UNIVERSIDAD DEL AZUAY

POSGRADOS

MAESTRÍA EN MATEMÁTICAS APLICADAS

ESTUDIO COMPARATIVO DE LA POBREZA EN EL ECUADOR EN LAS CIUDADES DE QUITO, GUAYAQUIL Y CUENCA, MEDIANTE EL USO DE LOS MODELOS LOGIT Y PROBIT

TRABAJO DE GRADUACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE MAGÍSTER EN MATEMÁTICAS APLICADAS

AUTORES

ING. IVÁN DANIEL LÓPEZ

ING. JUAN MANUEL MALDONADO

DIRECTOR

ECON. CARLOS CORDERO DÍAZ

CUENCA – ECUADOR

2017

ÍNDICE DE CONTENIDOS

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

RESUMEN

Desde una perspectiva social, el estudio y medición de la pobreza ha ido con el tiempo, desde métodos directos e indirectos hasta una combinación de éstos a través de la multidimensionalidad, pues en la realidad son una serie de variables las que determinan si un individuo es percibido como pobre o no pobre. En este trabajo, mediante el uso de los modelos Logit y Probit, se realiza una investigación donde se busca establecer la relación existente entre la variable dependiente pobreza y las variables independientes que modelan este fenómeno, enfocándose en las ciudades de Quito, Guayaquil y Cuenca. Las variables independientes se determinaron en base a la teorización de la pobreza, estudios previos y recomendaciones que varios investigadores hacen al respecto. Las dimensiones e indicadores que reflejen la multidimensionalidad del fenómeno y que permiten determinar la probabilidad de que un hogar sea pobre, se muestran en el desarrollo de este trabajo.

ABSTRACT

From a social perspective, the study and measurement of poverty has gone from direct and indirect methods to a combination of these through multidimensionality. In reality, a series of variables determine whether an individual is perceived as poor or not poor. In this work, an investigation was conducted using the Logit and Probit models to establish the relationship between the poverty dependent variable and the independent variables that model this phenomenon in Quito, Guayaquil and Cuenca. The independent variables were determined based on the theorization of poverty, previous studies and recommendations from several researchers. The dimensions and indicators that reflected the multidimensionality of the phenomenon and allowed to determine the probability of a household being poor were shown in the development of this work.

UNIVERSIDAD DEL
AZUAY
Dpto. Idiomas

Translated by:

Ing. Paúl Arpi

INTRODUCCIÓN

La pobreza ha acompañado al ser humano desde su aparición, lo cual se aprecia en pasajes bíblicos y en obras artísticas en las que las condiciones precarias de la mayoría contrastan con los banquetes de las clases señoriales. Definir la pobreza es un fenómeno complejo en donde las diferentes fuentes y organismos coinciden en relacionarla con la carencia o ausencia de algo para mantener un nivel de vida aceptable. En el interés del hombre por estudiar este fenómeno social se han ido construyendo conceptos, clasificaciones, causas-consecuencias y metodologías a fin de que los estados dirijan sus esfuerzos en post de disminuir los indicadores de presencia de pobreza.

Estudiar la pobreza implica conceptualizarla, clasificarla y medirla. Para la medición de la pobreza se han clasificado los métodos en directos, indirectos y combinados los cuales son resultado de una evolución en el espacio-tiempo en la manera de observar el fenómeno.

Los Métodos Indirectos están asociados a los recursos o ingresos para satisfacer las necesidades básicas y se describen el de Consumo Calórico y el del Costo de las Necesidades Básicas. El primero de ellos basado en el establecimiento de una línea de pobreza en base al consumo energético de una persona para sus actividades diarias y el segundo en una canasta de alimentos, bienes y servicios necesarios para la supervivencia

En los Métodos Directos, el de las Necesidades Básicas Insatisfechas despega en la década de los ochenta de la mano de la Comisión Económica para América Latina (CEPAL), con objeto de optimizar la información disponible de los censos de población y vivienda planteando indicadores relacionados con necesidades básicas estructurales (nutrición, salud, vivienda, educación, servicios públicos, oportunidad de empleo) que se requieren para el bienestar individual. En la década de los noventa el economista pakistaní Mahbub ul Haq introduce el Índice de Desarrollo Humano (IDH) basado en la media geométrica de tres dimensiones: Esperanza de Vida, Educación y Riqueza considerando a la persona y sus capacidades y no solo el crecimiento económico del país. Los índices de Pobreza por Capacidad y de Pobreza Humana descartan en la valoración los ingresos y consideran variables como el conocimiento, respeto y comunicación.

En cuantos a los Métodos Combinados que integran los dos anteriores, permiten tener una visión holística del fenómeno siendo el Índice de Pobreza Multidimensional de Sabina Alkire y James Foster el que describe la causalidad ya que sólo los ingresos, la producción o lo que se consume no refleja la complejidad de la Pobreza.

La descripción de las herramientas matemáticas se aborda en el capítulo dos, clasificando y desagregando las técnicas de análisis cuantitativo para el estudio de la pobreza, haciendo énfasis en los modelos Logit y Probit que por la distribución estadística que siguen son los recomendados para el estudio de los fenómenos sociales, en éste caso, la pobreza en las ciudades de Quito Guayaquil y Cuenca.

En el capítulo tres, en base a la Encuesta Nacional de Empleo, Desempleo y Subempleo del Instituto Nacional de Estadísticas y Censos, indicadores propuestos en la Constitución de la República del Ecuador, y el uso de los métodos y modelos matemáticos de medición de la pobreza (Logit y Probit), se definen las variables relevantes para la medición de la pobreza y se plantean modelos probabilísticos en base a estas variable para poder estudiar las diferencias en cuanto a los factores que inciden en este fenómeno en las tres principales ciudades del país.

CAPÍTULO 1

CONCEPTUALIZACIÓN DE LA POBREZA Y MÉTODOS DE MEDICIÓN

La pobreza es un fenómeno social con el que ha lidiado la humanidad desde sus inicios, si bien este fenómeno no siempre ha estado presente en la misma medida en todas las sociedades, su estudio ha sido de interés desde hace varios siglos atrás. Los primeros estudios respecto a la pobreza se remontan hacia el año 1668 cuando el estadístico inglés Gregory King dividió a la población entre aquellos que colaboraban a “incrementar la riqueza del reino” y los que “disminuían la riqueza del reino”, si bien este fue el primer acercamiento al estudio de la pobreza que contenía cierta formalidad no es hasta finales del siglo XIX que estudios más formales y científicos comenzaron a aparecer, se menciona a Charles Booth como el primero en intentar realizar una medición matemática de la pobreza en la década de 1890 al concebir un mapa de pobreza para la ciudad de Londres, de aquí en adelante una serie de autores han presentado una gran variedad de teorías y enfoques para la medición para el fenómeno. Es en los años cuarenta cuando se internacionaliza el concepto de pobreza al aparecer este por primera vez en los informes del Banco Mundial, a partir de estos informes el estudio de la pobreza adquiere una significativa relevancia al considerarse un problema palpable y que afectaba a una buena parte de la población mundial desencadenando un estudio más profundo por parte de gobiernos e instituciones alrededor de todo el mundo.

1.1 DEFINICIÓN Y CLASIFICACIÓN DE LA POBREZA

El fenómeno de la pobreza es complejo, diversos autores y entidades lo han conceptualizado como la carencia de recursos o condiciones que permitan a los individuos mantener un estatus de vida mínimo el cual es relativo dependiendo del espacio y tiempo, por lo que, el término pobreza es polisémico y tiene varias acepciones.

El Diccionario de la Real Academia de la Lengua define la pobreza como: “Necesidad, estrechez, carencia de lo necesario para el sustento de la vida”, implica la ausencia de condiciones y/o recursos para satisfacer las necesidades básicas de vivienda, alimentación, educación, salud, servicios, entre otros.

La Comisión Económica para América Latina y el Caribe (CEPAL) presenta un enfoque más social y establece que la pobreza es: "…la situación de aquellos hogares que no

logran reunir, en forma relativamente estable, los recursos necesarios para satisfacer las necesidades básicas de sus miembros..."y para su cálculo se genera un umbral de pobreza en base al ingreso necesario para satisfacer dichas necesidades básicas explicitadas en el costo de una canasta básica.

Para la Organización de las Naciones Unidas (ONU) la pobreza no solo depende de ingresos monetarios, en su definición incorpora variables relacionadas con el acceso a servicios básicos, para la ONU la pobreza es: “la condición caracterizada por una privación severa de necesidades humanas básicas, incluyendo alimentos, agua potable, instalaciones sanitarias, salud, vivienda, educación e información”. Bajo este mismo esquema y años más tarde el Programa de Naciones Unidas para el Desarrollo (PNUD) del mismo organismo internacional, incorpora a la definición de pobreza aspectos políticos y de participación, definiéndola como: “la incapacidad de las personas de vivir una vida tolerable. Entendiendo como vida tolerable el mantener una vida larga y saludable, con educación, libertad política, respeto de los derechos humanos, seguridad personal, acceso al trabajo productivo y bien remunerado y participación en la vida comunitaria.”

Es interesante el aporte de Paul Spicker (2009), economista británico experto en políticas sociales, al interpretar la pobreza en términos como: privación, necesidad, limitación de recursos, ausencia de titularidades, carencia de seguridad, exclusión, dependencia, bajo nivel de vida, desigualdad, posición económica y clase social. Spicker clasifica las definiciones de pobreza desde tres puntos de vista, como un concepto material, como una situación económica y como un juicio moral.

Para Spicker desde un concepto material la pobreza se clasifica en objetiva o subjetiva. La pobreza objetiva hace referencias a hechos o circunstancias que no cumplen las personas y que han sido prefijados por el investigador, se determinan indicadores como el nivel de ingresos, gastos, equipamiento de la vivienda, acceso a medicinas, entre otros. La pobreza subjetiva o pobreza por insatisfacción, es aquella expresada por el individuo independientemente de su situación económica y refleja su autopercepción de inconformidad con su estatus viéndose a sí mismo como pobre.

Desde el punto de vista de la situación económica la pobreza puede ser absoluta o relativa.

Atkinson y Bourguignon (2000) con base en Sen (1976) consideran a la pobreza absoluta como una pobreza de carencia, es pobre quien no tiene lo necesario para su subsistencia y no puede cubrir sus necesidades mínimas; este concepto es el más aplicado en los países no desarrollados o en desarrollo ya que en éstos es más difícil obtener la renta mínima para subsistir. El inconveniente es determinar esas necesidades mínimas para la canasta de subsistencia ya que depende de cada país, región geográfica, cultura, temporalidad, etc. La Cumbre Mundial de Desarrollo Social de la ONU (1995) definió la pobreza absoluta como: "condición caracterizada por una privación severa de las necesidades humanas básicas, que incluyen alimentos y agua potable, así como condiciones de higiene, salud, vivienda, educación e información".

La pobreza relativa o pobreza como exclusión no es tan crítica como la anterior, hace referencia a que la persona se encuentra en condiciones de inferioridad respecto a los de su entorno, por lo que, a quien se considera pobre en un país y en determinado momento, podría no serlo en otro lugar y tiempo. Esta noción está más cerca del concepto oficial de pobreza relativa de la Unión Europea (1984): "aquella persona cuyos recursos (materiales, culturales y sociales) son tan limitados que los excluyen del mínimo nivel de vida aceptable en el Estado Miembro en el que viven". Ante esto, la pobreza relativa puede distribuirse por niveles según su impacto y dependiendo el umbral de pobreza considerado, así López-Aranguren (2005) establece:

- *Pobreza en general:* por debajo del 50% de la medida central de recursos.
- *Precariedad económica:* situación entre el 50 y el 35%.
- *Pobreza moderada:* situación entre el 35 y el 25%.
- *Pobreza grave o severa:* situación entre el 25 y el 15%.
- *Pobreza extrema:* situación por debajo del 15% de la media.

