijado, o bien partiendo del valor de una canasta de bienes y servicios, empleando el método del costo de las necesidades básicas.

La línea de pobreza basada en la canasta de alimentos se denomina habitualmente “absoluta”, en contraposición a las líneas de “pobreza relativa”, puesto que se construye a partir de los requerimientos calóricos y nutricionales que aseguran un adecuado funcionamiento físico de la persona. A su vez, la determinación de la línea de pobreza sobre la base del comportamiento de un grupo de referencia introduce un criterio de adecuación al nivel de vida que existe en cada país y época. En consecuencia, la línea de pobreza definida de esta manera incluye implícitamente el costo de bienes y servicios necesarios para atender a los requerimientos de participación social.

¹¹ (Parra, 2015)

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

La canasta básica de alimentos se define sobre una determinada base correspondiente a un criterio normativo central: tal que su valor permita comprar los alimentos necesarios para alcanzar los requerimientos de energía recomendados por la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) y la Organización Mundial de la Salud (OMS) (sobre la base del promedio calculado según la edad, sexo y nivel de actividad física de las personas). Junto con ello, se tiene en consideración que la canasta básica de alimentos presente una estructura adecuada en lo que respecta a la participación de proteínas, carbohidratos y grasas.

A cuyo valor denominado línea de indigencia, se añade el monto requerido por los hogares, de tal forma que pueda satisfacer las necesidades básicas no alimentarias, con la finalidad de estimar el valor total de la línea de pobreza. Para ello, se realiza la respectiva multiplicación de la línea de indigencia por un factor que toma valores distintos según se refiera a áreas urbanas o rurales. En cuanto al valor de las líneas de indigencia y pobreza ambos se van actualizando cada año en base a la variación acumulada del índice de precios al consumidor ¹²(IPC).

3.2.2 Enfoque Multidimensional.

El término pobreza es un fenómeno complejo y multidimensional, razón por la cual existen múltiples definiciones y maneras de medirla, en Europa la Estrategia de Lisboa 2020 estableció una medida multidimensional de la pobreza común para los países europeos denominada ERPE (personas En Riesgo de Pobreza y/o Exclusión) o las siglas en inglés AROPE (At Risk of Poverty and/or Exclusión).

Según la Estrategia Europa 2020 se consideran personas en riesgo de pobreza y/o exclusión social a la población que se encuentra en alguna de las tres situaciones que se definen a continuación. Personas que viven en riesgo de pobreza (por debajo del 60% de la mediana del ingreso equivalente o por unidad de consumo), personas que sufren privación material severa (4 de los 9 ítems definidos) o personas que viven en hogares sin empleo o con baja intensidad en el empleo (sus miembros en edad de trabajar lo hicieron menos del 20% del total de su potencial de trabajo durante el año anterior al de la entrevista).

En caso de estar incluidas en dos o tres condiciones, las personas se contabilizan solo una vez. El Índice de Pobreza Multidimensional (IPM) elaborado por la ONU (UNDP, 2014) y que desde el 2010 sustituye a los anteriores Índices de Pobreza Humana, y el desarrollado por la CEPAL (2014), son dos ejemplos de indicadores de recuento ajustados (Alkire y Foster, 2008) que también responden a dicho enfoque multidimensional. Se obtienen mediante la siguiente operativa:

¹² IPC: Indicador de Medida Estadística que refleja la evolución del conjunto de precios de los bienes y servicios básicos que consume una determinada población en un respectivo territorio.

- Índice de Recuento (H), o porcentaje de personas con al menos dos carencias (% de población multidimensionalmente pobre).

$$(2) H = q/n$$

Donde q es el número de pobres y n la población total.

- Índice de Intensidad (A), que se basa en calcular el respectivo porcentaje medio de privaciones experimentadas por las personas pobres (o la cantidad promedio de carencias que experimentan los pobres).

$$(3) A = |c(k)|/qd$$

Suponiendo que k es un número entero entre 1 y d. Definimos el vector censurado de recuento de privaciones $c(k)$ de la siguiente manera:

Si $c_i(k) > K$, entonces $c(k) = c_i$, o el recuento de privaciones de la persona i
Si $c_i(k) < K$, entonces $c_i(k) = 0$.

Cabe destacar que $c_i(k)/d$ representa el porcentaje de posibles privaciones sufridas por una persona pobre i, y, por lo tanto, el promedio de la proporción de las privaciones entre los pobres está dado por A.

- Índice de Recuento Ajustado (M0), que combina la información sobre la incidencia (H) y la intensidad de la pobreza (A) al multiplicar ambos indicadores.

En CEPAL (2014) se considera una integración mediante la insuficiencia de recursos (hogares con ingresos “per cápita” insuficientes para cubrir las líneas de pobreza alimentarias y no alimentarias) y la carencia de algunos bienes patrimoniales, que CEPAL integra como dimensión del Estándar de Vida. Las dimensiones, indicadores, umbrales y pesos de su índice multidimensional figuran en la Tabla siguiente.

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

Tabla 3.1: Índice de Pobreza Multidimensional

Dimensiones	Indicadores de Privación personas que viven en:	Ponderación (%)	
VIVIENDA		22,2	
Precariedad de los materiales	Viviendas con piso de tierra o con techo o muros con materiales precarios (desechos, cartón, latas, caña, palma, paja, otros materiales).	7,4	
Hacinamiento	Hogares con tres o más personas por cuarto, en áreas rurales y urbanas.	7,4	
Tenencia insegura de la Vivienda	Hogares que i) habitan viviendas ocupadas ilegalmente, o ii) residen en viviendas cedidas o prestadas.	7,4	
SERVICIOS BÁSICOS		22,2	
Carencia de fuentes de agua mejoradas	Áreas Urbanas: Hogares que obtienen agua de alguna de las siguientes fuentes: Red Pública fuera del terreno; Pozos no protegidos o sin bomba a motor; Fuentes Móviles(aljibe,carro tanque,aguatero, entre otros);	7,4	
	Áreas Rurales: Hogares que obtienen agua de alguna de las siguientes fuentes: Pozos no protegidos o con bomba manual; Fuentes Móviles(aljibe,carro tanque,aguatero, entre otros); Agua Embotellada, o río,quebrada , lluvia y otros		
	Áreas Urbanas: Hogares que obtienen agua de alguna de las siguientes fuentes: con evacuación no conectada a red de alcantarillado o fosa séptica;; con baño compartido, o que no disponen de servicio higiénico;		7,4
	Áreas Rurales: Hogares que obtienen agua de alguna de las siguientes fuentes: que no disponen de servicio higiénico; con baño compartido, o con evacuación sin tratamiento a la superficie, río o mar;		
Carencia de Energía	Hogares que no tienen servicio eléctrico o que usan leña, carbón o desechos como combustible para cocinar.	7,4	
ESTÁNDAR DE VIDA		22,2	
Insuficiencia de recursos	Hogares con ingresos per cápita insuficientes para cubrir sus necesidades alimentarias 14,8 y no alimentarias.	14,8	
Carencia de bienes duraderos	Hogares que no cuentan con ninguno de los siguientes bienes: i) vehículo, ii) refrigerador y iii) lavadora.	7,4	
EDUCACIÓN		22,2	
Inasistencia a Escuela	Hogares donde al menos un niño o adolescente (entre 6 y 17 años) no asiste a un establecimiento educativo.	7,4	
Rezago Escolar	Hogares donde al menos un niño o adolescente (entre 6 y 17 años) está rezagado en el sistema educativo en más de dos años de acuerdo a su edad.	7,4	
Logro educativo infuficiente	Hogares donde ninguna persona de 20 años o más alcanzó un nivel educativo mínimo, entendiéndose por ello lo siguiente: Personas de entre 20 y 59 años: no cuentan con el primer ciclo de la educación Secundaria completo, y	7,4	
	Personas de 60 años o más: no cuentan con educación primaria completa.		
EMPLEO Y PROTECCIÓN SOCIAL		11,1	
Desocupación	Hogares donde al menos una persona de entre 15 y 65 años de edad está en alguna de las siguientes situaciones: desempleada; empleada sin remuneración, o es un trabajador desalentado	7,4	
	Hogares donde se cumplen todas las siguientes situaciones: ninguna persona cuenta con algún tipo de seguro de salud contributivo;		
	ninguna persona está afiliada a un sistema de previsión social contributivo, y ninguna persona tiene ingresos por pensiones o jubilaciones.		
Precariedad de la protección social		3,7	

Fuente: Elaboración por Francisco Javier Parra a partir de Panorama Social de América Latina 2014. UN. CEPAL. 2014. Santiago de Chile: Naciones Unidas, 2015. Panorama Social de América Latina. ISSN: 1020-515.

¹³ Cabe destacar que el Índice de Pobreza Multidimensional (IPM) diseñado por la ONU, es también una medida de privaciones que pueden afectar las condiciones de vida de determinadas personas, básicamente mostrando en aquel índice la índole y la intensidad de pobreza a nivel individual en tres aspectos básicos: educación, salud (sanidad) y nivel de vida en 10 indicadores.

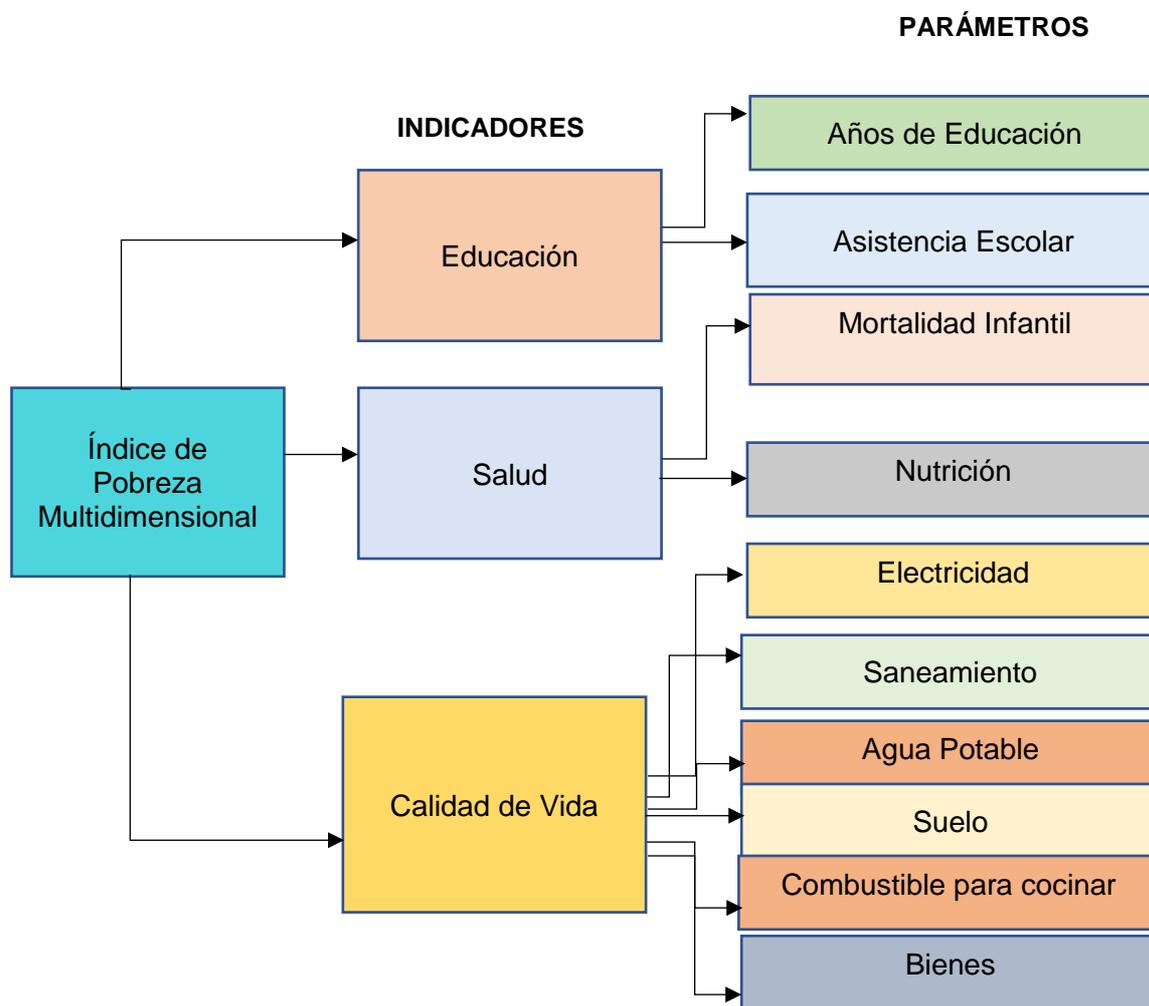
El Índice de Pobreza Multidimensional se recopila en base a 10 parámetros o indicadores agrupados en esas 3 dimensiones, cada dimensión tiene el mismo peso dentro del indicador (1/3), así como cada parámetro dentro de la dimensión. La ponderación de los parámetros depende de la dimensión a la que pertenecen, los de educación y salud se ponderarán con 1/6 y los de nivel de vida con 1/18.

