



PROYECTO DE GRADO

Presentado ante la ilustre UNIVERSIDAD DE LOS ANDES como requisito parcial para
obtener el Título de INGENIERO DE SISTEMAS

DISEÑO DE UN INDICADOR DE POBREZA QUE CLASIFIQUE A LOS HOGARES COMO POBRES O NO POBRES DE UNA FORMA IDÓNEA USANDO ANÁLISIS ESTADÍSTICOS Y MINERÍA DE DATOS.

Por

Br. Fatima Lucia Domingues Mendoza

Tutor: Francisco Barillas.

Noviembre 2019.

©2019 Universidad de Los Andes, Mérida, Venezuela.

C.C. Reconocimiento

**DISEÑO DE UN INDICADOR DE POBREZA QUE CLASIFIQUE A LOS
HOGARES COMO POBRES O NO POBRES DE UNA FORMA IDÓNEA USANDO
ANÁLISIS ESTADÍSTICOS Y MINERÍA DE DATOS.**

Br. Fatima Lucia Domingues Mendoza

Proyecto de Grado – Investigación de Operaciones – 92 Páginas

Resumen: En Venezuela, la pobreza es un fenómeno social que afecta considerablemente la calidad de vida de la sociedad y además de esto preocupa a los diferentes organismos competentes encargados de tomar decisiones para lidiar con esta problemática. El Instituto Nacional de Estadísticas (INE) es el encargado de llevar a cabo la medición de pobreza, lo realiza de dos maneras: Línea de pobreza y Necesidades Básicas Insatisfechas. Debido a que el proceso de medición tiene un grado de dificultad bastante elevado, tiende a no representar adecuadamente la realidad. Por ello este trabajo propone el diseño de un Indicador de pobreza, a través de análisis de correspondencias múltiples y árboles de decisión, considerando involucrar una nueva variable de estudio en el nuevo indicador, que me permite clasificar a la población en 3 grupos: hogares no pobres, hogares con pobreza relativa y hogares con pobreza extrema. Para ello se utilizó la base de datos proporcionada por el Instituto Nacional de Estadísticas (INE) 2011, abarcando todo el territorio nacional.

Palabras clave: Línea de Pobreza, Necesidades Básicas Insatisfechas, Instituto Nacional de Estadísticas (INE), Indicador de Pobreza.

Índice

	Pág.
RESUMEN	2
DEDICATORIA	3
AGRADECIMIENTO	4
INTRODUCCIÓN	11

Capítulo I

1.1 Planteamiento del problema	13
1.2 Objetivos del estudio	14
1.2.1 Objetivo General	14
1.2.2 Objetivos Específicos	14
1.3 Justificación	15
1.4 Alcance	16
1.5 Metodología	16
1.6 Estructura del documento	18

Capítulo II

2.1 Pobreza	19
2.2 Antecedentes	20
2.3 Perspectivas de la pobreza	22
2.4 El problema de la pobreza	23
2.5 La medición de la pobreza	24

2.6 La medición de la pobreza en Venezuela	25
2.7 Enfoque directo	25
2.7.1 Necesidades Básicas Insatisfechas	25
2.8 Enfoque indirecto	26
2.8.1 Línea de pobreza	26
2.9 Enfoque mixto	27
2.10 Análisis de Correspondencia Múltiple	28
2.10.1 Tabla disyuntiva completa	29
2.10.2 Tabla de contingencia de Burt	30
2.10.3 Interpretación	31
2.11 Arboles de Decisión	32
2.11.1 Notación	32
2.11.2 Forma Gráfica	33
2.11.3 División de un nodo	34
2.11.4 Nodos terminales	34
2.11.5 Índice de asociación y medidas de impureza	34
2.11.5.1 El índice de información o entropía	34
2.11.5.2 El índice de Gini	35
2.11.5.3 El índice de “Towing”	35
2.11.6 Poda del árbol	35
2.11.6.1 Poda de error reducido	36
2.11.6.2 Poda costo-complejidad	36
2.11.6 Reglas de parada	36

2.11.6.1 Pureza del nodo	37
2.11.6.2 Cota de profundidad	37
2.11.6.3 Umbral de soporte	37

Capítulo III

3.1 Metodología a utilizar	38
----------------------------	----

Capítulo IV

4.1 Análisis de Correspondencia Múltiple	47
4.2 Hogares Pobres	53
4.2.1 Generación de un árbol de regresión para estimar proporción de HP	53
4.2.2 Generación árbol de regresión_1	53
4.2.3 Evaluación del modelo Árbol_1	56
4.2.4 Prediccion_1	57
4.2.5 Bagging	59
4.2.6 Identificación de los predictores más importantes	60
4.2.7 Predicción con el modelo Baggin_2	62
4.3 Hogares en Pobreza Extrema	64
4.3.1 Generación del árbol de regresión_3	65
4.3.2 Evaluación del modelo Árbol de regresión HPE	66
4.3.3 Poda del árbol_3	67
4.3.4 Predicciones del modelo árbol podado	68
4.3.5 Generación del modelo por el método de Bagging	70
4.3.6 Identificación de los predictores más importantes	71

4.4 Propuesta de un nuevo indicador para medir pobreza	79
4.5 Comparación del nuevo indicador propuesto con los métodos de LP y NBI	80

Capítulo V

5.1 Conclusiones	83
5.2 Recomendaciones	85
REFERENCIAS	86
ANEXOS	88

www.bdigital.ula.ve

Índice de Tablas

Tabla 2.1 Enfoque mixto	27
Tabla 2.2 Frecuencia en Análisis de Correspondencia Múltiple	29
Tabla 2.3 Tabla disyuntiva completa buscada	30
Tabla 2.4 Notación de Árbol de Decisión	33
Tabla 4.1 Numero de municipios por estado	41
Tabla 4.2 Mínimo, cuartiles, mediana, media y máximo para el número total de hogares por municipio	41
Tabla 4.3 Variables de Estudio	42
Tabla 4.4 Estadísticas descriptivas	44
Tabla 4.5 Porcentaje de variabilidad en cada dimensión	47
Tabla 4.6 Valores propios, porcentaje de varianza explicado marginal y acumulado por cada dimensión	48
Tabla 4.7 Relación de correlación entre cada variable y cada dimensión	50
Tabla 4.8 Valor de la Canasta Alimentaria Normativa	75
Tabla 4.9 Número de Hogares Pobres según Línea de Pobreza	75
Tabla 4.10 Número de Hogares Pobres según Necesidades Básicas Insatisfechas	76
Tabla 4.11 Indicador FGT	76
Tabla 4.12 Parámetros de la distribución Gamma (Proporción de HP)	77
Tabla 4.13 Parámetros de la distribución Gamma (Proporción de HPE)	78
Tabla 4.14 Número de Hogares Pobres según indicador propuesto	80
Tabla 4.15 Porcentaje de pobreza según los métodos NBI, LP e Indicador propuesto	81

Índice de Gráficos

Gráfico 4.1 Histogramas	46
Gráfico 4.2 Porcentaje de varianza explicada en cada dimensión	50
Gráfico 4.3 Dim 1 vs Dim 2	51
Gráfico 4.4 Dim 1 vs Dim 3	52
Gráfico 4.5 Árbol de regresion_1	55
Gráfico 4.6 Evolución del error a medida que se incrementa los nodos	57
Gráfico 4.7 Predicciones vs los valores reales	58
Gráfico 4.8 Evolución del error respecto al número de árboles	60
Gráfico 4.9 Importancia de las variables	61
Gráfico 4.10 Predicciones modelo bagging vs valores reales	63
Gráfico 4.11 Comparación del error con ambos modelos	64
Gráfico 4.12 Evolución del error a medida que se incrementan los nodos	67
Gráfico 4.13 Predicciones modelo árbol podado vs valores reales	70
Gráfico 4.14 Evolución del error respecto al número de árboles	71
Gráfico 4.15 Importancia de las variables	72
Gráfico 4.16 Predicción de hogares pobres extremos vs valores reales	74
Gráfico 4.17 Comparación del error (Modelo de árbol podado y modelo bagging)	74
Gráfico 4.18 Distribución de probabilidad de la proporción de hogares pobres	77
Gráfico 4.19 Distribución de probabilidad de la proporción de hogares pobres extremos	78
Gráfico 4.20 Propuesta de nuevo indicador	79

Introducción

Históricamente, el estudio científico de la pobreza se remonta a comienzos del siglo XX. Atkinson (1987) citado por Feres J y Mancero X (2001), señala que antes de esa fecha se habían realizado algunas estimaciones sobre pobreza, pero que fue Booth entre 1892 y 1897 “el primero en combinar la observación con un intento sistemático de medición de la extensión del problema” elaborando un mapa de pobreza de Londres.

En contextos generales, esta investigación se enmarca en el estudio de la economía de la pobreza, concretamente en el análisis de la acepción multidimensional de la misma, para el caso específico español. El enfoque moderno de la pobreza se acerca más al concepto de bienestar en el que influyen numerosos factores, no sólo el nivel de renta.

La pobreza es un fenómeno social que se refiere a la incapacidad de las personas de vivir una vida tolerable. Entre los aspectos que la componen se mencionan llevar una vida larga y saludable, tener educación y disfrutar de un nivel de vida decente disfrutando de servicios básicos, además de otros elementos como la libertad política, el respeto de los derechos humanos, la seguridad personal, el acceso al trabajo productivo y bien remunerado y la participación en la vida comunitaria. No obstante dada la natural dificultad de medir algunos elementos constituyentes de la “calidad de vida”, el estudio de la pobreza se ha restringido a los aspectos cuantificables y generalmente materiales.

Existen diversas formas de medir los fenómenos sociales, tales como la educación, la salud, la pobreza, el desempleo, etc. Lo más importante es que debe tenerse presente que estos indicadores no son el hecho en sí, sino una forma de medirlos para tratar de comprenderlos. La naturaleza de los indicadores puede variar entre ellos según las variables empleadas en su construcción. En el caso de la pobreza antes de realizar su análisis se debe primero antes que nada definirla. Una vez establecidos los aspectos que abarca el término “pobreza”, su medición requiere de esos indicadores cuantificables para poder arrojar o dar conclusiones acerca de la misma.

En el caso del territorio venezolano, el organismo oficial encargado de realizar el proceso de evaluación de la pobreza es el Instituto Nacional de Estadística (INE), que lleva a cabo la medición mediante dos métodos: la Línea de Pobreza (basado en datos relacionados a los ingresos que poseen los hogares venezolanos) y las Necesidades Básicas Insatisfechas, donde se entiende como “necesidades básicas” el acceso sostenible a servicios como educación, agua potable, vivienda, entre otros.

Dado que la pobreza está asociada a las carencias en varios aspectos de la vida (como los anteriormente mencionados) y que los métodos utilizados por el INE no logran representar la realidad, este trabajo tiene como objetivo proponer un indicador de pobreza que esté acorde a la situación actual venezolana.

www.bdigital.ula.ve

Capítulo I

1.1 Planteamiento del problema

Ante la notable preocupación que genera la pobreza a nivel mundial, se han desarrollado metodologías para su definición, medición y análisis. En el caso del territorio venezolano, las estadísticas oficiales sobre pobreza son reportadas por el Instituto Nacional de Estadística (INE), ente que se basa en dos métodos para calcular los indicadores de pobreza: el método de Línea de Pobreza (LP) y el de Necesidades Básicas Insatisfechas (NBI).

En Venezuela existen varios métodos para estudiar el problema de la pobreza solo que los mismos presentan ciertas deficiencias que muchas veces nos afectan y no nos permiten tener cifras más exactas acerca de este fenómeno social.

Al método de la Línea de la Pobreza se le hacen algunas críticas ya que supone que si se tiene un ingreso superior a la línea establecida están cubiertas todas las necesidades esenciales, lo que conlleva a una concepción individualista de las necesidades; el método considera que la satisfacción de las necesidades básicas depende solamente del ingreso o del consumo corriente, cuando son más las variables que determinan la satisfacción como, digamos, los derechos de acceso a bienes y servicios gubernamentales (agua y drenaje, por ejemplo); y solo mide una dimensión de la pobreza, la de la potencialidad de satisfacer las necesidades básicas mediante el ingreso, sin considerar el nivel de satisfacción de esas necesidades.

Por otra parte, el método de NBI (Necesidades Básicas Insatisfechas) permite medir el número de hogares que no ha satisfecho alguna necesidad básica, pero no necesariamente mide la pobreza. Esto se debe, en primer lugar, a que no existe una forma única y establecida de relacionar el número de necesidades básicas insatisfechas con la condición de pobreza, lo que implica que la clasificación final en pobres y no pobres es arbitraria queda entregada enteramente al criterio del investigador. En segundo lugar, el número de personas pobres siempre aumenta ante la adición de nuevos indicadores; en una situación extrema, el uso de

un número suficientemente grande de indicadores podría originar que casi toda la población sea clasificada como pobre. Un tercer problema del método de NBI, en tanto, consiste en la elección de ponderadores para cada indicador, al momento de agregar las distintas necesidades entre sí.

El Índice FGT tiene un parámetro α que especifica la sensibilidad del índice y para valores mayores de $\alpha=2$ el índice se hace muy difícil de interpretar.

Toda esta problemática puede traer como consecuencia obtener resultados de pobreza muy distintos dependiendo del indicador que utilicemos, lo cual puede ser crucial en la obtención y análisis de resultados de cualquier trabajo de investigación que este referido a temas de pobreza.

En este proyecto de investigación la idea será indagar a través de análisis estadísticos multivariantes, cuál de los métodos de pobreza clasifica mejor a un hogar como pobre o no pobre, es decir cual o cuales métodos usan un mejor criterio de clasificación, y como objetivo final proponer un nuevo indicador de pobreza que esté acorde o ajustado, el cual permita representar la situación actual en Venezuela.

1.2 Objetivos

1.2.1 Objetivo General

Diseñar un indicador de pobreza que clasifique a los hogares como pobres o no pobres de una forma idónea.

1.2.2 Objetivos Específicos

- Calcular los indicadores de pobreza (necesidades insatisfechas, línea de pobreza, índice FGT) con la base de datos suministrada por INE.
- Clasificar los hogares como pobres o no pobres en base a los indicadores calculados anteriormente.
- Comparar los resultados entre los indicadores calculados anteriormente.
- Diseñar un indicador de pobreza en base a las fallas conocidas de los indicadores usados.
- Comparar los resultados del indicador nuevo con los indicadores ya conocidos.

1.3 Justificación

La pobreza es uno de los problemas con más relevancia a nivel mundial, ya que influye directamente en el desarrollo de la sociedad y la forma en que los individuos se relacionan. Es por ello el propósito de contribuir con el avance de conocimientos sobre los determinantes del mismo, esfuerzos realizados a través de estudios, análisis y acuerdos internacionales, donde se enfatiza la disminución de la pobreza en general y la erradicación de la pobreza extrema. En un estudio reciente el Banco Mundial (BM) propone el objetivo de ponerle fin a la pobreza incluyendo medidas orientadas a reducir el nivel de desigualdad, los investigadores del banco identificaron seis estrategias de alto impacto para aumentar los ingresos de los pobres, aumentar acceso a los servicios básicos y mejorar sus perspectivas de desarrollo a largo plazo.

Como podemos observar es de gran preocupación el problema de la pobreza y de aquí la importancia de estudiar a fondo este fenómeno social, para lo cual este proyecto de investigación, será dedicado a proponer una solución a algunas de las debilidades y defectos que tienen algunos de los indicadores de pobreza usados en Venezuela a través de la aplicación y el cálculo de Medidas de Acuerdo haciendo diversas combinaciones con líneas de pobreza relativas y distintas restricciones al método de NBI respecto al número de variables a considerar, para así determinar en qué punto ambas medidas concuerdan en mayor proporción o porcentaje en la clasificación de los hogares como pobres o no pobres.

Para el caso del territorio venezolano, los métodos usados oficialmente para medir la pobreza son el de Línea de Pobreza y el de Necesidades Básicas Insatisfechas, que como se comentó previamente presentan algunas limitaciones, con lo cual se genera la interrogante: ¿la información proporcionada por estos métodos realmente representa la situación de pobreza en Venezuela?

Dada la multidimensionalidad del fenómeno, se debe contar con información amplia y suficiente de los factores o dimensiones que indicen sobre el mismo, tales como economía, educación, condiciones de vivienda, servicios y la salud.

A través de la medición de la pobreza podemos saber la situación actual en la cual se encuentra el país desde ese punto de vista social y así evitar andar a ciegas y que el sector

público conduzca el barco del país sin brújula. Es por ello que se propone el diseño de un indicador de pobreza para representar adecuadamente los aspectos relacionados a este fenómeno.

1.4 Alcance

En este trabajo de investigación se cumplen importantes metas como lo es obtener medidas de acuerdo entre las necesidades básicas insatisfechas y las líneas de pobreza. Al medir el acuerdo entre estos dos evaluadores de pobreza se determina que ambas medidas tienen muchas diferencias en cuanto a la clasificación de hogares como pobres relativos, pobres extremos o no pobres, lo cual es un aspecto que deja mucho que pensar en cuanto la confiabilidad de utilizar un indicador u otro para determinar los niveles de pobreza.

A través del análisis de correspondencias múltiples utilizando como variables de entrada aquellas que determinan las líneas de pobreza y las necesidades básicas insatisfechas se logra detallar que la alta dependencia económica en un hogar, la adecuación del mismo, los ingresos y la escolaridad del jefe del hogar son variables muy importantes al momento de medir pobreza. Es por ello que se propone el diseño de un indicador de pobreza que permita estudiar y analizar la pobreza en Venezuela, observando cuatro aspectos principales: Condición económica, nivel de educación, condiciones de salud y asistencia sanitaria, y nivel de vida en general (descrita por las condiciones y características de las viviendas y el acceso a servicios públicos como privado).

1.5 Metodología

Esta tesis comienza con el cálculo de indicadores de pobreza: La línea de pobreza y las necesidades básicas insatisfechas (NBI), tomando en cuenta la base de datos proporcionada por el Instituto Nacional de Estadística (INE) 2011.

Para el caso de la Línea de Pobreza según INE, se designa a los ingresos en esta escala ingresos en canasta alimentaria per cápita o ICAPC y la Canasta Alimentaria Normativa (CAN), en donde tenemos $ICACP = (\text{ingreso total del hogar per cápita}) / (\text{CAN per cápita})$. Ahora definimos el Índice de Línea de Pobreza (ILP) como: $ILP = 0$ si $ICAPC \geq 2$ (Hogares no pobres), $ILP = 1$ si $2 > ICACP \geq 1$ (Hogares en pobreza relativa), $ILP = 2$ si $ICAPC < 1$ (Hogares en pobreza extrema).

El Indicador de pobreza Necesidades Básicas Insatisfechas (NBI) define seis indicadores binarios de Necesidades Básicas, asignándole uno (1) en el hogar si satisface lo siguiente:

1. Hogares en Viviendas Inadecuadas (IVI). Este indicador considera los hogares que habitan en viviendas improvisadas (ranchos y ranchos campesinos), en casa de vecindad y en otra clase (tráileres o remolques, embarcaciones, carpas, cuevas, etc.)
2. Hogares en Hacinamiento Crítico (IHC). Este indicador mide el grado de hacinamiento en los cuartos de la vivienda que los integrantes de un hogar utilizan para dormir. Se incluyen en esta categoría a aquellos hogares que tienen más de tres personas por cuarto para dormir.
3. Hogares en Viviendas sin Servicios Básicos (IFS). A través de este indicador se mide la inaccesibilidad de los hogares al agua potable y a la eliminación de excretas, aquí se toma en consideración si el hogar carece al menos de uno de los dos servicios.
4. Hogares con Niños en Escolaridad (7 a 12 Años) que No asisten a la Escuela (IAE). Este indicador considera los hogares con niños en edad escolar (7 a 12 años) que no asisten a la escuela.
5. Escolaridad del Jefe del Hogar (IEJ). Este se refiere a la educación del jefe del hogar, cuyo jefe del mismo tiene menos de cuatro grados de escolaridad.
6. Hogares con Alta Dependencia Económica (IDE). Este indicador está referido a las condiciones de empleo de los integrantes del hogar. En él se contemplan aquellos hogares con más de tres personas por ocupado.

Ahora definimos el Índice de NBI como:

$$\text{NBI} = \text{IVI} + \text{IHC} + \text{IFS} + \text{IAE} + \text{IEJ} + \text{IDE}$$

Este índice indica el número de necesidades básicas insatisfechas por hogar, con rango entre cero (0) y seis (6). Finalmente se constituye el indicador NBI (INBI) como:

$\text{INBI} = 0$ si $\text{NBI} = 0$ (hogares no pobres), $\text{INBI} = 1$ si $\text{NBI} = 1$ (hogares en pobreza relativa), $\text{INBI} = 2$ si $\text{NBI} \geq 2$ (hogares en pobreza extrema).