Como se describe, las distribuciones son graduales partiendo de una situación generalista hasta una de pobreza extrema.

Los citados son los conceptos tradicionales de pobreza, empero su estudio se ha profundizado aparecen otras acepciones que la matizan para mayor precisión en su descripción, según (López-Aranguren, 2005):

Pobreza estática: Presenta un cuadro o fotografía en un determinado momento.

Pobreza dinámica: Mide la duración de la pobreza a través del tiempo, implica obtener información sobre cuánto tiempo permanecen los individuos pobres en esa situación y cuáles son sus trayectorias de entrada y salida. Este análisis permite indagar sobre las causas que empujan a caer en la pobreza, proponer medidas para facilitar su salida y conocer el impacto de las políticas públicas. A su vez, se clasifica en pobreza transitoria y pobreza permanente, según sea la duración de la situación de escasez de recursos.

Pobreza integral: Hace referencia tanto a la escasez de ingresos como a la dificultad de acceso a los servicios sociales que facilitan la cobertura de las necesidades vitales básicas.

Nueva pobreza: Involucra las transformaciones provocadas por las innovaciones tecnológicas, industriales y otras circunstancias sociales o económicas.

Pobreza unidimensional: Es el estudio de la pobreza desde la perspectiva de una única variable objeto de análisis. Sobresale el análisis dela renta ya sea desde el punto de vista del gasto o del ingreso.

Pobreza multidimensional: Estudia a la pobreza de manera más completa ya que incorpora otros factores además del meramente monetario. Considera el estado de la vivienda, salud, educación, empleo, relaciones sociales, libertades, etc. Esta acepción es la más compleja de todas pues incorpora un mayor número de variables explicativas del fenómeno, por tanto, la precisión del mismo mejora, ayudando a describir y explicar de manera más eficiente el fenómeno de la pobreza.

Finalmente, para Spicker la pobreza desde el punto de vista de juicio moral está implícita en la autocensura de los individuos de considerarse a sí mismos como pobres o no pobres.

1.2 CAUSAS Y PERSPECTIVAS

Debido a la complejidad del fenómeno es fácil confundir si una variable es causa o consecuencia de la pobreza y de cualquiera de los conceptos o tipos de pobreza que se estudie, las causas pueden ser estructurales o coyunturales y en dependencia de factores económicos, demográficos, sociales y culturales.

Las causas de la pobreza en función de factores económicos se refieren a la falta de activos monetarios, políticos, ambientales y de infraestructura que limitan la adquisición de

bienes y servicios adecuados para la producción, es decir, considera la propiedad y distribución inequitativa de los medios y resultados de producción, puesto que considerar el crecimiento económico medido por el producto interno bruto per cápita no toma en cuenta la desigual distribución entre los miembros. Esto podría atribuirse a un sistema capitalista en el que el sentido de acumulación del mismo olvida los derechos de las personas excluyéndolas de la generación de bienestar con lo que se genera desigualdad entre propietario y operario; desigualdad y pobreza tienen el mismo objeto de estudio, pero no son sinónimos por lo que cada uno debe remitirse a su metodología de análisis (Mogollón Plazas & Solano Parra, 2012).

Las relaciones sociales en América Latina debido a la colonización han sido de explotación y desigualdad estructural por causas étnicas, y bajo este contexto se expresa que "la aparición de la pobreza como fenómeno colectivo tiene su primer momento cuando las formas de dominio y explotación coloniales rompen con los anteriores sistemas de reciprocidad y de inscripción de las comunidades y las familias. Actualmente, factores sociales que inciden a la presencia de pobreza son la urbanización sin límites por el auge de personas a las ciudades industrializadas con lo que se abandona el campo y se asientan en condiciones de insalubridad, hacinamiento y falta de servicios básicos. Los conflictos político-religiosos aúpan la pobreza ya que deben abandonar su forma de vida establecida debiendo incluso traspasar fronteras y adquirir la calidad de refugiados." (Álvarez, 2009)

Según el Fondo de Población de la Organización de las Naciones Unidades, nacen cada año alrededor de ochenta millones de personas, con lo que se estima que para el 2050 la accesibilidad a alimentación, vivienda y salud va ser compleja considerando que por los avances científicos y médicos se ha incrementado el promedio de vida. Los desastres naturales generan la perdida de bienes materiales, colapsan los servicios básicos y desmoronan la situación emocional de los individuos, presentando situaciones de pobreza coyuntural que deben ser corregidas con políticas a corto y mediano plazo.

Se ha relacionado también a la pobreza con la cultura, ya que "…al referirse a un solo estilo de vida, el cual es compartido por poblaciones y personas en situación de pobreza que conviven en contextos sociales e históricos específicos, la cultura de la pobreza tiende a reproducirse al ser transmitida entre generaciones, aspecto éste que dificulta su erradicación". (Phélan, 2006). En términos de Oscar Lewis, citado por Barbieri y Castro,

la cultura de la pobreza es "aquella que tiene su propia estructura y lógica, un modo de vida que pasa de generación en generación. No sólo es un problema de privación y desorganización, es un término que significa la ausencia de algo. Es una cultura en el sentido antropológico tradicional en la medida que proporciona a los seres humanos un esquema de vida, un conjunto listo a dar soluciones a problemas humanos y que desempeña así una función significativa de adaptación" (Barbieri & Castro, 2001). De esta manera la cultura de la pobreza asume aspectos negativos de la población, prejuicios y estereotipos, vinculados a la inferioridad de clase, etnia, estructura familiar, desempleo, migración interna e internacional, entre otros factores, lo que causa exclusión y falta de participación dentro de la sociedad. En este sentido el enfoque cultural culpa a los mismos pobres de ser pobres al reproducir sus condiciones culturales generación tras generación.

El poder político del estado, determina los objetivos prioritarios de la sociedad, como se produce y distribuye la riqueza, la promoción de los derechos (económicos, sociales, culturales, ambientales, …) y las relaciones de poder entre los actores sociales. Intentar disminuir la pobreza implica dejar de defender sectores tradicionalmente privilegiados, aperturar la participación democrática y la organización popular, con objeto de que las libertades, pluralismo, inclusión, tolerancia y solidaridad eviten trastornos a nivel mundial. Políticamente, por parte de los países se han aplicado paliativos a las situaciones de pobreza: subsidios, pensiones, sanidad gratuita, que calman momentáneamente la situación, pero que seguirá latente en el tiempo mientras no se aborde desde lo estructural que implicaría, entre otros, la disminución de la brecha entre los más ricos y el resto de la población ya que en palabras de Winnie Byanyima directora ejecutiva de Oxfam Internacional “no es justo que 62 personas posean la misma riqueza que la mitad de la población mundial “ y de estas tan sola 9 sean mujeres, implica eficiencia en la gestión de recursos, puesto que en promedio por cada euro que se destina a la asistencia social se requiere de dos a tres para moverlo. (Josep Miró -Fundación Caritas-, 2005)

1.3 MÉTODOS DE MEDICIÓN DE LA POBREZA

Sintetizando, la clasificación de la pobreza puede ser: objetiva o subjetiva, absoluta o relativa, unidimensional o multidimensional y dependiendo de esto se determina un método de medición directo, indirecto o combinado.

El método de medición de la pobreza directo también se lo denomina de las Necesidades Básicas Insatisfechas (NBI o de los indicadores sociales) y bajo la conceptualización de Sen citada por Fresneda (2007) ubica a la pobreza en el campo de las realizaciones para la satisfacción de las necesidades básicas; mientras que al método indirecto también conocido como método del Ingreso o del Consumo ubica a la pobreza en el campo de las capacidades de las personas para satisfacer sus necesidades básicas. Una combinación de los dos métodos citados, genera la tercera forma de medición de la pobreza.

1.3.1 MÉTODOS DIRECTOS DE MEDICIÓN DE LA POBREZA

El investigador fija un estándar bajo el cual al individuo se lo considera pobre si no cumple con una o varias de las condiciones mínimas establecidas para la medición.

El método directo relaciona el bienestar del individuo con el consumo efectivamente realizado y no con la capacidad con que podría hacerlo.

Dentro de los métodos directos de medición de la pobreza se encuentra el de las Necesidades Básicas Insatisfechas (NBI), el Índice de Desarrollo Humano (IDH), el Índice de Pobreza de Capacidad (IPC) y el Índice de Pobreza Humana (IPH).

1.3.1.1 Método de las Necesidades Básicas Insatisfechas

En la década de los ochenta la Comisión Económica para América Latina (CEPAL), con objeto de optimizar la información disponible de los censos de población y vivienda, plantea este método directo de medición de la pobreza en el que se considera un conjunto de indicadores relacionados con necesidades básicas estructurales (nutrición, salud, vivienda, educación, servicios públicos, oportunidad de empleo) que se requieren para el bienestar individual.

Bajo esta medición, se denomina en condiciones de pobreza extrema a quienes evidencian dos o más indicadores de carencia y se determina como población pobre a quienes presentan una sola necesidad básica insatisfecha.

Como ventajas del método se tiene:

- Se aprovecha la información disponible de censos y datos de población lo cual implica disminución de costos en la recolección de la información.

- Al contar con una base de datos de censos poblacionales, puede apreciarse su evolución a través del tiempo y realizar comparaciones geográficas.
- Permite focalizar la pobreza a través de mapas por región para el establecimiento de políticas públicas.
- Diagnostica el estado de vida de los individuos independientemente de los ingresos que posea.

Como desventajas:

- Todos los indicadores tienen el mismo peso de incidencia.
- Se considera a igual nivel de pobreza a individuos que carezcan de uno, dos, o más indicadores.
- Ubica en el mismo nivel de pobreza a un hogar en el que uno, dos, o más niños no asisten a la escuela.
- No diferencia la ubicación del hogar de zona urbana o rural, con lo que se sobreestimaría la pobreza en las zonas rurales.

En nuestro país, el Ministerio Coordinador de Desarrollo Social a través del Sistema Nacional de Indicadores Sociales y Económicos, determina que un hogar es pobre por necesidades básicas insatisfechas si presenta una de las siguientes situaciones:

- Características físicas inadecuadas: piso de tierra, piedra; paredes de cartón, estera; techo de lata o cualquiera material que evidencie condiciones no apropiadas para el hábitat humano.
- Servicios públicos inadecuados o carencia de los mismos: energía eléctrica, agua, conexión hidrosanitaria o pozo séptico.
- Alta dependencia económica: tres o más personas dependen de una fuente de ingresos proporcionada por un jefe de hogar con menos de dos años de escolaridad.
- Inasistencia a la educación básica de uno o más niños entre seis a doce años de edad.
- Hacinamiento crítico: tres o más personas por espacio habitacional.

1.3.1.2 Índice de Desarrollo Humano (IDH)

En base a las ideas de Amartya Sen, el economista pakistaní Mahbub ul Haq plantea otra perspectiva para clasificar a los países en base al desarrollo humano, esto como un proceso por el que una sociedad mejora las condiciones de vida de sus ciudadanos a través de un incremento de los bienes con los que puede cubrir sus necesidades básicas y complementarias y de la creación de un entorno en el que se respeten los derechos humanos de todos ellos.

Este índice fue desarrollado en la década de los noventa por el Programa de las Naciones Unidad para el Desarrollo (PNUD) que considera a la persona y sus capacidades y no sólo el crecimiento económico del país, ya que dos países con el mismo valor de ingresos per cápita obtienen diferentes resultados en el índice de desarrollo humano debido a las políticas gubernamentales establecidas.