- **Educación** (ponderación de los parámetros 1 y 2 de 1/6).
 - Años de educación: Si ningún miembro del hogar ha completado cinco años de nivel educativo.
 - Niños escolarizados: Al menos un niño en edad escolar en el hogar no asiste a la escuela hasta el grado 8.
- **Asistencia Sanitaria-Salud** (ponderación de los parámetros 3 y 4 de 1/6).
 - Mortalidad infantil: Si al menos un niño ha muerto en la vivienda.
 - Nutrición: Si al menos un adulto o niño en la vivienda se encuentra desnutrido.
- **Calidad de Vida** (ponderación de los parámetros 5 al 10 de 1/18).
 - Electricidad: Sin acceso a Servicio de Electricidad.
 - Saneamiento: Sin acceso en la vivienda a un baño con condiciones suficientes o si su baño es compartido con otros hogares (según la definición MDG).
 - Agua potable: Si en la vivienda no tiene acceso a agua potable o la misma está a más de 30 minutos caminando desde el hogar ida y vuelta (Definición MDG).
 - Suelo: Si el piso de la vivienda es de tierra, arena o estiércol.
 - Combustible para cocinar: Si el combustible principal de uso en la vivienda es de leña, carbón o estiércol.
 - Bienes: Si en la vivienda no se dispone de uno de los siguientes artículos: radio, televisión, teléfono, bicicleta, moto, o refrigerador, y no posee un automóvil o camioneta.

¹³ (OPHI, 2018)

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

Diagrama 3.2. Componentes del IPM



Fuente: Elaboración Propia a partir de los Indicadores proporcionados por el Programa de las Naciones Unidas para el desarrollo.

4 METODOLOGÍA

4.1 EL MATCHING ESTADÍSTICO: METODOLOGÍA PARA EL CÁLCULO DE LA TASA DE POBREZA EN ECUADOR.

Partiendo de los orígenes de la metodología aplicada en el estudio , el cual cuenta con una definición más reciente por parte de científicos informáticos alrededor de la década de los 90, cuya finalidad se orienta en el aprendizaje computacional por medio de la inteligencia artificial.

El Matching Estadístico esta basado en un enfoque de modelos para proporcionar información estadística conjunta con variables e indicadores recopilados a través de dos o más fuentes, cuyas unidades de análisis provienen de una misma población y poseen variables en común pero no se superponen.

Básicamente se trata de un proceso de imputar a unos individuos (receptores) una determinada información para algunas variables a partir de la recopilación de aquella información proveniente de otros individuos (donantes), a los que se les han observado algunas características comunes y que se relacionan con la información que se quiere estimar (Leulescu y Agafitei, 2013).

Las diferentes bases de datos empleadas se sitúan en relación a través de una serie de variables comunes, que deben cumplir algunos requisitos básicos en cuanto a su distribución y deben ser similares en su estructura.

La fusión de ficheros, parte del respectivo estudio del conjunto de datos, determinando aquellas variables comunes, junto con una armonización de las mismas, realizando una recodificación con la finalidad de que haya una coincidencia de ambos parámetros en cada base, utilizando el cuestionario respectivo tanto del CENSO como el de la ECV.

La metodología solo será implementada para aquellas variables en las que puedan armonizarse, tal que presenten una similitud en su definición, descartando aquellas en la que no es posible , por lo que al compartir una misma población se requiere que sigan una distribución similar.

Seguido, se busca generar un modelo a partir de esas variables comunes para tratar de estimar el índice en cada provincia junto con nuestra variable explicativa, aquella que ha venido incluida en la ECV-2014 , descartando su inclusión del CENSO.

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

Para la correcta aplicación del Matching Estadístico entre la ECV-2014 y el CENSO-2010, siguiendo las recomendaciones metodológicas se han establecido los siguientes pasos:

- Estudio de las bases de datos e identificación de variables comunes a ambos conjuntos de datos.
- Recodificación y armonización de las variables comunes cuando ha sido necesario. Evaluación individual de dichas variables, comprobando que su distribución es similar.
- Formulación de diferentes modelos explicativos o funciones de enlace de la variable objetivo Y en la base de datos de la ECV (2014).
- Evaluación de la capacidad explicativa de cada modelo y elección del más adecuado.
- Estimación de la variable objetivo en el CENSO NACIONAL DE POBLACIÓN Y VIVIENDA a partir del modelo seleccionado.

4.1.1 Variables Comunes entre el Censo de Población y Vivienda 2010 con la Encuesta de Condiciones de Vida del año 2014 para Ecuador.

Partiendo del estudio para la realización del correspondiente Matching Estadístico en ambas encuestas, se han seleccionado un conjunto de variables comunes (covariables), teniendo en cuenta las dimensiones expuestas anteriormente, las cuales pueden ser de gran uso con la finalidad de obtener la respectiva tasa de pobreza en los hogares a nivel provincial en Ecuador.

- Sexo del Jefe(a) del Hogar.
- Estado Civil del Jefe(a) del Hogar.
- Origen Etnico del Jefe(a) del Hogar.
- Tipo de Vivienda: Casa o Villa/Departamento/Mediagua/Choza/Covacha.
- Vía de Acceso Principal a la Vivienda: Calle: Adoquinada/Empedrada/Tierra.
- Material Predominante del Techo, Paredes o Piso de la Vivienda.
- Estado del Material predominante del Techo, Paredes o Piso de la Vivienda.
- Número de dormitorios que posee la Vivienda.
- Combustible principal para cocinar.
- Principial medio para obtener Agua la vivienda.
- Principial medio para obtener Energía Eléctrica la vivienda.
- Acceso de la vivienda a Servicio de Telefonía Convencional.
- Acceso de la vivienda a Servicio de Internet.
- Acceso de la vivienda a Servicio de Televisión por Cable.
- Tipo de Conexión de red del Servicio Higienico.
- Nivel de Instrucción del Jefe(a) del Hogar.

4.1.2 Tratamiento de los Datos

Una vez realizado el correspondiente análisis de las variables comunes a los ficheros de hogares de la ECV-2014 y del CENSO-2010. Un primer paso corresponde a la respectiva armonización de las definiciones de dichas variables tanto en la encuesta como para el censo. Algunos de los criterios sobre los que se ha basado la armonización se especifican en la siguiente tabla.

TABLA 4.1. Armonización de variables comunes identificadas en la ECV-2014 junto con el CENSO-2010.

Variables Comunes en ECV-2014 y CENSO 2010	Estructura
TIPO DE VIVIENDA	<p>Recodificada para armonizarla con la ECV 2014. Finalmente cuenta con 7 Niveles:</p> <ol style="list-style-type: none"> 1.- Casa o Villa. 2.- Departamento en Casa o Edificio. 3.-Cuarto (s) en casa de inquilinato. 4.-Mediagua. 5.-Rancho. 6.- Covacha. 7.- Choza.
ACCESO PRINCIPAL A LA VIVIENDA	<ol style="list-style-type: none"> 1.- Calle o carretera adoquinada, pavimentada o de concreto. 2.- Calle o carretera empedrada. 3.-Calle o carretera lastrado o de tierra . 4.-Camino, sendero, chaquiñán. 5.-Río / mar / lago. . 6.- Otro
MATERIAL DEL TECHO	<ol style="list-style-type: none"> 1.-Hormigón (losa, cemento). 2.- Asbesto (eternit, eurlolit). 3.-Zinc. 4.-Teja. 5.-Palma, Paja u Hoja. 6.- Otros Materiales.
MATERIAL DE LAS PAREDES	<ol style="list-style-type: none"> 1.-Hormigón. 2.- Ladrillo o bloque. 3.-Adobe o tapia 4.-Madera. 5.-Caña revestida o bahareque. 6.- Caña no revestida. 7.-Otros materiales
COMBUSTIBLE PARA COCINAR...	<p>Recodificada para armonizarla con la ECV 2014. Finalmente cuenta con 4 Niveles</p> <ol style="list-style-type: none"> 1.- Gas. 2.-Electricidad . 3.-Leña o Carbon. 4.-Otro.
SERVICIO HIGIENICO O ESCUSADO DE LA VIVIENDA	<p>Recodificada para armonizarla con la ECV 2014. Finalmente cuenta con 5 Niveles</p> <ol style="list-style-type: none"> 1.- Conectado a red pública de alcantarillado . 2.-Conectado a pozo séptico. 3.-Conectado a pozo ciego. 4.-Letrina. 5.-No Tiene.

Fuente: Elaboración Propia en base al cuestionario proporcionado por la ECV-2014 y el Censo-2010

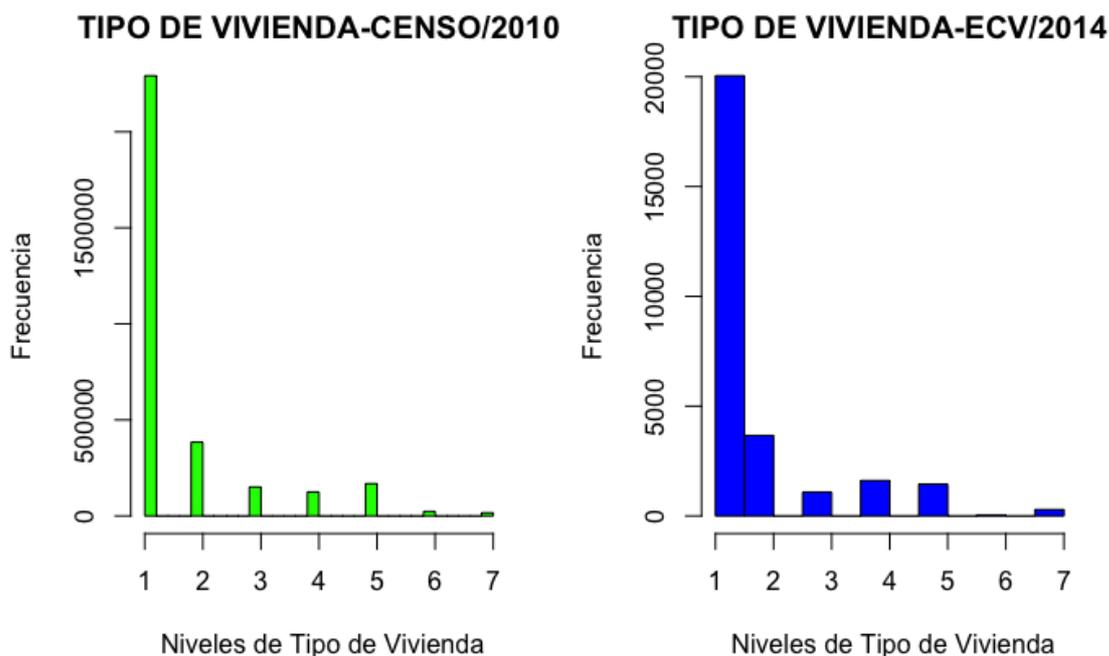
ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

Cabe destacar que algunas variables como en la cuestión, en caso de desempleo, el Jefe de Hogar a qué se dedica si es Rentista, Estudiante, Jubilado etc. fue omitida del presente análisis debido a una mayor proporción de Datos Nulos, tanto para la ECV como para el Censo.

En el respectivo proceso de recodificación, se incurrió en varias transformaciones de las variables de la ECV para que sigan el esquema de las variables con respecto al Censo, así como la generación de variables nuevas. Por una parte, se han creado un conjunto de variables tipo dummy, como ETNIA1 que indica si el jefe de hogar es de origen étnico Afrodescendiente o Afroecuatoriano, así como otras de tipo cuantitativa que indica el Número de personas en edad de trabajar del hogar que no se encuentran en el mercado laboral, por otra parte también se ha creado una variable que indique el Número de niños y niñas en edad de entre 5 a 14 años en la vivienda que no asisten a un centro de Educación Básica, Número de personas en el Hogar que cuentan con algún Nivel Educativo, entre otras variables creadas en el programa estadístico R.

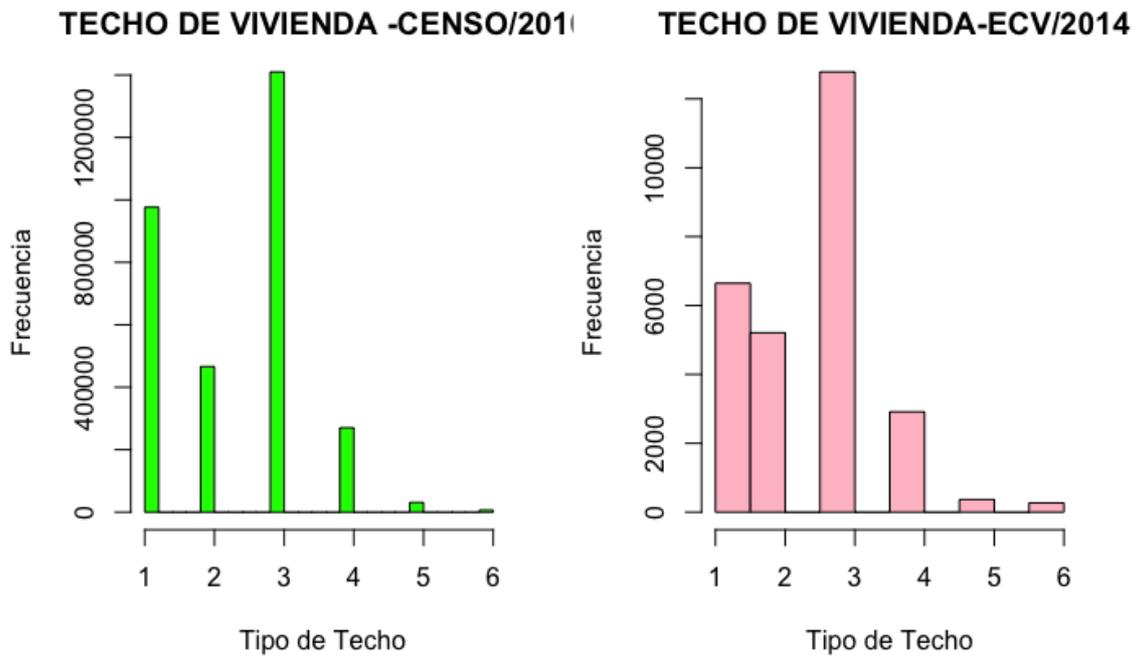
Un requisito particular del presente Matching Estadístico, se fundamenta en que dichas variables seleccionadas deben distribuirse de manera similar en ambos conjuntos de datos. Se evaluó la similitud entre variables elegidas para nuestro estudio a través de un respectivo análisis gráfico.