1.6 Estructura del documento

Este trabajo consta de 5 capítulos estructurados de la siguiente manera:

Capítulo 1: Introducción. Se plantea el problema, objetivos, justificación y alcance del proyecto. Se presenta brevemente la metodología empleada en el trabajo.

Capítulo 2: Marco Teórico. Se describen los antecedentes de investigaciones de pobreza, así como las bases teóricas para su comprensión, análisis y medición. Se explican los fundamentos de la metodología para la construcción del Indicador de Pobreza.

Capítulo 3: Marco Metodológico. Se explica cómo se realizó el estudio para resolver el problema planteado y cuáles fueron los softwares (programas) de requerimientos.

Capítulo 4: Análisis de Resultados. Se analizan los resultados obtenidos de cada uno de los métodos usados en el desarrollo del proyecto.

Capítulo 5: Conclusiones y Recomendaciones. Se describen los principales hallazgos encontrados en la investigación y se plantean algunas recomendaciones del estudio realizado.

Capítulo II

Marco Teórico

2.1 Pobreza

El término de pobreza es un fenómeno complejo, heterogéneo, multidimensional que tiene muchas definiciones que puede variar con los valores y actitudes de la sociedad, a través de una serie de similitudes.

En su acepción más simple y partiendo de la definición más básica, la pobreza de acuerdo con el Diccionario de la Real Academia Española en su actualización 2018, es un sustantivo que determina la falta o escasez. La pobreza se refiere a una situación o forma de vida que surge como producto de la carencia o privación de los recursos para satisfacer las necesidades básicas humanas, tales como alimentación, vivienda, educación, salud, agua potable, que tienen incidencia sobre las personas.

Sen (1992) afirma que la pobreza no es falta de riqueza o ingreso, sino de capacidades básicas.

Deleeck et al (1992) afirma que “La pobreza no se limita a una dimensión, como por ejemplo el ingreso; se manifiesta en todas las dimensiones de la vida como la vivienda, la educación y la salud”.

Spicker (1993) señala que la pobreza se refiere no sólo a privaciones sino a privaciones sufridas durante un período de tiempo.

Según la ONU (1995), 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 sólo de ingresos monetarios sino también del acceso a servicios”.

Según PNUD (1997) define a la pobreza como: “Incapacidad de las personas de vivir una vida tolerable”.

Paul Spicker (1999) reconoce once posibles formas de identificar esta palabra: como necesidad, estándar de vida, insuficiencia de recursos, carencia de seguridad básica, falta de titularidades, privación múltiple, exclusión, desigualdad, clase, dependencia y padecimiento inaceptable.

Según el Banco Mundial (2000), la pobreza “es un fenómeno multidimensional, que incluye incapacidad para satisfacer las necesidades básicas, falta de control sobre los recursos, falta de educación y desarrollo de las destrezas, deficiente salud, desnutrición, falta de vivienda, acceso limitado al agua y a los servicios sanitarios, vulnerabilidad a los cambios bruscos, violencia y crimen, falta de libertad política y de expresión.

Gary Fields (2001) define la pobreza como la incapacidad de un individuo o una familia para disponer de los recursos suficientes para satisfacer sus necesidades básicas.

(Haughton & Khandker, 2009), mencionan que la pobreza es la privación de bienestar de manera pronunciada, es decir, la falta de acceso a capacidades básicas para funcionar en la sociedad y de un ingreso adecuado para enfrentar necesidades de educación, salud, seguridad, empoderamiento y derechos básicos.

Las definiciones antes mencionadas describen diferentes acepciones de la pobreza, ya que engloba una gran parte de diferentes de punto de vista que existe sobre esta temática y que son utilizados en el desarrollo de este trabajo.

2.2 Antecedentes

En el caso de la medición de la pobreza existen gran cantidad de indicadores, el cual cada uno de ellos presenta diferentes rasgos de un mismo fenómeno, los principales indicadores de pobreza utilizados en Venezuela son: Línea de Pobreza, Necesidades Básicas Insatisfechas, Índice FGT.

Las líneas de pobreza van asociadas a los conceptos de pobreza elegidos; de este modo, podemos diferenciar tantos como acepciones hayamos distinguido. Es un indicador coyuntural, pues se basa en la medición del ingreso, el cual es una variable que puede cambiar muy rápidamente. Mide el poder adquisitivo con respecto a la canasta básica y la canasta alimentaria.

En el caso de las necesidades básicas insatisfechas, es un método directo para identificar carencias críticas en una población y caracterizar la pobreza, tomando en consideración indicadores relacionados con las necesidades básicas estructurales (vivienda, educación, salud, infraestructura pública, etc.) que se requiere para evaluar el bienestar individual.

Hernán Méndez (1959) citado por Huerta J, realizó una versión modificada del Índice de Graffar (versión Méndez) con el fin de utilizarlo en Venezuela para clasificar a la población en estratos sociales y considera cuatro variables: La Procedencia del Ingreso, la Profesión del Jefe del Hogar, El nivel de Instrucción de la Madre y las Condiciones de Alojamiento.

Foster, Greer y Thorbecke (1984) citados por Fernández A y Martín G (1994), propusieron el índice FGT que tiene la capacidad de medir tres aspectos fundamentales como lo son la cantidad de pobres que existen en una región, la brecha que existe entre los pobres y la severidad de la pobreza.

Sen (1992), indica que el concepto de *pobreza* se construye a partir de las capacidades, es decir, de lo que la gente puede hacer y define a la pobreza como *la ausencia de capacidades básicas que le permiten a cualquier individuo insertarse en la sociedad, a través del ejercicio de su voluntad*. En consecuencia, la pobreza no es cuestión de escaso bienestar, sino de incapacidad para conseguir bienestar precisamente debido a la ausencia de medios.

Vásquez, Ramírez y Camardiel (1997) citado por Huerta J, desarrollaron un Índice Sintético de Pobreza que Considera 7 variables: Las Condiciones Estructurales de la Vivienda, el Equipamiento del Hogar, el Acceso a Servicios, la Ocupación de los Miembros del Hogar y la Educación de los Miembros del Hogar.

El PNUD (Programa de las Naciones Unidas para el Desarrollo) citado por Huerta J, a comienzos de los 90 construyó el Índice de Desarrollo Humano. Este indicador Pretende medir el grado promedio de desarrollo en cuanto a las capacidades humanas básicas. Considera tres variables que son la Salud (esperanza de vida), la educación (logro educativo) y el Ingreso (PIB per cápitá).

INE (2004), publicó un artículo de Condiciones de vida y la pobreza en Venezuela en donde estudian la heterogeneidad de la pobreza en el país y hacen una comparación entre el método de Necesidades básicas insatisfechas y la línea de pobreza.

Ato, Benavente y López (2006), realizaron un análisis comparativo de tres enfoques para evaluar el acuerdo entre observadores que originan tres tipos diferentes de medidas: descriptivas, basadas en modelos log lineales.

2.3 Perspectivas de la pobreza

La pobreza puede ser entendida como un problema de múltiples dimensiones, tanto de un modelo económico como social, asumido en un conglomerado específico y en un momento determinado por los distintos actores, instituciones económicas y políticas, que producen en la sociedad la existencia de sectores sociales que quedan excluidos de los beneficios totales o parciales de ese modelo. Desde esta perspectiva la pobreza es el resultado de complejos procesos, muchos de los cuales son imperceptibles.

Otro perspectiva de la pobreza, es el de pobreza humana, esta se refiere a la privación en cuanto a tener una vida larga y saludable; poder acceder al conocimiento; alcanzar un nivel de vida decente y a acceder a la participación (PNUD, 2000).

En otra perspectiva la pobreza se plantea como un problema moral, más que económico. "La teoría económica contempla solamente aquellas interacciones humanas que involucran el intercambio monetario y de bienes, de hecho ignora gran parte de la existencia humana. La economía no tiene en cuenta el amor, la familia, la cultura, la salud, la espiritualidad, el medio ambiente o cualquier otra cosa que haga la vida rica y significativa" (WORLD FAITHS DEVELOPMENT COUNCIL, 1999).

Con base al PNUD la pobreza tiene tres perspectivas: La del ingreso, la de las necesidades básicas y la de las capacidades. En la perspectiva del ingreso una persona es pobre solo cuando su nivel de ingreso es inferior a la línea de pobreza que se ha definido. Muchos países han adoptado líneas de pobreza de ingreso para vigilar los progresos hechos en la reducción de la incidencia de la pobreza. Con frecuencia se define la línea de pobreza en términos de tener ingreso suficiente para comprar una cantidad determinada de alimentos.

Desde la perspectiva de las necesidades básicas la pobreza es la privación de los medios materiales para satisfacer en medida mínimamente aceptable las necesidades humanas, incluidos los alimentos. Este concepto de privación va mucho más allá de la falta de ingreso privado: incluye la necesidad de servicios básicos de salud y educación y otros servicios esenciales que la comunidad tiene que prestar para impedir que la gente caiga en la pobreza.

Desde la perspectiva de capacidad la pobreza representa la ausencia de ciertas capacidades básicas para funcionar, una persona que carece de la oportunidad para lograr unos niveles mínimamente aceptables de esos funcionamientos. Los funcionamientos pertinentes a este análisis pueden variar desde los físicos, como estar bien nutrido, estar vestido y vivir en forma adecuada, evitar la morosidad prevenible, hasta logros sociales más complejos, como participar en la vida de la comunidad. El criterio de la capacidad concilia los conceptos de pobreza absoluta y relativa, ya que la privación relativa de ingreso y productos puede provocar una privación absoluta de capacidad mínima (PNUD 1997)

Existen dos definiciones básicas y distintas que se deben considerar. La pobreza absoluta y la pobreza relativa. La pobreza absoluta *se define como la situación en la cual no están cubiertas las necesidades básicas del individuo, es decir, existe carencia de bienes y servicios básicos*, son estos: nutrición, salud, vestido y vivienda. La pobreza relativa *sitúa el fenómeno de la pobreza en la sociedad objeto de estudio. Desde esta perspectiva se considera que una persona es pobre cuando se encuentra en una situación de clara desventaja, económica y socialmente, respecto al resto de personas de su entorno. Esta concepción de la pobreza está muy ligada a la noción de desigualdad*”. (INE, 2006)

2.4 El problema de la pobreza

Considerando diferentes visiones sobre la pobreza, han surgido otras que enfocan, en conjunto, la pobreza como concepto material, como situación económica, como condición social y como juicio moral, en general como un problema de toda ciencia que tiene en el centro de estudio al hombre.

La literatura sobre el tema ofrece, al menos, dos tipos fundamentales de definición. La primera enfatiza en las necesidades básicas insatisfechas y la segunda en la incapacidad de satisfacer tales necesidades.

El caso de las necesidades básicas insatisfechas se resume en que es pobre aquella persona que no logra satisfacer sus necesidades básicas de alimentación, vivienda, vestido y educación. La segunda definición, se concentra en la incapacidad de la persona para generar los medios necesarios para satisfacer dichas necesidades (CEPAL/CLAD/SELA, 1996).

Actualmente, la lucha contra la pobreza está basada en seis estrategias: desarrollo y nutrición en la primera infancia, cobertura universal de salud, acceso universal a educación de calidad, transferencias monetarias a familias pobres, infraestructura rural, en particular caminos y electrificación, tributación progresiva y equitativa. (Banco mundial, 2016).

2.5 La medición de la Pobreza

El término de “pobreza” es multidimensional, no existe un enfoque autosuficiente para dimensionarlo es por ello que existen diferentes abordajes para interpretar y medir la pobreza.

Según (CEPAL, 2015) En general... Para obtener una medida de pobreza se requiere de: medición monetaria del bienestar del hogar (consumo o ingreso) y línea de pobreza en ese mismo espacio.

Ante esta variedad de interpretaciones, existe consenso y hay similitud entre los investigadores, es por ello que para medirla es necesario:

- Definir un concepto simple y observable.
- Especificar una variable que refleje el bienestar.
- Determinar el valor de una línea de pobreza general.
- Identificar las personas cuyos niveles de consumo se encuentran por debajo y por encima de estas líneas.
- Realizar una recolección con todos los datos obtenidos anteriormente.

Los principales métodos para medir la pobreza tienen varios enfoques, entre lo que destacan el directo, indirecto y mixto. Para fines de este trabajo de investigación, a continuación se definirán y se resaltarán sus características más importantes.

2.6 La medición de la Pobreza en Venezuela

El Instituto Nacional de Estadísticas (INE), es el encargado de medir la pobreza en los hogares venezolanos, además de hacer un reporte del mismo. Esta medición la realiza a través de dos métodos fundamentales, el de la línea de pobreza (LP) que relaciona el ingreso del hogar, conjunto de alimentos y costos básicos prioritarios de educación y salud, todos estos elementos mencionados antes son los de la canasta básica, por otro lado el método de necesidades básicas insatisfechas (NBI), define los déficit de la satisfacción de las necesidades básicas a través de indicadores que expresan el nivel mínimo debajo del cual se considera insatisfecha su necesidad.

2.7 Enfoque directo

En el enfoque directo, no es solo una alternativa metodológica al método de líneas de pobreza, sino que representa una conceptualización distinta de la pobreza (Sen, 1981). Observa directamente las condiciones de vida de la población, es decir, una persona pobre es aquella que no satisface una o varias necesidades básicas. Este método relaciona el bienestar con el consumo efectivamente realizado.

2.7.1 Necesidades básicas insatisfechas

La medición de la pobreza a través de este método requiere de la elección de características de los hogares que “además de representar alguna dimensión importante de la privación, también se encuentre suficientemente asociadas con las situaciones de pobreza como para representar las demás carencias que configuran tales situaciones” (INDEC, 1984).

Las variables normalmente usadas para determinar las Necesidades Básicas Insatisfechas son las siguientes:

- Hacinamiento: hogares con más de tres personas por cuarto.
- Vivienda: paredes, piso, techo y tipo de vivienda.
- Disponibilidad de servicios como electricidad, agua potable y servicios sanitarios.
- Asistencia escolar: hogares que tienen al menos un niño en edad escolar que no asiste a la escuela.
- Capacidad de subsistencia económica del jefe de hogar, según nivel de escolaridad y dependencia de los ocupados.

Entre las ventajas del método de NBI encontramos la utilización de datos provenientes de las organizaciones competentes del mismo y así construir “mapas de pobreza”, además que este método radica en su fácil aplicación. Sin embargo tiene alguna desventaja: se encuentra el hecho de que no capta adecuadamente situaciones que pueden caracterizarse como de pobreza reciente, en las que los hogares siguen satisfaciendo sus necesidades aunque sus recursos hayan disminuido drásticamente (Feres y Mancero, 2001), no considera la intensidad de la pobreza, es decir, es igualmente pobre un hogar que no haya satisfecho al menos una necesidad como al que no haya logrado satisfacer cuatro o cinco de dichas necesidades, no se reconocen las diversidades pues las variables utilizadas para determinar necesidades básicas insatisfechas, son consideradas como valores promedios, independiente del país, región, cultura, familia, género u otras características personales.

2.8 Enfoque Indirecto

En el enfoque indirecto, clasificara como pobres a aquellas personas que no cuenten con los recursos suficientes para satisfacer sus necesidades básicas. Este método evalúa el bienestar a partir de la capacidad para realizar consumos.

2.8.1 Línea de pobreza

Establece el ingreso o gasto mínimo que permite mantener un nivel de vida adecuado, según ciertos estándares elegidos (Olavarriá, 2001). Dicho de otro modo, la línea de pobreza permite distinguir quienes son considerados pobres en una sociedad de aquellos que no lo son.

Según (Ravallion, 1998), indica que una línea de pobreza puede ser definida como el valor, en términos monetarios, del bienestar económico de un individuo en un lugar y momento dado.

El INE (2011a) indica que una de las formas para construir esta línea de pobreza, consiste en estimar primero el costo de una canasta básica de alimentos cuyo contenido calórico y proteico permita satisfacer un nivel mínimo de requerimientos nutricionales (generalmente entre 2100 y 2300 Calorías por persona por día), para luego añadirle el costo de la canasta básica no alimentaria, la cual comprende a otros bienes y servicios relacionados con la vivienda, vestido y educación, entre otros.

El costo de la canasta básica de alimentos equivale a lo que se denomina Línea de Pobreza Extrema, y el costo de la canasta total se refiere a la Línea de Pobreza Absoluta. El nivel de bienestar individual, expresado en valor monetario, generalmente es aproximado por el ingreso o el consumo familiar per cápita, en el marco de medición de la Línea de Pobreza (INE, 2011a).

La ventaja principal de las líneas de pobreza reside en la simplicidad para la identificación de los pobres y no pobres, además, una vez realizado esto, la posibilidad de cuantificar la extensión de la pobreza y poder hacer comparaciones en el tiempo. Sin embargo, como limitante o desventaja se encuentra en las normativas de los que parte para la construcción de la canasta básica alimentaria y el hecho de que tome únicamente el ingreso o el gasto como medidas del bienestar.

2.9 Enfoque Mixto

El enfoque mixto, propone la integración de los criterios de la línea de pobreza y necesidades básicas insatisfechas, ambos métodos se complementan, el primero siendo como un indicador indirecto enfocados en aspectos económicos y que apunta a la capacidad de satisfacción de necesidades básicas, y el segundo siendo un indicador directo enfocado en los aspectos estructurales de la pobreza. La complementariedad de ambos métodos lleva a la definición de cuatro tipologías presentadas en el cuadro de la siguiente tabla.

Tabla 2.1 Enfoque Mixto

Método NBI → Método LP ↓	Hogares con NBI (1 o más)	Hogares sin NBI
Hogares bajo LP	Hogar con pobreza crónica	Hogar con pobreza reciente
Hogares sobre LP	Hogar con carencia inercial	Hogar no pobre

Fuente: INE

1. Los hogares que se encuentran por debajo de la línea de pobreza y presenta una o más necesidad básica insatisfecha son denominados “Hogar con pobreza crónica”.
2. Los hogares que se encuentran por encima de la línea de pobreza y presenta una o más necesidad básica insatisfecha son denominados “Hogar con carencia inercial”
3. Los hogares que se encuentran por debajo de la línea de pobreza y satisfacen las necesidades básicas insatisfechas son denominados “Hogar con pobreza reciente”.
4. Los hogares que se encuentran por encima de la línea de pobreza y satisfacen las necesidades básicas insatisfechas son denominados “Hogar no pobre”.

2.10 Análisis de Correspondencia Múltiple

El análisis de correspondencia múltiple (ACM) es una generalización del análisis de correspondencia simple (ACS). Es una técnica estadística que estudia las relaciones de 3 o más variables categóricas a partir de una muestra de individuos. De hecho permite descubrir grandes tablas lógicas (compuesta de ceros y unos), como por ejemplo, las que resultan de la codificación de una encuesta. Las filas de estas tablas son, en general, individuos u observaciones y las columnas son las modalidades de las variables nominales (en el ejemplo de la encuesta serían las modalidades de respuesta a cada una de las preguntas). El propósito es representar gráficamente la estructura de relaciones de dos o más variables cualitativas mediante mapas de posicionamiento” (Vivanco, 1999).

En muchas ocasiones se dispone de la información correspondiente a un conjunto de individuos sobre los que se miden una serie de variables cualitativas que verifican las características de que las modalidades de cada variable son mutuamente excluyentes y exhaustivas. Normalmente, se dice que estas variables están dispuestas bajo forma disyuntiva completa. Si se denomina a “s” el número de variables cualitativas y “n” al número de individuos, la tabulación de estos datos bajo codificación condensada da como resultado una tabla R de n filas y s columnas, cuyo término general R_{iq} designa la modalidad de la variable q que presenta el sujeto i.

Tabla 2.2 Frecuencia en Análisis de Correspondencia Múltiple

R($n \times s$)		Q					
		1	2	...	q	...	S
I	1	r_{11}	r_{12}	...	r_{1q}	...	r_{1s}
	2	r_{21}	r_{22}	...	r_{2q}	...	r_{2s}
	
	i	r_{i1}	r_{i2}	...	r_{iq}	...	r_{is}
	
	n	r_{n1}	r_{n2}	...	r_{nq}	...	r_{ns}

Fuente: Fernández (2011).

Denominando p_q al número de modalidades de la variable q, se tiene que: $r_{iq} \leq p_q$. Sin embargo, esta tabla no es explotable ya que las sumas en filas y en columnas no tienen sentido. Por ello, resulta necesario efectuar una recodificación de la tabulación de las variables.