El Índice de Desarrollo Humano (IDH) es un indicador de los logros medios obtenidos en las tres dimensiones fundamentales del desarrollo humano: tener una vida larga y saludable, adquirir conocimientos y disfrutar de un nivel de vida digno, aplicándose la media geométrica de estos tres factores.

Salud es el nivel de vida larga y saludable que se evalúa según la esperanza de vida al nacer, que es la media de la cantidad de años que vive una población nacida en un determinado año con la tasa de mortalidad vigente como constante.

Se denomina Educación o Conocimiento a los años promedio de escolaridad de los adultos mayores a 25 años y los años esperados de escolaridad de los niños en edad escolar.

La dimensión del nivel de Vida Digna o Riqueza, se mide conforme al PIB per cápita PPA (paridad de poder adquisitivo) que homogeniza los precios de los productos a una moneda común a fin de determinar objetivamente la real producción de bienes y servicios por país independientemente del precio a que se oferten.

El cálculo del Índice de Desarrollo Humano considera tres índices:

1. El Índice de Esperanza de Vida (IEV):

$$IEV = \frac{E_i - 20}{\max E_i - 20}$$

Donde se considera 20 años el valor mínimo de esperanza de vida, y E_i la esperanza de vida en cada país.

2. El índice de Educación (IE):

$$IE = \frac{\sqrt{IAPE \times IAEE}}{\max \sqrt{IAPE \times IAEE}}$$

Donde:

IAPE es el índice de años promedio de educación de los mayores de 25 años.

$$IAPE = \frac{APE_i}{\max APE_i - 0}$$

APE representa los años promedio de educación de los ciudadanos mayores a 25 años.

IAEE es el índice de años esperados de escolaridad de los niños de 6 a 12 años.

$$IAEE = \frac{AEE_i}{\max AEE_i - 0}$$

AEE representa los años esperados de escolaridad de los niños de 6 a 12 años.

El número mínimo de años de educación y los de escolaridad es cero.

3. El índice de riqueza (IR):

$$IR = \frac{\ln(IPIB) - \ln(100)}{\ln(40000) - \ln(100)}$$

Donde $IPIB$ es el índice del producto interno bruto y los valores máximos y mínimo son de 40000 y 100.

El Índice de Desarrollo Humano es la media geométrica de los tres índices: índice de esperanza de vida (IEV), índice de educación (IE) y el índice de riqueza (IR):

$$IDH = \sqrt[3]{IEV \times IE \times IR}$$

Como ventajas del método se cita:

- Implica otros factores de una vida digna, a más de lo estrictamente económico.
- Permite comparaciones a nivel mundial ya que las variables son accesibles de todos los países.
- Los cálculos matemáticos son relativamente simples.
- Permite tener una visión general del país.

Desventajas del IDH:

- Considera variables relativas a los países y no al ser humano.
- No toma en cuenta la desigualdad en la distribución del ingreso.
- Al trabajar con promedios, pueden existir grupos étnicos, culturales, de género, de raza, que difieren radicalmente del promedio nacional.
- Refleja solo una parte del desarrollo humano pues ignora desigualdades, seguridad, libertades, empoderamiento.

1.3.1.3 Índice de Pobreza de Capacidad (IPC)

El Índice de Pobreza de Capacidad (*IPC*) "es un índice simple compuesto de tres indicadores que reflejan el porcentaje de la población con deficiencias de su capacidad en tres aspectos básicos del desarrollo humano: tener una vida saludable con buena alimentación, tener capacidad de procreación en condiciones de seguridad y saludables, y estar alfabetizado y poseer conocimientos". (Valdivieso 2011).

Este *IPC* deja de lado los ingresos para considerar la falta de capacidades y privaciones no solo en lo material sino en aspectos como el conocimiento, respeto y comunicación.

Los indicadores de las tres dimensiones consideradas son:

- Porcentaje de niños menores de cinco años con déficit de peso (desnutrición crónica), que refleje el no tener una vida saludable y buena alimentación.
- Porcentaje de partos que no reciben atención de personal capacitado, que representa la incapacidad de procreación en condiciones seguras y saludables.

- Porcentaje de mujeres de 15 años o más que son analfabetas, que refleja la carencia de conocimientos.

Como ventajas del método se pueden mencionar:

- Simplicidad en los cálculos y recogimiento de la información.
- Presta atención en la capacidad de las personas para acceder a bienes y servicios.
- No se consideran los ingresos para su medición.

Y como desventajas:

- Es una medida incompleta del bienestar, ya que elimina el ingreso, pero no considera otras variables.
- Al no considerar los ingresos, se relega una variable de importancia.
- Al igual que el *IDH* trabaja con promedios y se puede caer en generalizaciones que no consideran a los subgrupos.

1.3.1.4 Índice de Pobreza Humana (IPH)

Este método fue desarrollado en 1997 y buscaba superar algunos inconvenientes que presentaban otros índices, como la posibilidad de comprar los niveles de pobreza internacionalmente.

Al igual que los dos métodos anteriores, se basa en las restricciones o limitaciones de los individuos para acceder a salud (nivel de vida larga y saludable), educación y buen nivel de vida. Al igual que el Índice de Pobreza por Capacidad no considera los ingresos; y los indicadores de las dimensiones que toma en cuenta se parametrizan para países desarrollados y en desarrollo, estos indicadores son:

- Porcentaje de población estimada que morirá antes de los 40 años de edad (60 años en los países desarrollados).
- Porcentaje de adultos analfabetos.
- Porcentaje de población con acceso a agua potable.
- Porcentaje de desnutrición en niños menores de cinco años.
- En el caso de países desarrollados, se reemplaza el porcentaje de acceso a agua potable y de desnutrición por el porcentaje de la población que vive por debajo

del umbral de pobreza (50% de la mediana del ingreso disponible del grupo familiar) y la tasa de desempleo a largo plazo.

Como ventajas del *IPH* se tiene:

- El método va más allá de los simples indicadores económicos.
- Es un indicador que permite comparaciones a nivel mundial.
- Considera el contexto de la pobreza para su cálculo.

Desventajas del *IPH* se citan:

- Exclusión de algunos indicadores justificándose por su dificultad en la medición o carencia de los mismos, entre estos: falta de libertad política, incapacidad de tomar decisiones, seguridad, participación en la comunidad, sostenibilidad, equidad intergeneracional.
- Al trabajar con promedios nacionales en educación, salud y nivel de vida, es difícil identificar los grupos particulares que sufren las carencias.

El *IPH* se calcula de forma separada para los países desarrollados y en transición (*IPH-1*) y para los países que la Organización para la Cooperación y Desarrollo Económico ha clasificado como países de ingresos altos (*IPH-2*) con el fin de mejorar las diferencias de medición que se puedan evidenciar entre los dos grupos.

1.3.2 MÉTODOS INDIRECTOS DE MEDICIÓN DE LA POBREZA

A los métodos indirectos de medición de la pobreza también se los denomina métodos de medición de los Ingresos en donde se define como pobre a aquel que carece de recursos o de la capacidad para alcanzar a satisfacer las necesidades básicas.

Puede medirse desde un enfoque absoluto o relativo. El enfoque de pobreza absoluta la asocia con privación, miseria y hace referencia a la no cobertura de las necesidades fisiológicas de las personas: alimentos, salud, vestimenta, educación, agua potable, vivienda, información. A la pobreza relativa se la asocia con desigualdad o desventaja, varía en el tiempo y en el espacio, ya que depende de los estándares prefijados en cada sociedad en función de su geografía y ya que al pasar los años van apareciendo nuevas necesidades.

Como métodos indirectos de medición de la pobreza están el método de Consumo Calórico y el método del Costo de las Necesidades Básicas.

1.3.2.1 Método del Consumo Calórico

A partir de estudios nutricionales sobre la actividad física de las personas, se determina un consumo de energía para las personas medido en calorías. Es un método indirecto porque mide el potencial de obtener ciertas calorías y no el consumo efectivo de las mismas.

En esta metodología, hay dos mecanismos para determinar la línea de pobreza. La primera es seleccionar una muestra de hogares con un consumo calórico cercano al calculado y emplear el promedio de sus ingresos como línea de pobreza. La segunda opción es realizar una regresión entre consumo calórico e ingresos y con la relación encontrada evaluar el ingreso necesario para consumir las calorías calculadas.

La ventaja de este método indirecto es que requiere de menor información y es de fácil disponibilidad ya que se basa en parámetros exclusivamente alimentarios y accesibles en el ente rector de salud, empero debe considerarse, que no necesariamente un criterio nutricional es criterio de bienestar ya que la relación consumo calórico-gasto tiene diversos comportamientos debido a preferencias alimentarias, práctica de deportes, precios, entre otros.

Dentro de este contexto, Ravallion (1998) analiza que, respecto al mismo alimento y gastos iguales, los hogares urbanos tienen gustos más caros que los rurales por lo que cada caloría consumida sería más costosa e induciría a pensarse que los hogares urbanos son más pobres.

1.3.2.2 Método del Costo de las Necesidades Básicas

Consiste en determinar una canasta de alimentos, bienes y servicios básicos de consumo y la línea de pobreza será el gasto necesario para adquirirla.

La pobreza, calculada de ésta manera mide la privación de las personas en la satisfacción de las necesidades básicas, es decir, un consumo por debajo de la canasta de alimentos, bienes y servicios que satisfacen las necesidades básicas; en pobreza extrema se ubicaría

la población que no alcanza ni siquiera acceder a una canasta que le permita cubrir los requerimientos mínimos nutricionales.

En nuestro país, el Instituto Nacional de Estadísticas y Censos (INEC) es el ente que proporciona las fuentes oficiales de datos para establecer la línea de pobreza que se basará en información de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) del segundo quimestre del año 2016 y la Encuesta Nacional de Condiciones de Vida 2014 que permiten determinar el Índice de Precios al Consumidor (IPC).

La ventaja de establecer una línea de pobreza es que permite monitorear el comportamiento de reducción o ampliación de la misma y el volumen de población en cada segmento (no pobre, pobre, pobreza extrema).

Como desventajas se cita el que se desconoce el contexto y cultura de cómo cada hogar distribuye sus ingresos en alimentos, bienes y servicios ignorando las diferencias al interior de la casa.

1.3.3 MÉTODOS COMBINADOS DE MEDICIÓN DE LA POBREZA

Al aplicar un método combinado se busca la complementariedad de uno indirecto con uno directo, así pues, el método indirecto basado en el umbral de pobreza considera al ingreso como medio para la satisfacción de varias necesidades básicas, mientras que por el método directo de las necesidades básicas se observa la disponibilidad y acceso a los mismos. Existen dos métodos combinados el Integrado y el Multidimensional.

1.3.3.1 Método Integrado de Medición de la Pobreza

El principal tratadista del método Julio Boltviniky argumenta que el bienestar de hogares y personas depende de seis variables: ingreso corriente, acceso a bienes y servicios gratuitos, los activos no básicos y la capacidad de endeudamiento del hogar, el tiempo libre y el disponible para trabajo doméstico, educación y reposo, patrimonio familiar (propiedad o derecho de uso de activos que proporcionan servicios de consumo básico) y los conocimientos y destrezas de las personas.