GRÁFICO 4.2. Histograma de Variables Armonizadas: Tipo de Vivienda



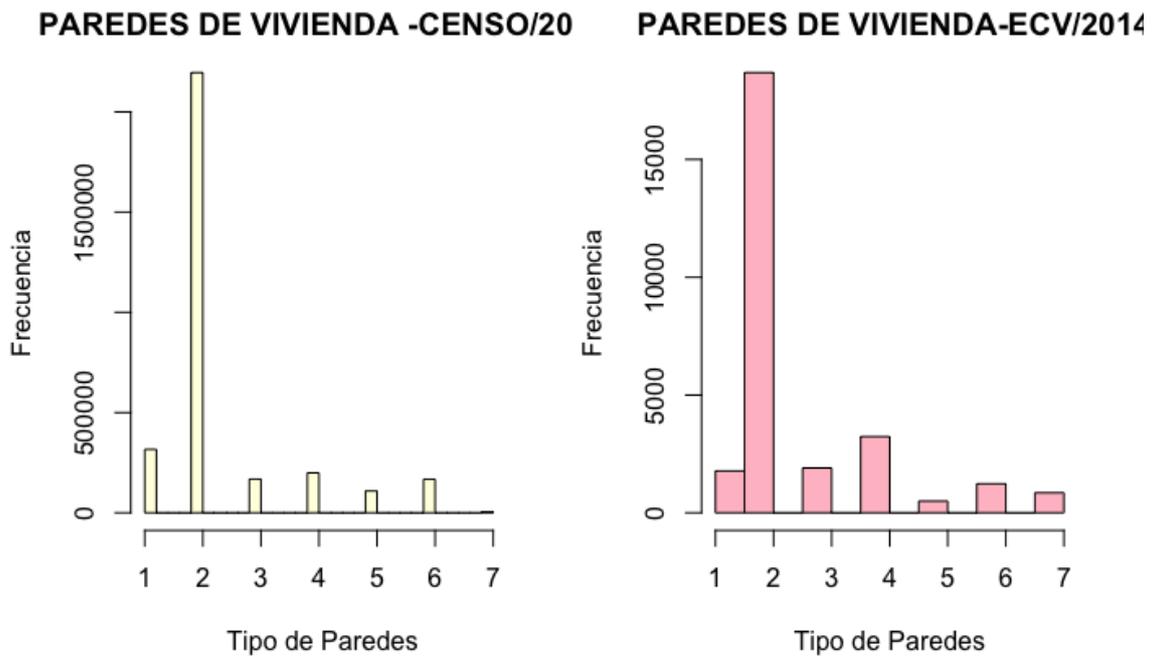
Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

GRÁFICO 4.3. Histograma de Variables Armonizadas: Techo de la Vivienda



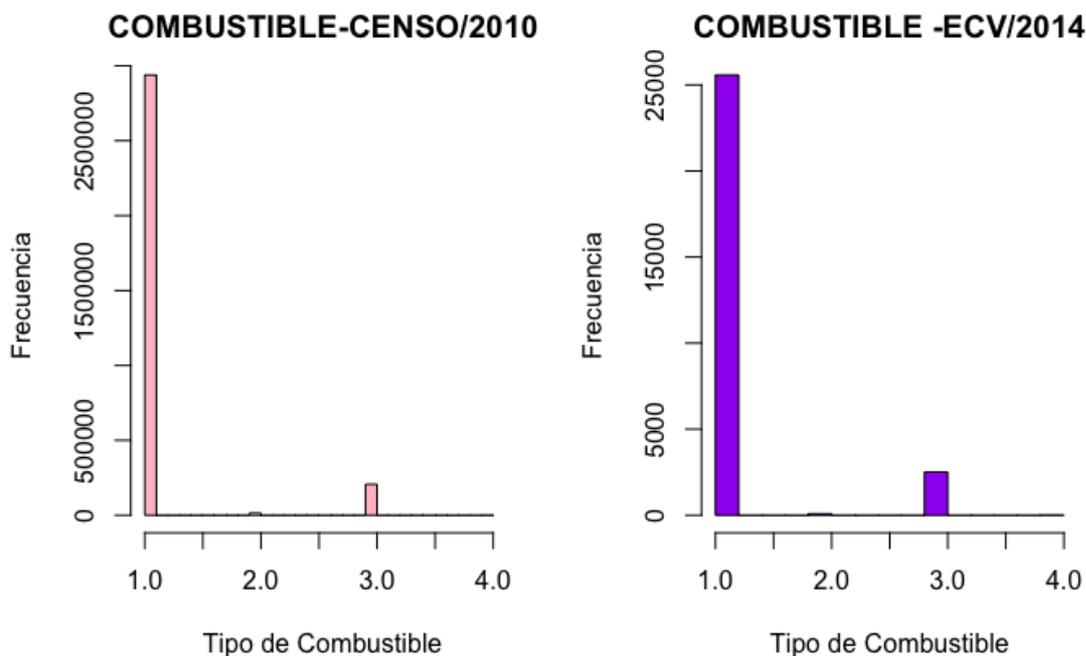
Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

GRÁFICO 4.4 Histograma de Variables Armonizadas: Paredes de Vivienda



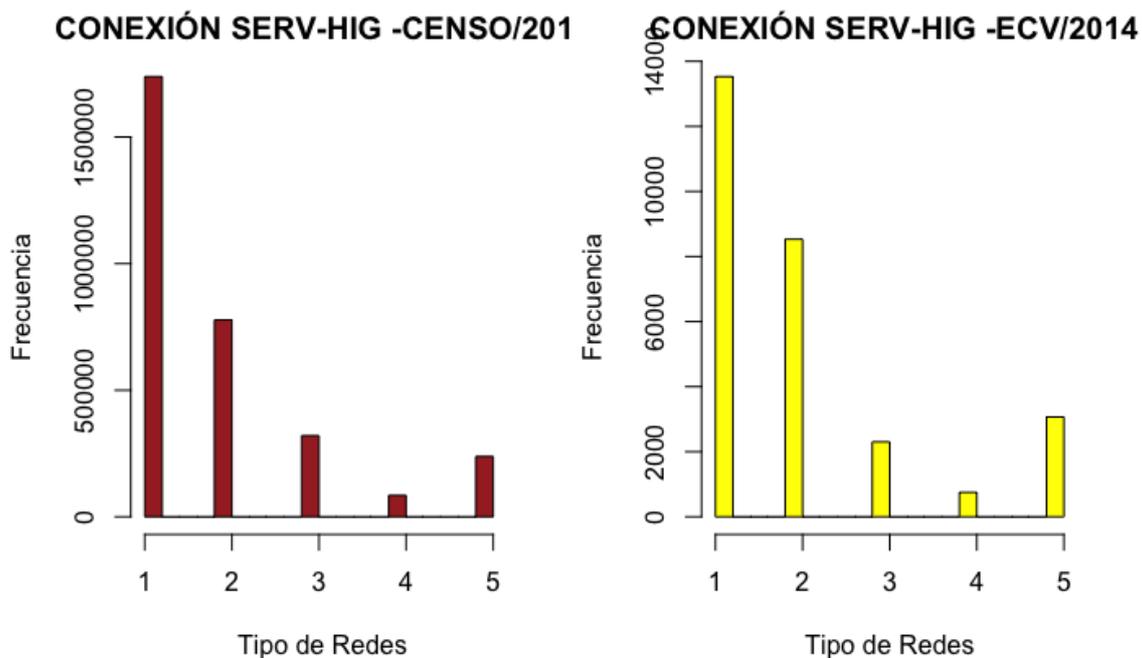
Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

GRÁFICO 4.5 Histograma de Variables Armonizadas: Tipo de Combustible



Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

GRÁFICO 4.6 Histograma de Variables Armonizadas: Tipos de Redes de SERV.HIG



Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

4.1.3 Selección de Variables y Aplicación de las Técnicas de Fusión.

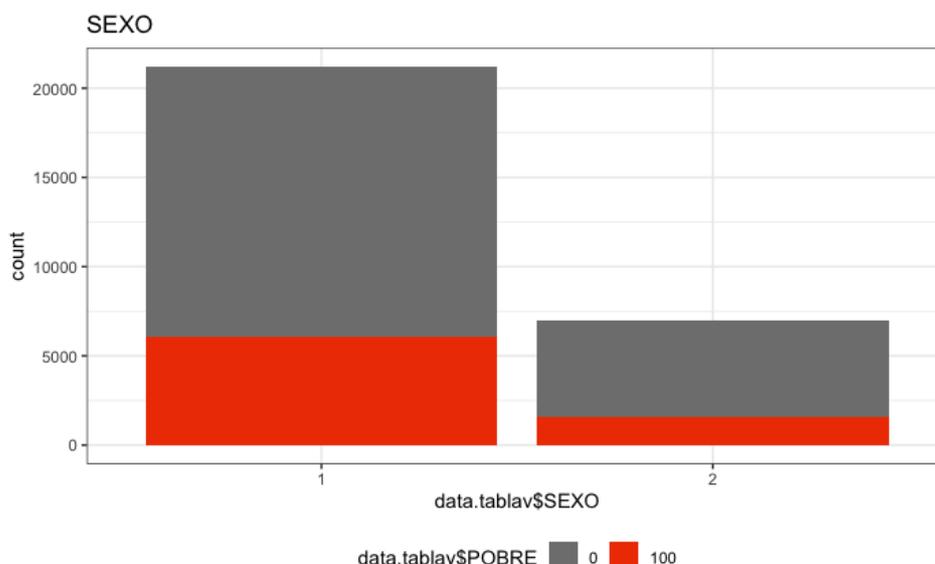
Para poder reducir el conjunto amplio de variables comunes se ha realizado un análisis basado en la preselección, utilizando un procedimiento automático, acompañado con una técnica estadística la cual trata de estudiar si la distribución de porcentajes con respecto a una determinada variable se repiten por igual en las categorías de la otra variable. Debido a que se estudia una variable en función de otra, por lo que mediante el análisis a mas diferencia de porcentaje en cada fila entre las categorías de las variables (en columna) mas fuerte es la relación entre las variables , caso contrario si no hay diferencias de respuestas es decir si los porcentajes en una misma fila son muy similares es que no hay relación entre dichas variables, se hizo el respectivo análisis para el conjunto de variables cualitativas que disponemos , las cuales se detallan a continuación:

SEXO DEL JEFE DE HOGAR

1. Masculino
2. Femenino

SEXO	0	100
1	0.72	0.28
2	0.78	0.22

GRÁFICO 4.7 Análisis Proporcional de Distancia entre Categorías de la Variable Sexo del Jefe(a) del Hogar.



Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

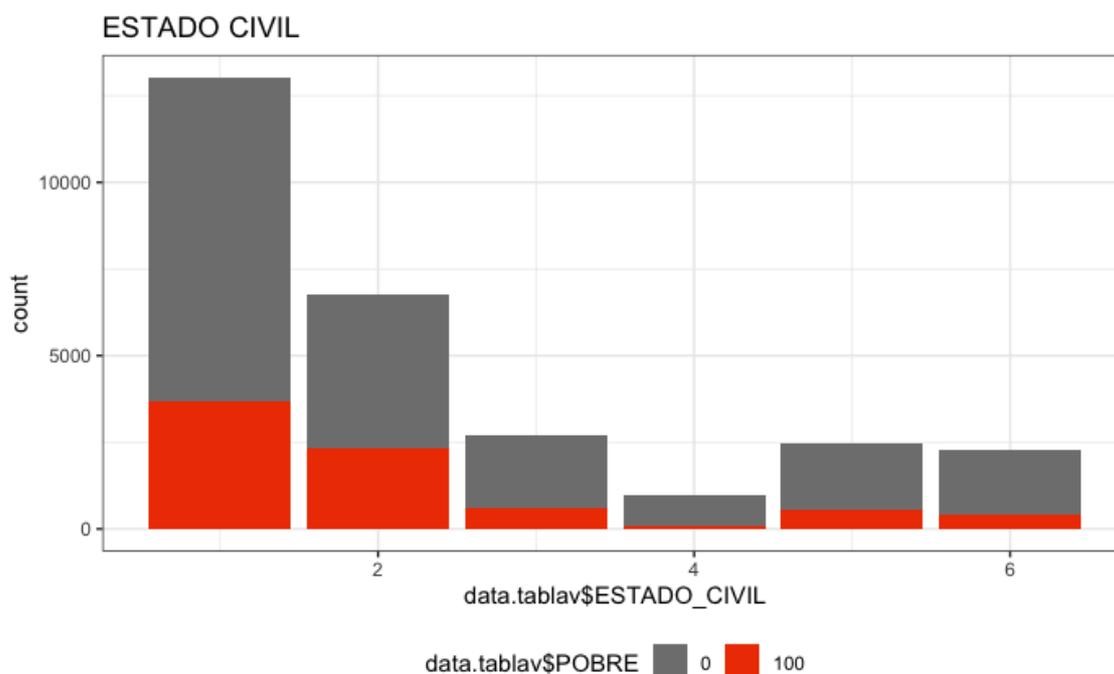
Se puede observar en cuanto a la distancia de las respectivas proporciones de hogares pobres y ricos de acuerdo al Sexo del Jefe de Hogar mediante el respectivo análisis que dicha variable no discrimina el ser rico o pobre debido a que presentan similitud de proporciones a nivel vertical.