2.10.1 Tabla disyuntiva completa

El resultado de la recodificación de la tabla condensada anterior es una tabla lógica, es decir, compuesta de ceros y unos a la que se le denomina tabla disyuntiva completa. Para representar la forma en que se obtiene esta tabla, vamos a designar por I al conjunto de los n individuos y por p al número total de modalidades de las s variables. Se tiene, por lo tanto, que:

$$p = \sum_{q=1}^s p_q$$

(1)

A partir de la tabla de datos condensada R, se construye la tabla Z de n filas y p columnas que describe, mediante una codificación lógica, las s modalidades que presenta cada uno de los n individuos. Por tanto, la tabla Z es el resultado de la yuxtaposición de s subtablas:

$$Z = [z_1, z_2, \dots, z_q, \dots, z_s] \quad (2)$$

Esta es la tabla disyuntiva completa buscada, y su término general se escribe $Z_{ij} = 1$ o $Z_{ij} = 0$ según si el sujeto i presenta la modalidad j de la variable q, o no.

2.3 Tabla disyuntiva completa buscada

Z(n × p)				Z ₁			...			Z _s			Total
		1	...	j ₁	...	p ₁	...	1	...	j _s	...	p _s	
		1	j	p	
I	1	Z ₁₁	...	1	...	0	...	0	...	1	...	Z _{1p}	Z _{1..} = s
	2	Z ₂₁	...	0	...	0	...	0	...	0	...	Z _{2p}	Z _{2..} = s

	i	Z _{i1}	...	0	...	1	...	1	...	0	...	Z _{ip}	Z _{i..} = s

	n	Z _{n1}	...	1	...	0	...	1	...	0	...	Z _{np}	Z _{n..} = s
	Total	Z _{.1}	Z _{.j}	Z _{.p}	Z = ns

Fuente: Fernández (2011).

Las marginales en la fila de la tabla disyuntiva son constantes e iguales al número de variables s:

$$Z_{i..} = \sum_{j=1}^p z_{ij} = s. \quad (3)$$

Las marginales en columna corresponden al número de individuos que presentan la modalidad j de la variable q:

$$z_{.j} = \sum_{i=1}^n Z_{ij} \quad (4)$$

La suma de las marginales da el efectivo total Z de la tabla:

$$Z = \sum_{i=1}^n \sum_{j=1}^p z_{ij} = ns \quad (5)$$

2.10.2 Tabla de contingencia de Burt

A partir de la tabla disyuntiva completa Z, se construye la tabla simétrica B de orden (p × p) que recoge los cruces dos a dos de todas las variables:

$$B = Z'Z \quad (6)$$

A la matriz B se le denomina tabla de contingencia de Burt asociada a la tabla disyuntiva completa Z. El término general de B se escribe:

$$b_{ij'} = \sum_{j=1}^n z_{ij} z_{ij'} \quad (7)$$

B es una yuxtaposición de las tablas de contingencia. Las marginales son, para todo $j \leq p$:

$$b_j = \sum_{j=1}^p b_{ij'} = sZ_{.j}$$

Y el efectivo total será:

$$b = s^2 n \quad (9)$$

2.10.3 Interpretación

Según Fernández (2011), existen algunas reglas que ayudan a entender y contextualizar los resultados que se obtienen al realizar un ACM:

- 1. Proximidad entre individuos en términos de parecido:** Dos individuos se parecen si tienen casi las mismas modalidades. Es decir, dos individuos están próximos si han elegido globalmente las mismas modalidades.
- 2. Proximidad entre modalidades de variables diferentes en términos de asociación:** Son cercanos puesto que globalmente están presentes en los mismos individuos. Es decir, dos modalidades están próximas si han sido elegidas globalmente por el mismo conjunto de individuos.
- 3. Proximidad entre modalidades de una misma variable en términos de parecido:** (a) Son excluyentes por construcción. (b) Si son cercanas es porque los individuos que las poseen presentan casi el mismo comportamiento en las otras variables.

4. Al observar la representación gráfica, una modalidad estará más alejada del origen de coordenadas a medida que tenga menor número de frecuencia.

2.11 Árboles de decisión

Son técnicas utilizadas con gran enfoque en estadísticas, minería de datos y aprendizaje automático que consisten en estudiar grandes masas con el fin de descubrir patrones no triviales. En los modelos de árboles cuando la variable dependiente es continua son árboles de regresión y cuando la variable dependiente es de tipo cualitativo son árboles de clasificación. En esencia, se trata de dar con un esquema de múltiples bifurcaciones, anidadas en forma de árbol, de manera que siguiendo cada una de las ramas obtengamos al final una predicción para la clase de pertenencia (clasificación) o para el valor que toman (regresión) los individuos que cumplen con las exigencias en las distintas bifurcaciones.

Los árboles de decisión se construyen mediante un algoritmo conocido como segmentación recursiva. Existen tres procedimientos principales para dicha construcción: CHAID (Chi-Square Automatic Interaction Detector), QUEST (Quick Unbiased Efficient Statistical Tree) y CART (Classification And Regression Trees).

Si Y es una variable respuesta y las p variables predictores son x_1, x_2, \dots, x_p , donde las X son tomadas fijas y Y es una variable aleatoria, el problema estadístico es establecer una relaciones entre Y y las X de tal forma que sea posible predecir Y basado en los valores de X . Matemáticamente, se quiere estudiar la probabilidad condicional de la variable aleatoria Y . $P[Y=y | x_1, x_2, \dots, x_p]$ o una función de su probabilidad tal como la esperanza condicional. $E[Y | x_1, x_2, \dots, x_p]$, según se trate de un árbol de clasificación o de regresión.

2.11.1 Notación

Los árboles de decisión, son grafos que constan de nodos de decisión representados por un cuadrado, nodos de probabilidad representados por un círculo y ramas representadas por líneas o una línea cruzada por otras dos.

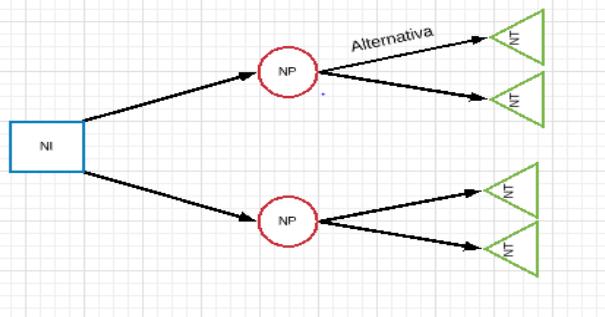
2.3 Tabla de Notación

Símbolo	Nombre	Descripción
	Nono de Decisión	Indica que se debe tomar una decisión
	Nodo de Probabilidad	Muestra varios resultados
	Alternativa o Ramificaciones	Cada línea de ramificación indica un posible resultado
	Alternativa Rechazada	Indica una alternativa o un resultado que no se debe tomar en cuenta
	Nodo Terminal	Indica un resultado definitivo

Fuente: Elaboración Propia

2.11.2 El árbol de decisión obtiene la siguiente forma gráfica

www.udigital.edu.ve



Fuente: Elaboración Propia

Al comienzo del árbol de decisión se encuentra un Nodo Inicial que no es apuntado por ninguna flecha, dos nodos internos que muestra resultados inciertos y tres nodos terminales que están en el tercer nivel. La completa homogeneidad de los nodos terminal es un ideal alcanzado en el análisis de datos real, de esta manera, el objetivo del algoritmo de segmentación recursiva es hacer las variables resultantes en los nodos terminales tan homogéneas como sea posible. Una medida cuantitativa de la homogeneidad es la noción de impureza.

2.11.3 División de un nodo

Para dividir el nodo inicial en nodos homogéneos, se debe seleccionar entre los rangos de todas las variables predictoras el valor de división que más se acerque al límite de pureza para cada uno de los hijos. La finalidad es que si el nodo A se divide en A1 y A2, la pureza de ambos debe ser mayor que la del nodo A, o a su vez, su impureza menor. Impureza que se mide por la mínima probabilidad, la entropía o el índice de Gini.

2.11.4 Nodos terminales

El proceso de segmentación recursiva continúa hasta que el árbol sea saturado en el sentido de que los nodos descendientes no se puedan partir en una división adicional. El número total de divisiones permitidas disminuye a medida que aumentan los niveles del árbol. Todo nodo que no se pueda dividir es llamado nodo terminal.

2.11.5 Índice de asociación y medidas de impureza

En la creación de árboles de decisión existen diversos criterios que son utilizados para la creación de subdivisiones de conjunto de datos, de tal forma que, con cada nueva subdivisión que se realice mejore la variable criterio.

Existen diferentes criterios que se pueden utilizar para la selección de cada nueva subdivisión, entre ellos podemos encontrar:

2.11.5.1 El índice de información o entropía

$$i(t) = \sum_j p(j|t) \ln p(j|t) \quad (10)$$

El objetivo es encontrar la ecuación que maximice a A(T)

$$\Delta i(t) = - \sum_{j=1}^k p(j|t) \ln p(j|t), \quad (11)$$

donde $j = 1, \dots, k$ es el número de clases de la variable respuesta categórica y

$p(j|t)$ la probabilidad de clasificación correcta para la clase j en el nodo t .

2.11.5.2 El índice de Gini

$$i(t) = \sum_{i \neq j} p(j|t)p(i|t) \quad (12)$$

Encontrar la partición que maximice $\Delta i(t)$.

$$\Delta i = - \sum_{j=1}^k [p_j(t)]^2, \quad (13)$$

2.11.5.3 El índice “Towing”

Busca dos clases que juntas formen más del 50% de los datos, el cual estas definen dos “Súper Categorías” en cada división para las cuales la impureza es definida por el índice Gini. Aunque el índice Towing produce arboles más balanceados.

Seleccione la partición s , que maximice

$$\frac{PLPR}{4} \left[\sum_j |p(j|t_L) - p(j|t_R)| \right]^2, \quad (14)$$

Donde T_L y T_R representan los nodos hijos izquierdo y derecho respectivamente, P_L y P_R representan la proporción de observaciones en t que pasaron a T_L y a T_R en cada caso.

2.12 Poda del Árbol

El árbol obtenido es generalmente sobre ajustado por tanto es podado, hasta encontrar el tamaño adecuado del árbol. La poda busca un balance entre el aumento del error de clasificación y la reducción del tamaño del árbol. Los subárboles a reemplazar son elegidos examinando cada nodo interno del árbol completo, reemplazando el subárbol por una hoja y calculando el error de clasificación que genera el árbol podado sobre un conjunto de prueba. Entre las técnicas de comparación para determinar el árbol óptimo tenemos:

2.12.1 Poda de error reducido: Comienza a recorrer el árbol construido de las hojas hacia la raíz y en cada nodo interno se calcula el error de clasificación, posteriormente se reemplaza el nodo por una hoja y se calcula el error de clasificación con el árbol podado, si este error es menor que el calculado con el árbol completo, se reemplaza el nodo por la hoja, de lo contrario, se queda el árbol como estaba.

2.12.2 Poda costo-complejidad: Este proceso se realiza en dos pasos. En el primer paso se genera, una familia de árboles denotada por $T = \{T_0, \dots, T_k\}$, donde T_{i+1} es obtenido podando el árbol T_i . T_0 es el árbol completo y T_k es un árbol representado solo por el nodo raíz. Para construir el árbol T_1 , se calcula el parámetro α para cada nodo interno t de T_0 . Aquellos nodos internos que tengan el mínimo valor de α serán reemplazados por hojas, la clase asignada a esa nueva hoja será la de mayor frecuencia en los objetos asociados a ese nodo. Este procedimiento se efectúa hasta que se construye el árbol T_k . α se define como:

$$\alpha = \frac{e(t) - \sum_{l \in L_t} e(l)}{n(|L_t| - 1)} \quad (15)$$

Donde, t es el nodo interno a reemplazar, n es el número de objetos del conjunto utilizado para obtener el error de clasificación, que pasan por el nodo t en el recorrido que hacen del árbol, $e(t)$ es el número de errores si t es reemplazado por una hoja y L_t es el conjunto de hojas que tiene el subárbol con raíz t .

El segundo paso es encontrar el árbol de la familia T con el menor error de clasificación, utilizando un conjunto de prueba o realizando validación cruzada con el CE, este árbol elegido caracterizar al CE.

2.13 Reglas de parada

Tratan de predecir si es necesario seguir construyendo el árbol por la rama actual o no. Estas reglas también son llamadas reglas de pre-poda porque reduce la complejidad del árbol durante su construcción.

Lo tradicional es detener la construcción cuando se ha llegado a un nodo puro, entendiendo por nodo puro aquel que contiene ejemplos de una única clase. Además se

puede utilizar otros criterios de parada que no son mencionados anteriormente, a continuación se describen algunos:

2.13.1 Pureza del nodo: suele utilizarse un umbral de pureza para detener la construcción del árbol de decisión cuando la ramificación no suponga una disminución significativa de la impureza del mismo.

2.13.2 Cota de profundidad: se puede predeterminar con anterioridad una cota de profundidad para no construir arboles muy complejos. Cuando un nodo se halle a más de cierta profundidad se detiene el proceso de generación.

2.13.3 Umbral de soporte: cuando nos encontramos con un nodo que posea menos de X ejemplos podemos paralizar el proceso de construcción del árbol debido a la falta de información, ya que no es fiable su clasificación avalada por menos de X entrenamientos.

Capítulo III

Metodología

3.1 Metodología a utilizar

En este capítulo se describe cómo resolver el problema planteado y como dar respuesta a los objetivos propuestos en este trabajo de investigación.

Esta tesis comienza tomando como referencia las medidas de pobreza aportadas por el INE en el año 2011. Estos resultados nos muestran el número de hogares pobres y pobres extremos por municipio en cada uno de los estados de Venezuela, además de también mostrar el número de hogares en hacinamiento crítico, hogares con alta dependencia económica, hogares con niños que no asisten a la escuela, hogares en viviendas sin servicios básicos, hogares en viviendas inadecuadas, luego se obtiene la proporción de hogares con la deficiencia acorde a las variables aquí mencionadas, y posteriormente se categorizan las variables proporcionales por cuartiles, a estas variables categorizadas se les realiza un análisis de correspondencias múltiples incluyendo la variable estado o entidad.

Luego del análisis de correspondencias múltiples se procede a realizar un árbol de regresión o de clasificación con el fin de comparar los resultados obtenidos con la técnica de análisis de correspondencias.

A continuación se procede con el cálculo de indicadores de pobreza: La línea de Pobreza y las Necesidades Básicas Insatisfechas, tomando como referencia de base de datos del INE en el año 2011, y de esta forma comparar con los resultados de pobreza proporcionados por el INE.

Para el caso de la Línea de Pobreza según INE (2001), se designa a los ingresos en esta escala ingresos en canasta alimentaria per cápita o ICAPC y la Canasta Alimentaria Normativa (CAN), en donde tenemos $ICACP = (\text{ingreso total del hogar per cápita}) / (\text{CAN per cápita})$. Ahora definimos el Índice de Línea de Pobreza (ILP) como: $ILP = 0 \text{ si } ICAPC \geq 2$

(Hogares no pobres), ILP=1 si $1 \leq \text{ICACP} < 2$ (Hogares en pobreza relativa), ILP=2 si $\text{ICACP} < 1$ (Hogares en pobreza extrema).

La CAN se fija en este proyecto de investigación de acuerdo a la página del INE (<http://www.ine.gov.ve/documentos/Economia/CanastaNormativaAlimentaria/html/CANMes.html>), tomando como referencia los valores mensuales para el año 2011.

El otro Indicador de pobreza considerado es el de Necesidades Básicas Insatisfechas (NBI). Para ello definimos seis indicadores binarios de Necesidades Básicas, asignándole uno (1) en el hogar si satisface lo siguiente:

1. Hogares en Viviendas Inadecuadas (IVI). Este indicador considera los hogares que habitan en viviendas improvisadas (ranchos y ranchos campesinos), en casa de vecindad y en otra clase (tráileres o remolques, embarcaciones, carpas, cuevas, etc.)
2. Hogares en Hacinamiento Crítico (IHC). Este indicador mide el grado de hacinamiento en los cuartos de la vivienda que los integrantes de un hogar utilizan para dormir. Se incluyen en esta categoría a aquellos hogares que tienen más de tres personas por cuarto para dormir.
3. Hogares en Viviendas sin Servicios Básicos (IFS). A través de este indicador se mide la inaccesibilidad de los hogares al agua potable y a la eliminación de excretas, aquí se toma en consideración si el hogar carece al menos de uno de los dos servicios.
4. Hogares con Niños en Escolaridad (7 a 12 Años) que No asisten a la Escuela (IAE). Este indicador considera los hogares con niños en edad escolar (7 a 12 años) que no asisten a la escuela.
5. Escolaridad del Jefe del Hogar (IEJ). Este se refiere a la educación del jefe del hogar, cuyo jefe del mismo tiene menos de cuatro grados de escolaridad.
6. Hogares con Alta Dependencia Económica (IDE). Este indicador está referido a las condiciones de empleo de los integrantes del hogar. En él se contemplan aquellos hogares con más de tres personas por ocupado.

Ahora definimos el Índice de NBI como:

$$\mathbf{NBI = IVI + IHC + IFS + IAE + IEJ + IDE}$$

Este índice indica el número de necesidades básicas insatisfechas por hogar, con rango entre cero (0) y seis (6). Finalmente se constituye el indicador NBI (INBI) como:

$INBI=0$ si $NBI=0$ (hogares no pobres), $INBI=1$ si $NBI=1$ (hogares en pobreza relativa), $INBI=2$ si $NBI \geq 2$ (hogares en pobreza extrema).

Finalmente se propone un nuevo indicador de pobreza que esté acorde para mostrar la realidad.

www.bdigital.ula.ve

Capítulo IV

Análisis de Resultados

Tabla 4.1 Números de municipios por estado

ENTIDAD	AMAZONAS	ANZOÁTEGUI	APURE	ARAGUA
	7	21	7	18
BARINAS		BOLÍVAR	CARABOBO	COJEDES
	12	11	14	9
DELTA AMACURO	DEPEND	FEDERALES	DISTRITO CAPITAL	FALCÓN
	4	1	1	25
GUÁRICO		LARA	MÉRIDA	MIRANDA
	15	9	23	21
MONAGAS	NUEVA ESPARTA		PORTUGUESA	SUCRE
	13	11	14	15
TÁCHIRA		TRUJILLO	VARGAS	YARACUY
	29	20	1	14
ZULIA				
	21			

Fuente: INE 2011.

El estado con mayor cantidad de municipios es el estado Táchira, con un total de 29 municipios, seguido por el estado Falcón con 25 municipios. El estado con menor cantidad de municipios son los estados Distrito Capital, Dependencias Federales y Vargas. Para un total de 336 municipios en todo el territorio nacional.

Tabla 4.2 Mínimo, cuartiles, mediana, media y máximo para el número total de hogares por Municipio

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
339	4723	8695	20577	19683	530054

Fuente: Elaboración Propia

El número mínimo de hogares en Venezuela para el año 2011 es de 339 y corresponde al estado Amazonas municipio Autónomo Río Negro, el máximo es de 530.054 hogares y corresponde al Distrito Capital municipio Libertador, el número promedio de hogares es de 20.577 con una mediana de 8.695 hogares.

Tabla 4.3 Variables de Estudio

ENTIDAD	25 estados, tomando en cuenta a las dependencias federales.
MUNICIPIOS	336 municipios en Venezuela para el año 2011.
HP	Número de hogares pobres por municipio en cada estado de Venezuela para el año 2011.
PHP	Proporción de hogares pobres, se obtiene dividiendo el número de hogares pobres por municipio, entre el número de total de hogares en el respectivo municipio.
HPE	Número de hogares pobres extremos por municipio en cada estado de Venezuela para el año 2011
PHPE	Proporción de hogares pobres extremos, se obtiene dividiendo el número de hogares pobres extremos por municipio, entre el número de total de hogares en el respectivo municipio.
AE	Hogares con Niños en Escolaridad (7 a 12 Años) que No asisten a la Escuela.
PAE	Proporción de hogares con niños que no asisten a la escuela.
DE	Hogares con Alta Dependencia Económica. Este indicador está referido a las condiciones de empleo de los integrantes del hogar. En él se contemplan aquellos hogares con más de tres personas por ocupado.

PDE	Proporción de hogares con alta dependencia económica.
HVSB	Hogares en Viviendas sin Servicios Básicos. A través de este indicador se mide la inaccesibilidad de los hogares al agua potable y a la eliminación de excretas, aquí se toma en consideración si el hogar carece al menos de uno de los dos servicios.
PHVS	Proporción de hogares en viviendas sin servicios básicos.
HVI	Hogares en Viviendas Inadecuadas. Este indicador considera los hogares que habitan en viviendas improvisadas (ranchos y ranchos campesinos), en casa de vecindad y en otra clase (tráileres o remolques, embarcaciones, carpas, cuevas, etc.).
PHVI	Proporción de hogares en viviendas inadecuadas.
HHC	Hogares en Hacinamiento Crítico (IHC). Este indicador mide el grado de hacinamiento en los cuartos de la vivienda que los integrantes de un hogar utilizan para dormir. Se incluyen en esta categoría a aquellos hogares que tienen más de tres personas por cuarto para dormir.
PHHC	Proporción de hogares en hacinamiento crítico.