En función de factores estructurales y coyunturales para la presencia de pobreza, Katzman formula una tipología estableciendo cuatro categorías: no pobreza, pobreza inercial,

pobreza reciente y pobreza crónica que se explica mediante una matriz de doble entrada entre el método directo e indirecto:

Tabla 1: Tipología de la Pobreza

Método Indirecto → *Método Directo↓*	*Ingresos por debajo de la línea de pobreza*	*Ingresos por encima de la línea de la pobreza*
Insatisfacción de por lo menos una necesidad básica	Pobreza crónica	Pobreza inercial
Satisfacción de todas las necesidades básicas	Pobreza reciente	No pobreza

Fuente: Katzman

Este método, desglosa las variables de bienestar de las Necesidades Básicas Insatisfechas en indicadores de logro; considera una canasta de alimentos, bienes y servicios básicos elementales y a más considera otra dimensión denominada “pobreza de tiempo” referente a los adultos del hogar que trabajan más de 48 horas semanales.

Al combinarse el método indirecto del umbral de la pobreza con el método de las necesidades básicas insatisfechas, se cuenta con la ventaja de que en la medición no sólo se considera los ingresos sino un conjunto de indicadores sociales que reflejan en mayor grado la realidad de los hogares que al trabajarse con sólo uno de los métodos.

A pesar de la ventaja de combinar los métodos, se hereda las deficiencias de los mismos, como por ejemplo la exclusión de ciertas variables de estudio (seguridad, libertades…) y las diferencias propias internas a cada hogar en función de su número de integrantes.

1.3.3.2 Método Multidimensional de Medición de la Pobreza

La multi-causalidad de la pobreza, el contexto geográfico y las características socio-culturales y económicas de las poblaciones, hacen que al decurrir el tiempo aparezcan

nuevas metodologías para su medición, ya que sólo los ingresos, la producción o lo que se consume no refleja la complejidad del fenómeno.

Los principales tratadistas del método son Sabina Alkire y James Foster, que para el 2010 dan rigor teórico y matemático a la metodología al asignar ponderaciones a los indicadores de las variables consideradas en el Índice de Desarrollo Humano dentro de la teoría de Amartya Sen.

En cuanto a lo teórico, de Sen se considera que para medir el bienestar y la pobreza se debe evaluar lo referente a las capacidades de las personas en lugar de los bienes o recursos que posean, por lo que las políticas públicas deben enfocarse en garantizar las capacidades de las personas, capacidades entendidas como una habilidad no desarrollada o efectivamente realizada, es decir una capacidad potencial que se puede incrementar aumentando la libertad. Sen articula tres conceptos: *capacidades, funcionamientos y bienes primarios.* La *capacidad* se refiere a las opciones asequibles a una persona entre los cuales puede elegir lo que razonablemente valora. Los *funcionamientos* son las realizaciones, esto es, las capacidades de ser o hacer, elegidas por cada persona de entre las "*n*" combinaciones que se le presentan. Los *bienes primarios*, por su parte, son convertidos mediante el uso que se les da en algo valorado (Groppa, 2004). Sen no establece las dimensiones ni variables que deben plantearse para la metodología, pero si argumenta que éstas deben salir de la misma sociedad mediante el debate público, ya que hay diferenciación en el bienestar de las personas en función de sus características personales, clima y costumbres sociales, medio ambiente, preferencias, etc.; es decir, las capacidades a las que se refiere dependerán de cada sociedad y hay que evaluarlas en su contexto y realidad.

En cuanto a la ponderación matemática asignada, se recomienda una distribución equitativa para cada dimensión a evaluar, y de igual manera para el número de indicadores en cada dimensión, así, si se evaluará las tres dimensiones del *IDH* (salud, educación y riqueza) cada uno de ellos tendría un peso del 33,33% y como dentro de educación tenemos dos variables (escolaridad de los mayores a 25 años y años esperados de escolaridad para niños) cada indicador representaría el 16,66%. El puntaje máximo de pobreza multidimensional es 100% y se considera a una persona multidimensionalmente pobre cuando los indicadores ponderados de las carencias sumen, por lo menos, un 33,33%. Los autores dejan al criterio del investigador en función de su respaldo teórico y práctico la ponderación que den a los

indicadores y dimensiones, ya que lo establecido por ellos y luego por otros autores "puede ser vista como un juicio de valor que debe estar abierto al debate y escrutinio público" (Alkire & Foster 2007), sin embargo, la referencia sería:

Tabla 2: Tipo de pobreza según índices multidimensionales

Índice	*Clasificación*
$i \geq 50\%$	Pobreza multidimensional extrema
$i \geq 33{,}33\%$	Pobreza multidimensional
$20\% \leq i \leq 33{,}33\%$	Vulnerabilidad o riesgo de caer en pobreza multidimensional

Fuente: Alkire & Foster

Básicamente, la metodología multidimensional de Alkire-Foster consiste en:

Plantear una matriz $n \times d$ que coteje la población y el número de dimensiones en estudio, donde n es el número de personas que componen la población por lo que será un número positivo mayor que 0 y d es el número de dimensiones en que se va a evaluar la pobreza, así por ejemplo si se considerara educación, salud y riqueza d tomaría el valor de 3, los subíndices i y j tomaran valores entre $i = 1,2,\dots,n$ y $j = 2,3,\dots,d$ debiendo ser como mínimo $d \geq 2$ para representar la multidimensional del análisis.

Cada vector-fila y_i muestra el logro del individuo i en las diferentes dimensiones, mientras que cada vector-columna y_j evidencia la distribución de logros de la dimensión j para un grupo de individuos. Con esto, se tendría la matriz Y con elementos y_{ij}, la función que contiene la información multidimensional de los individuos es $Y = \{y \in R_+^{nd} : n \geq 1, d \geq 2\}$

$$Y_{ij} = \begin{bmatrix} n_1 d_1 & n_1 d_2 & \cdots & n_1 d_j \\ n_2 d_1 & n_2 d_2 & \cdots & n_2 d_j \\ \vdots & \vdots & \ddots & \vdots \\ n_i d_1 & n_i d_2 & \cdots & n_i d_j \end{bmatrix}$$

Luego, se tiene un vector-fila Z con elementos z_j que representan el umbral de pobreza debajo del cual se considera a la persona pobre en la dimensión j, por tanto $z_j > 0$.

Los autores en su análisis y considerando a Bourguignon y Chakravarty (2003), utilizan la función $p: R_+^d \times R_{++}^d \to \{0,1\}$ que ejecuta un mapeo del vector de desempeño de la

persona i y del vector de línea de corte z en R_{++}^{d}a una variable indicador de manera tal que $p(y_{ij}, z) = 1$ si la persona i es pobre o $p(y_{ij}, z) = 0$ si la persona i no es pobre. De esto, se obtiene como resultado el conjunto $Z \subseteq \{1, \dots, n\}$ de personas que son pobres en y dado z, cotejando la información de cada persona con el umbral establecido de pobreza. Posteriormente el paso de agregación toma a p como dado y asocia la matriz, así como el vector de línea de corte z con un nivel general $M(y; z)$ de pobreza multidimensional. La relación funcional resultante $M: Y \times R_{++}^{d} \to R$ es el índice o medida de pobreza multidimensional.

La función $M = (p, M)$ representa la nueva medida multidimensional, en donde los datos se expresan en privaciones, y esta matriz se expresa por $g^0 = \lfloor g_{ij}^0 \rfloor$ y toma valores de 0 y 1. Cuando $y_{ij} < z_j$ toma el valor de 1 o caso contrario 0 g^0 seguirá siendo una matriz de tamaño $n \times d$. Partiendo de la matriz g^0 se construye un vector-columna c de recuento de privaciones o carencias, cuya observación *i-ésima* representa el número de privaciones sufridas por la persona i.

En la praxis, un breve ejemplo del método multidimensional de Alkire y Foster, se detalla a continuación:

Dada una matriz Y de dimensión $n \times d$ con elementos y_{ij}:

$$Y = \begin{bmatrix} 13 & 9 & 4 & 2 \\ 8 & 5 & 10 & 15 \\ 15 & 7 & 2 & 5 \\ 12 & 14 & 8 & 9 \end{bmatrix}$$

y un vector-fila Z con umbrales de pobreza z_j para cada dimensión:

$$Z = [13 \quad 12 \quad 3 \quad 8]$$

Se forma la matriz de carencias g^0, cotejando cada fila de la matriz Y con los umbrales del vector-fila Z, asignando 1 al estar dentro o bajo el límite establecido por el umbral o 0 en caso contrario, es decir, la matriz g^0 asigna 1 a la persona i pobre en la dimensión j, identificándose así a los pobres de la población. Para el presente caso:

$$g^0 = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

El vector-columna C que totaliza el número de carencias de los individuos, sería:

$$C = \begin{bmatrix} 3 \\ 2 \\ 3 \\ 1 \end{bmatrix}$$

Ahora, cada uno de los elementos c_i es el número de dimensiones en las que presentan privaciones los individuos, por ejemplo, el cuarto individuo tiene privación en una sola dimensión y al tratarse la metodología de multi-dimensionalidad debe considerarse como restricción $k \geq 2$, por lo que la matriz censurada $C(k)$ sería:

$$C(k) = \begin{bmatrix} 3 \\ 2 \\ 3 \\ 0 \end{bmatrix}$$

Realizado esto, se calcula la Incidencia de la Pobreza (*H*) con el ratio: número de personas en pobreza multidimensional sobre total de individuos estudiados; para el presente caso hay tres de cuatro personas lo que representaría:

$$H = \frac{q}{n} = \frac{3}{4} = 0.75$$

Como puede apreciarse, la incidencia de pobreza no cambiará así un individuo aumente o disminuya el número de dimensiones en que tiene privaciones, por lo que, es necesario trabajar con otro índice denominado Intensidad de Pobreza (*A*) que toma en cuenta el número de privaciones de un individuo en promedio. En el presente caso la Intensidad de Pobreza para cada individuo (A_i) y la Intensidad de Pobreza total estarían dados por:

$$A_1 = \frac{3}{4} = 0.75$$

$$A_2 = \frac{2}{4} = 0.50$$

$$A_3 = \frac{3}{4} = 0.75$$

$$A_4 = \frac{0}{4} = 0$$

$$A = \frac{\sum_{i=1}^{n} A_i}{j} = \frac{2}{3} = 0.667$$

La Intensidad de Pobreza o Promedio de la Proporción de Privaciones (*A*) es una medida de pobreza multidimensional que informa la cantidad de dimensiones en las que una persona pobre promedio tiene privaciones, en el presente caso, un pobre promedio tiene privación en el 66.7% de las dimensiones por lo que según la clasificación del modelo caería en pobreza multidimensional extrema según la referencia de Alkire & Foster citada anteriormente.

M_0 se define como la tasa de Recuento Ajustada y es la proporción de carencias que experimenta la población pobre sobre el máximo posible de carencias que podría experimentar toda la población, es decir, del vector columna $C(k)$ tenemos que la población pobre experimenta ocho privaciones y que dieciséis -en el ejemplo- es el máximo posible de carencias que podría sufrir toda la población, es decir $n \times d$ privaciones, lo que equivaldría al producto de *H* por *A*, así:

$$M_0 = \frac{\sum c_i}{n \times d} = H \times A = \frac{3}{4} \times \frac{2}{3} = \frac{6}{12} = 0.50$$

Que se interpreta que el 50% de la población en estudio es multidimensionalmente pobre.