ESTADO CIVIL DEL JEFE DE HOGAR

1. Casado/a
2. Unido/a
3. Separado/a
4. Divorciado/a
5. Viudo/a
6. Soltero/a

EST.CIVIL	0	100
1	0.72	0.28
2	0.66	0.34
3	0.78	0.22
4	0.93	0.07
5	0.78	0.22
6	0.81	0.19

GRÁFICO 4.8 Análisis Proporcional de Distancia entre Categorías de la Variable Estado Civil de Jefe(a) del Hogar.



Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

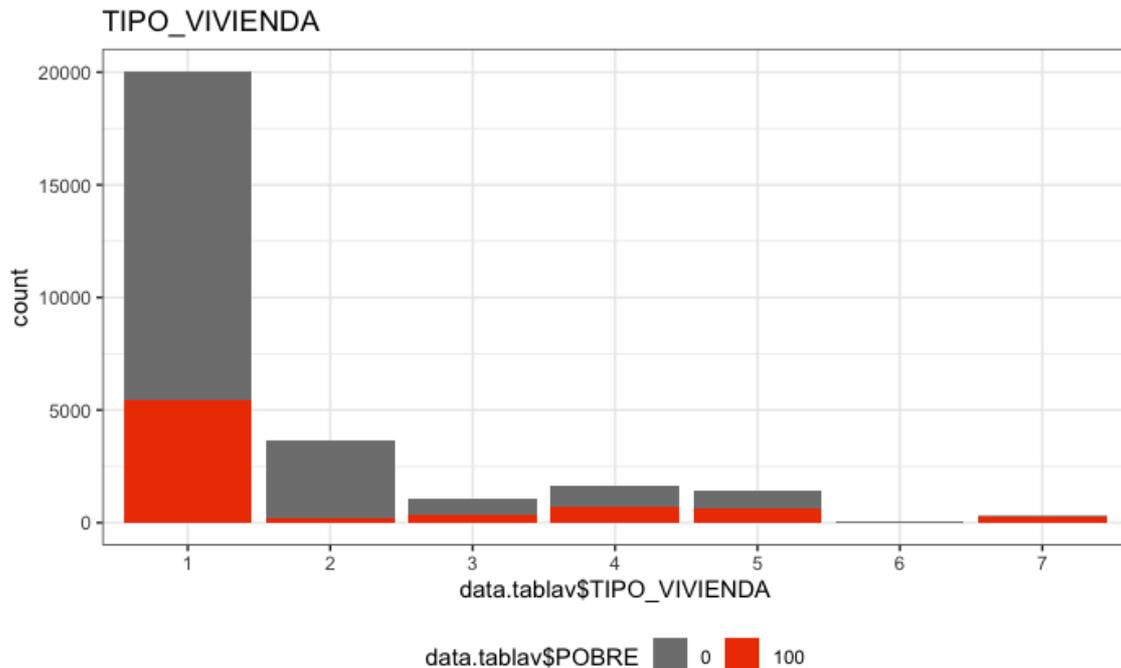
Se puede observar en cuanto a la distancia de las respectivas proporciones de hogares pobres y ricos de acuerdo al Estado Civil del Jefe de Hogar que el tipo: Unido/a, Separado/a, Divorciado/a, Viudo/a y Soltero/a en base a un análisis propio, se llega a la conclusión que si discrimina el ser rico o pobre debido a la no similitud que presenta.

TIPO DE VIVIENDA

1. Casa/Villa
2. Departamento en casa o edificio
3. Cuarto(s) en casa de inquilinato
4. Mediagua
5. Rancho
6. Covacha
7. Chozas

TIP.VIV	0	100
1	0.73	0.27
2	0.95	0.05
3	0.70	0.30
4	0.55	0.45
5	0.54	0.46
6	0.38	0.62
7	0.15	0.85

GRÁFICO 4.9 Análisis Proporcional de Distancia entre Categorías de la Variable Tipo de Vivienda.



Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

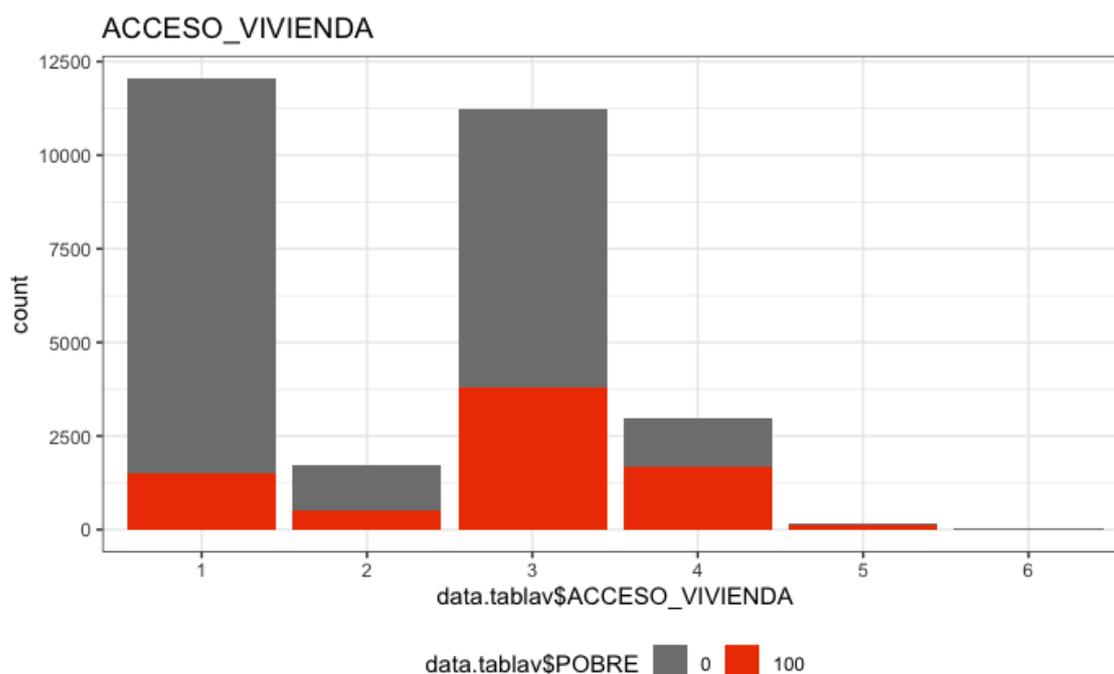
Se puede observar en cuanto a la distancia de las respectivas proporciones de hogares pobres y ricos de acuerdo al Tipo de Vivienda del Jefe de Hogar: Casa/Villa, Departamento en casa o edificio, Cuarto en casa de inquilinato, Mediagua, Rancho, Covacha, Choza en base a un análisis propio, se llega a la conclusión que si discrimina el ser rico o pobre debido a la no similitud que presenta.

ACCESO PRINCIPAL DE LA VIVIENDA

1. Carretera, calle pavimentada o adoquinada
2. Empedrado
3. Lastrado/Calle de Tierra
4. Sendero/chaquiñan
5. Río/Mar/Lago
6. Otro

ACCES.VIV	0	100
1	0.88	0.12
2	0.71	0.29
3	0.66	0.34
4	0.44	0.56
5	0.37	0.63
6	0.75	0.25

GRÁFICO 4.10 Análisis Proporcional de Distancia entre Categorías de la Variable Acceso Principal de Vivienda.



Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

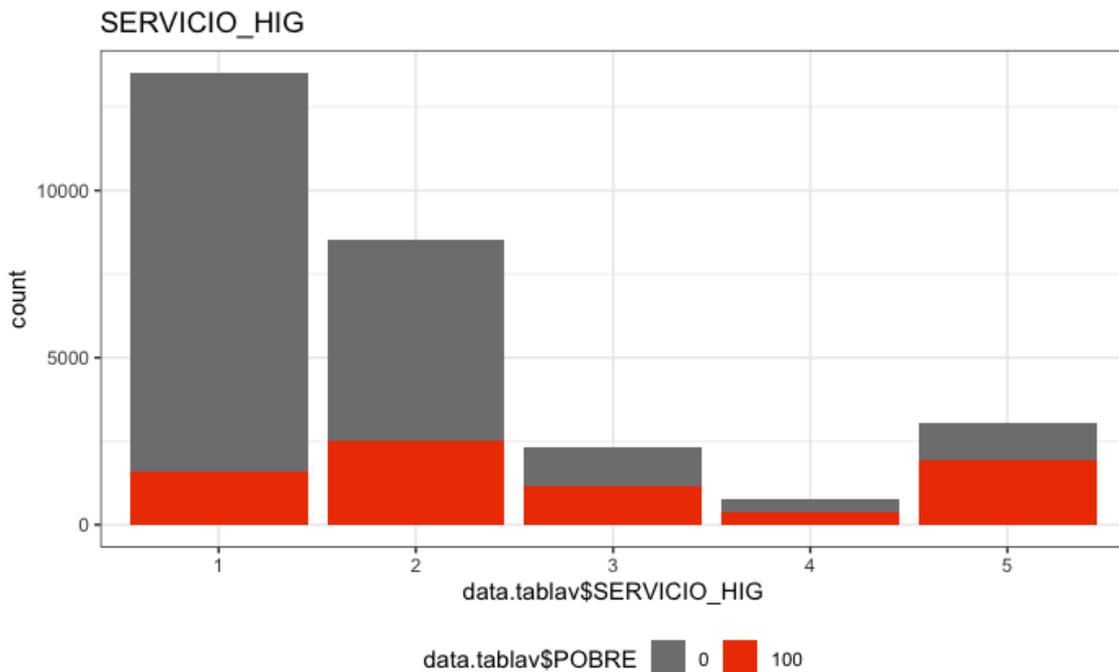
Se puede observar en cuanto a la distancia de las respectivas proporciones de hogares pobres y ricos de acuerdo al Acceso Principal de Vivienda: Carretera/calle pavimentada,Empedrado,Sendero,Rio,Otro en base a un análisis propio, se llega a la conclusión que si discrimina el ser rico o pobre debido a la no similitud que presenta.

CONEXIÓN A RED DE SERVICIO HIGIENICO

1. Inodoro y alcantarillado
2. Inodoro y pozo séptico
3. Inodoro y pozo ciego
4. Letrina
5. No tiene

SERV.HIG	0	100
1	0.88	0.12
2	0.71	0.29
3	0.50	0.50
4	0.47	0.53
5	0.36	0.64

GRÁFICO 4.11 Análisis Proporcional de Distancia entre Categorías de la Variable Conexión a Red de Servicio Higienico.



Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

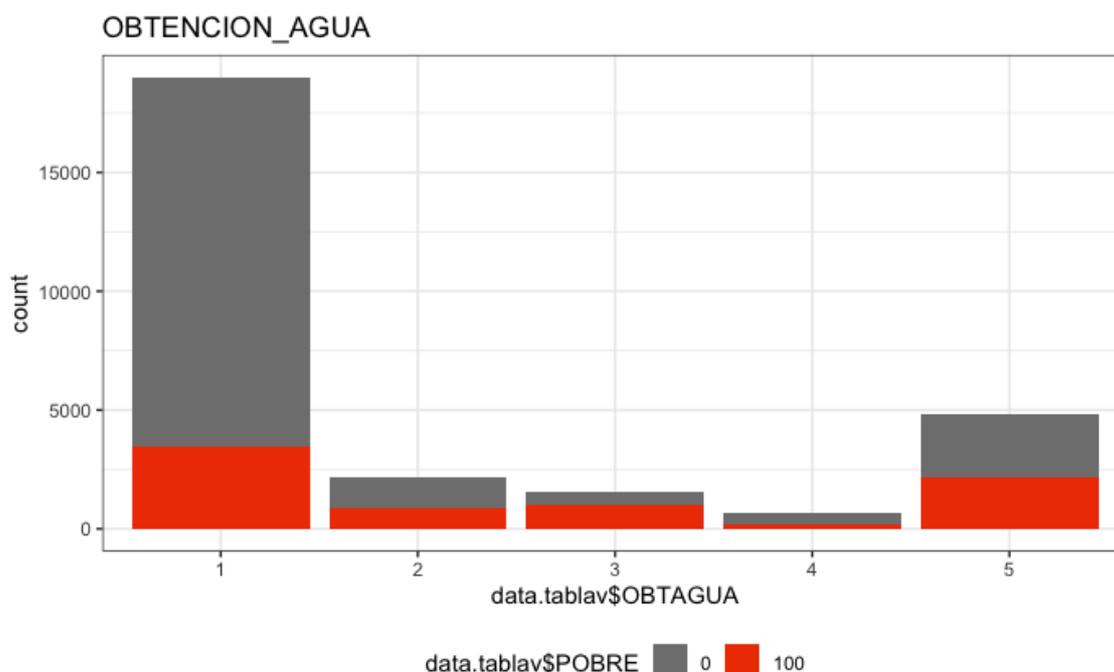
Se puede observar en cuanto a la distancia de las *respectivas* proporciones de hogares pobres y ricos de acuerdo a la conexión a red de Servicio Higiénico: Inodoro y alcantarillado, Inodoro y pozo séptico, Inodoro y pozo ciego, Letrina, No Tiene en base a un análisis propio, se llega a la conclusión que si discrimina el ser rico o pobre debido a la no similitud que presenta.