Fuente: Elaboración Propia.

Tabla 4.4 Estadísticas descriptivas

HP	PHP	HPE	PHPE
Min. : 104	Min. : 0.03306	Min. : 11.0	Min. : 0.002165
1st Qu.: 1286	1st Qu.: 0.21061	1st Qu.: 343.2	1st Qu.: 0.046813
Median : 2477	Median : 0.29216	Median : 722.0	Median : 0.081577
Mean : 5055	Mean : 0.29873	Mean : 1435.0	Mean : 0.092761
3rd Qu.: 5567	3rd Qu.: 0.36019	3rd Qu.: 1522.0	3rd Qu.: 0.115669
Max. : 85677	Max. : 0.83111	Max. : 25942.0	Max. : 0.552961
AE	PAE	DE	PDE
Min. : 5.00	Min. : 0.004604	Min. : 17.0	Min. : 0.00550
1st Qu.: 73.75	1st Qu.: 0.011528	1st Qu.: 289.5	1st Qu.: 0.03876
Median : 153.00	Median : 0.016489	Median : 575.5	Median : 0.06356
Mean : 339.54	Mean : 0.021402	Mean : 907.4	Mean : 0.07189
3rd Qu.: 369.25	3rd Qu.: 0.024096	3rd Qu.: 1026.5	3rd Qu.: 0.09883
Max. : 7746.00	Max. : 0.244922	Max. : 12047.0	Max. : 0.33094
HVSB	PHVS	HVI	PHVI
Min. : 2.0	Min. : 0.002638	Min. : 3.0	Min. : 0.001647
1st Qu.: 388.8	1st Qu.: 0.049909	1st Qu.: 327.2	1st Qu.: 0.045503
Median : 874.5	Median : 0.088145	Median : 839.0	Median : 0.096093
Mean : 1883.6	Mean : 0.123721	Mean : 1843.8	Mean : 0.105521
3rd Qu.: 2000.5	3rd Qu.: 0.157286	3rd Qu.: 2148.2	3rd Qu.: 0.144781
Max. : 50165.0	Max. : 0.737281	Max. : 29733.0	Max. : 0.501180
HHC	PHHC		
Min. : 15	Min. : 0.01040		
1st Qu.: 465	1st Qu.: 0.08307		
Median : 979	Median : 0.10569		
Mean : 2143	Mean : 0.11365		
3rd Qu.: 2180	3rd Qu.: 0.13244		
Max. : 53105	Max. : 0.54921		

Fuente: Elaboración Propia.

Podemos observar las estadísticas descriptivas por cada variable. El mínimo de hogares pobres en Venezuela para el año 2011 era de 104 hogares correspondiente al estado Táchira Municipio Simón Rodríguez, el municipio con menos cantidad de hogares dentro del estado Táchira con un total de 758 hogares de la muestra.

El promedio de hogares pobres en Venezuela para el año 2011 era de 5.055 hogares con una mediana de 2.477, el máximo de hogares pobres es de 85.677 correspondiente al estado Zulia municipio Maracaibo, el cual tiene un total de 343.969 hogares (el municipio con más hogares dentro del estado Zulia).

El número mínimo de hogares pobres extremos es de 11 hogares el cual corresponde al municipio Simón Rodríguez del estado Táchira, el promedio de hogares pobres extremos es de 1.435 con una mediana de 722 hogares, el máximo de hogares pobres extremos es de

25.942, correspondientes al municipio Iribarren estado Lara, un municipio con un número total de 249.827 hogares.

En los hogares con niños que no asisten a la escuela, el número mínimo es de 5 hogares pertenecientes al municipio Simón Rodríguez estado Táchira el cual tiene un total de 758 hogares, el máximo es de 7.746 hogares con niños que no asisten a la escuela perteneciente al municipio Maracaibo estado Zulia, el número promedio de hogares con niños en esta condición es de aproximadamente 340 con una mediana de 153 hogares.

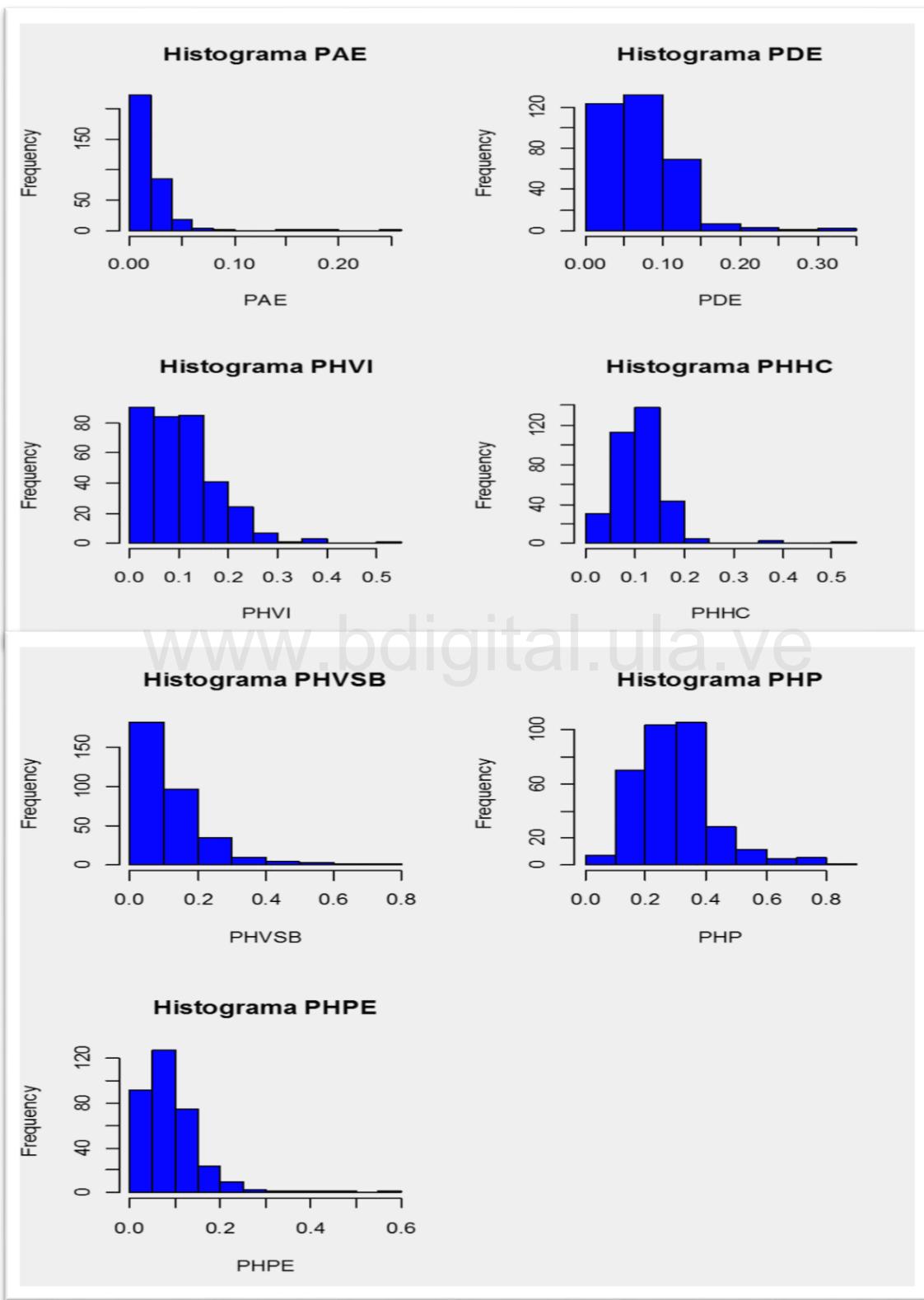
El número mínimo de hogares con alta dependencia económica es de 17 hogares pertenecientes a Dependencias Federales y el máximo es de 12.047 hogares pertenecientes al municipio Maracaibo estado Zulia, la media es de 907 hogares con esta condición y una mediana de 575.

El número mínimo de hogares en viviendas sin servicios básicos es de 2 pertenecientes al municipio Simón Rodríguez estado Táchira y el número máximo es de 50.165 pertenecientes al municipio Iribarren estado Lara, el municipio con más hogares dentro del estado Lara. El número promedio de hogares en esta condición es de 1.884 con una mediana de 875 hogares.

El número mínimo de hogares en viviendas inadecuadas es de 3 hogares pertenecientes al municipio Juan Vicente Campo Elías estado Trujillo, y el número máximo es de 29.733 hogares pertenecientes al municipio Iribarren estado Lara. El número promedio de hogares con esta condición es de 1.844 con una mediana de 839 hogares.

El número mínimo de hogares en hacinamiento crítico es de 15 hogares pertenecientes al municipio Simón Rodríguez estado Táchira y el número máximo es de 53.105 hogares pertenecientes al municipio Maracaibo estado Zulia.

Gráfico 4.1 Histogramas



Fuente: Elaboración Propia.

Podemos observar una leve asimetría positiva en el histograma para la proporción de hogares pobres en Venezuela. De hecho lo podemos corroborar con el coeficiente de asimetría, el cual arroja un valor positivo de skewness(PHP)= 1,055629.

En el histograma para la proporción de hogares pobres extremos se observa una asimetría positiva, skewness(PHPE)= 2.440733. Los datos están sesgados a la derecha.

En vista de que este análisis descriptivo no nos da información acerca de los estados en que hay mayor pobreza y pobreza extrema, ya que al ser un análisis básico queda predeterminado a la dimensión geográfica y al número de hogares existentes en un estado o municipio, por ejemplo, a mayor tamaño geográfico probablemente haya más hogares y por consiguiente más hogares pobres en comparación con estados o municipios más pequeños. Todo esto nos lleva a pensar en una técnica de análisis multivariantes como lo es el análisis de correspondencias múltiples.

4.1 Análisis de Correspondencias Múltiples

El análisis se ejecutó en 336 individuos (Total de municipios en Venezuela para el año 2011), descrito por 8 variables (ENTIDADDF, PHP, PHPE, PAE, PDE, PHVSB, PHVI, PHHC), categorizamos las variables cuantitativas por cuartiles para poder hacer el análisis incluyendo la variable entidad en el estudio, y así observar en que estados hay mayor pobreza y pobreza extrema, así como también ver que variables y que categorías de estas influyen más en que un hogar sea calificado como pobre o pobre extremo.

Tabla. 4.5 Porcentaje de variabilidad en cada dimensión

## `\$Dim 1` ## `\$Dim 1`\$quali ## R2 p.value ## PHPEC 0.8992445 4.821549e-165 ## PHPC 0.8917233 7.440039e-160 ## PHVSBC 0.6326289 7.451343e-72 ## PHVIC 0.5990911 1.441095e-65 ## PAEC 0.5472114 8.171339e-57 ## PHHCC 0.5122454 1.823400e-51 ## ENTIDADDF 0.5872784 2.521740e-46 ## PDEC 0.3381568 1.499421e-29	La proporción de hogares en pobreza extrema es explicada por el 89,92% de variabilidad en la dimensión 1. La proporción de hogares con alta dependencia económica (PDEC) es la variable que presenta un menor porcentaje con un 33,82%, por esa razón se encuentra ubicada en el último lugar.
--	---

## \$`Dim 2` ## \$`Dim 2`\$quali ## R2 p.value ## PHPEC 0.71388070 7.544808e-90 ## PHPC 0.67431595 1.595011e-80 ## PHVSBC 0.48850202 4.757376e-48 ## PHVIC 0.36541026 1.448656e-32 ## PHHCC 0.28476167 5.404917e-24 ## PAEC 0.23998098 1.187131e-19 ## ENTIDADDF 0.35866276 7.060109e-19 ## PDEC 0.08353125 2.236496e-06	Para la dimensión 2 las variables se encuentran ubicadas en el mismo orden de importancia que la para la dimensión 1, a excepción de que pone por encima PHHCC de PAEC, aquí se explica el 71,39% de variabilidad de la variable proporción de hogares en pobreza extrema (PHPEC) y solo el 8,35% para la variable proporción de hogares con alta dependencia económica (PDEC).
## \$`Dim 3` ## \$`Dim 3`\$quali ## R2 p.value ## PHPEC 0.45589843 1.310215e-43 ## PHPC 0.44279829 6.703227e-42 ## ENTIDADDF 0.41515821 2.029711e-24 ## PDEC 0.26329996 7.039841e-22 ## PHVSBC 0.11618769 6.330093e-09 ## PHVIC 0.11271012 1.197417e-08 ## PHHCC 0.07721783 6.738994e-06 ## PAEC 0.04145784 2.800192e-03	Para la dimensión 3 el orden de importancia de las variables cambia, se mantiene en primer lugar PHPEC, pero la EntidadF sube al tercer lugar con un 41,52% de variabilidad explicada, y la proporción de hogares con niños que no asisten a la escuela (PAEC) se encuentra en el último lugar con un 4,15% de variabilidad explicada.

Fuente: Elaboración Propia.

Tabla 4.6 Valores propios, porcentaje de varianza explicado marginal y acumulado por cada dimensión

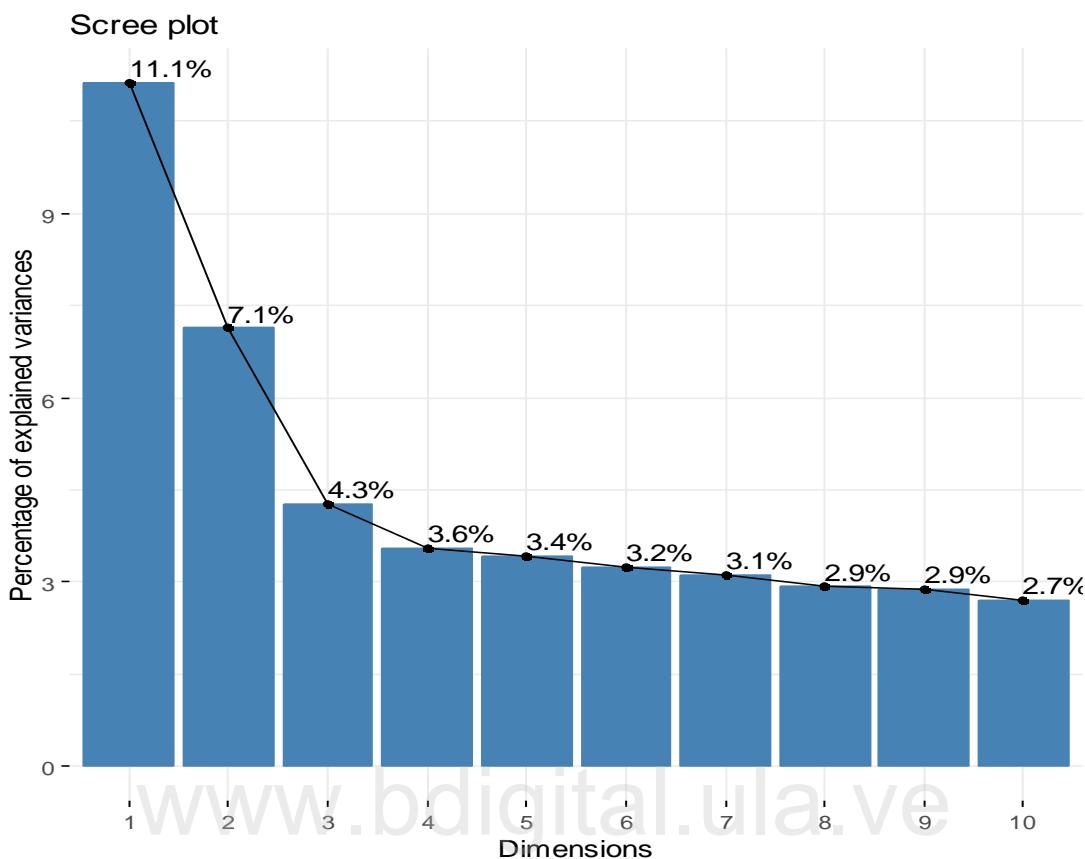
	eigenvalue	variance.percent	cumulative.variance.percent
## Dim.1	0.625947467	11.1279550	11.12795
## Dim.2	0.401130698	7.1312124	18.25917
## Dim.3	0.240591046	4.2771742	22.53634
## Dim.4	0.200106963	3.5574571	26.09380
## Dim.5	0.192101571	3.4151390	29.50894
## Dim.6	0.182565217	3.2456039	32.75454
## Dim.7	0.175302979	3.1164974	35.87104
## Dim.8	0.164563123	2.9255666	38.79661
## Dim.9	0.162282521	2.8850226	41.68163
## Dim.10	0.152226140	2.7062425	44.38787
## Dim.11	0.141394700	2.5136836	46.90155
## Dim.12	0.140699725	2.5013285	49.40288

## Dim.13	0.136744986	2.4310220	51.83390
## Dim.14	0.132623615	2.3577532	54.19166
## Dim.15	0.130601355	2.3218019	56.51346
## Dim.16	0.129303337	2.2987260	58.81219
## Dim.17	0.127686833	2.2699881	61.08217
## Dim.18	0.126078356	2.2413930	63.32357
## Dim.19	0.125440889	2.2300603	65.55363
## Dim.20	0.125067196	2.2234168	67.77704
## Dim.21	0.125016531	2.2225161	69.99956
## Dim.22	0.125000000	2.2222222	72.22178
## Dim.23	0.125000000	2.2222222	74.44400
## Dim.24	0.125000000	2.2222222	76.66623
## Dim.25	0.120829290	2.1480763	78.81430
## Dim.26	0.108045406	1.9208072	80.73511
## Dim.27	0.105074924	1.8679986	82.60311
## Dim.28	0.101812123	1.8099933	84.41310
## Dim.29	0.095004701	1.6889725	86.10207
## Dim.30	0.090795851	1.6141485	87.71622
## Dim.31	0.087080160	1.5480917	89.26431
## Dim.32	0.076892371	1.3669755	90.63129
## Dim.33	0.074300255	1.3208934	91.95218
## Dim.34	0.071658712	1.2739327	93.22612
## Dim.35	0.059591743	1.0594088	94.28553
## Dim.36	0.058082162	1.0325718	95.31810
## Dim.37	0.053415248	0.9496044	96.26770
## Dim.38	0.041065429	0.7300521	96.99775
## Dim.39	0.038005846	0.6756595	97.67341
## Dim.40	0.034991544	0.6220719	98.29548
## Dim.41	0.028559020	0.5077159	98.80320
## Dim.42	0.025769154	0.4581183	99.26132
## Dim.43	0.023163252	0.4117912	99.67311
## Dim.44	0.009636210	0.1713104	99.84442
## Dim.45	0.008751349	0.1555795	100.00000

Fuente: Elaboración Propia.

Las dos primeras dimensiones reúnen el 18,26% de varianza explicada por el ACM, y con la tercera dimensión se reúne al 22,54%. En las tres primeras dimensiones podemos observar el mayor porcentaje de variabilidad.

Gráfico 4.2 Porcentaje de varianza explicada en cada dimensión



Se puede observar en el gráfico que el mayor porcentaje de varianza explicada se encuentra entre la dimensión 1, la dimensión 2 y la dimensión 3, razón por la cual nos quedaremos con las tres primeras dimensiones.