CAPÍTULO 2

TÉCNICAS DE ANÁLISIS CUANTITATIVO PARA EL ESTUDIO DE LA POBREZA

Desde el punto de vista de la investigación cuantitativa el estudio de la pobreza ha generado un interés tan grande como el que existe desde el punto de vista económico, político o social, el problema radica en que las técnicas estadísticas más sencillas (univariables) no permiten explicar la complejidad del fenómeno de la pobreza, debido a que, como otros fenómenos sociales, la pobreza no responde al comportamiento de una sola variable y es necesario un enfoque multivariable donde se pueda establecer la relación entre varias variables para así poder entender la naturaleza del fenómeno. Se entiende por otro lado, que estudios como el de la pobreza son también de naturaleza cualitativa debido a que un individuo u hogar se lo puede clasificar como “pobre” o “no pobre” lo que conlleva al uso de variables de carácter binario o dicotómico; esta mezcla de variables tanto cualitativas como cuantitativas hacen que el grupo de técnicas estadísticas destinadas al estudio de este tipo de fenómenos se vea restringido a un grupo específico. Generalmente el análisis de regresión se ha vuelto un estándar para el estudio de los fenómenos sociales debido especialmente a su naturaleza multivariable, su poder explicativo, y su facilidad de interpretación; sin embargo, las estimaciones de regresión con una variable cualitativa dependiente pueden conducir a errores graves en la inferencia, por lo tanto, lo que se necesita es un conjunto de técnicas estadísticas que puedan hacer el trabajo de regresión multivariable pero que no estén sujetas a desventajas en presencia de variables dependientes cualitativas, afortunadamente en la actualidad, existen varios procedimientos de este tipo y el avance de las ciencias computacionales ha contribuido en el desarrollo de las actividades de análisis.

2.1 MODELADO ESTADÍSTICO

El uso de modelos es una práctica común del ser humano en un intento de comprender o predecir aspectos sobre la realidad que los rodea; en un contexto científico un modelo se expresa de forma matemática para describir un fenómeno y su construcción implica una mezcla arte y conocimientos según las palabras de McCullagh y Nelder (1989). Un modelo busca explicar la variación en una respuesta a partir de la relación entre una fuente

de variabilidad determinística y otra aleatoria, Judd y McClelland (1989) plantean esta idea mediante la expresión:

$$Datos = Modelo + Error$$

Los *Datos* hacen referencia al fenómeno que se está buscando explicar, es decir las observaciones bajo análisis o dicho de otra manera la variable dependiente, el *Modelo* es una función de varias variables predictoras (variables independientes) que busca explicar el fenómeno que se está estudiando (los datos), y dado que los modelos rara vez logran explicar un fenómeno de una manera 100% precisa se introduce el *Error* que representa la discrepancia entre los *Datos* y el *Modelo*, lo que se busaca de manera lógica es minimizar el *Error* y que el *Modelo* sea una representación lo más cercana posible de la realidad (*Datos*). Independientemente del fenómeno que se encuentre bajo estudio el modelado estadístico debe considerar los siguientes criterios o especificaciones para cualquier estudio: a) el *principio de parsimonia* (principio de Occam), hace referencia a la simplicidad del modelo tomando en cuenta que si bien la abstracción es parte implícita y necesaria en un modelo, este debe tratar de representar la realidad de una manera simple y sobria ante una situación compleja (Judd & McClelland, 1989), b) el *principio de bondad de ajuste:* toma en cuenta la inclusión de las variables independientes en el modelo con el propósito de una mejor representación de la variable dependiente y la consecuente disminución del error (Ruiz-Soler, Pelegrina & López-González, 2000), y c) *integración teórica*, el modelo debe ser consistente con la teoría y tener un buen nivel de predicción (Ato & Vallejo, 2007).

No obstante, se debe tener en cuenta que se pueden cometer errores de especificación por parte del investigador, según Rao y Miller (1971) estos errores están relacionados principalmente con: la omisión de variables con poder explicativo lo que causa sesgo en los parámetros estimados, la inclusión de variables extrañas en cuyo caso se obtienen parámetros insesgados pero que pueden influir en la adopción de una función para el análisis equivocada, y los errores de medición o uso de base de datos (Morales, 2001); para evitar estos problemas y lograr un modelo más parsimonioso se recomienda seguir las siguientes etapas:

1. Especificación del modelo teórico: Se debe determinar cuáles son las variables de interés y la relación que existe entre ellas, además se debe tener en consideración

la simplicidad del modelo (parsimonia) y que se tenga una representación lo suficientemente buena de la realidad que permita la minimización del error (bondad de ajuste).

2. Estimación de parámetros: Hace referencia al cálculo de los coeficientes a partir de los datos observados para determinar si el modelo es aceptable como una representación de la realidad.
3. Selección del modelo: Tomando en cuenta la de diferencia que pueda existir entre los datos observados y los datos ajustado se toma la decisión de optar por el modelo seleccionado o caso contrario rechazarlo.
4. Evaluación del modelo: En el contexto más general se refiere a la comprobación de los supuestos de normalidad, linealidad, homoscedasticidad e independencia
5. Interpretación del modelo: Se debe comprender y explicar las implicaciones respecto al modelo y a la situación bajo análisis con un criterio estadístico, lógico y sustantivo.
6. Retroalimentación: Tomando en cuenta los cinco pasos mencionados se toma la decisión de aceptar o no el modelo, en caso de no aceptar el modelo se debe volver a reiniciar el proceso si es necesario.

2.2 EL ANÁLISIS MULTIVARIABLE

El objetivo de un modelo de regresión múltiple consiste en calcular el valor esperado de una variable dependiente cualitativa *Y* dados los valores de un grupo de variables independientes o regresoras (tres o más) en el que se estiman los coeficientes de regresión con la finalidad de poder explicar el efecto que tendrá en la variable dependiente las variaciones que se puedan dar en las variables independientes. En palabras de Gujarati el análisis de regresión múltiple es “el análisis de regresión condicional sobre los valores fijos de las variables explicativas, y lo que obtenemos es el valor promedio o la media de *Y*, o la respuesta media de *Y* a los valores dados de las regresoras”, en el caso donde *Y* sea cualitativa (dicotómica) el propósito del análisis de regresión es encontrar la probabilidad de que un evento ocurra.

El origen de los modelos de regresión multivariable se remonta hacia finales del siglo XIX e inicios del siglo XX, sin embargo, dada la limitada capacidad de procesamiento de información que existía en ese entonces no es hasta la década de los ochenta que se

generaliza su uso debido en gran medida al desarrollo de los ordenadores y su aplicación en el análisis estadístico en numerosas disciplinas (López Roldan & Bachelli, 2015).

Algunas de las ventajas derivadas del uso del análisis multivariable y que fueron mencionadas por Catell (1966) son:

- Economía en el almacenamiento de datos.
- Mayor consistencia en la inferencia estadística.
- Desarrollo de conceptos teóricos más adecuados.
- Mayor precisión y perspectiva conceptual.

Debido a la variedad de métodos y técnicas existentes de análisis multivariable su clasificación resulta dificultosa, además están los varios enfoques que se pueden operar de forma simultánea y que incluso se pueden superponer, lo que dificulta aún más la tarea de clasificar las técnicas de análisis multivariable, desde el punto de vista de López Roldan y Bachelli (2015) se pueden considerar cuatro criterios para la clasificación de estas técnicas de análisis.

El primer criterio responde al modelo de análisis de la investigación y se establece tomando en cuenta su finalidad analítica-explicativa, dos tipos de técnicas se desprenden de este criterio.

- *Técnicas de análisis multivariable explicativo-causales:* Plantean hipótesis proposicionales simples o hipótesis de dependencia lineal.
- *Técnicas exploratorias y de estructuración:* Implican niveles elementales de categorización o contrastación, establecen estructuras de interrelación entre variables para identificar los factores más discriminantes de la realidad social.

Un segundo criterio hace referencia a un esquema algebraico y técnico-instrumental antes que a una finalidad analítica. Las técnicas enmarcadas en este grupo pueden ser aquellas que impliquen:

- Relación de dependencia entre las variables, donde se establece una diferencia entre las variables dependientes e independientes.

- Relación de interdependencia, o simple asociación de variables, donde todas las variables tienen la misma consideración, generalmente de variables independientes.

Un tercer criterio considera el número de variables a considerar para el análisis. Se pueden considerar dos situaciones.

- Primero, cuando se trata de modelos de relación de dependencia se identifica el número de variables dependientes que intervendrán en el análisis suponiendo previamente la existencia de dos o más variables independientes.
- Segundo, cuando se trata de modelos de interdependencia, se considera la relación que pueda existir entre más de dos variables.

Finalmente, los dos criterios anteriores se pueden relacionar con la naturaleza de las variables considerando que estas pueden ser de tipo cualitativo (numéricas, continuas o discretas) o cuantitativo (categóricas, nominales u ordinales).

Los criterios de clasificación expuestos no son los únicos que existen, pero tratan de categorizar y explicar de la forma más clara la complejidad que implica el análisis multivariable, la tabla 3, que se expone a continuación, muestra una clasificación de un buen número de técnicas de análisis multivariable considerando en lo posible los criterios ya mencionados, sin embargo, algunas de las técnicas que se muestran son difíciles de categorizarlas dentro de un grupo específico debido a la complejidad que implica su uso y posterior análisis.

Otras técnicas de análisis multivariable y que no corresponden a algunas de las clasificaciones mostradas en la Tabla 3 son:

- Análisis de Redes Sociales.
- Minería de Datos.
- Análisis de Redes Neuronales.
- Simulación Social.
- Análisis de Decisión Multicriterio.

Tabla 3: Clasificación de técnicas de análisis multivariable

<table>
<tr><td colspan="2" rowspan="2">Análisis de relaciones de Interdependencia V ↔ V</td><td>Variables Cuantitativas</td><td colspan="3">Análisis Factorial Exploratorio
Análisis de Componentes Principales
Análisis de Clasificación
Análisis de Escalamiento Multidimensional Métrico</td></tr>
<tr><td>Variables Cualitativas</td><td colspan="3">Análisis de Tablas de Contingencia Multidimensionales
Análisis Log-Lineal
Análisis de Clasificación
Análisis de Escalamiento Multidimensional No Métrico
Análisis de Clases latentes</td></tr>
<tr><td rowspan="10">Análisis de relaciones de Dependencia VI → VD</td><td rowspan="6">Una Variable Dependiente</td><td rowspan="3">VD Cuantitativa</td><td>VI Cuantitativas</td><td colspan="2">Análisis de Regresión Lineal Múltiple
Análisis de Regresión No Lineal Múltiple</td></tr>
<tr><td>VI Cualitativas</td><td colspan="2">Análisis de Varianza Múltiple
Análisis de Segmentación
Análisis Conjunto</td></tr>
<tr><td>VI Cualitativas y Cuantitativas</td><td colspan="2">Análisis de Covarianza</td></tr>
<tr><td rowspan="3">VD Cualitativa</td><td>VI Cuantitativas</td><td colspan="2">Análisis Discriminante</td></tr>
<tr><td>VI Cualitativas</td><td colspan="2">Análisis Log-Lineal Logit
Análisis de Segmentación
Análisis Conjunto</td></tr>
<tr><td>VI Cualitativas y Cuantitativas</td><td colspan="2">Análisis de Regresión Ordinal
Análisis de Regresión Lineal Múltiple con variables ficticias
Análisis de Regresión Logística
Análisis de Regresión Probit</td></tr>
<tr><td rowspan="4">Dos o más Variables Dependientes</td><td rowspan="3">VD Cuantitativas</td><td>VI Cuantitativas</td><td colspan="2">Análisis de Modelos de Ecuaciones Estructurales</td></tr>
<tr><td>VI Cualitativas</td><td colspan="2">Análisis Multivariable de Varianza</td></tr>
<tr><td>VI Cualitativas y Cuantitativas</td><td colspan="2">Análisis de Covarianza Múltiple
Análisis de Correlación Canónica</td></tr>
<tr><td rowspan="2">VD Cualitativas</td><td>VI Cuantitativas</td><td colspan="2">Análisis de Discriminación Múltiple</td></tr>
<tr><td>VI Cualitativas</td><td colspan="2">Análisis Log-Lineal Logit Múltiple</td></tr>
</table>

V: variable VD: variable dependiente VI: variable independiente.