OBTENCIÓN DE SERVICIO DE AGUA PÚBLICA

1. Red Pública
2. Pozo
3. Río vertiente o acequia
4. Carro repartidor/triciclo
5. Otro.

OBT.AGUA	0	100
1	0.82	0.18
2	0.60	0.40
3	0.36	0.64
4	0.76	0.24
5	0.55	0.45

GRÁFICO 4.12 Análisis Proporcional de Distancia entre Categorías de la Variable Obtención de Servicio de Agua Potable.

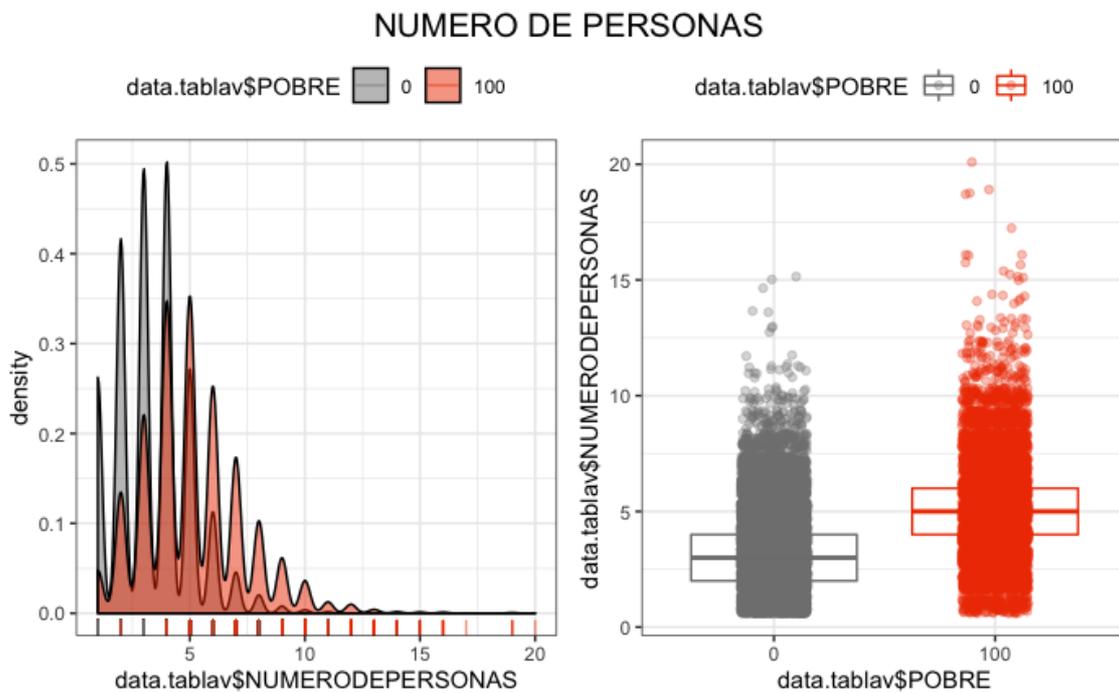


Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

Se puede observar en cuanto a la distancia de las respectivas proporciones de hogares pobres y ricos de acuerdo a la Obtención de Servicio de Agua Potable: Red Pública, Río vertiente o acequia, Otro en base a un análisis propio, se llega a la conclusión que si discrimina el ser rico o pobre debido a la no similitud que presenta

Para el respectivo análisis en la selección del conjunto de variables cuantitativas que influyen en ser pobre o no, se realiza el estudio de la distribución mediante diagrama de cajas para observar el comportamiento a través de sus dispersiones.

GRÁFICO 4.13 Representación Gráfica de Variable Número de Personas en el Hogar.

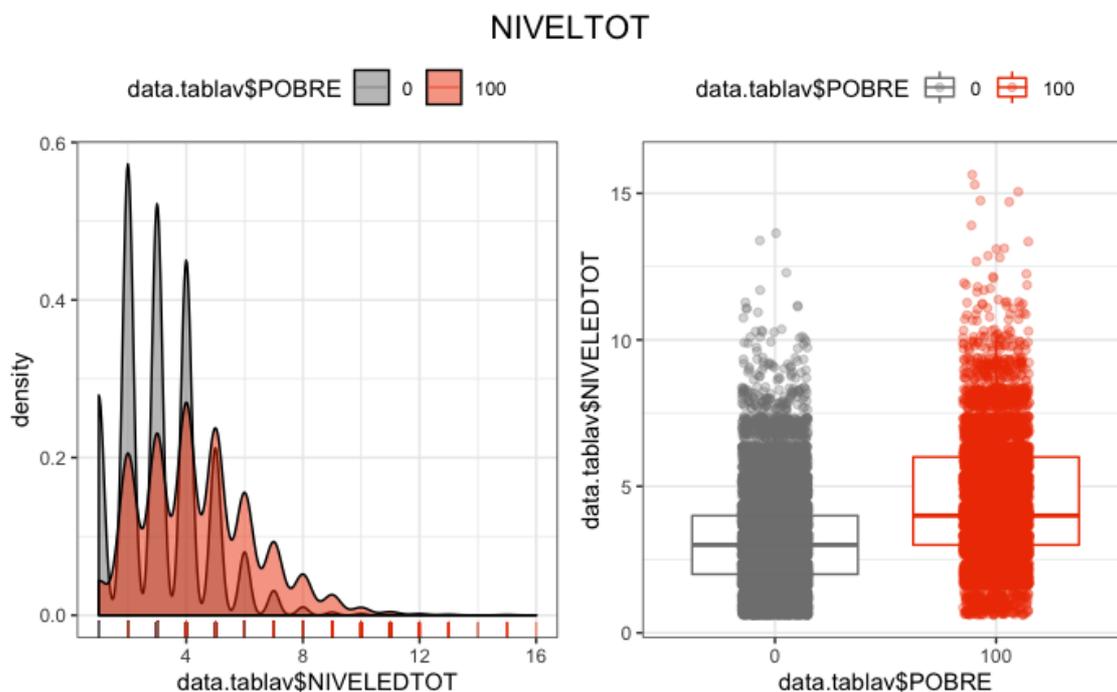


Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

Según el comportamiento del Gráfico se puede llegar a la conclusión que la variable NUMERODEPERSONAS está relacionada con la variable dependiente ya que presenta distintas dispersiones en el correspondiente diagrama de cajas..

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

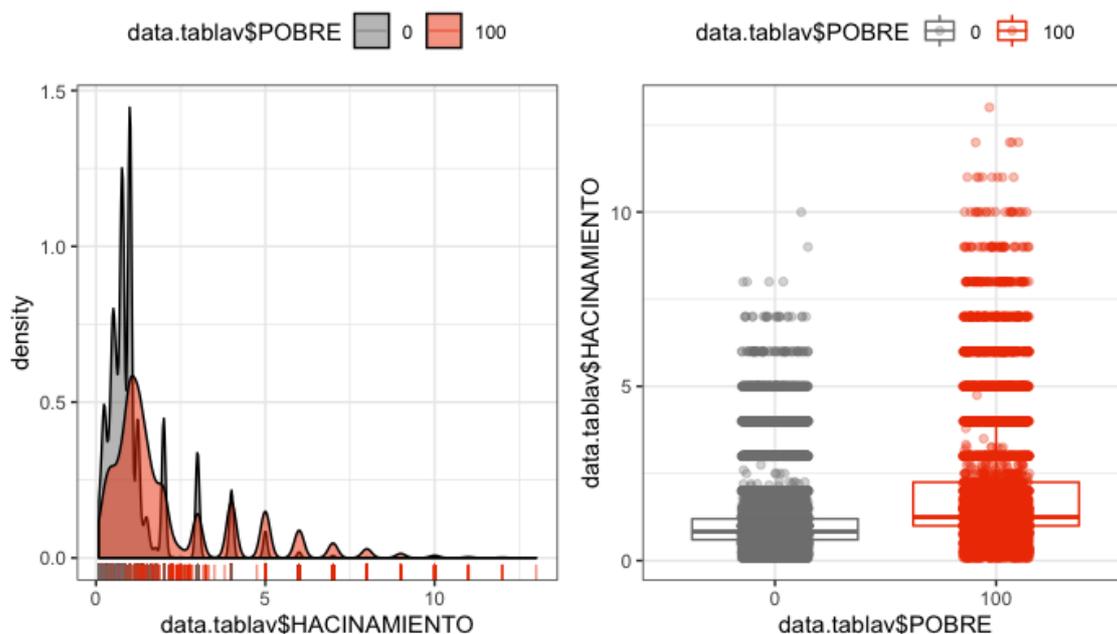
GRÁFICO 4.14 Representación Gráfica de Variable NIVELEDTOT: Total de personas en el Hogar que poseen algún Nivel Educativo .



Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

Según el comportamiento del Gráfico se puede llegar a la conclusión que la variable NIVELEDTOT está relacionada con la variable dependiente ya que presenta distintas dispersiones en el correspondiente diagrama de cajas.

GRÁFICO 4.15 Representación Gráfica de Variable HACINAMIENTO.



Fuente: Elaboración Propia en base al Cuestionario proporcionado por la ECV-2014 y el Censo-2010 mediante uso del Lenguaje de Programación Estadístico "R".

Según el comportamiento del Gráfico se puede llegar a la conclusión que la variable HACINAMIENTO está relacionada con la variable dependiente ya que presenta distintas dispersiones en el correspondiente diagrama de cajas.

Por lo que se ha seleccionado un conjunto de variables explicativas que formarán parte de nuestro respectivo modelo

SEXO: Dicotómica: Sexo del Jefe(a) de Hogar.

ESTADO_CIVIL1: Dicotómica: Estado Civil del Jefe(a) del Hogar Tipo: 2. Unido,3. Separado,4.Divorciado,5.Viudo,6.Soltero.

EDUCACION_SUP: Dicotómica: Nivel de Instrucción del Jefe de Hogar Tipo: 5.Secundaria,7.Educación Media,8.Post Bachillerato,9.Superior,10.Postgrado.

ETNIA1: Dicotómica: Etnia del Jefe de Hogar Tipo: 2.Afroecuatoriano o Afrodescendiente.

ACCESO_VIVIENDA1: Acceso a Vivienda de Tipo: 1.Carretera, Calle Pavimentada o Adoquinada ,2. Empedrado,6.Otro.

ACCESO_VIVIENDA2: Acceso a Vivienda de Tipo: 4.Sendero, 5.Rio/Mar/Lago

TIPO_VIVIENDA1: Dicotómica: Tipo de Vivienda: 1. Casa o Villa, 2.Departamento en casa o edificio,3. Cuarto en casa de inquilinato.

TIPO_VIVIENDA2: Dicotómica: Tipo de Vivienda: 4.Mediagua,5.Rancho, 6.Covacha, 7.Choza.

TECHO_1: Dicotómica: Techo de Tipo: 1.Hormigon/Losa o Cemento, 5.Palma/Paja/Hoja.

PAREDES_1: Dicotómica: Paredes de Tipo: 1.Hormigon.

PISO_1: Dicotómica: Piso de Tipo: 1.Duela/Parquet/Tabloncillo, 6 .Tierra.

OBTAGUA_1: Obtención del agua en la vivienda por tipo: 1.Red Pública.

OBTAGUA2: Obtención del agua en la vivienda por tipo: 3. Rio Vertiente o Acequia , 5.Otro.

SERV_HIG1: Conexión de red de Servicio Higienico: 1.Tipo Inodoro y Alcantarillado, 2. Inodoro y Pozo Séptico.

SERV_HIG2: Conexión de red de Servicio Higienico: 3.Tipo Inodoro y Pozo Ciego, 4.Letrina, 5. No Tiene.

COMBUSTIBLE1: Combustible que usa la vivienda para cocinar de tipo: 1. Gas, 2.Electricidad.

INTERNET: Dicotómica: Viviendas que disponen de Servicio de Internet.

TELF_CONV: Dicotómica: Viviendas que disponen de Servicio de Telefonía Convencional.

LECTOT: Total de personas en la vivienda que no saben leer ni escribir.

SERV_ELECT1: Dicotómica: Viviendas que disponen de Servicio Eléctrico mediante red de Energía Pública.

EDAD1: Dicotómica Edad del Jefe de Hogar de mayor o igual a 65 años.

NUMERODEPERSONAS: Total de personas que se encuentran en la vivienda.

NIVELEDTOT: Número de personas en la vivienda que dispone de algún Nivel Educativo.

HACINAMIENTO: Número de personas que son residentes en la vivienda / Número de Dormitorios disponibles.

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

4.1.4 Técnicas y Regularización del Mejor Modelo.

¹⁴Puede darse el caso de que alguna o muchas variables usadas en el modelo no estén asociadas con la variable dependiente lo cual provocaría una sub-especificación o sobre-especificación. Incluir variables irrelevantes añade complejidad innecesaria al modelo, por lo que el proceso de depuración de variables puede mejorar la interpretabilidad de los resultados.