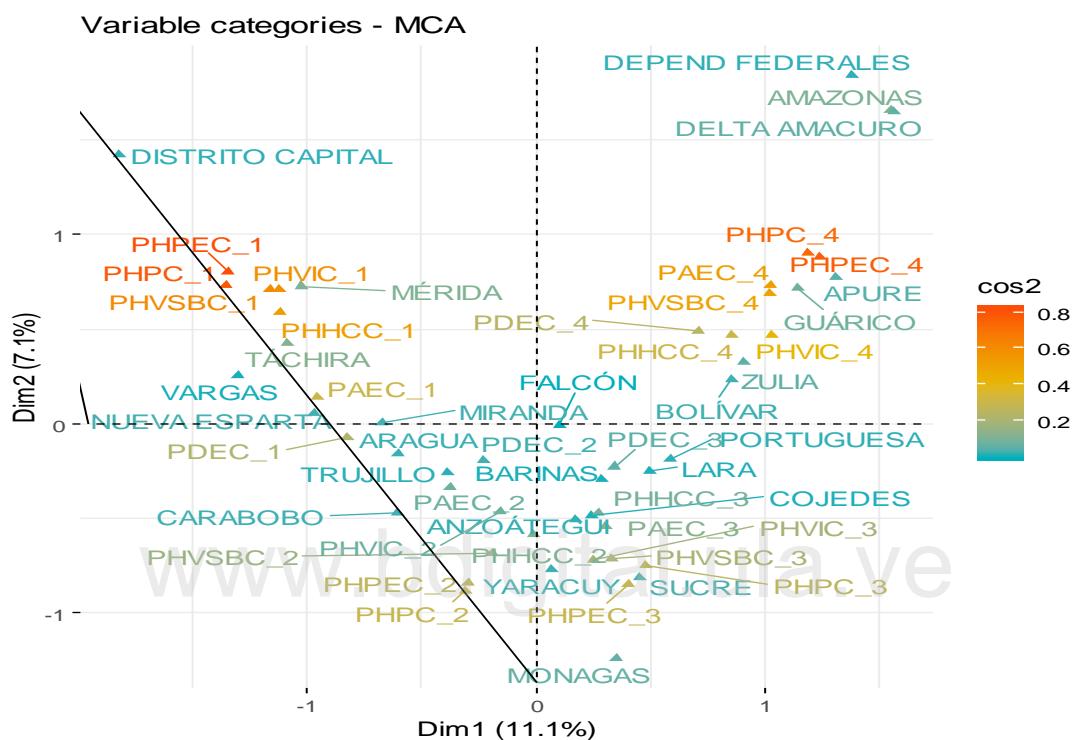
Tabla 4.7 Relación de correlación entre cada variable y cada dimensión.

##	Categorical	variables			(eta2)
		Dim.1	Dim.2	Dim.3	
##	ENTIDADDF	0.587	0.359	0.415	
##	PDEC	0.338	0.084	0.263	
##	PHVSBC	0.633	0.489	0.116	
##	PHVIC	0.599	0.365	0.113	
##	PHHCC	0.512	0.285	0.077	
##	PAEC	0.547	0.240	0.041	
##	PHPC	0.892	0.674	0.443	
##	PHPEC	0.899	0.714	0.456	

Fuente: Elaboración Propia.

Cada una de las variables presenta mayor correlación con la dimensión 1, es decir una alta asociación lineal entre cada variable y el eje 1, más sin embargo con las otras dos dimensiones también hay una asociación, a excepción de la variable PHHCC con la dimensión 3, su correlación es muy baja, es igual a 0,077.

Gráfico 4.3 Dim. 1 vs Dim. 2



Fuente: Elaboración Propia.

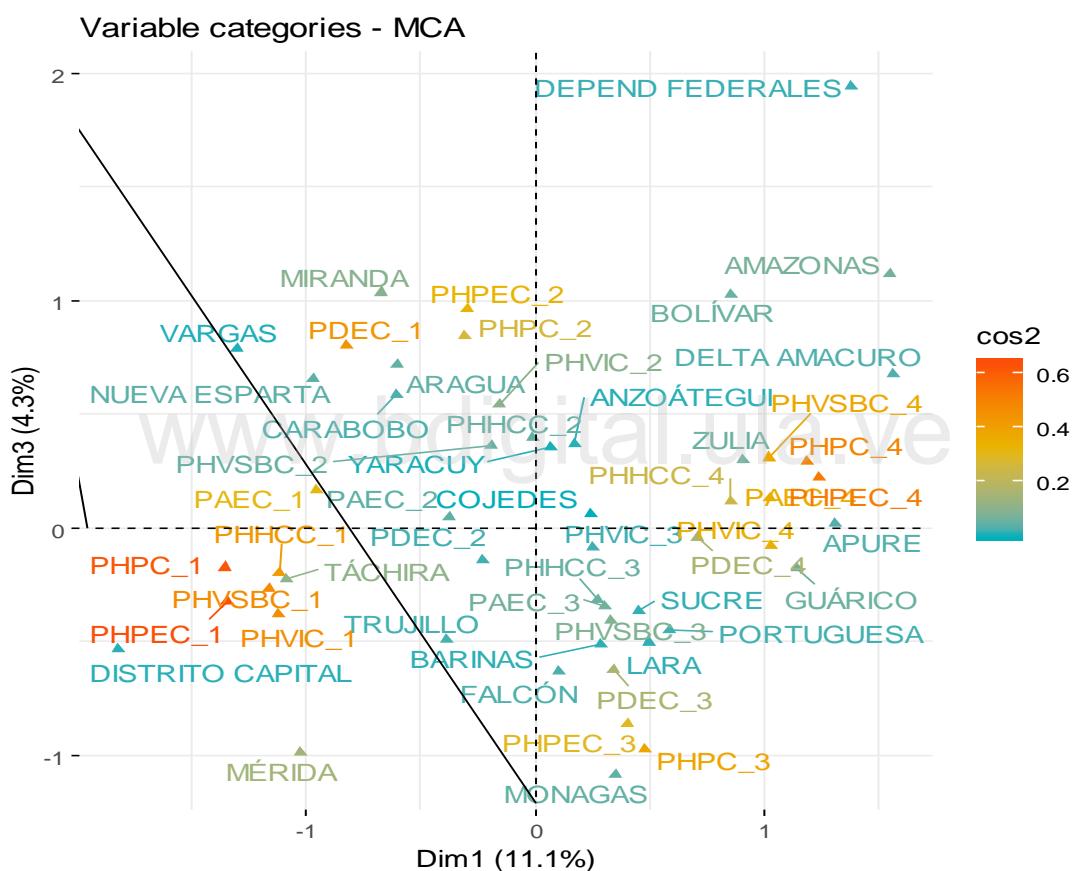
El gráfico muestra la disposición de las modalidades en el primer plano factorial. Se observa que el primer eje factorial (Dim1) están dispuestos aquellos hogares con menor proporción de hogares tanto pobres como pobres extremos, la proporción de hogares en viviendas sin servicios básicos se encuentra en el nivel 1 (hay menor cantidad de hogares con esta deficiencia), además también ocurre igual con las variables hacinamiento crítico, hogares en viviendas inadecuadas y hogares con alta dependencia económica, las cuales también se encuentran en el nivel 1, se puede observar que están los estados Mérida, Táchira y Vargas.

Del lado derecho se puede observar que están los estados Bolívar, Zulia, Apure, Guárico, Amazonas, Delta Amacuro y Dependencias Federales con una proporción de hogares pobres y pobres extremos en el nivel 4, con proporción de hogares en viviendas sin

servicios básicos, hogares en viviendas inadecuada, hacinamiento crítico, alta dependencia económica y hogares con niños que no asisten a la escuela en el nivel 4 también.

Los cosenos cuadrados indican la asociación entre la modalidad y el eje, un valor de estos cercano a 1.0 indica un ángulo de la modalidad con el respectivo eje próximo a 0. Las categorías de las variables entre rojo y naranja son la que presentan mayor asociación con los ejes 1 y 2.

Gráfico 4.4 Dim. 1 vs Dim. 3



Fuente: Elaboración Propia.

El tercer eje factorial (dimensión 3), agrupa los estados con menor pobreza, como Mérida, Táchira Trujillo, Distrito Capital y Vargas, y en contra parte los estados con mayor pobreza y pobreza extrema son Guárico, Zulia, Yaracuy, Delta Amacuro, Amazonas, Bolívar, Apure y las Dependencias Federales, con una proporción de hogares en alta dependencia económica, hogares en hacinamiento crítico, hogares en viviendas inadecuadas,

hogares con niños que no asisten a la escuela y hogares en viviendas sin servicios básicos, todas en el nivel 4.

4.2 Hogares Pobres

4.2.1 Generación de un árbol de regresión para estimar proporción de hogares pobres

Creamos nuestro conjunto de entrenamiento tomando un 70% de los datos en forma aleatoria

```
set.seed(1649)
```

```
hogaresentrena <- sample_frac(datoshogares, .7)
```

Obtenemos el conjunto de datos complementarios para la muestra de prueba, es decir, el 30% restante.

```
hogaresprueba <- setdiff(datoshogares, hogaresentrena)
```

4.2.2 Generación árbol de regresión_1

```
arbol_1 <- rpart(formula = PHP ~ ., data = hogaresentrena, method="anova")  
arbol_1
```

```
## n= 235  
##  
## node), split, n, deviance, yval  
##      * denotes terminal node  
##  
## 1) root 235 3.57149700 0.2897990  
##  2) PHVSB< 0.1386948 168 1.02122900 0.2374538  
##  4) PHHC< 0.07082446 41 0.10551000 0.1427076  
##  8) PDE< 0.06550847 28 0.03198555 0.1201149 *  
##  9) PDE>=0.06550847 13 0.02844942 0.1913689 *  
##  5) PHHC>=0.07082446 127 0.42884960 0.2680412  
## 10) PDE< 0.06298458 70 0.13167540 0.2328812  
## 20) PHVSB< 0.05912047 22 0.02256208 0.1899381 *
```

```

## 21) PHVSB>=0.05912047 48 0.04994812 0.2525635 *
## 11) PDE>=0.06298458 57 0.10436720 0.3112200
## 22) PAE< 0.02899711 48 0.04517166 0.2987789 *
## 23) PAE>=0.02899711 9 0.01214197 0.3775726 *
## 3) PHVSB>=0.1386948 67 0.93570420 0.4210525
## 6) PHVSB< 0.4090129 59 0.25119910 0.3854649
## 12) PHVSB< 0.2106126 35 0.08654281 0.3538553
## 24) PAE< 0.01821624 14 0.02687696 0.3127512 *
## 25) PAE>=0.01821624 21 0.02024315 0.3812580 *
## 13) PHVSB>=0.2106126 24 0.07868626 0.4315622 *
## 7) PHVSB>=0.4090129 8 0.05870769 0.6835109 *

```

Lo anterior muestra el esquema de nuestro árbol de regresión. Cada inciso nos indica un nodo y la regla de clasificación que le corresponde. Siguiendo estos nodos, podemos llegar a las hojas del árbol.

R nos muestra información correspondiente a cada rama del árbol: criterio de división, número de observaciones en cada rama antes de la división (n), el error o deviance(RSS), y la media de la predicción final para cada rama (yval). Con asteriscos se indican los nodos terminales.

La altura de las ramas indica la efectividad de la división en cuanto a reducción del RSS.

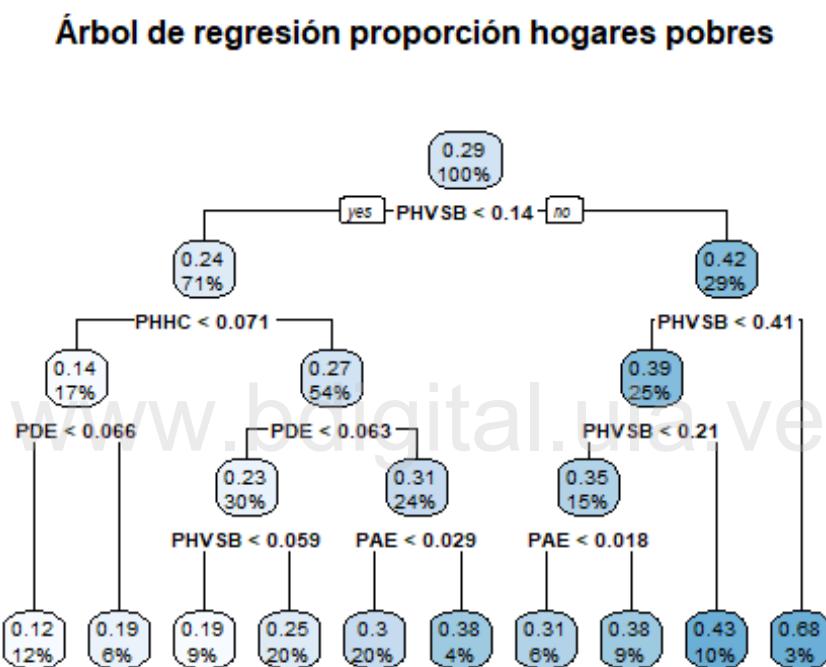
Por ejemplo, el primer nodo contiene todas las observaciones de la muestra de entrenamiento (235), con un RSS de 3.57149700 y una predicción para la proporción promedio de hogares pobres de 0.2897990. La primera división está definida por la condición PHVSB< 0.1386948 (proporción de hogares en viviendas sin servicios básicos), habiendo 168 observaciones que cumplen esta condición. El RSS para este grupo es de 1.02122900, siendo la proporción promedio de hogares pobres de este grupo igual a 0.2374538.

El modelo toma en cuenta según orden de importancia la proporción de hogares en viviendas sin servicios básicos (PHVSB), en segundo lugar la proporción de hogares en hacinamiento crítico (PHHC), en tercer lugar la proporción de hogares con alta dependencia

económica (PDE) y por último la proporción de hogares con niños que no asisten a la escuela (PAE). El modelo no considera las variables proporción de hogares en viviendas inadecuadas (PHVI), los estados ni los municipios.

Todo lo anterior resulta mucho más claro si lo visualizamos, así que creamos una gráfica usando nuestro modelo con la función rpart.plot() de rpart.plot.

Gráfico 4.5. Árbol de regresión_1



Fuente: Elaboración Propia.

En el grafico podemos observar que si $\text{PHVSB} > 0,14$ la proporción promedio de hogares pobres es igual a 0,42 (esto para un 29% de la muestra de entrenamiento), el otro 71% lo envía a la condición $\text{PHVSB} < 0,14$). Este 71% a su vez se subdivide en 17% y 54% dependiendo de la condición $\text{PHHC} < 0,071$, si es afirmativo toma el 17% y la proporción promedio de hogares pobres es igual a 0,14, si $\text{PHHC} > 0,071$ toma el 54% y la proporción promedio de hogares pobres es igual a 0,27.

Después de todo esto, debemos evaluar nuestro modelo para ver si las estimaciones para la proporción son confiables.

4.2.3 Evaluación del modelo Árbol_1

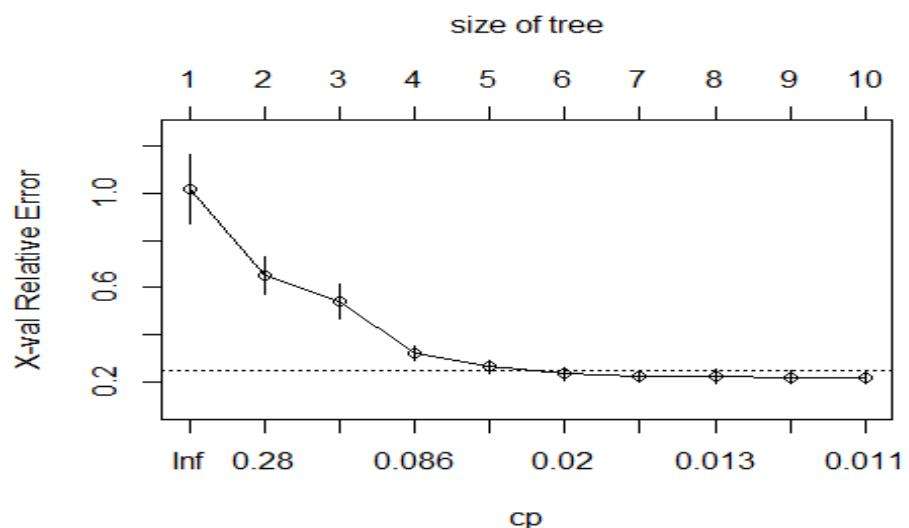
```
printcp(arbol_1)

##
## Regression tree:
## rpart(formula = PHP ~ ., data = hogaresentrena, method = "anova")
##
## Variables actually used in tree construction:
## [1] PAE  PDE  PHHC PHVSB
##
## Root node error: 3.5715/235 = 0.015198
##
## n= 235
##
##          CP    nsplit rel error xerror  xstd
## 1 0.452069      0 1.00000 1.01715 0.143700
## 2 0.175220      1 0.54793 0.65416 0.078866
## 3 0.136321      2 0.37271 0.54382 0.073984
## 4 0.053985      3 0.23639 0.32300 0.031465
## 5 0.024071      4 0.18241 0.26607 0.027753
## 6 0.016566      5 0.15833 0.23485 0.027884
## 7 0.013175      6 0.14177 0.22545 0.026099
## 8 0.012621      7 0.12859 0.22462 0.028684
## 9 0.011038      8 0.11597 0.22084 0.028406
## 10 0.010000      9 0.10493 0.21998 0.027327
```

Observando la columna xerror, el error siempre va disminuyendo y esta tendencia no cambia por lo que no será necesario podar el árbol.

Verifiquemos esto con el siguiente gráfico.

Gráfico 4.6 Evolución del error a medida que se incrementan los nodos



Fuente: Elaboración Propia.

A partir de las ramificaciones de cortes vemos cómo se comporta el error y podemos detallar que el modelo en ningún momento incrementa el error (siempre desciende).

Veamos que tan bien hace predicciones nuestro modelo.

4.2.4 Prediccion_1

```
prediccion_1 <- predict(arbol_1, newdata = hogaresprueba)
prediccion_1

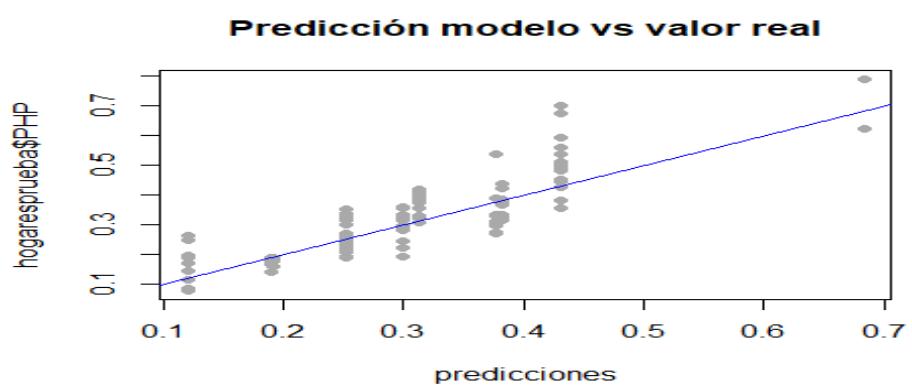
##    1         2         3         4         5         6         7
## 0.4315622 0.3812580 0.4315622 0.4315622 0.2525635 0.3812580 0.3127512
##    8         9        10        11        12        13        14
## 0.3812580 0.3812580 0.2525635 0.3127512 0.2525635 0.3127512 0.1899381
##    15        16        17        18        19        20        21
## 0.1201149 0.2987789 0.3775726 0.1899381 0.1201149 0.3775726 0.2987789
##    22        23        24        25        26        27        28
## 0.3812580 0.2525635 0.4315622 0.2525635 0.4315622 0.2525635 0.2525635
##    29        30        31        32        33        34        35
## 0.2525635 0.1899381 0.2525635 0.2525635 0.4315622 0.2525635 0.3812580
##    36        37        38        39        40        41        42
## 0.6835109 0.3812580 0.2987789 0.1899381 0.2525635 0.3775726 0.3775726
```

```

##   43      44      45      46      47      48      49
## 0.4315622 0.3127512 0.3127512 0.3812580 0.3127512 0.4315622 0.3812580
##   50      51      52      53      54      55      56
## 0.4315622 0.2987789 0.3775726 0.1201149 0.2987789 0.1201149 0.2987789
##   57      58      59      60      61      62      63
## 0.1913689 0.1201149 0.2525635 0.2525635 0.2525635 0.3127512 0.4315622
##   64      65      66      67      68      69      70
## 0.1899381 0.2525635 0.2525635 0.2987789 0.2987789 0.3775726 0.2525635
##   71      72      73      74      75      76      77
## 0.1201149 0.4315622 0.1899381 0.2987789 0.2987789 0.3127512 0.3127512
##   78      79      80      81      82      83      84
## 0.2987789 0.3127512 0.1201149 0.2987789 0.1913689 0.1913689 0.1913689
##   85      86      87      88      89      90      91
## 0.1201149 0.1913689 0.3775726 0.2987789 0.1201149 0.1899381 0.3775726
##   92      93      94      95      96      97      98
## 0.2525635 0.2525635 0.3127512 0.2987789 0.3775726 0.3775726 0.4315622
##   99     100     101
## 0.4315622 0.6835109 0.4315622

```

Gráfico 4.7 Predicciones vs los valores reales



Fuente: Elaboración Propia.

En el eje de las abscisas se encuentran las predicciones y en el eje de las ordenadas los valores reales para la muestra de prueba. Mientras mejor se ajuste la nube de puntos a la línea recta, mejores son las predicciones.

```
test.MSE.arbol_1 <- mean((prediccion_1 - hogaresprueba$PHP)^2)
test.MSE.arbol_1
## [1] 0.004947057
```

El test MSE correspondiente a este árbol de regresión simple árbol_1 es de 0,004947057 unidades, lo cual equivale a decir que las predicciones del modelo se alejan de los valores reales en $\sqrt{0,004947057} = 0,07034$ unidades, en promedio.