Fuente: López González & Bachelli

Como se mencionó al inicio de este capítulo y haciendo un contraste con la tabla de clasificación de técnicas de análisis multivariable se puede ubicar al estudio de la pobreza en la categoría donde se trabaja con una variable dependiente de tipo cualitativo mientras las variables independientes son una mezcla de variables de tipo cuantitativo y cualitativo, es decir que las técnicas que podrían ser utilizadas para el estudio de este fenómeno social son: el Análisis de Regresión Ordinal, el Análisis de Regresión Lineal Múltiple con variables ficticias, el Análisis de Regresión Logística y el Análisis de Regresión Probit; más adelante se describirán estos tres últimos.

2.3 MODELO DE REGRESIÓN LINEAL MULTIVARIABLE GENERALIZADO

El modelo de regresión múltiple es una extensión del modelo de regresión simple (modelo con regresor único) que se estudia como modelo básico de regresión, en el cual se han incluido variables adicionales que cumplen el papel de regresoras; este modelo permite estimar el efecto que tiene sobre una variable dependiente aleatoria Y_i con observaciones independientes $(y_1, y_2, \dots, y_N)$ la variación de una variable X_{i1} mientras se mantiene constante el resto de variables regresoras $(X_{i1}, X_{i2}, \dots, X_{ik})$. Un modelo de regresión lineal con K variables independientes $(1, \dots, K)$ está dado por:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK} + \varepsilon_i \qquad [2.1]$$

O de forma más general y utilizando la notación de sumatoria:

$$Y_i = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ik} + \varepsilon_i \qquad [2.2]$$

Pare al caso β_0 hace referencia al termino independiente o constante, β_k es el coeficiente de la variable independiente o explicativa X_{ik} (coeficiente de la pendiente de X_{ik}) y ε_i es el error o componente aleatorio no percibido (el residual), el subíndice i denota la *i-esima* observación de la muestra de tamaño N y se considera a la combinación $\beta_k X_{ik}$ como el componente sistemático del modelo.

Al considerar el modelo lineal expuesto se debe tomar en cuenta que la linealidad se puede dar de distintas maneras, según lo cual se obtienen diferentes modelos; es decir, a un modelo se lo considera lineal si lo es en sus parámetros, independientemente si las variables explicativas cumplen o no la condición de linealidad, si el modelo es lineal tanto

en sus variables explicativas como en sus parámetros se asemejará al modelo general ya expuesto, caso contrario si el modelo es lineal en sus parámetros pero no en sus variables se habla de un modelo de *m-ésimo* orden con Km variables independientes y $Km + 1$ parámetros, estos modelos son susceptibles de ser linealizados mediante la inclusión de componentes de interacción y la conversión de sus variables explicativas, de forma general estos modelos presentan la siguiente formulación:

$$Y_i = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ik} + \sum_{k=1}^{K} \beta_k X_{ik}^2 + \cdots + \sum_{k=1}^{K} \beta_k X_{ik}^m + \varepsilon_i$$

Finalmente, si el modelo formulado no es lineal ya sea en sus parámetros y/o variables explicativas este puede adoptar diferentes formulaciones. Se puede asumir que la *i-ésima* observación y_i es una realización de la variable aleatoria Y_i cuyo valor esperado está dado por $E(Y_i) = \mu_i$, si se asume que el vector de observaciones se encuentra sobreentendido el modelo anteriormente descrito en una forma matricial se podría expresar mediante la siguiente ecuación:

$$E(Y) = \mu = \sum_{k=1}^{K} \beta_k x_k \qquad [2.4]$$

Se pude introducir el término η para obtener un modelo más generalizado lo que permite vincular la función $\sum_{k=1}^{K} \beta_k X_k$ con μ en una forma no necesariamente lineal, en general se puede expresar η como un predictor lineal producido por K variables independientes $(x_1, x_2, \dots, x_k)$ y parámetros desconocidos. Independiente del tipo de modelo el conjunto de variables explicativas siempre produce η como un predictor lineal de Y, la relación entre η y las variables independientes está dado por:

$$\eta = \sum_{k=1}^{K} \beta_k x_k \qquad [2.5]$$

Sin embargo la relación entre η y μ debe ser especificada, ya que esta relación permite distinguir ciertos modelos lineales generalizados de otros, esta relación se especifica mediante una denominada función de enlace $\eta = g(\mu)$ y su formulación depende en gran medida del tipo de estudio que se esté realizando; los tipos de funciones de enlace más

comunes para $g(\mu)$ y que se utilizan en la investigación social se mencionan a continuación (Liao, 1994):

1. *Lineal*: donde $\eta = \mu$, por lo cual el modelo se asemeja a la ecuación 2.5 .
2. *Logit*: $\eta = \log[\mu/(1-\mu)]$, aplicando esta función en 2.5 se obtiene un modelo Logit que genera un una respuesta con variable de tipo binario.
3. *Probit:* $\eta = \Phi^{-1}(\mu)$, donde Φ^{-1} representa la inversa de la función de distribución acumulativa normal estándar; de manera similar, esta función de enlace especifica un modelo Probit que produce una variable de tipo binaria como resultado.
4. *Logarítmica:* $\eta = \log(\mu)$, normalmente, se utiliza el logaritmo natural y se especifica un modelo de regresión de Poisson.
5. *Logit multinomial:* $\eta_j = \log(\mu_i/\mu_J)$ donde *j* representa la *j-ésima* observación en $1, \ldots, J$ categorías de respuestas; esta función se entiende como una extensión del modelo Logit en el cual J es igual a 2 lo que genera un modelo Logit multinomial donde la variable de resultado bajo estudio es de carácter politómico.

La elección del tipo de función de enlace o modelo estadístico depende en gran medida de la distribución de los datos y el conocimiento que tenga el investigador de la naturaleza de los mismos; específicamente, la distribución del componente aleatorio en Y (la parte que no puede ser sistemáticamente explicada por variables x) determina la función de enlace y el tipo de modelo lineal generalizado. En los modelos lineales generalizados la distribución del componente aleatorio proviene de una familia exponencial, a los cuales pertenecen la distribución normal, la distribución exponencial, la distribución de Poisson y algunas otras.

La tabla 4 muestra una clasificación más detalla de los modelos lineales generalizados, se puede observar como dentro de la categoría de los modelos lineales generalizados el modelo lineal es el más elemental, en este caso el componente aleatorio del modelo resulta tener una distribución normal, este hecho implica gran importancia ya que según sea la distribución de los errores serán las distribuciones condicionadas de los valores pronosticados.

Tabla 4: Principales modelos lineales generalizados

Naturaleza de la Variable de Respuesta	**Componente**		**Función de enlace**	**Modelo Lineal**
	Sistemático	**Aleatorio**		
Numérica cuantitativa	Numérico	Normal	Identidad	Regresión lineal
	Categórico	Normal	Identidad	ANOVA o de diseño experimental
	Mixto	Normal	Identidad	ANCOVA o de diseño experimental con variables concomitantes
Categoría binaria				
No agrupada	Mixto	Binomial (1) Bernoulli	Logit	Regresión logística
Agrupada (frecuencias)	Categórico	Binomial (n)	Logit Probit	Análisis Logit Regresión Probit
Categórica polinómica				
No agrupada	Mixto	Multinomial	Logit generalizado	Regresión logística multinomial
Agrupada (frecuencias)	Categórico	Multinomial	Logit generalizado	Análisis Logit multinomial
Recuento	Mixto	Poisson	Logarítmica	Regresión de Poisson
Frecuencia	Categórico	Poisson	Logarítmica	Análisis loglineal

Fuente: López González & Ruiz Soler

Las funciones de enlace del tipo Probit y logarítmica están basadas en la distribución binomial, muchos de los fenómenos que se estudian en el campo de las ciencias sociales responden a este tipo de distribución y usan estos modelos para su estudio, situaciones como pobre o no pobre, vivo o muerto, o en general cualquier situación que plantee la ocurrencia o no ocurrencia de algún evento son situaciones que se estudian con estos modelos.

La función de enlace de tipo logarítmica sigue una distribución de Poisson, esta distribución es utilizada usualmente en las ciencias sociales para estudiar variables del

tipo cuantitativo como los votos que pueden existir en una elección, el número de turistas que pierden su vuelo por año u otras situaciones que no resultan ser tan comunes.

Por último, la función de enlace multinomial se presenta en situaciones donde los resultados se pueden dar a partir de un grupo de más de dos opciones, como el momento en que un estudiante de bachillerato debe elegir una carrera universitaria o cuando una familia debe elegir un destino para sus vacaciones, este tipo de elecciones se pueden considerar como situaciones politómicas.

2.4 INTERPRETACIÓN DE LOS PARÁMETROS ESTIMADOS

De la ecuación 2.5 se sabe que la relación de cada x en η se da de una forma siempre lineal, por lo que interpretar los parámetros estimados como efectos lineales del predictor η resulta común para para todos los modelos lineales generalizados; una característica importante que se debe tener en cuenta y que es común de todos los modelos lineales generalizados es que cada estimación representa el efecto parcial que tiene un coeficiente mientras las demás variables se mantienen controladas, a esto es lo que se conoce como el criterio del *ceteris paribus* (permaneciendo todo lo demás constante) y se debe tener presente ya que se aplica a todos los modelos lineales generalizados de forma similar a cuando se interpreta las estimaciones que se realizan con un modelo de regresión lineal simple.

Liao en su obra “Interpreting Probability Models” identifica cinco formas a través de las cuales se puede interpretar los parámetros estimados para un modelo de probabilidad, (1) la interpretación del signo de los parámetros estimados y su significancia estadística, (2) los valores predichos o transformados de η dado un conjunto de valores en las variables explicativas, (3) el efecto marginal de una variable explicativa sobre η o sobre una transformación de η, (4) las probabilidades estimadas dado un conjunto de valores en las variables explicativas y (5) el efecto marginal de una variable explicativas sobre la probabilidad de un evento.

1. *La interpretación del signo de los parámetros estimados y su significancia estadística.*

   Este método de interpretación puede ser usado en cualquier modelo de probabilidad ya que no depende del tipo de función de enlace que se utilice en el modelo, además

es el más utilizado en muchos estudios y un importante número de investigadores lo prefiere, primero debido a su simplicidad, y segundo debido a que es utilizado al momento de aplicar el método de estimación de los mínimos cuadrados ordinarios.

Establecido un cierto nivel de significancia un signo positivo para uno de los parámetros estimados indica la probabilidad que la respuesta o cierto evento aumente a medida que aumenta la presencia de x mientras los otros parámetros permanecen constantes, de igual manera, la presencia de un signo negativo sugiere que una respuesta o evento decrece según decrece el nivel de presencia de x. Sin embargo, se debe considerar que esta interpretación no es la más acertada debido a que no se conoce como puede afectar en la probabilidad de ocurrencia de un evento el incremento o decremento de x; no obstante la ventaja de aplicar este método radica en su facilidad y que es indiferente al tipo de función de enlace que se esté aplicando en el modelo .

2. *Los valores estimados de η o transformados de η dado un conjunto de valores en las variables explicativas.*

Este tipo de método no es de uso amplio a pesar de que la estimación de la probabilidad de un evento versus la probabilidad de no ocurrencia dado un modelo Logit o Probit es relativamente sencilla, dada esta circunstancia no se profundizará en la teoría relacionada a este tema, sin embargo, si se desea profundizar a cerca de este tema se puede revisar los trabajos realizados por Liao y Aldrich.