Para ello, se realiza un ejercicio de selección automática con el método ¹⁵“stepAIC” en R el cual trata de buscar un modelo cuyo AIC sea pequeño, ya que en ese caso habría una verosimilitud muy grande y pocos parámetros.

TABLA 4.16. Resultados de Técnicas de Selección Automática del Mejor Modelo en R.

Step: AIC=20162.28

data.pobre1\$POBRE ~ AREA + SEXO + HACINAMIENTO + NUMERODEPERSONAS + NIVELEDTOT + EDUCACION_SUP + ACCESO_VIVIENDA1 + ACCESO_VIVIENDA2 + TIPO_VIVIENDA1 + TECHO_1 + PAREDES_1 + PISO_1 + OBTAGUA_1 + OBTAGUA_2 + SERV_HIG1 + COMBUSTIBLE1 + INTERNET + TELF_CONV + SERV_ELECT1 + LECTOT

	Df	Deviance	AIC
<none>		20120	20162
+ ETNIA1	1	20119	20163
+ ESTADO_CIVIL1	1	20120	20164
+ EDAD1	1	20120	20164
- SERV_ELECT1	1	20132	20172
- PISO_1	1	20133	20173
- ACCESO_VIVIENDA2	1	20136	20176
- ACCESO_VIVIENDA1	1	20142	20182
- SEXO	1	20145	20185
- OBTAGUA_1	1	20154	20194
- TECHO_1	1	20157	20197
- PAREDES_1	1	20159	20199
- OBTAGUA_2	1	20190	20230
- TIPO_VIVIENDA1	1	20193	20233
- EDUCACION_SUP	1	20233	20273
- HACINAMIENTO	1	20255	20295
- NIVELEDTOT	1	20295	20335
- SERV_HIG1	1	20300	20340
- COMBUSTIBLE1	1	20318	20358
- LECTOT	1	20328	20368
- AREA	1	20415	20455
- TELF_CONV	1	20459	20499
- INTERNET	1	20492	20532
- NUMERODEPERSONAS	1	21221	21261

¹⁴ (Gil, 2018)

¹⁵ La función stepAIC extrae el AIC asociado a los diferentes modelos que resultan de añadir o quitar del modelo actual cada una de las variables, y se elige el modelo cuyo valor de AIC sea menor. El proceso se repite hasta que la inclusión o eliminación de alguna variable incremente el valor del AIC en vez de reducirlo.

Call: glm(formula = data.pobre1\$POBRE ~ AREA + SEXO + HACINAMIENTO + NUMERODEPERSONAS + NIVELEDTOT + EDUCACION_SUP + ACCESO_VIVIENDA1 + ACCESO_VIVIENDA2 + TIPO_VIVIENDA1 + TECHO_1 + PAREDES_1 + PISO_1 + OBTAGUA_1 + OBTAGUA_2 + SERV_HIG1 + COMBUSTIBLE1 + INTERNET + TELF_CONV + SERV_ELECT1 + LECTOT, family = "binomial", data = data.pobre1)

Coefficients:

(Intercept)	AREA	SEXO	HACINAMIENTO	NUMERODEPERSONAS
-0.7175	-0.7695	0.2135	0.1588	0.9017
NIVELEDTOT	EDUCACION_SUP	ACCESO_VIVIENDA1	ACCESO_VIVIENDA2	TIPO_VIVIENDA1
-0.3925	-0.4408	-0.1932	0.2121	-0.4341
TECHO_1	PAREDES_1	PISO_1	OBTAGUA_1	OBTAGUA_2
-0.3132	-0.6061	0.1917	0.3381	0.4958
SERV_HIG1	COMBUSTIBLE1	INTERNET	TELF_CONV	SERV_ELECT1
-0.5988	-0.8704	-1.6425	-0.9998	-0.3250
LECTOT				
0.3757				
Degrees of Freedom: 28184 Total (i.e. Null); 28164 Residual				
Null Deviance: 32840				
Residual Deviance: 20120 AIC: 20160				

Fuente: Elaboración Propia en base al Método de Selección al Mejor Modelo haciendo uso de la función "stepAIC" mediante el Lenguaje de Programación Estadístico "R".

Por lo que en base a los datos obtenidos mediante la Selección Automática al Mejor Modelo se puede observar que el programa da como resultado aquel modelo con un menor valor AIC de 20160 en el cual se excluye la variable ETNIA1 junto con la variable ESTADO_CIVIL1, TIPO_VIVIENDA2, SERV_HIG2, EDAD1.

Dicho método aplica un Modelo Lineal Generalizado cuya utilidad se basa de extensión en cuanto a los modelos lineales que permiten utilizar distribuciones no normales de los errores (Binomiales, Poisson, Gamma, etc.) y varianzas no constantes.

En los modelos de respuesta Binaria o Dicotómica, se tiene que la variable de respuesta Y puede tomar dos valores, codificándolos usualmente como 1 para la categoría de interés y 0 para la otra.

La distribución de Y es una Bernoulli cuya esperanza es:

$$E [Y] = P [Y = 1] = p \quad (0 < p < 1)$$

Si tenemos una variable X, posible predictora de la variable Y, entonces la distribución condicional de Y sobre un valor de X = x, también sigue una distribución de Bernoulli de forma que la esperanza condicionada de Y sobre X = x es :

$$E [Y | X = x] = P [Y = 1 | X = x] = p(x)$$

y la varianza condicionada.

$$\text{Var}[Y | X = x] = p(x) \cdot (1 - p(x))$$

4.1.4.1 Métodos de Clasificación

La clasificación supervisada es una tarea que mas frecuentemente es llevada a cabo por los denominados Sistemas Inteligentes. Por lo tanto, un gran número de paradigmas desarrollados bien por la Estadística (Regresión Logística) o bien por la Inteligencia Artificial (Redes Neuronales, Árboles de Decisión) son capaces de realizar tareas propias de clasificación.

Dicho paso previo a aplicar un método de clasificación, se encarga de realizar una partición a un amplio conjunto de datos, en dos conjuntos de menor tamaño, que serán utilizados con la respectiva finalidad: entrenamiento y test.

El respectivo conjunto de datos que forman parte para el entrenamiento, se basa en la estimación de los parámetros del modelo y en cuanto al conjunto de datos de test se emplea para verificar como se comporta el modelo estimado.

Cabe recalcar que cada registro que forma parte de nuestra base de datos, debe aparecer en cada subconjunto utilizando un procedimiento de muestreo: muestreo aleatorio simple o muestreo estratificado, lo primordial es entrenar el respectivo modelo con un conjunto de datos independiente de los que hemos usado con el test

Como resultado de aplicar un método de clasificación, se incurrirán en 2 errores, en el posible caso de una variable binaria que toma valores 0 y 1, habrá valores que se clasifiquen incorrectamente como en el caso de que valores 0 se clasifiquen incorrectamente como uno y 1 que se clasifiquen incorrectamente como cero.

Para la implementación de los respectivos métodos utilizados se parte de una muestra de entrenamiento la cual recoge el 60% de los casos, mientras que la muestra de test recoge el 40% restante.

Los Métodos No Paramétricos utilizados en el presente trabajo se detallan a continuación.

4.1.4.1.1 Árboles de Decisión.

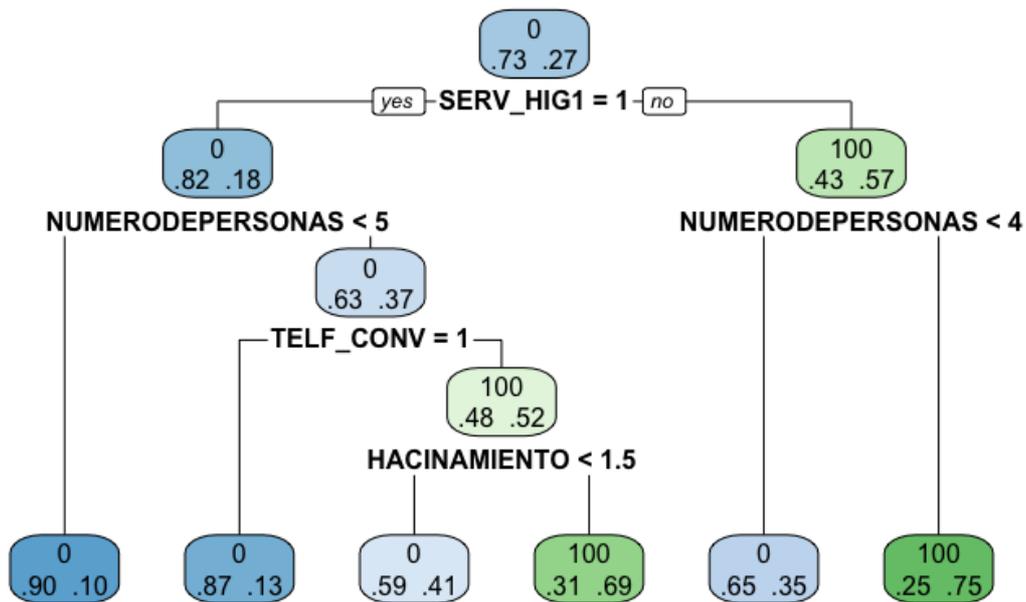
Dicho Método es apropiado cuando hay un número elevado de datos, siendo una de sus ventajas su carácter descriptivo que permite entender e interpretar fácilmente las decisiones tomadas por el modelo, revelando formas complejas en la estructura de datos que no se pueden detectar con los métodos convencionales de regresión.

Los Árboles de Decisión o de clasificación son un modelo surgido en el ámbito del aprendizaje automático (Machine Learning) y de la inteligencia Artificial que, partiendo de una base de datos, crea diagramas de construcciones lógicas que nos ayudan a resolver problemas.

A esta técnica también se la denomina segmentación jerárquica. Es una técnica explicativa y descomposicional que utiliza un proceso de división secuencial, iterativo y descendente que, partiendo de una variable dependiente, forma grupo homogéneos definidos específicamente mediante combinaciones de variables independientes en las que se incluyen la totalidad de los casos recogidos en la muestra.

Para la construcción del árbol se ha utilizado el package “rpart”. Recursive Partitioning and Regression Trees del respectivo programa estadístico R.

GRÁFICO 4.17. Árbol de Decisión



Fuente: Elaboración Propia mediante uso del package “rpart” en el Lenguaje de Programación Estadístico “R”.

De manera general, lo que se encarga de realizar en el presente algoritmo, es encontrar una variable independiente que mejor separa nuestros datos en determinados grupos, dicha separación puede ser expresada con una regla. A cada regla le corresponde un determinado nodo.

Dada nuestra variable objetivo la cual toma dos niveles: 0 (rico) y 100 (pobre). Se puede observar que la variable que mejor separa nuestros datos es SERV_HIG1 y la regla resultante es SERV_HIG1=1, esto se interpreta como el conjunto de datos para los que esta regla es verdadera, tienen mas probabilidad de pertenecer al grupo de hogares ricos caso contrario pertenecen al grupo de hogares pobres.

Una vez realizado este proceso, los datos son separados (particionados) en grupos a partir de la respectiva regla obtenida y así sucesivamente se repite la iteración hasta que no es posible obtener una mejor separación, dado este evento el algoritmo se detiene denominándolo nodo terminal u hoja.

Una de las principales ventajas de este método son su interpretabilidad, pues nos da un conjunto de reglas a partir de las cuales se pueden tomar decisiones. Este es un algoritmo que no tiene mucha demanda en poder de cómputo en comparación con procedimientos más sofisticados y, a pesar de ello, que tiende a dar buenos resultados de predicción para muchos tipos de datos.

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

TABLA 4.18. Matriz de Confusión

TestPredRpart	0	100
0	7668	1600
100	559	1447

Fuente: Elaboración Propia mediante uso del package "rpart" en el Lenguaje de Programación Estadístico "R".

Se puede observar en base a los resultados, del cual se reporta 7668 casos en los que se ha hecho la clasificación de la categoría rico de manera correcta y 1600 casos en los que se ha clasificado de manera errónea dicha categoría asignándole la categoría pobre y así sucesivamente para la otra categoría.

De lo que se obtuvo un 80.84% en cuanto a la Evaluación de la efectividad del modelo seleccionado.

4.1.4.1.2 Redes Neuronales

Las Redes Neuronales Artificiales son aquellos modelos computacionales inspirados en las neuronas biológicas, y que están conformadas por un conjunto de unidades de cómputo básico (neuronas) las cuales están conectadas entre ellas de múltiples maneras.

Estas conexiones estarán definidas por unos pesos los cuales determinarán la fuerza o importancia de dichas conexiones, y durante el proceso de aprendizaje o entrenamiento de la red, serán estos pesos los que se ajustarán con el fin de producir la salida adecuada según la entrada que se aplique a la red.

El objetivo es obtener la definición de red con mejor capacidad explicativa.