4.2.5 Bagging

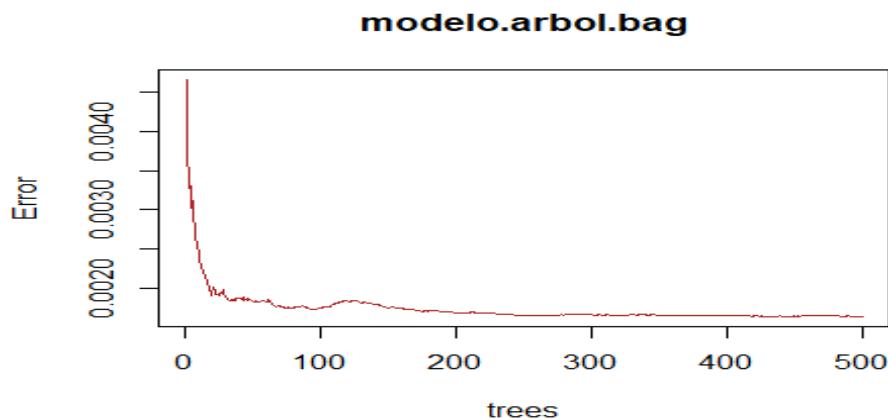
Ajuste del modelo

Para intentar mejorar la capacidad predictiva del árbol obtenido en el paso inicial, aplicaremos el método de bagging. El conjunto de datos cuenta con 8 variables, 7 predictores y 1 variable respuesta. Por tanto, para aplicar bagging todos los predictores han de considerarse en cada partición. El número de pseudo-árboles que se generan por defecto son 500 (no se recomienda usar un número pequeño, para asegurar que todas o casi todas las observaciones tengan participación):

```
modelo.arbol.bagging
##
## Call:
## randomForest(formula = PHP ~ ., data = hogaresentrena, mtry = 7,      importance =
## TRUE, ntree = 500)
##       Type of random forest: regression
##                 Number of trees: 500
## No. of variables tried at each split: 7
##
##       Mean of squared residuals: 0.001643498
##       % Var explained: 89.19
```

El porcentaje de la varianza explicada por el modelo aplicando bagging es del 89,19%. El Mean of squared residuals se corresponde con el out-of-bag-MSE, que puede entenderse como un test error. Su evolución respecto al número de árboles puede representarse de la siguiente manera:

Gráfico 4.8 Evolución del error respecto al número de árboles



Fuente: Elaboración Propia.

El error se estabiliza al alcanzar los 200 árboles aproximadamente.

4.2.6 Identificación de los predictores más importantes

Al combinar múltiples árboles perdemos la posibilidad de obtener una representación gráfica como en el caso de un árbol único. Sin embargo, podemos utilizar la función importance() para obtener la importancia de cada predictor:

```
importance(modelo.arbol.bag)

##          %IncMSE IncNodePurity
## ENTIDADADF 9.3572519  0.1667323
## MUNICIP -0.7095705  0.1297607
## PAE     16.8223645  0.1096165
## PDE     32.7108088  0.1733797
## PHVSB  60.6882546  2.3857957
```

```
## PHVI 24.8056770 0.2712562
## PHHC 23.2382593 0.3072628
```

Para cada predictor se devuelven dos valores:

- %IncMSE: disminución media de la precisión de las predicciones sobre las muestras cuando la variable dada se excluye del modelo.
- IncNodePurity: medida de la disminución total de impureza de los nodos (medida por el training RSS) que resulta de la división de la variable dada.

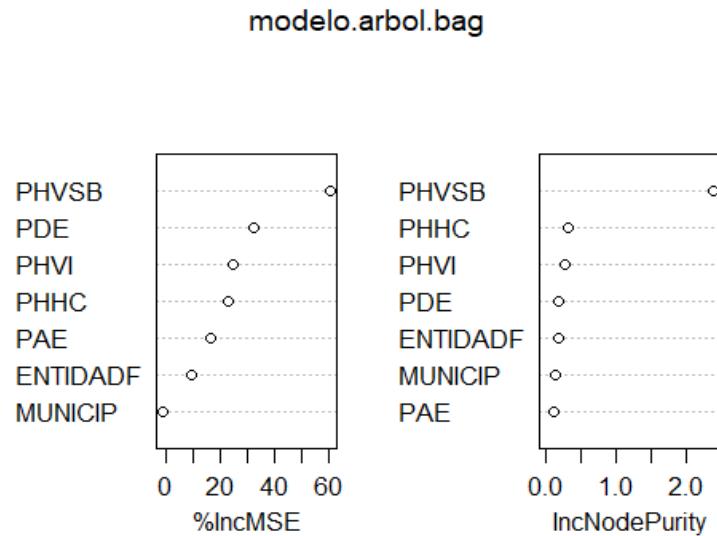
Los predictores más importantes se corresponderán a aquellos con mayor %IncMSE y IncNodePurity.

En este caso las variables más importantes son PHVSB, PDE, PHVI, PHHC, PAE, la entidad y el municipio son las menos importantes según este modelo.

Visualicémoslo gráficamente

```
varImpPlot(modelo.arbol.bag)
```

Gráfico 4.9 Importancia de las variables



Fuente: Elaboración Propia.

Sin duda alguna la variable más relevante es la proporción de hogares en viviendas sin servicios básicos (PHVSB) al igual que en el árbol 1 o árbol simple, más sin embargo en el árbol simple no se incluyó la proporción de hogares en viviendas inadecuadas (PHVI), mientras que en este modelo obtenido por el método de Bagging si se incluye como una variable importante (de tercer lugar). La entidad y el municipio nuevamente vuelven a quedar descartadas (al igual que en el modelo generado con el árbol simple).

4.2.7 Predicción con el modelo Baggin_2

```
prediccion_2 <- predict(modelo.arbol.bag, newdata = hogaresprueba)
```

```
prediccion_2
```

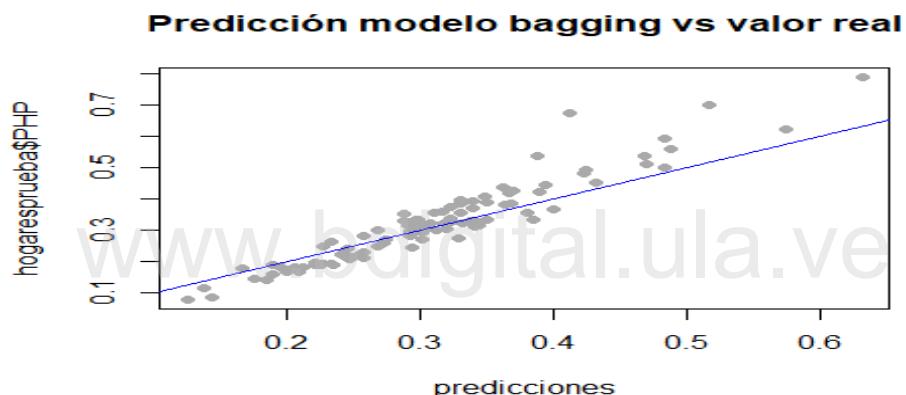
```
##      1      2      3      4      5      6      7
## 0.4325786 0.3998285 0.5174608 0.3810127 0.2909511 0.3636001 0.3183519
##      8      9     10     11     12     13     14
## 0.3448883 0.3889082 0.3223598 0.3407918 0.2748714 0.3310929 0.2004339
##     15     16     17     18     19     20     21
## 0.2269665 0.2934601 0.3887914 0.2118462 0.2213448 0.3014811 0.3174986
##     22     23     24     25     26     27     28
## 0.3621952 0.2573407 0.3938040 0.2682496 0.4694570 0.2452667 0.2881678
##     29     30     31     32     33     34     35
## 0.2879047 0.2087907 0.2410405 0.2727323 0.4242400 0.2947981 0.3322616
##     36     37     38     39     40     41     42
## 0.5751445 0.3844027 0.2554083 0.1959525 0.2459383 0.3294029 0.3409000
##     43     44     45     46     47     48     49
## 0.4125027 0.3227971 0.3305445 0.3389024 0.3670841 0.4678883 0.3683381
##     50     51     52     53     54     55     56
## 0.4841820 0.3079434 0.3413615 0.1253330 0.2572210 0.1385146 0.2334114
##     57     58     59     60     61     62     63
## 0.2062986 0.1447944 0.2469953 0.2451885 0.2281802 0.3492541 0.3700331
##     64     65     66     67     68     69     70
## 0.1857238 0.2484797 0.2684354 0.2998257 0.3022250 0.3503434 0.3117607
##     71     72     73     74     75     76     77
```

```

##  0.2099844  0.4233291  0.2036038  0.3114242  0.2919134  0.3399005  0.3306438
##    78        79        80        81        82        83        84
##  0.2974527  0.3229020  0.2220212  0.2920702  0.1901893  0.1971780  0.1891479
##    85        86        87        88        89        90        91
##  0.2330507  0.1673857  0.3378836  0.2964808  0.1758077  0.2252970  0.3017854
##    92        93        94        95        96        97        98
##  0.2354090  0.2571902  0.3042582  0.2983200  0.3496566  0.3200658  0.4884891
##                99                  100
##  0.4829535 0.6317408 0.3639889

```

Gráfico 4.10 Predicciones modelo Bagging vs valores reales



Fuente: Elaboración Propia.

Observamos que la nube de puntos se ajusta mejor a la línea recta con el método de Bagging, en comparación con el modelo obtenido por el árbol 1 o árbol simple.

```

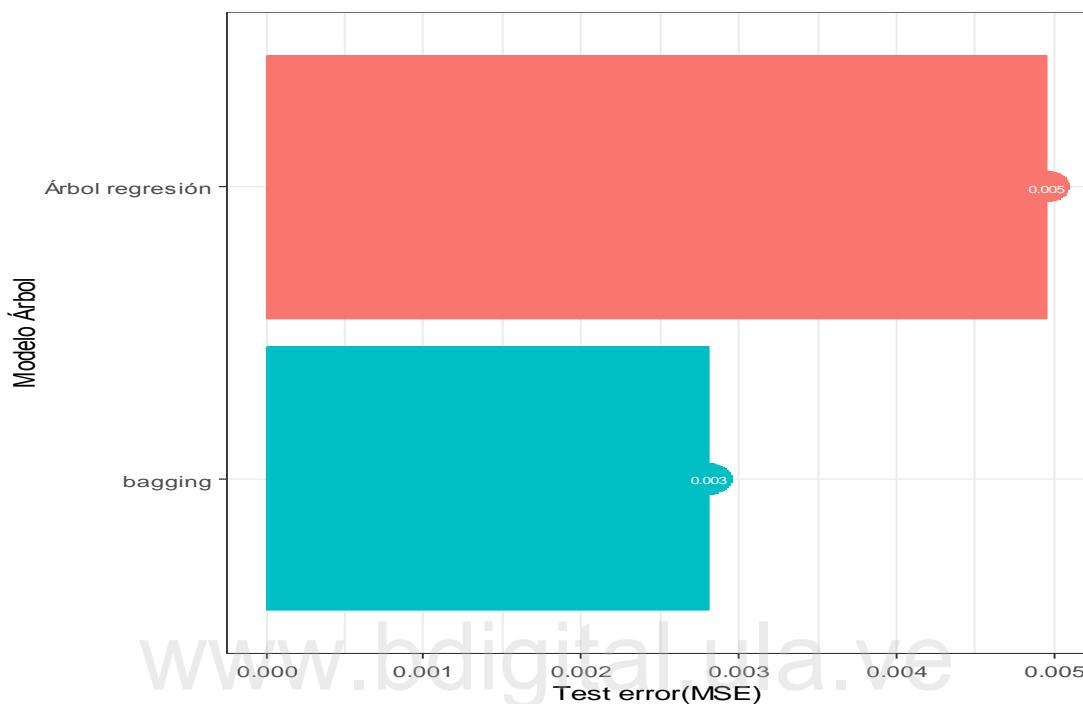
test.MSE.arbol_2      <-      mean((prediccion_2      -      hogaresprueba$PHP)^2)
test.MSE.arbol_2
## [1] 0.002811441

```

Se ha conseguido reducir el test error de 0,004947057 del árbol simple a 0,002811441 con el método de bagging. Las predicciones del modelo Bagging se alejan de los valores reales en $\sqrt{0,002811441} = 0,05302$ unidades, en promedio.

Veamos la comparación del error para ambos modelos (modelo de árbol simple y modelo Bagging) gráficamente.

Gráfico 4.11 Comparación del error con ambos modelos



Fuente: Elaboración Propia.

Con el gráfico observamos que la reducción del error es significativa, por lo tanto las predicciones con el método de Bagging son más confiables que las obtenidas con el árbol simple.

4.3 Hogares en Pobreza Extrema

```
datoshogares2<-
data.frame(PHPE,ENTIDADDF,MUNICIP,PAE,PDE,PHVSB,PHVI,PHHC)
```

Generación de un árbol de regresión para estimar proporción de hogares pobres extremos.

Creamos nuestro conjunto de entrenamiento tomando un 70% de los datos en forma aleatoria y luego el 30% restante para la muestra de prueba.

```
hogaresentrena2 <- sample_frac(datoshogares2, .7)
hogaresprueba2 <- setdiff(datoshogares2, hogaresentrena2)
```

4.3.1 Generación del árbol de regresión_3

```
arbol_3

## n= 235
##
## node), split, n, deviance, yval
##      * denotes terminal node
##
## 1) root 235 1.010163000 0.08670352
## 2) PHVSB< 0.4090129 227 0.472420700 0.07875663
## 4) PHVI< 0.1170569 151 0.147505800 0.05644718
## 8) PHVSB< 0.05141247 59 0.013281280 0.02798438 *
## 9) PHVSB>=0.05141247 92 0.055773910 0.07470049
## 18)
ENTIDADF=ARAGUA,BOLÃVAR,CARABOBO,MÃ‰RIDA,MIRANDA,MONAGA
S,NUEVA ESPARTA,PORTUGUESA,SUCRE,TÃCHIRA,TRUJILLO 49 0.016535290
0.06179668 *
## 19)
ENTIDADF=ANZOÃTEGUI,APURE,COJEDES,FALCÃN,GUÃRICO,LARA,YAR
ACUY,ZULIA 43 0.021782340 0.08940484 *
## 5) PHVI>=0.1170569 76 0.100440500 0.12308200
## 10) PHVSB< 0.1602364 52 0.025282880 0.10550500
## 20) PAE< 0.03406567 44 0.012463000 0.09934705 *
## 21) PAE>=0.03406567 8 0.001974718 0.13937360 *
## 11) PHVSB>=0.1602364 24 0.024283750 0.16116550
## 22) PHVI< 0.1871321 14 0.006192722 0.14246120 *
## 23) PHVI>=0.1871321 10 0.006336131 0.18735140 *
## 3) PHVSB>=0.4090129 8 0.116630000 0.31219640 *
```

Por ejemplo, el primer nodo contiene todas las observaciones de la muestra de entrenamiento (235), con un RSS de 1,010163000 y una predicción para la proporción promedio de hogares pobres extremos de 0.08670352. La primera división está definida por la condición PHVSB<0.4090129 (proporción de hogares en viviendas sin servicios básicos), habiendo 227 observaciones que cumplen esta condición. El RSS para este grupo es de 0,472420700, siendo la proporción promedio de hogares pobres extremo de este grupo igual a 0.07875663.

El modelo toma en cuenta según orden de importancia la proporción de hogares en viviendas sin servicios básicos (PHVSB), en segundo lugar la proporción de hogares en viviendas inadecuadas (PHVI), en tercer lugar la ENTIDAD y por último la proporción de hogares con niños que no asisten a la escuela (PAE).

4.3.2 Evaluación del modelo Árbol de regresión hogares pobres extremos

```
printcp(arbol_3)

##
## Regression tree:
## rpart(formula = PHPE ~ ., data = hogaresentrena2, method = "anova")

##
## Variables actually used in tree construction:
## [1] ENTIDADF PAE    PHVI   PHVSB

##
## Root node error: 1.0102/235 = 0.0042986

##
## n= 235

##
##      CP nsplit rel error xerror  xstd
## 1 0.416876   0  1.00000 1.01490 0.26571
## 2 0.222216   1  0.58312 0.68089 0.14549
## 3 0.077661   2  0.36091 0.44425 0.11137
## 4 0.050362   3  0.28325 0.35912 0.11146
## 5 0.017281   4  0.23289 0.31877 0.10972
```

```

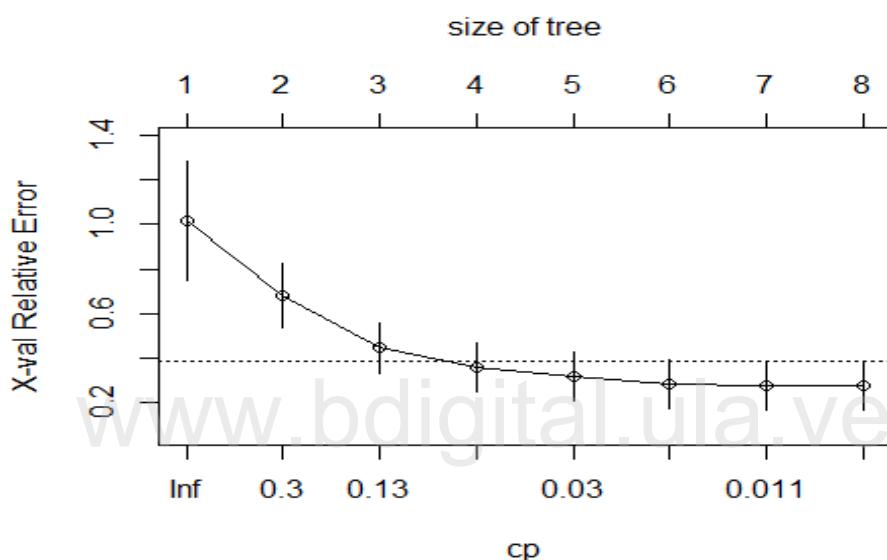
## 6 0.011637 5 0.21560 0.28015 0.10940
## 7 0.010736 6 0.20397 0.27378 0.10942
## 8 0.010000 7 0.19323 0.27447 0.10953

```

Observando la columna xerror, el error siempre va disminuyendo hasta que llega al valor 0,27378 y luego aumenta a 0,27447, lo que nos indica que será necesario podar el árbol.

Verifiquemos esto con el siguiente gráfico.

Gráfico 4.12. Evolución del error a medida que se incrementan los nodos



Fuente: Elaboración Propia.

4.3.3 Podar del árbol_3

```

parbol3<-prune(arbol_3,
cp=arbol_3$cptable[which.min(arbol_3$cptable[, "xerror"]),"CP"])
parbol3

## n= 235
##
## node), split, n, deviance, yval
##      * denotes terminal node
##
## 1) root 235 1.010163000 0.08670352

```

```

## 2) PHVSB< 0.4090129 227 0.472420700 0.07875663
## 4) PHVI< 0.1170569 151 0.147505800 0.05644718
## 8) PHVSB< 0.05141247 59 0.013281280 0.02798438 *
## 9) PHVSB>=0.05141247 92 0.055773910 0.07470049
## 18)
ENTIDADF=ARAGUA,BOLÃVAR,CARABOBO,MÃRIDA,MIRANDA,MONAGA
S,NUEVA ESPARTA,PORTUGUESA,SUCRE,TÃCHIRA,TRUJILLO 49 0.016535290
0.06179668 *

## 19)
ENTIDADF=ANZOÃTEGUI,APURE,COJEDES,FALCÃN,GUÃRICO,LARA,YAR
ACUY,ZULIA 43 0.021782340 0.08940484 *

## 5) PHVI>=0.1170569 76 0.100440500 0.12308200
## 10) PHVSB< 0.1602364 52 0.025282880 0.10550500 *
## 11) PHVSB>=0.1602364 24 0.024283750 0.16116550
## 22) PHVI< 0.1871321 14 0.006192722 0.14246120 *
## 23) PHVI>=0.1871321 10 0.006336131 0.18735140 *
## 3) PHVSB>=0.4090129 8 0.116630000 0.31219640 *

```

En este modelo con el árbol podado se excluye la proporción de hogares con niños que no asisten a la escuela.