3. *El efecto marginal de una variable explicativa sobre η o sobre una transformación de η.*

Independientemente del modelo de probabilidad que se esté utilizando una estimación de β_k genera un efecto marginal de la correspondiente variable explicativa x_k en η; tal como en un modelo de regresión clásico estos efectos marginales son lineales por lo que su interpretación resulta sencilla. Como en un modelo lineal ordinario se puede decir que el incremento de una unidad en x_k generaría una cantidad β_k estimada de incremento o decremento en η mientras los demás parámetros se mantienen constantes.

Sin embargo, tal interpretación del efecto marginal, aunque sencilla, no es tan útil porque η no es directamente interpretable (o Y no es interpretable en términos de η) a menos que se esté trabajando con una identidad, como en el caso del modelo lineal clásico. A menudo, es más significativo interpretar el efecto de un β_k estimado en una transformación de η en lugar de en η mismo.

4. *Las probabilidades estimadas dado un conjunto de valores en las variables explicativas.*

En un análisis de regresión clásico, un investigador podría desear calcular los valores medios previstos de Y condicionados un conjunto de valores para las variables x. Del mismo modo, es posible que se desee calcular las probabilidades predichas a partir de modelos de probabilidad como el Logit y el Probit, sujetos a un conjunto de valores para las variables x. Las probabilidades predictas son intuitivamente atractivas porque estas probabilidades dan una idea de que tan posible es que ciertos tipos de individuos tomen cierto curso de acción (o tengan ciertas actitudes).

En el marco de modelos lineales generalizados, las probabilidades predichas a menudo se derivan mediante el cálculo de los valores de μ, utilizando ciertos valores de las variables x. Entonces se convierte en una cuestión de expresar μ, que representa la probabilidad en muchos modelos lineales generalizados, como una función de las variables x, lo que se especifica por la función de enlace de cada modelo de probabilidad. Existen excepciones, en el modelo de regresión de Poisson, por ejemplo, μ no está directamente relacionado con la probabilidad de un evento, sino con la expectativa de contar eventos. Esto sugiere que con el modelo de regresión de Poisson las probabilidades predichas se calculan de manera diferente y que la interpretación en términos de μ se da por el conteo de los eventos esperados. Por lo tanto, se debe calcular las probabilidades predichas asociadas basadas en la función de enlace específico para cada modelo.

5. *El efecto marginal de una variable explicativa sobre la probabilidad de un evento.*

Esta última forma de interpretación combina rasgos de las dos formas anteriores, el efecto marginal sobre η o una transformación de η y las probabilidades estimadas dado un conjunto de valores en las variables explicativas. Se puede interpretar el

efecto marginal de x_k, expresado por el β_k estimado en la probabilidad de un evento, en lugar de en η o una transformación de η.

Al igual que con el cálculo de las probabilidades predichas, se necesita determinar el nivel de cada una de las variables x para calcular el efecto marginal. Además, la función de enlace determina la fórmula específica para calcular el efecto marginal sobre la probabilidad en un modelo de probabilidad en particular.

Las cinco técnicas mencionadas no son exclusivas y existen algunas variaciones para estos métodos de interpretación, en econometría, por ejemplo, se suele utilizar la interpretación de las elasticidades lo cual resulta ser un tipo de efecto marginal.

2.5 CONSIDERACIONES PARA EL CÁLCULO DE LOS PARÁMETROS ESTIMADOS

Antes de poder estimar los parámetros que caracterizan al modelo lineal generalizado es necesario estableces un conjunto de supuestos denominados hipótesis del modelo lineal generalizado. Cabe recalcar que algunos de estos supuestos son similares a los que se consideran para el modelo clásico de regresión lineal, sin embargo, se deben tomar en cuenta ciertas atenciones para su aplicación en el modelo lineal generalizado, ya que existen ciertas diferencias en la relación entre los valores ajustados de μ y el predictor lineal η entre el modelo lineal generalizado y el modelo lineal clásico. Estas diferencias se concretan en la función de enlace y en la distribución que ésta debe seguir.

Hipótesis 1: Linealidad en los parámetros. Establece la linealidad en los parámetros en la relación que existe entre las variables independientes y la variable dependiente del modelo, esta relación de linealidad no aplica a las variables del modelo pudiendo existir o no una relación lineal entre estas estas. Mientras que el modelo lineal clásico se produce una relación de identidad entre los valores ajustados y el predictor lineal, $\eta = \mu$, en el modelo lineal generalizado la linealidad se establece en la escala del predictor lineal pero no en la escala de los valores ajustados, por lo que, entre los valores ajustados y los valores predichos media una función que los relaciona, la función de enlace: $\eta = g(\mu)$. Las tres funciones siguientes, cumplen con la hipótesis de linealidad en los parámetros, sin embargo, la primera es la única estrictamente lineal.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K$$

$$Y = \beta_0 + \beta_1 e^{x_1} + \beta_2 e^{x_2} + \cdots + \beta_K e^{x_K}$$

$$Y = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \cdots + \beta_K \ln x_k$$

Un modelo en donde no se cumple la hipótesis de linealidad en los parámetros sería, por ejemplo:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2^2 x_2 + \cdots + \beta_K^n x_K$$

Existen ciertos modelos donde la linealidad de los parámetros se puede obtener mediante una transformación de η, sin embargo la transformación no es siempre posible como por ejemplo en el caso anterior.

Hipótesis 2: Especificación correcta. Esta hipótesis supone que las variables independientes "x" escogidas para ser incluidas en el modelo son aquellas relevantes para explicar la variable dependiente "y".

Hipótesis 3. Grados de libertad positivos. Los grados de libertad en un modelo de regresión se definen cono la diferencia que existe entre el número de observaciones (n) y el número de parámetros por estimar (K); esta hipótesis supone que los grados de libertad del modelo debe ser mayor o igual a cero, o dicho de otra forma supone que, el número de observaciones debe ser como mínimo igual al número de parámetros a estimar; en la práctica siempre será preferible poseer más observaciones que parámetros a estimar.

Hipótesis 4: Parámetros constantes. Esta hipótesis supone que los parámetros se mantienen constantes en el tiempo o según el individuo.

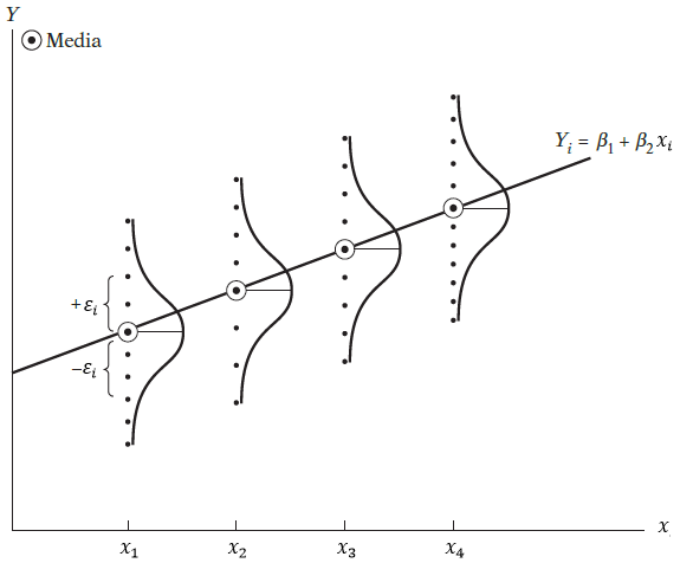
Hipótesis 5: Independencia lineal entre las variables explicativas Esta hipótesis implica que cada variable independiente contiene información adicional sobre la dependiente que no está contenida en otra, dicho de otra forma esta hipótesis implica que no existe una relación lineal exacta entre ninguna de las variables "x" del modelo, no cumplirse esta hipótesis supondría que alguna de las variables explicativas es redundante. A esta hipótesis también se la conoce con el nombre de principio de no colinealidad.

Hipótesis 6: Represores no estocásticos. Esta hipótesis implica que los valores que toman las variables explicativas "x" pueden considerarse fijos en muestras repetidas. Además, se debe considera que los valores de "x" deben ser independientes del término del error lo cual significa que la covarianza entre cada variable "x" y ε_i debe ser cero.

Hipótesis 7. Hipótesis referentes al error. Esta hipótesis se divide a su vez en tres; (1) esperanza nula en todo instante de tiempo, (2) varianza constante u homoscedasticidad y (3) ausencia de autocorrelación.

(1) Esperanza nula en todo instante de tiempo: Esta hipótesis supone que $E(\varepsilon_i) = 0, \forall i = 1,2,\dots,n$; es decir, dado el valor de x_i la media del término del componente aleatorio ε_i es cero. Esta hipótesis sostiene que los factores que no han sido incluido en el modelo de forma explícita y que, por consiguiente se encuentran integrados en ε_i no perturban el valor de la media de Y, en otras palabras se podría decir que valores positivos de ε_i se ven cancelados por valores negativos de ε_i de manera que el efecto sobre Y es igual a cero, además esto sugiere que las perturbaciones idealmente deberían estar normalmente distribuidas; gráficamente esta hipótesis se representa con la siguiente figura.

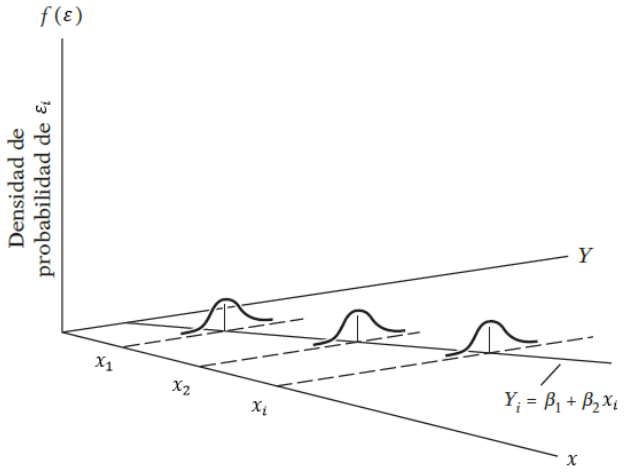
Figura 1. Distribución condicional de las perturbaciones ε_i.



Fuente: Gujarati, 2010

(2) Varianza constante u homoscedasticidad: Esta hipótesis implica que $\text{var}(\varepsilon_i) = E(\varepsilon_i{}^2) = \sigma^2, \forall i = 1,2,\dots,n$; dicho de otra manera, la varianza de ε_i para cada x_i es una cantidad constante igual a σ^2, lo que significa que las poblaciones de Y correspondientes a diferentes valores de x tienen igual varianza. En la figura 2 esto se puede apreciar al notar que la variación en torno a la línea de regresión es la misma para todos los valores de x.

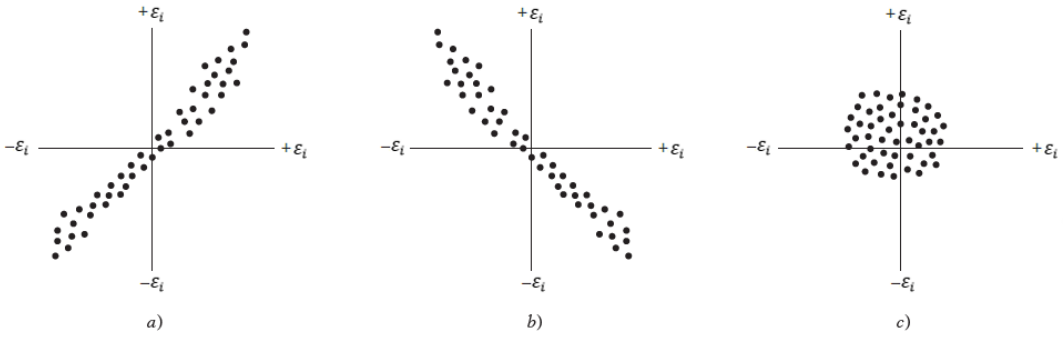
Figura 2. Homoscedasticidad



Fuente: Gujarati, 2010

(3) *Ausencia de autocorrelación:* Esta hipótesis comprende que $\text{cov}(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i, \varepsilon_j) = 0, \forall i, j = 1,2, \dots, n, i \neq j$, en otras palabras, dados dos valores x_i y x_j $(i \neq j)$ la correlación entre ε_i y ε_j $(i \neq j)$ cualquiera es cero. Si no hay autocorrelación, eventos pasados no ayudan a predecir comportamientos futuros y los errores son completamente aleatorios y no previsibles; es decir las observaciones se muestran de manera independiente. La figura 3a) y b) son ejemplos donde se puede observar un patrón para ε por lo que existe correlación serial o autocorrelación, lo que requiere esta hipótesis es que dichas correlaciones estén ausentes, la figura 3.6c) es un ejemplo de esta caso.