TABLA 4.19. Matriz de Confusión

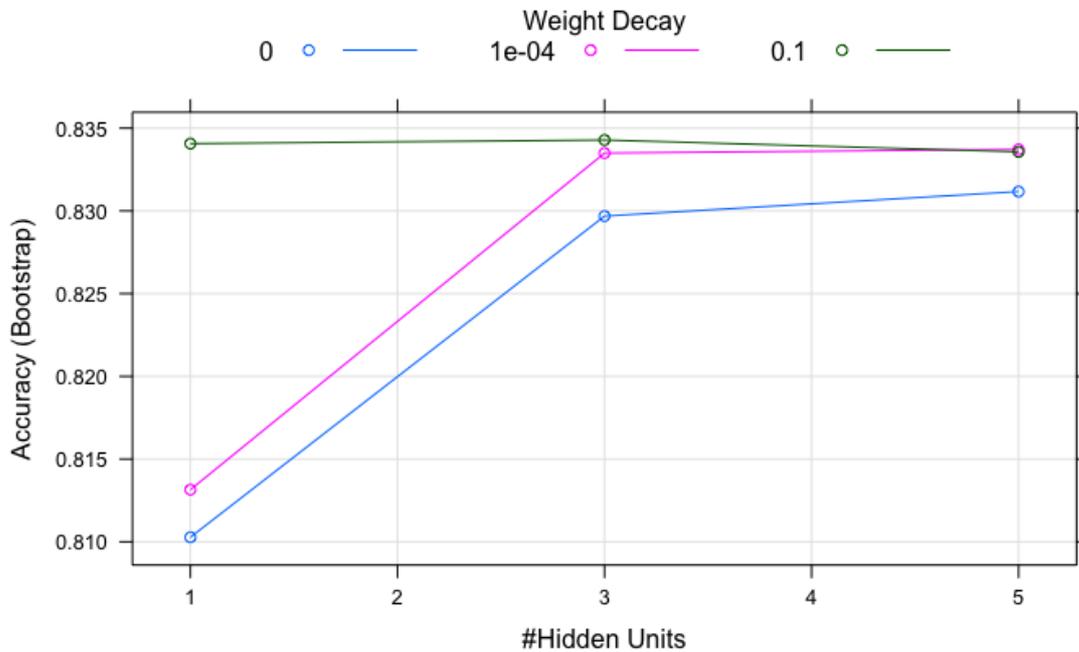
	Real	
Asignado	0	100
0	7495	1101
100	742	1935

Fuente: Elaboración Propia mediante uso del package "rpart" en el Lenguaje de Programación Estadístico "R".

Por lo que aplicando dicho método se obtuvo una asignación de 7495 de manera correcta correspondiente a la categoría de ser rico mientras una asignación de 1101 de manera errónea asignándole otra categoría.

Por lo que el modelo seleccionado arroja una efectividad del 83.65% de aciertos

GRÁFICO 4.20. Redes Neuronales



Fuente: Elaboración Propia mediante uso del package "neuralnet" en el Lenguaje de Programación Estadístico "R".

4.1.4.1.3 Curva de ROC

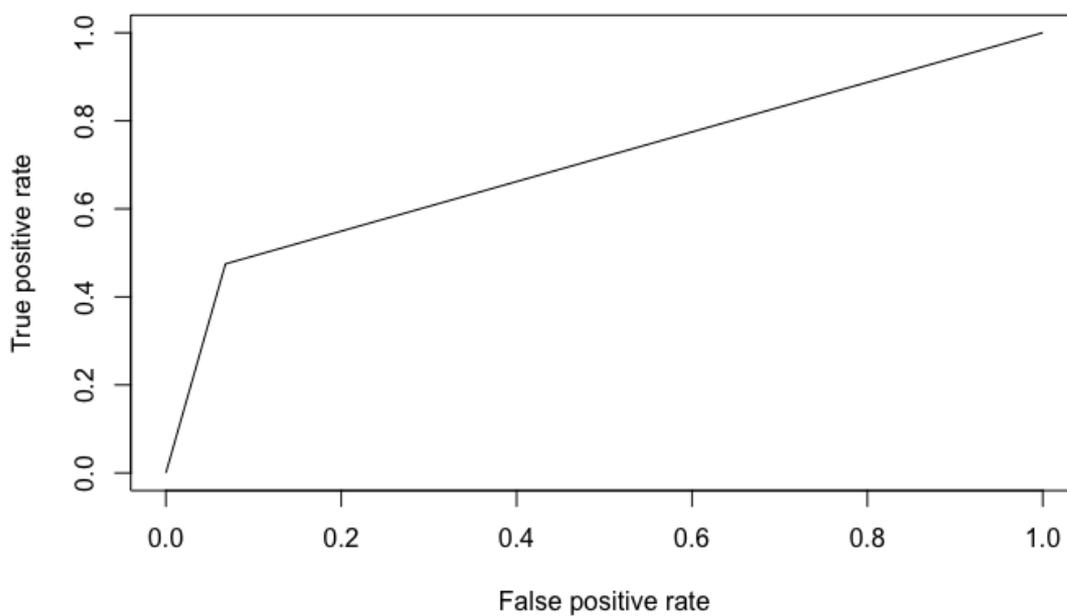
Un método para evaluar los clasificadores alternativos a la métrica expuesta es la Curva ROC (Receiver Operating Characteristic). La Curva ROC es una representación gráfica del rendimiento del clasificador que muestra la distribución de fracciones de verdaderos positivos y de falsos positivos.

En la Teoría de detección de señales, una curva ROC (acrónimo de Receiver Operating Characteristic, o Característica Operativa del Receptor) es una representación gráfica de la sensibilidad frente a la especificidad para un sistema clasificador binario según se varía el umbral de discriminación.

Otra interpretación de este gráfico es la representación de la razón o ratio de verdaderos positivos (VPR = Razón de Verdaderos Positivos) frente a la razón o ratio de falsos positivos (FPR = Razón de Falsos Positivos) también según se varía el umbral de discriminación (valor a partir del cual decidimos que un caso es un positivo).

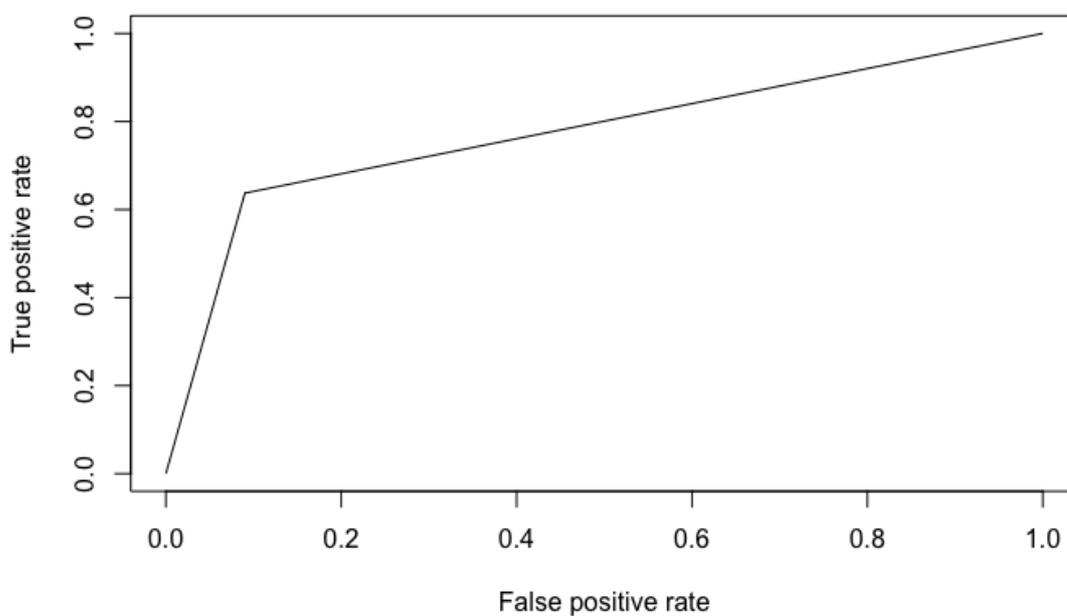
ROC también puede significar Relative Operating Characteristic (Característica Operativa Relativa) porque es una comparación de dos características operativas (VPR y FPR) según cambiamos el umbral para la decisión.

GRÁFICO 4.21. Curva de ROC para Árbol de Decisión.



Fuente: Elaboración Propia mediante uso del package "ROCR" en el Lenguaje de Programación Estadístico "R".

GRÁFICO 4.22. Curva de ROC para Redes Neuronales.



Fuente: Elaboración Propia mediante uso del package "ROCR" en el Lenguaje de Programación Estadístico "R".

Información contenida en la curva:

Si la prueba fuera perfecta para el estudio, es decir, sin solapamiento, hay alguna región en la que cualquier punto de corte tiene sensibilidad y especificidad iguales a 1: la curva sólo tiene el punto (0,1).

Caso contrario si la prueba fuera inútil: se muestra que la sensibilidad (verdaderos positivos) es igual a la proporción de falsos positivos, la curva sería la diagonal de (0,0) a (1,1).

En cuanto a los métodos paramétricos se han incurrido en:

4.1.4.1.4 Regresión Logarítmica.

El problema que se presentan en los modelos probabilísticos lineales en cuanto a la existencia de predicciones establecidas fuera rango (negativas o mayores que uno), es debido a que utilizan una función de probabilidad que depende linealmente de las variables explicativas (X), que se resolverían acotando dicha distribución de probabilidad. El modelo Logit en concreto utiliza, para ello, la función de distribución logística:

Debido a que la función de distribución logística no tiene forma lineal, el modelo Logit se estima de forma diferente, así en vez de minimizar las sumas de las diferencias al cuadrado entre los valores observados y los estimados por el modelo, el carácter no lineal de los modelos Logit requiere la utilización del método de Máxima Verosimilitud para ser estimado, maximizando la verosimilitud de que un suceso tenga lugar, aunque se podría estimar por MCO mediante una transformación logarítmica de los datos (Gujarati, 1997).

Por lo que se realiza un análisis exploratorio de las variables disponibles, realizando la respectiva regresión entre la variable que queremos explicar (POBRE) y el conjunto de covariables seleccionadas. Los resultados obtenidos de la regresión paso por paso mediante uso del programa estadístico R son los que se detallan a continuación:

ESTIMACIÓN DEL ÍNDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

TABLA 4.23. Resultados de la Aplicación de la Regresión Logarítmica en “R”.