4.3.4 Predicciones del modelo árbol podado

```

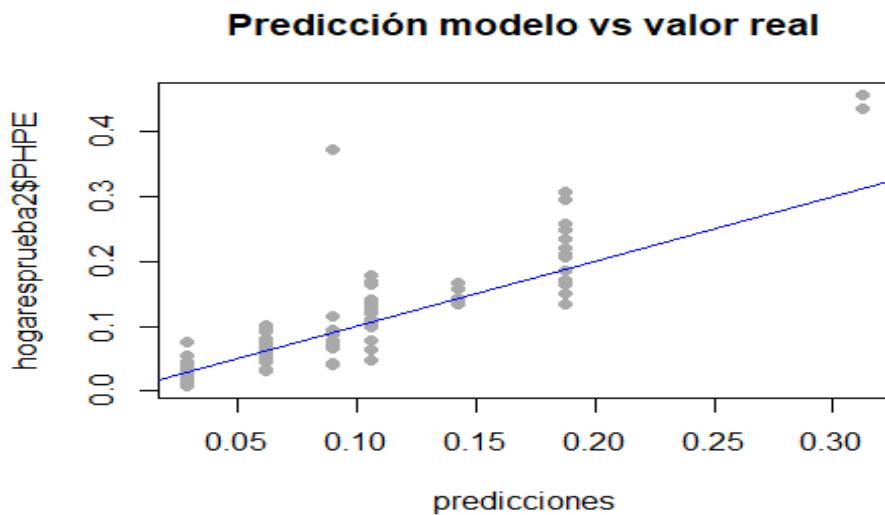
prediccion_4      <-      predict(parbol3,      newdata      =      hogaresprueba2)
prediccion_4

##          1          2          3          4          5          6
## 0.14246124  0.14246124  0.08940484  0.08940484  0.08940484  0.18735139
##          7          8          9          10         11         12
## 0.08940484  0.08940484  0.10550498  0.10550498  0.06179668  0.10550498
##         13         14         15         16         17         18
## 0.18735139  0.02798438  0.10550498  0.06179668  0.10550498  0.02798438
##         19         20         21         22         23         24
## 0.02798438  0.06179668  0.10550498  0.14246124  0.06179668  0.18735139

```

##	25	26	27	28	29	30
##	0.10550498	0.18735139	0.06179668	0.10550498	0.10550498	0.02798438
##	31	32	33	34	35	36
##	0.06179668	0.10550498	0.18735139	0.10550498	0.08940484	0.31219644
##	37	38	39	40	41	42
##	0.14246124	0.08940484	0.02798438	0.08940484	0.08940484	0.08940484
##	43	44	45	46	47	48
##	0.18735139	0.10550498	0.14246124	0.10550498	0.18735139	0.18735139
##	49	50	51	52	53	54
##	0.14246124	0.18735139	0.08940484	0.10550498	0.02798438	0.02798438
##	55	56	57	58	59	60
##	0.02798438	0.02798438	0.02798438	0.02798438	0.06179668	0.06179668
##	61	62	63	64	65	66
##	0.06179668	0.18735139	0.14246124	0.02798438	0.06179668	0.06179668
##	67	68	69	70	71	72
##	0.10550498	0.10550498	0.10550498	0.06179668	0.06179668	0.18735139
##	73	74	75	76	77	78
##	0.02798438	0.10550498	0.10550498	0.10550498	0.10550498	0.10550498
##	79	80	81	82	83	84
##	0.10550498	0.06179668	0.06179668	0.02798438	0.02798438	0.02798438
##	85	86	87	88	89	90
##	0.06179668	0.02798438	0.06179668	0.06179668	0.02798438	0.02798438
##	91	92	93	94	95	96
##	0.06179668	0.06179668	0.08940484	0.08940484	0.10550498	0.10550498
##	97	98	99	100	101	
##	0.02798438	0.18735139	0.18735139	0.31219644	0.08940484	

Gráfico 4.13 Predicciones modelo árbol podado vs valores reales



Fuente: Elaboración Propia.

```
test.MSE.arbol_4 <- mean((prediccion_4 - hogaresprueba2$PHP)^2)
```

```
test.MSE.arbol_4
```

```
## [1] 0.002045094
```

Las predicciones del modelo árbol podado se alejan de los valores reales en $\sqrt{0,002045094} = 0,045222715$ unidades, en promedio.

4.3.5 Generación del modelo por el método de Bagging

```
modelo.arbol.bag2 <- randomForest(PHPE ~ ., data = hogaresentrena2,
mtry = 7,
importance = TRUE,
ntree = 500)
modelo.arbol.bag2

##
## Call:
## randomForest(formula = PHPE ~ ., data = hogaresentrena2, mtry = 7, importance =
TRUE, ntree = 500)
```

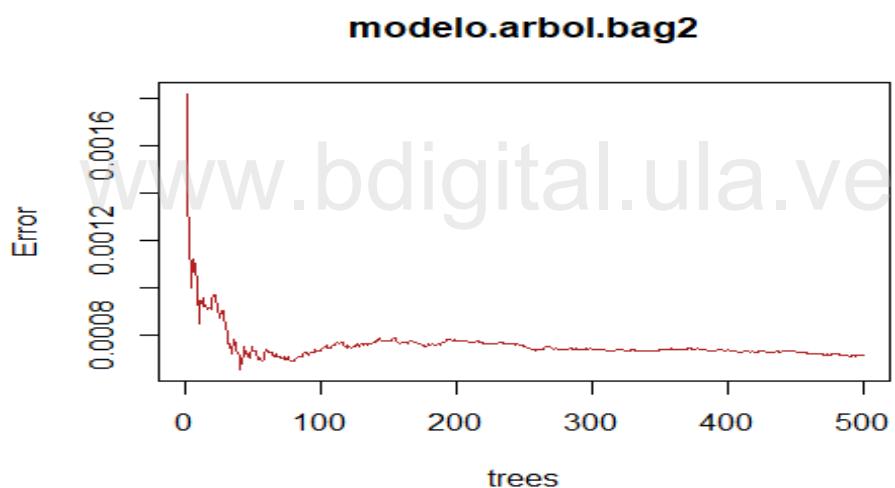
```

##      Type of random forest: regression
##      Number of trees: 500
## No. of variables tried at each split: 7
##
##      Mean of squared residuals: 0.0007143806
##      % Var explained: 83.38

```

El porcentaje de la varianza explicada por el modelo aplicando bagging es del 83,38%. El Mean of squared residuals se corresponde con el out-of-bag-MSE, que puede entenderse como un test error. Su evolución respecto al número de árboles lo vemos en el gráfico siguiente:

Gráfico 4.14 Evolución del error respecto al número de árboles



Fuente: Elaboración Propia.

El error se estabiliza al alcanzar los 300 árboles aproximadamente.

4.3.6 Identificación de los predictores más importantes

```

importance(modelo.arbol.bag2)

##      %IncMSE IncNodePurity
## ENTIDADADF 14.609827  0.24304309
## MUNICIP  2.923435   0.03537000
## PAE     12.145115   0.07532529

```

```

## PDE    8.014288  0.04021091
## PHVSB  28.515245  0.29733538
## PHVI   50.336437  0.20141611
## PHHC   10.455946  0.11056566

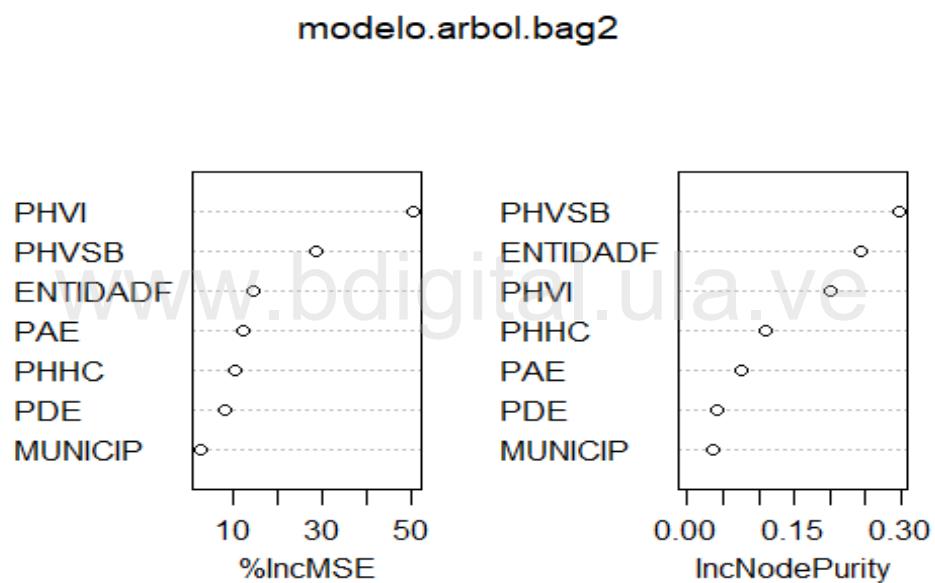
```

En este caso las variables más importantes son PHVI, PHVSB, ENTIDADDF, PHHC y las menos importantes son PDE y MUNICIPIO según este modelo.

Visualicémoslo gráficamente

```
varImpPlot(modelo.arbol.bag2)
```

Gráfico 4.15 Importancia de las variables



Fuente: Elaboración Propia.

```

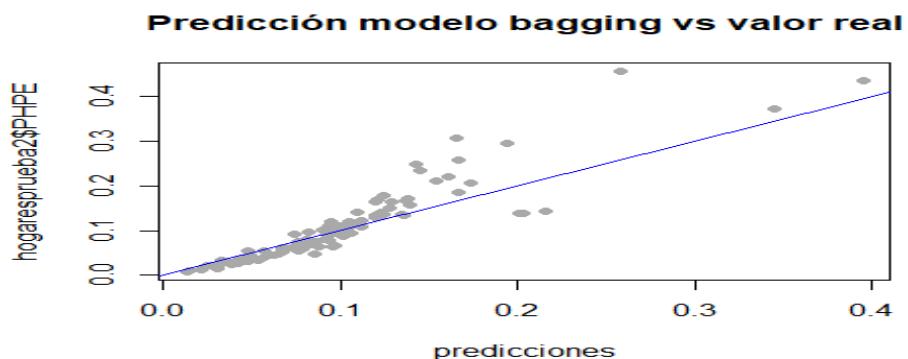
prediccion_5    <-    predict(modelo.arbol.bag2,    newdata    =    hogaresprueba2)
prediccion_5

##          1          2          3          4          5          6
## 0.21538864  0.20052663  0.34541852  0.08703038  0.09427589  0.12844066
##          7          8          9         10         11         12
## 0.09384766  0.09688232  0.13501922  0.12014881  0.09082186  0.08812572
##         13         14         15         16         17         18
## 0.12308944  0.04175945  0.08584763  0.07830942  0.12123000  0.04910155

```

##	19	20	21	22	23	24
##	0.06757798	0.06674261	0.09452072	0.13957029	0.07906410	0.16605635
##	25	26	27	28	29	30
##	0.09840755	0.17376555	0.07666873	0.09359605	0.10319614	0.04218086
##	31	32	33	34	35	36
##	0.06532012	0.08365506	0.14261318	0.10466177	0.09186795	0.39500527
##	37	38	39	40	41	42
##	0.20343087	0.04802821	0.02528470	0.05670858	0.08235328	0.10638565
##	43	44	45	46	47	48
##	0.16504152	0.09974119	0.12362383	0.12388464	0.13794288	0.14501414
##	49	50	51	52	53	54
##	0.13537889	0.16086089	0.10160620	0.09522835	0.01296930	0.04789935
##	55	56	57	58	59	60
##	0.01428405	0.03864650	0.03868611	0.02166320	0.06752875	0.07362359
##	61	62	63	64	65	66
##	0.06279810	0.12753394	0.13702575	0.03162185	0.06812298	0.07606192
##	67	68	69	70	71	72
##	0.09225087	0.09370999	0.12399777	0.09926773	0.04792238	0.15445815
##	73	74	75	76	77	78
##	0.03307989	0.11154782	0.09695140	0.11973452	0.11950140	0.09936268
##	79	80	81	82	83	84
##	0.10991590	0.05712170	0.08062462	0.02249372	0.02794412	0.03082585
##	85	86	87	88	89	90
##	0.08517458	0.02687001	0.08229287	0.07443080	0.03474453	0.04501036
##	91	92	93	94	95	96
##	0.07457310	0.05290837	0.07502938	0.09277066	0.09018142	0.11190963
##	97	98	99	100		101
##	0.07679294	0.16630988	0.19368123	0.25799906	0.10719669	

Gráfico 4.16 Predicción de hogares pobres extremos vs valores reales



Fuente: Elaboración Propia.

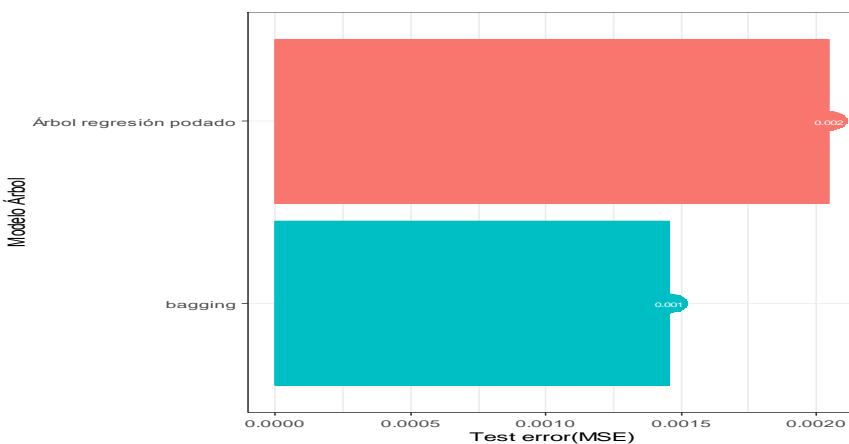
La nube de puntos se ajusta a la línea recta.

```
test.MSE.arbol_5 <- mean((prediccion_5 - hogaresprueba2$PHPE)^2)
test.MSE.arbol_5

## [1] 0.001452999
```

Se ha conseguido reducir el test error de 0,002045094 del árbol podado a 0,001452999 con el método de bagging. Las predicciones del modelo Bagging se alejan de los valores reales en $\sqrt{0,001452999} = 0,038118223$ unidades, en promedio.

Gráfico 4.17 Comparación del error (modelo de árbol podado y modelo Bagging).



Fuente: Elaboración Propia.

Se reduce significativamente el error de predicción con el método de Bagging.

Tabla 4.8 Valor de la Canasta Alimentaria Normativa

Valor de la canasta alimentaria normativa, según mes 2008-2014
(Montos en Bolívares)

Meses	2008		2009		2010		2011		2012		2013		2014	
	(Monto)	Δ%	(Monto)	Δ%	(Monto)	Δ%	(Monto)	Δ%	(Monto)	Δ%	(Monto)	Δ%	(Monto)	Δ%
Enero	749,30	--	915,40	4,37	1.102,58	2,96	1.423,99	3,87	1.764,12	1,31	2.175,66	4,34	3.640,55	9,51
Febrero	772,27	3,07	917,61	0,24	1.117,24	1,33	1.445,99	1,54	1.772,01	0,45	2.174,43	-0,06	3.730,48	2,47
Marzo	776,29	0,52	923,00	0,59	1.158,93	3,73	1.452,20	0,43	1.769,05	-0,17	2.266,04	4,21	3.946,64	5,79
Abril	763,95	(1,59)	918,56	(0,48)	1.271,68	9,73	1.458,00	0,40	1.781,18	0,69	2.411,93	6,44	4.189,54	6,15
Mayo	791,02	3,54	903,69	(1,62)	1.269,68	(0,16)	1.486,59	1,96	1.811,89	1,72	2.620,62	8,65	4.448,56	6,18
Junio	803,65	1,60	906,22	0,28	1.298,76	2,29	1.518,23	2,13	1.831,55	1,09	2.737,07	4,44	4.741,50	6,58
Julio	851,09	5,80	973,88	7,47	1.312,35	1,05	1.576,11	3,81	1.822,32	-0,50	2.779,21	1,54	4.876,99	2,86
Agosto	820,03	(3,65)	995,92	2,26	1.330,79	1,41	1.603,98	1,77	1.835,28	0,71	2.915,28	4,90	5.391,35	10,55
Septiembre	813,42	(0,81)	1.029,01	3,32	1.334,67	0,29	1.635,40	1,96	1.881,96	2,54	3.054,84	4,79	5.741,06	6,49
Octubre	827,46	1,73	1.053,73	2,40	1.353,27	1,39	1.667,96	1,99	1.936,98	2,92	3.161,37	3,49	5.978,76	5,01
Noviembre	853,82	3,19	1.067,92	1,35	1.359,37	0,45	1.710,41	2,55	1.989,09	2,69	3.347,28	5,88	6.382,62	6,75
Diciembre	877,04	2,72	1.070,89	0,28	1.370,93	0,85	1.741,29	1,81	2.085,22	4,83	3.324,41	-0,68		

Nota: Incluye diez áreas Metropolitanas del país: Caracas, Maracaibo, Maracay, Valencia, Barquisimeto, Ciudad Guayana, Maturín, Mérida, San Cristóbal, Barcelona-Pto. La Cruz y un Dominio denominado Resto Nacional conformado por 72 localidades entre ciudades medianas, pequeñas y áreas rurales del país.

Fuente:<http://www.ine.gov.ve/documentos/Economia/CanastaNormativaAlimentaria/html/CANMes.html>

El valor promedio de la CAN para el año 2011 es de 1560,0125 BsF.

Tabla 4.9 Número de Hogares Pobres Según la Línea de Pobreza

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	No pobre	5.215.312	75,4	75,4	75,4
	Pobre relativo	1.216.498	17,6	17,6	93,0
	Pobre extremo	482.147	7,0	7,0	100,0
	Total	6.913.957	100,0	100,0	

Fuente: Elaboración Propia.

Tabla 4.10 Número de Hogares Pobres Según Necesidades Básicas Insatisfechas

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos No pobre	3.196.216	46,2	46,2	46,2
Pobre relativo	2.037.214	29,5	29,5	75,7
Pobre extremo	1.680.527	24,3	24,3	100,0
Total	6.913.957	100,0	100,0	

Fuente: Elaboración Propia.

Tabla 4.11 Indicador FGT

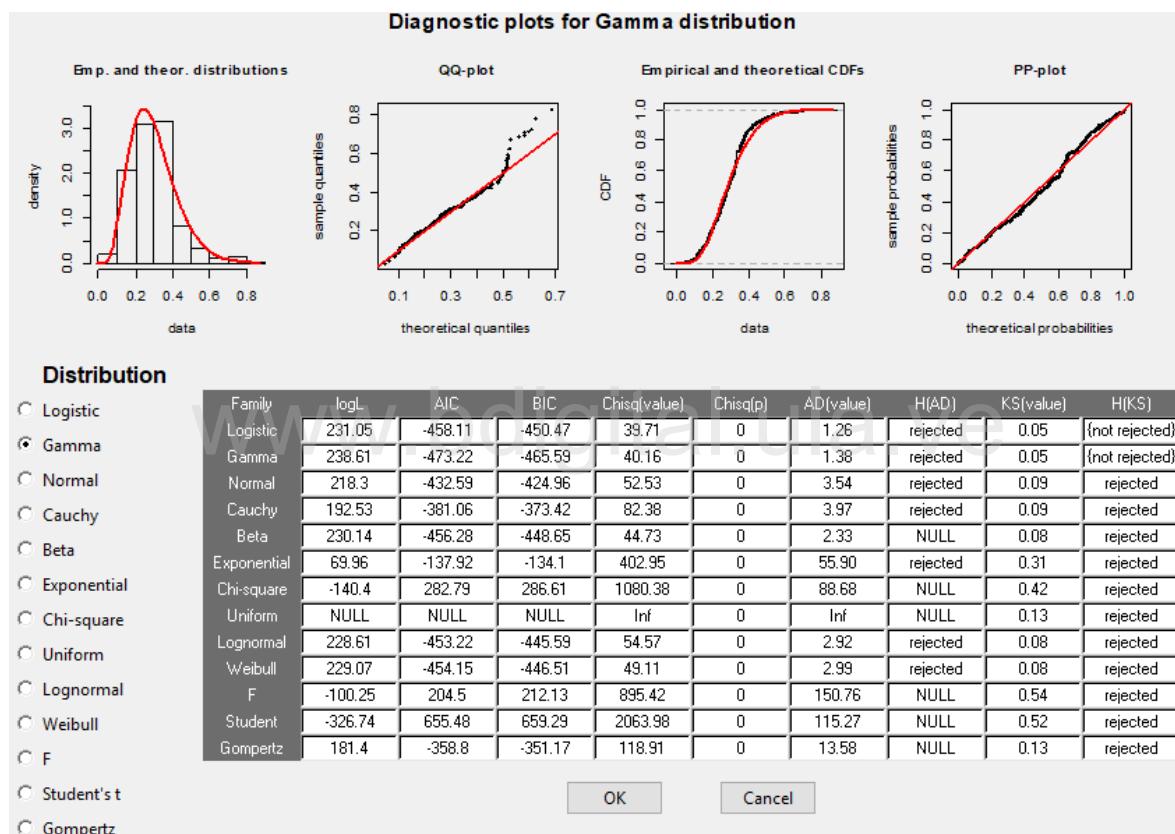
Índice FGT para $\alpha = 2$

ESTADOS	Índice FGT2	ESTADOS	Índice FGT2
Distrito Capital	0,053416568	Lara	0,067464637
Amazonas	0,113473782	Mérida	0,040167678
Anzoátegui	0,10257875	Miranda	0,042425256
Apure	0,132445679	Monagas	0,094077846
Aragua	0,0654656778	Nueva Esparta	0,080119826
Barinas	0,053321476	Portuguesa	0,250975126
Bolívar	0,056435678	Sucre	0,1563702627
Carabobo	0,057986545	Táchira	0,06756278
Cojedes	0,182456087	Trujillo	0,06896267
Delta Amacuro	0,095421207	Yaracuy	0,116526278
Falcón	0,161207382	Zulia	0,180487272
Guárico	0,317134598	Vargas	0,044616896
Dependencias Federales	0,250789871		

Fuente: Elaboración Propia.

Este índice FGT será mayor cuanto más grande sea la desigualdad entre los pobres. Si observamos los resultados el estado Guárico es quien presenta un mayor índice indicándonos que es el que muestra mayor desigualdad entre hogares pobres, seguido a este se encuentran los estados Portuguesa, Dependencias Federales, Zulia, Falcón y Sucre. Con menor desigualdad de pobreza entre hogares se encuentran los estados Mérida, Miranda, el Distrito Capital, Vargas, Táchira y Trujillo.

Gráfico 4.18 Distribución de probabilidad de la proporción de hogares pobres. (Distribución Gamma)



Fuente: Elaboración Propia.

Tabla 4.12 Parámetros de la distribución Gamma (proporción de hogares pobres)

```

Chosen continuous distribution is: Gamma (gamma)
Fitted parameters are:
  shape      rate
  5.565065 18.630112

```

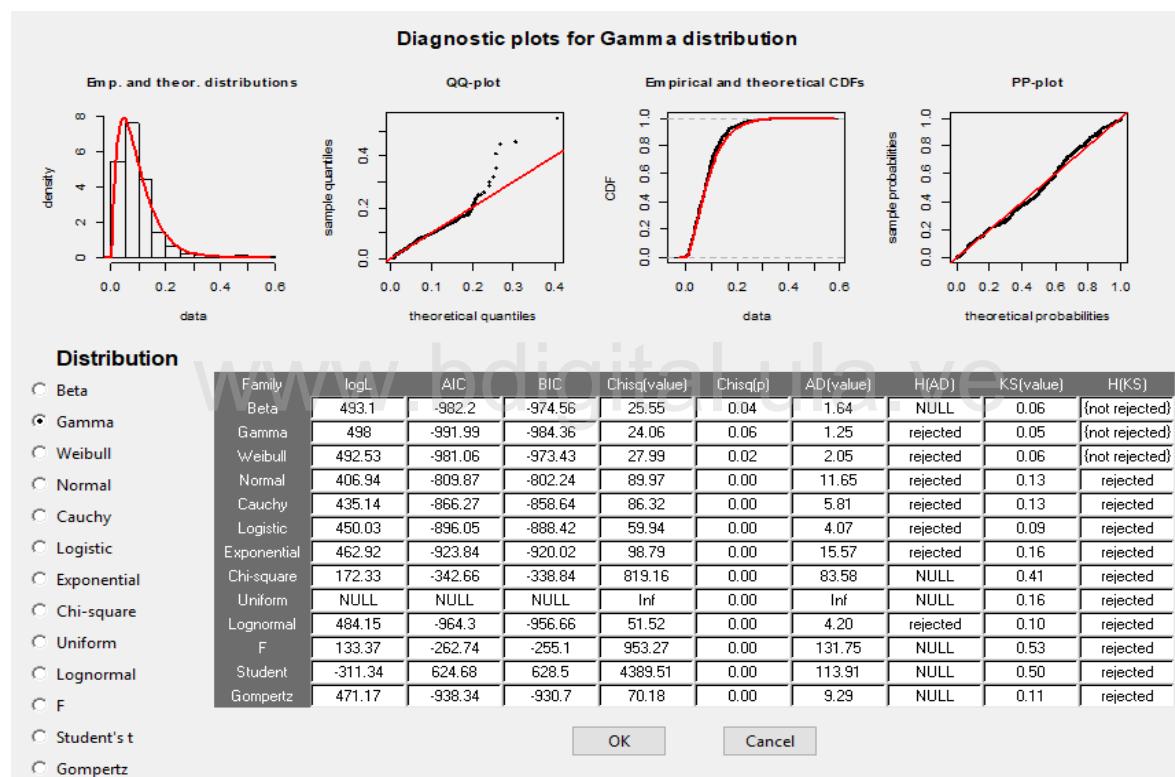
Fuente: Elaboración Propia.

La proporción de hogares pobres sigue una distribución Gamma con parámetros 5,56 y 18,63.

La media de la distribución Gamma para la proporción de hogares pobres es de $5,57*1/18,63=0,30$.

Hay una proporción promedio de 0,30 hogares pobres, lo que en términos porcentuales representa un 30%.

Gráfico 4.19 Distribución de probabilidad de la proporción de hogares pobres extremos. (Distribución Gamma)



Fuente: Elaboración Propia.

Tabla 4.13 Parámetros de la distribución Gamma (proporción de hogares pobres extremos)

```

Chosen continuous distribution is: Gamma (gamma)
Fitted parameters are:
    shape      rate
    1.920574 20.706272
  
```

Fuente: Elaboración Propia.

La proporción de hogares pobres extremos sigue una distribución Gamma con parámetros 1,92 y 20,76.

La media de la distribución Gamma para la proporción de hogares pobres extremos es de $1,92*1/20,71=0,09$

Hay una proporción promedio de 0,09 hogares pobres extremos, lo que en términos porcentuales representa un 9%.

4.4 Propuesta de un nuevo indicador para medir pobreza

Construir un nuevo indicador que vincule o que relaciona tanto a las variables o factores económicos por el método de necesidades básicas insatisfechas y con el ingreso per cápita del hogar.

La idea es relacionar tanto los factores económicos incorporados por el método de necesidades básicas insatisfechas así como también el ingreso de un hogar, y de esta forma unir todas estas variables para determinar pobreza y representar lo más cercano a la realidad.

Por ejemplo el ingreso máximo per cápita registrado en los hogares para el año 2011 fue de 3700 BsF, la media obtenida por la distribución Gamma (distribución asociada a la proporción de hogares pobres) es de 0,30 (proporción promedio de hogares pobres). La proporción 1-0,30 corresponde al complemento de la proporción media.

Multiplicando

$0,30*3700=1110$ Bs F (Extremo Inferior)

$(1-0,30)*3700=2590$ Bs F (Extremo Superior)

Gráfico 4.20 Propuesta de Nuevo Indicador



Fuente: Elaboración Propia.

A = Distribución Asociada a la PHP * Ingreso Máximo Per Cápita.

B = Complemento de la Distribución Asociada a la PHP * Ingreso Máximo Per Cápita.

Finalmente comparamos el ingreso per cápita de cada hogar, si está por debajo de 1110 Bs F el hogar se considera pobre extremo, si esta entre 1110 Bs F y por debajo de 2590 Bs F el hogar se considera pobre, si es mayor o igual a 2590 BS F el hogar se considera no pobre.

Tabla 4.14 Número de Hogares Pobres Según Indicador Propuesto

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos				
No pobre	2.321.117	33,57	33,57	33,57
Pobre relativo	2.630.214	38,04	38,04	71,61
Pobre extremo	1.962.626	28,39	28,39	100,0
Total	6.913.957	100,0	100,0	

Fuente: Elaboración Propia.

Los resultados obtenidos por este nuevo indicador, muestran mayor porcentaje de hogares pobres y pobres extremos en comparación con los resultados obtenidos por el método NBI y la línea de pobreza, el cual permite estas más cerca de la realidad, ya que permite tomar en cuenta todas las variables en este nuevo indicador.

4.5 Comparación del nuevo indicador propuesto con los métodos de LP y NBI

Para finalizar este capítulo, se comparan los resultados obtenidos del nuevo indicador con los datos reportados por el Instituto Nacional de Estadística (INE) con respecto a medición de pobreza. El INE realiza dicha medición a través de dos métodos Línea de Pobreza (LP) y Necesidades Básicas Insatisfechas (NBI). Para ello se elaboró la siguiente tabla:

Tabla 4.15 Porcentaje de pobreza según los métodos NBI, LP e Indicador propuesto

	% NBI	%LP	%IP
No pobre	46,2	75,4	33,57
Pobre relativo	29,5	17,6	38,04
Pobre Extremo	24,3	7	28,39
Total	100	100	100

Fuente: Elaboración Propia.

Se observa que para el territorio venezolano existe una diferencia muy importante en hogares pobres extremos del método de LP con respecto al nuevo indicador propuesto, ya que el nuevo indicador presenta un valor de 28,39%, mientras que para línea de pobreza no supera el 10%.

De este mismo modo, se puede observar que para hogares no pobres del método de NBI y LP con respecto al nuevo indicador, presentan una diferencia muy significativa, ya que en el nuevo indicador están asociados variables del método de NBI y el ingreso per cápita del hogar, el cual esto permite tener una amplia visión de estos hogares. Siendo así, resulta claro que se redujo considerablemente hogares no pobres con este nuevo indicador.

Cabe considerar, por otra parte que pobreza relativa para LP presenta un valor de 17,6 y para NBI un valor de 29,5%, mientras que para el nuevo indicador un valor de 38,04%, el cual se puede observar que tiene un aumento mayor al 10% respecto a cada uno de ellos.

Capítulo V

Conclusiones y Recomendaciones

5.1 Conclusiones

En la realización de este trabajo se logró diseñar una alternativa para la medición de pobreza en Venezuela, ya que este fenómeno en la actualidad es muy complejo para obtener resultados con exactitud, toma en cuenta datos que no son de fácil acceso. A través del uso de Análisis de Correspondencias Múltiples y Arboles de Decisión se logró conocer cuáles son las variables y aspectos más importantes para su medición, esto nos lleva a realizar una fusión del método de Necesidades Básicas Insatisfechas y Línea de Pobreza, ya que obtuvimos resultados que pudieran adaptarse más a la realidad.

De acuerdo a lo observado en los resultados de Análisis de Correspondencia Múltiple conseguimos que las variables de estudio más importantes para la clasificación de los hogares son Hogares en Viviendas sin Servicios Básicos (HVSB) y Hogares en Viviendas Inadecuadas (HVI), lo que nos indica que son variables que pudieron haber sido tomadas en cuenta en la construcción del nuevo indicador.

Por otro lado, en el uso de Arboles de Decisión observamos y corroboramos que el uso de la variable Hogares en Viviendas sin Servicios Básicos (HVSB) es una de las variables más importantes en el estudio de la pobreza, lo que se interpreta como una variable de mucha importancia para la clasificación de hogares en el territorio nacional.

A su vez la construcción del nuevo indicador lleva implícito el uso de los métodos de Necesidades Básicas Insatisfechas y Línea de Pobreza, lo que conlleva a sumarle una nueva variable que llamamos Ingreso Máximo. Sin embargo, se pudo observar que no se debe rechazar ninguna variable ya que la propuesta de esta alternativa es poder utilizarla en la actualidad.

Finalmente sin olvidar mencionar la variable de Ingreso Máximo, es de suma importancia, ya que a su vez a través de ella de alguna manera podemos observar cuál es su

disponibilidad para el acceso a la canasta alimentaria y a los servicios básicos, lo que corrobora nuevamente que son variables indispensables en nuestro estudio.

El nuevo indicador da como resultado la clasificación de hogares en tres grupos: hogares con pobreza extrema, hogares con pobreza relativa y hogares no pobres. En el cual la clasificación se traduce en una forma más cercana a lo que observamos día a día en Venezuela.

Se concluyó por lo antes mencionado que el nuevo indicador abarca una gama mucho más amplia y suficiente de factores que determinen e incidan de una manera significativa en la pobreza.

www.bdigital.ula.ve

5.6 Recomendaciones

En este proyecto de investigación surgen importantes aspectos que resaltar como lo son:

Tomar en cuenta la selección de las variables para determinar la pobreza, ya que determinar la pobreza en Venezuela no es igual que determinar pobreza en cualquier otro país latinoamericano o europeo.

Obtener información actualizada, es decir, una base de datos más reciente (esperando que el INE o el BCV apliquen una nueva encuesta o censo).

Inclusión de otros aspectos que pueden influir en el fenómeno de la pobreza, como lo son seguridad, acceso a planes para obtención de nuevas viviendas, programas sociales, entre otros.

Ampliar el periodo de estudio para probar el uso del nuevo indicador y analizar si se pueden obtener resultados confiables.

En cuanto a las técnicas utilizadas como lo fueron Análisis de Correspondencias Múltiples y Arboles de Decisión para obtener las variables más importantes, queda a criterio del investigador cual usar, en lo particular en este proyecto ambas coincidieron con los resultados y se pudieron interpretar de forma conjunta.

Referencias

- Banco Mundial, (2019). Pobreza. Recuperado de <https://www.bancomundial.org/es/topic/poverty/overview#1>
- Bazán Ojeda, Abigail; Quintero Soto, Ma. Luisa; Hernández Espitia, Aurea Leticia (2011). Evolucion del concepto de pobreza y el enfoque multidimensional para su estudio. Universidad Autonoma del Estado de Mexico. Recuperado de <https://www.redalyc.org/pdf/401/40118420013.pdf>
- Carlos Poza Lara (2008). Pobreza multidimensional, Madrid.
- Coa, M. (2012). Propuesta para la medición de la Pobreza en Venezuela mediante modelos de clases latentes (Tesis de Maestría). Universidad de Los Andes, Mérida, Venezuela.
- Erro, L (2008). Arboles de decisión para grandes conjuntos de datos. Coordinación de ciencias computaciones, INAOE.
- Fernández A., Martin G. (1994). Algunas consideraciones sobre los índices de pobreza de Foster, Greer y Thorbecke. Departamento de Estadística y Econometría. Universidad de Málaga.
- Instituto Nacional de Estadísticas e Informática, (1993). Mapa de Necesidades Básicas Insatisfechas de los Hogares a Nivel Distrital. Recuperado de <http://proyectos.inei.gob.pe/web/biblioineipub/bancopub/est/lib0068/n00.htm>
- Iglesias, O (2016). Banco Mundial: 6 estrategias para acabar con la pobreza. Recuperado de <https://www.fundacionsistema.com/banco-mundial-6-estrategias-para-acabar-con-la-pobreza/>
- Inmaculadasol, (2017). Pobreza Absoluta vs. Pobreza Relativa Recuperado de <https://inmaculadasol.com/2017/02/06/pobreza-absoluta-vs-pobreza-relativa/>
- Instituto Nacional de Estadísticas, (2003). Identificación y Medición de la Pobreza en el estado Mérida. Mérida, Venezuela.
- Instituto Nacional de Estadística. La pobreza y su medición. Recuperado de <https://www.ine.es/daco/daco42/sociales/pobreza.pdf>

- Maya, E (2018). Los arboles de decisión como herramienta para el análisis de riesgos de los proyectos. (Maestría). Universidad EAFIT. Recuperado de https://repository.eafit.edu.co/bitstream/handle/10784/12980/Elena_MayaLopera_2018.pdf?sequence=2&isAllowed=y
- Riutort, M. (1999). Pobreza, Desigualdad y Crecimiento Económico en Venezuela. Universidad Católica Andrés Bello, Instituto de Investigaciones Económicas y Sociales. Departamento de Investigaciones Económicas. Caracas, Venezuela.
- Rojo, J (2006). Arboles de clasificación y regresión. Laboratorio de Estadísticas, Instituto de Economía y Geografía. Consejo Superior de Investigaciones Científicas, Madrid. Recuperado de http://humanidades.cchs.csic.es/cchs/web_UAE/tutoriales/PDF/AnswerTree.pdf
- Torres, E. (2012). Evolución de la pobreza en Venezuela. Un estudio desde los enfoques estático y dinámico (tesis doctoral). Universidad de los Andes, Mérida, Venezuela.

www.bdigital.ula.ve

www.bdigital.ula.ve

Anexos

Anexo 1. Programa en R para Análisis de Correspondencia Múltiple

```
rm(list=ls(all=TRUE))

ls()

setwd("C:/Users/fatima/Desktop/Tesis Fatima")

dir()

acmpobreza<-read.csv2("Datosacm.csv",header=T)

attach(acmpobreza)

summary(acmpobreza)

View(acmpobreza)
```

Anexo 2. Programa en R para las estadísticas descriptivas

```
# Creamos una base de datos para las estadísticas descriptivas.

datosestd<-data.frame(HP,PHP,HPE,PHPE,AE,PAE,DE,PDE,HVSB,PHVSB,HVI,PHVI,HHC,PHHC)

datosestd

attach(datosestd)

View(datosestd)

summary(datosestd)
```

Anexo 3. Programa en R para Histogramas

```
win.graph()

par(mfrow=c(2,2))

hist(PAE, main='Histograma PAE', col='blue')

hist(PDE, main='Histograma PDE', col='blue')

hist(PHVI, main='Histograma PHVI', col='blue')

hist(PHHC, main='Histograma PHHC', col='blue')
```

```

win.graph()
par(mfrow=c(2,2))
hist(PHVS, main='Histograma PHVS', col='blue')
hist(PHP, main='Histograma PHP', col='blue')
hist(PHPE, main='Histograma PHPE', col='blue')

```

Anexo 4. Programa en R de categorización y cuartiles de la variable PHP

```

# Obtengo los cuartiles de la variable PHP
qp = quantile(PHP, probs = seq(0, 1, .25))
qp
# Categorizar la variables PHP
PHPC<-PHP
PHPC[PHP<=qp[2]]<-"1"
PHPC[PHP>qp[2] & PHP<=qp[3]]<-"2"
PHPC[PHP>qp[3] & PHP<=qp[4]]<-"3"
PHPC[PHP>qp[4]]<-"4"
PHPC
# Frecuencias
table(PHPC)

```

Anexo 5. Programa en R para gráficos de la función densidad de PHP y PHPE

Graficos función de densidad de PHP y PHPE

```

plot(density(PHP), xlab = "Proporción", ylab = "y", las=1, main = "")
plot(density(PHPE), xlab = "Proporción", ylab = "y", las=1, main = "")

```

Anexo 6. Programa en R de categorización y cuartiles de la variable PHPE

```
# Obtengo los cuartiles de la variable PHPE
qe = quantile(PHPE, probs = seq(0, 1, .25))

qe

# Categorizar la variables PHPE

PHPEC<-PHPE

PHPEC[PHPE<=qe[2]]<-"1"
PHPEC[PHPE>qe[2] & PHPE<=qe[3]]<-"2"
PHPEC[PHPE>qe[3] & PHPE<=qe[4]]<-"3"
PHPEC[PHPE>qe[4]]<-"4"

PHPEC

# Frecuencias

table(PHPEC)
```

Anexo 7. Programa en R para Arboles de Regresión

```
rm(list=ls(all=TRUE))

ls()

setwd("E:/Respaldo/Escritorio/Tesis Fatima")

dir()

hogares<-read.csv2("VariablesIP.csv",header=T)

attach(hogares)

summary(hogares)

View(hogares)
```

Anexo 8. Programa en R para distribución de los datos

```
library(rriskDistributions)

# Hogares pobres

res1<-fit.cont(PHP)

# Hogares pobres extremos

res2<-fit.cont(PHPE)
```

Anexo 9. Programa en R para conjuntos de entrenamientos y pruebas

```
# Usamos la función sample_frac() de dplyr para obtener un subconjunto de nuestros
# datos

# que consiste en 70% del total de ellos. Usamos también set.seed() para que sea
# reproducible
set.seed(1649)

hogaresentrena <- sample_frac(datoshogares, .7)

# Con setdiff() de dplyr, obtenemos el subconjunto de datos complementario al de
# entrenamiento

# para nuestro conjunto de prueba, esto es, el 30% restante.

hogaresprueba <- setdiff(datoshogares, hogaresentrena)
```

```
# Usamos la función rpart de rpart para entrenar nuestro modelo. Esta función nos pide una  
# formula para especificar la variable objetivo de la clasificación. La formula que usaremos  
# es tipo ~ ., la cual expresa que intentaremos clasificar Rendimiento usando a todas las demás  
# variables como predictoras.
```

Anexo 9. Programa en R para grafico del árbol

```
modelo <- c("Árbol regresión", "bagging")  
test.MSE <- c(test.MSE.arbol_1, test.MSE.arbol_2)  
comparacion <- data.frame(modelo = modelo, test.MSE = test.MSE)
```

www.bdigital.ula.ve