Call:

```
glm(formula = POBRE ~ AREA + SEXO + HACINAMIENTO + NUMERODEPERSONAS +
  NIVELEDTOT + EDUCACION_SUP + ACCESO_VIVIENDA1 + ACCESO_VIVIENDA2 +
  TIPO_VIVIENDA1 + TECHO_1 + PAREDES_1 + PISO_1 + OBTAGUA_1 +
  OBTAGUA_2 + SERV_HIG1 + COMBUSTIBLE1 + INTERNET + TELF_CONV +
  SERV_ELECT1 + LECTOT, family = "binomial", data = data.pobre1)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8872	-0.5718	-0.2397	0.2354	3.4649

Coefficients:

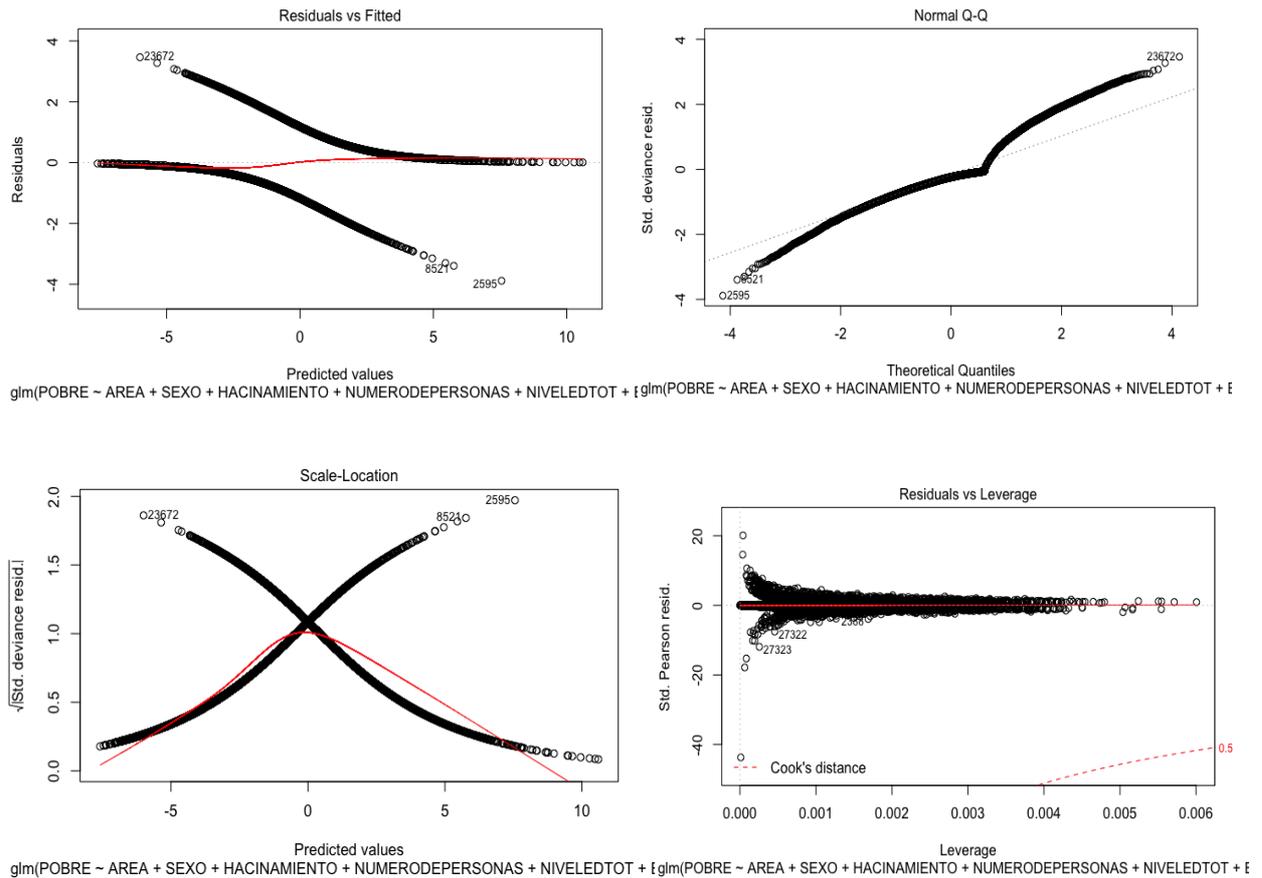
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.71751	0.13508	-5.312	1.08e-07 ***
AREA	-0.76950	0.04527	-16.997	< 2e-16 ***
SEXO	0.21350	0.04289	4.978	6.43e-07 ***
HACINAMIENTO	0.15885	0.01382	11.496	< 2e-16 ***
NUMERODEPERSONAS	0.90169	0.02814	32.040	< 2e-16 ***
NIVELEDTOT	-0.39246	0.02982	-13.162	< 2e-16 ***
EDUCACION_SUP	-0.44078	0.04178	-10.551	< 2e-16 ***
ACCESO_VIVIENDA1	-0.19316	0.04133	-4.673	2.97e-06 ***
ACCESO_VIVIENDA2	0.21214	0.05407	3.924	8.71e-05 ***
TIPO_VIVIENDA1	-0.43414	0.05082	-8.542	< 2e-16 ***
TECHO_1	-0.31321	0.05251	-5.965	2.45e-09 ***
PAREDES_1	-0.60609	0.10135	-5.980	2.23e-09 ***
PISO_1	0.19173	0.05368	3.571	0.000355 ***
OBTAGUA_1	0.33808	0.05890	5.740	9.48e-09 ***
OBTAGUA_2	0.49583	0.05968	8.308	< 2e-16 ***
SERV_HIG1	-0.59883	0.04460	-13.428	< 2e-16 ***
COMBUSTIBLE1	-0.87037	0.06224	-13.983	< 2e-16 ***
INTERNET	-1.64248	0.09669	-16.988	< 2e-16 ***
TELF_CONV	-0.99979	0.05703	-17.531	< 2e-16 ***
SERV_ELECT1	-0.32499	0.09503	-3.420	0.000626 ***
LECTOT	0.37569	0.02626	14.304	< 2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
 (Dispersion parameter for binomial family taken to be 1)
 Null deviance: 32843 on 28184 degrees of freedom
 Residual deviance: 20120 on 28164 degrees of freedom
 AIC: 20162
 Number of Fisher Scoring iterations: 6

Fuente: Elaboración Propia mediante uso del Modelo GLM en el Lenguaje de Programación Estadístico “R”.

Por lo que se puede observar en base a los resultados proporcionados de la regresión que variables como hogares que disponen de servicio de INTERNET, tienen probabilidad negativa de ser pobre, de la misma forma ocurre también con hogares que disponen de servicio de TELF_CONV, también cabe destacar que variables como SEXO, HACINAMIENTO, NUMERODEPERSONAS, ACCESOVIENDA_2, entre otras que se detallan en la presente tabla se relaciona positivamente, a un nivel de significación de 5% dichos parametros se presentan estadísticamente significativas.

GRÁFICO 4.24 Medidas de Influencia



Fuente: Elaboración Propia mediante uso de la función plot aplicada sobre un objeto GLM en el Lenguaje de Programación Estadístico "R".

El primer gráfico, empezando por arriba y de izquierda a derecha, se muestran los residuos de la devianza frente al predictor lineal (transformaciones logit). El segundo compara los residuos de la devianza estandarizados con los cuantiles de una normal estándar. El tercer gráfico es de nuevo un gráfico de los residuos frente al predictor lineal, pero tomando la raíz cuadrada del valor absoluto de los residuos de la devianza estandarizados. El cuarto gráfico es el más útil para detectar valores influyentes, ya que compara los residuos estandarizados de Pearson con los hat values y además muestra líneas de contorno para las distancias de cookC.

ESTIMACIÓN DEL INDICE DE POBREZA MULTIDIMENSIONAL A NIVEL PROVINCIAL PARA ECUADOR

Mediante la estimación del modelo logit con glm, asignándole la respectiva categoría pobre, aquel que cumpla la condición de que presente una estimación mayor a 0.5 utilizando las variables antes seleccionadas, se realiza un conteo para ver los resultados obtenidos.

TABLA 4.25. Matriz de Confusión

	0	1
0	19040	1553
100	3082	4510
[1]	0.8355508	

Fuente: Elaboración Propia mediante uso de la función logit en el Lenguaje de Programación Estadístico "R".

De lo cual el modelo predictivo acierta en un 83.55% de los casos

Seguido del análisis exploratorio, se ha obtenido las estimaciones a partir del modelo seleccionado previamente para la ECV-2014, de la misma manera se realizó las predicciones mediante uso de la función "predict" en R a partir de las variables seleccionadas del respectivo modelo para el Censo de Población y Vivienda 2010 y posteriormente se realiza una comparación con los datos reales obtenidos de la ECV-2014 a nivel regional.

TABLA 4.26 Resultados de las correspondiente Estimaciones de la ECV del modelo seleccionado junto con predicciones para el CENSO-2010

PROVINCIA	ORIGINAL_ECV	ESTIMACIONES_ECV	PREDICCIONES_CENSO
AZUAY	0.1730175	0.149896266	0.20403718
BOLIVAR	0.3877095	0.385569335	0.36648202
CAÑAR	0.2598930	0.241639698	0.27264370
CARCHI	0.3170732	0.164719626	0.19785639
COTOPAXI	0.4025157	0.315266486	0.34341243
CHIMBORAZO	0.4455100	0.324017467	0.32859852
EL ORO	0.1438042	0.095386210	0.24008637
ESMERALDAS	0.3670653	0.294883721	0.32703776
GUAYAS	0.1791888	0.156099242	0.26556421
IMBABURA	0.3045113	0.205323194	0.23630748
LOJA	0.3069647	0.269827586	0.27881352
LOS RIOS	0.2650120	0.202312139	0.29885384
MANABI	0.2823529	0.260834671	0.34700332
MORONA SANTIAGO	0.5964432	0.509668508	0.39272203
NAPO	0.4544199	0.383216783	0.37176885
PASTAZA	0.4055944	0.372191011	0.31917959
PICHINCHA	0.1301939	0.082053073	0.09161271
TUNGURAHUA	0.2675325	0.168549088	0.18339287
ZAMORA CHINCHIPE	0.3786408	0.337988827	0.33901996
GALAPAGOS	0.0000000	0.007692308	0.12227488
SUCUMBIOS	0.3566879	0.286821705	0.29198787
ORELLANA	0.3414966	0.328767123	0.37425052
SANTO DOMINGO DE LOS TSACHILAS	0.2038540	0.095336788	0.21147792
SANTA ELENA	0.2229437	0.176602925	0.38248278

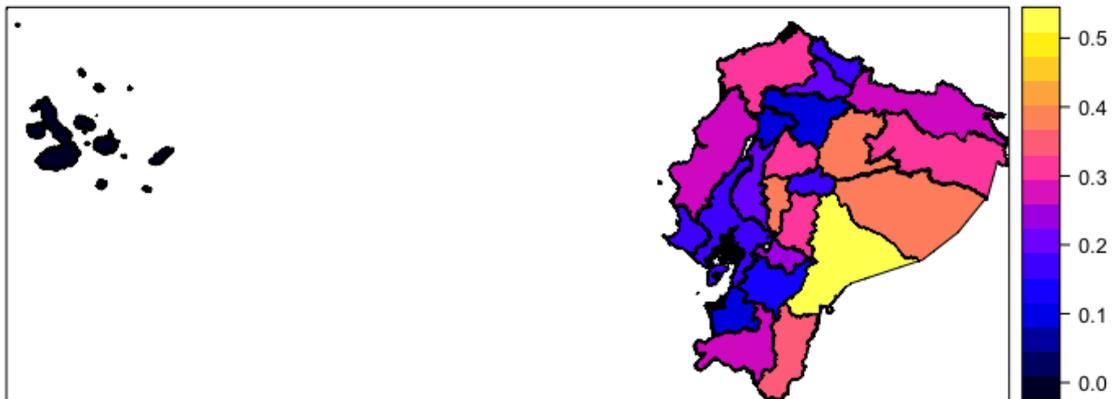
Fuente: Elaboración Propia mediante uso de la función "predict" aplicada sobre un objeto GLM en el Lenguaje de Programación Estadístico "R".

Por lo que se puede observar en base a los resultados proporcionados que la provincia con un mayor Índice de Pobreza se reporta en Morona Santiago seguido de Napo. Cabe destacar que la región que presenta un índice relativamente menor de pobreza es la provincia de Galápagos, esto se debe a que dicha provincia es considerada como uno de los principales destinos turísticos de miles de personas alrededor del mundo, permitiendo generar una Sostenibilidad Económica para los habitantes de aquella provincia.

4.1.5 Cartografía

Para la representación gráfica se hizo uso de la función “hcmmap” la cual permite crear mapas dinámicos , solo se selecciona el respectivo mapa que se desea de la colección highmaps para usarlo con nuestros datos y su correspondiente personalización, haciendo uso de la Cartografía con las dimensiones del territorio, disponible en la página web del Instituto de Estadística y Censo (INEC) o Instituto Geográfico Militar, y uniendo con la información recopilada a través de la función “ssplot” se obtiene el correspondiente mapa del territorio Ecuatoriano, incluyendo el Índice de Pobreza a nivel regional.

GRÁFICO 4.27. Cartografía de la Estimaciones para la ECV-2014 a nivel regional.



Fuente: Elaboración Propia mediante uso de la función “ssplot” en el Lenguaje de Programación Estadístico “R”.

5 CONCLUSIONES

Se ha empleado la técnica del estadístico matching learning, obteniendo la respectiva estimación mediante un Modelo de Regresión Logístico, el cual permite obtener el índice de pobreza a nivel regional en Ecuador con datos recopilados de la ECV-2014 junto con datos del CENSO-2010 proporcionados por el INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSO (INEC).

En cuanto a la implementación metodológica se ha hallado algunas dificultades para la aplicación del correspondiente matching.

El primer paso al momento de seleccionar variables comunes, se ha de tener en consideración que la presencia de pequeñas diferencias en el diseño de las preguntas, puede interpretarse en importantes divergencias de los datos, ocasionando como resultado el no poder utilizar las respectivas variables para el correspondiente proceso.

La armonización de las definiciones de las variables comunes ha permitido una mejor integración entre conjuntos de datos. Por otra parte, la presencia de variables que disponen de un alto porcentaje de casos nulos supone una dificultad añadida.

Se ha hecho uso de la variable pobreza proporcionada por la ECV-2014 la cual indica si un hogar es pobre asignándole un valor de 100 y 0 en caso contrario, la cual permite realizar la estimación junto con un conjunto de variables (Sexo del Jefe de Hogar, Hacinamiento, Acceso a Vivienda, Tipo de Vivienda, entre otras.) anteriormente seleccionadas, para poder hallar el Índice de Pobreza. Haciendo uso de modelos tanto Paramétricos como No Paramétricos de los cuales, para el presente propósito, se optó por utilizar la correspondiente estimación mediante el Modelo Paramétrico. Por otra parte, se hizo uso de la función predict del Lenguaje de Programación Estadístico R para poder obtener la respectiva predicción en cuanto a pobreza utilizando las mismas variables para el amplio conjunto de datos proporcionados por el Censo-2010 y de esta manera poder obtener el correspondiente índice a nivel regional.

Por lo que en base a los resultados obtenidos se puede observar que la provincia con un mayor índice de pobreza a nivel nacional se reporta en MORONA SANTIAGO tanto para los resultados obtenidos de la estimación en la ECV como en los resultados obtenidos de las predicciones para el CENSO, en contraste con la provincia de GALÁPAGOS la cual esta asociada a un índice relativamente menor.

6 REFERENCIAS BIBLIOGRÁFICAS

United Nations Development Programme: Human Development Reports: Multidimensional Poverty Index (MPI).

FRITSCH, S. and GUENTHER, F. neuralnet: Training of Neural Networks. R package version 1.33, 2016. <https://CRAN.R-project.org/package=neuralnet>

Estimación de la Tasa de Pobreza en Cantabria mediante Matching Estadístico, Autores: Pablo Lobete López, Francisco J.Parra Rodriguez.

Curso de Estadística con R: Autor: Francisco Parra Rodriguez -ICANE

CAÑADAS, L. Regresión logística: tratamiento computacional con R. Universidad de Granada, España, 2013.

Estimación de la Tasa de Pobreza en Argentina en Área Pequeña. Autor: Francisco J.Parra Rodriguez-ICANE.

Panorama Social de América Latina 2013. UN. CEPAL. 2013. Santiago de Chile: Naciones Unidas, 2014. Panorama Social de América Latina. ISSN: 1020-5152. http://repositorio.cepal.org/bitstream/handle/11362/35904/S2013868_es.pdf?sequence=1.

Ánalysis Geográfico de la Pobreza y Desigualdad por consumo en Ecuador. Más allá del Nivel Provincial, Autores: Esteban Cabrera, Andrea Molina Vera, Mary Alexander Sharman, Lorena Moreno, Facundo Cuevas.