Figura 3. Patrones de correlación entre las perturbaciones



Fuente: Gujarati, 2010

Ahora bien, existen ciertos autores que consideran que no es importante la veracidad de estas hipótesis sino las predicciones basadas en estos supuestos. Independientemente si se coincide o no con este punto de vista se debe considerar que la inclusión de los supuestos tiene una intención más allá de reproducir la realidad exactamente que es la de facilitar el desarrollo de la materia en pasos graduales.

Como se mencionó al inicio de esta sección, algunos de estos supuestos no se aplican de forma estricta al modelo lineal generalizado, por ejemplo, mientras que en el modelo lineal clásico el componente aleatorio debe distribuirse normalmente, en el modelo lineal generalizado el componente aleatorio no sigue obligatoriamente una distribución normal sino que utiliza cualquier distribución del tipo exponencial y, en consecuencia, las distribuciones de los valores pronosticados del criterio no serán normales necesariamente. Por otro lado, la condición de homoscedasticidad en el componente aleatorio del modelo clásico de regresión permite que las distribuciones condicionadas de los valores pronosticados de la variable de respuesta también lo sean; dado que en el modelo lineal generalizado los errores pueden seguir cualquier distribución de la familia exponencial, la condición de homoscedasticidad no es imprescindible para la distribución de los errores. Estas diferencias además obligan a tener que utilizar un método distinto al de los mínimos cuadrados para la estimación de los parámetros, estas diferencias se hacen más evidentes en algunos los modelos de respuesta cualitativa como los que se menciona a continuación.

2.6 MODELOS CON VARIABLE DEPENDIENTE DICOTÓMICA

En ocasiones ciertos fenómenos que se desean estudiar no son de naturaleza continua sino discreta, por ejemplo, cuando se desea analizar la condición de pobreza, la variable de respuesta, o regresada, sólo puede adquirir dos valores, 1 si la persona se clasifica como “pobre” o 0 si se clasifica como “no pobre”, en otras palabras, la variable regresada es de carácter dicotómico o binario, y es importante notar que a la final se trata de una variable de carácter cualitativo. Los modelos donde la variable dependiente Y es cualitativa tienen por objetivo encontrar la probabilidad de que un evento ocurra. Por lo que, a los modelos de regresión con respuestas cualitativas se los conocen como modelos de probabilidad.

El enfoque del modelo lineal generalizado discutido previamente puede ser utilizado para describir modelos donde la variable regresada es de carácter dicotómico. Este enfoque asume una variable de respuesta Y^* definida por la relación:

$$Y^* = \sum_{k=1}^{K} \beta_k x_k + \varepsilon \qquad [2.6]$$

En la práctica, Y^* no se observa, y ε está distribuido simétricamente con una media cero y tiene función de distribución acumulativa definida como $F(\varepsilon)$. Lo que se observa es una variable dummy Y, es decir, una realización de un proceso binomial, definido por:

$$Y = \begin{cases} 1 & \text{si } Y^* > 0 \\ 0 & \text{caso contarrio} \end{cases}$$

De la relación planteada se puede escribir:

$$\begin{aligned} \Pr(Y = 1) &= \Pr\left(\sum_{k=1}^{K} \beta_k x_k + \varepsilon > 0\right) \\ \Pr(Y = 1) &= \Pr\left(\varepsilon > -\sum_{k=1}^{K} \beta_k x_K\right) \\ \Pr(Y = 1) &= 1 - F\left(-\sum_{k=1}^{K} \beta_k x_K\right) \end{aligned} \qquad [2.7]$$

Donde F es una función de distribución acumulativa de ε. Se puede ver η, el predictor lineal generalizado, como el componente sistemático en Y^*, y ε como el componente aleatorio en Y^*. La forma funcional de F depende de la asunción hecha sobre la distribución de ε en la ecuación 2.6. Obviamente, la distribución de ε determina la función de enlace de un modelo lineal generalizado, otra de las formas de representar F.

Los modelos de regresión de respuesta binaria son: (1) El modelo lineal de probabilidad, (2) El modelo Logit, (3) El modelo Probit y (4) El modelo Tobit. A continuación, se procederá a describir brevemente los tres primeros ya que son los que se consideran como los más adecuados para el estudio de fenómenos sociales como el de la pobreza, una descripción más a detalle se puede revisar en Aldrich & Nelson (1984) o en Liao (1994).

2.6.1 Modelo lineal de probabilidad

El modelo de lineal de probabilidad se asemeja mucho al modelo lineal de regresión común, pero con la diferencia que la variable dependiente es binaria o dicotómica, en este modelo el valor esperado de Y_i dado x_i puede entenderse como la probabilidad condicional de que el evento ocurra dado x_i, esto es, $\Pr(Y_i = 1|x_i)$. Ahora, si la probabilidad de que un evento ocurra ($Y_i = 1$) se denota por P_i, la probabilidad de que el evento no ocurra ($Y_i = 0$) estará denotada por $1 - P_i$ por lo que se concluye que Y_i sigue una distribución de Bernoulli, de lo cual se deriva la siguiente expresión.

$$E(Y_i = 1|x_i) = \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i = P_i$$

Partiendo del hecho de que el modelo lineal de probabilidad se asemeja mucho al modelo clásico de regresión lineal se podría pensar que el método de mínimos cuadrados ordinarios es adecuado para la estimación de los parámetros del modelo, desafortunadamente este hecho no resulta cierto y se deben considerar ciertos aspectos.

Primeramente, si bien el método de los mínimos cuadrados ordinarios no requiere que las perturbaciones sigan una distribución normal, por cuestiones de inferencia estadística se consideraba el cumplimiento de este supuesto, sin embargo, esta condición de normalidad para ε_i ya no se mantiene en el modelo lineal general de probabilidad debido a que ε_i solo pude tomar dos valores al igual que Y_i, es decir que ε_i también sigue una distribución de Bernoulli o también llamada binomial.

En segundo lugar, para el modelo de regresión lineal clásico se consideraba que la media del término del componente aleatorio ε_i era cero y que no existía autocorrelación entre ε_i y ε_j $(i \neq j)$, es decir se suponía que $E(\varepsilon_i) = 0, \forall i = 1,2, \dots, n$ y que $\text{cov}(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i, \varepsilon_j) = 0, \forall i, j = 1,2, \dots, n\,, i \neq j$; si bien estos dos supuestos se pueden seguir sosteniendo, ya no es posible sostener la hipótesis referente a la homoscedasticidad de las perturbaciones, esto ocurre debido a que la varianza para una distribución de Bernoulli es una función de la media que se maneja en este tipo de distribución la cual resulta ser igual a la probabilidad de éxito de un evento, por lo que la varianza de los errores deja de ser homoscedástica para convertirse en heteroscedástica. Las estimaciones realizadas mediante el método de los mínimos cuadrados ordinarios dejan de ser eficientes en

presencia de heteroscedasticiad por lo que es necesario la aplicación de otros métodos para superar esta situación.

Tercero, debido a que un modelo lineal de probabilidad mide la probabilidad que ocurra un evento Y_i dado x_i esta probabilidad necesariamente debe estar en el rango de entre 0 y 1, sin embargo, no se puede garantizar que $\widehat{Y}_i$ cumpla con esta condición ya que el método de estimación de los mínimos cuadrados ignora esta restricción y es aquí donde radica el verdadero problema de tratar de estimar los parámetros del modelo lineal de probabilidad mediante esta técnica. Existen algunas alternativas para subsanar este inconveniente, primero si la probabilidad estimada es un número negativo se supondrá que la probabilidad estimada es igual a 0, y segundo si la probabilidad es mayor que 1 se supondrá que la probabilidad estimada tomará el valor posible más alto, es decir 1. Los modelos Logit y Probit que se analizarán en secciones siguientes garantizan que las probabilidades estimadas se encuentren efectivamente en el rango de entre 0 y 1.

Por último, el coeficiente de determinación (R^2) calculado mediante el método de los mínimos cuadrados ordinarios tiene un valor limitado en los modelos de respuesta dicotómica, esto se debe a que el valor de $\widehat{Y}_i$ únicamente puede ser 0 o 1 por lo que todos los valores posibles $\widehat{Y}_i$ deberían encontrarse sobre el eje de las x ($\widehat{Y}_i$=0) o en la recta correspondiente a $\widehat{Y}_i$=1 por lo cual no se espera existe un modelo de lineal de probabilidad que se ajuste a dicha dispersión. Por estas razones, Aldrich & Nelson argumentan que "debe evitarse el coeficiente de determinación como estadístico de resumen en modelos con variable dependiente cualitativa".

Hasta antes del desarrollo de herramientas informáticas que permitiesen estimar de forma rápida y práctica los coeficientes resultantes del uso de los modelos Logit o Probit, los modelos lineales de probabilidad gozaban de gran aceptación debido a la simplicidad de sus cálculos.

2.6.2 El modelo Logit

Cuando se asume que el componente aleatorio para una variable de respuesta sigue una distribución binomial se puede asumir además que se trabajará con una distribución del tipo logística (acumulativa), de manera que se puede utilizar un modelo del tipo Logit

para el análisis de los datos, y la función de enlace a utilizar sería la correspondiente a este tipo de modelo, es decir

$$\eta = \log[\mu/(1-\mu)].$$

Al aplicar esta función de enlace a la ecuación 2.6, se especifica un modelo Logit que produce una variable de resultado binaria. El modelo Logit por lo general toma dos formas. Puede expresarse en términos del logaritmo de la razón de las probabilidades de un evento (en términos de Logit), en esta forma el modelo se especifica como:

$$\log\left[\frac{\Pr(Y=1)}{1-\Pr(Y=1)}\right] = \sum_{k=1}^{K} \beta_k x_k \qquad [2.8]$$

Debido a que se está modelando $\Pr(Y=1)$, μ se convierte en la probabilidad esperada de que Y sea igual a 1. Usando la ecuación 2.7 se puede transformar la ecuación 2.8 en una especificación del modelo Logit de probabilidad reemplazando la función de distribución acumulativa general y F, con una función de distribución acumulativa específica y L, que representa la distribución logística.

$$\Pr(Y=1) = 1 - L\left(-\sum_{k=1}^{K} \beta_k x_k\right) = L\left(\sum_{k=1}^{K} \beta_k x_k\right) = \frac{e^{\sum_{k=1}^{K} \beta_k x_k}}{1 + e^{\sum_{k=1}^{K} \beta_k x_k}} \qquad [2.9]$$

La ecuación 3.4 representa la probabilidad de que ocurra un evento. Para la no ocurrencia del evento, la probabilidad es solamente 1 menos la probabilidad de que ocurra el evento.

Debido a las dos formas de modelos Logit que se han expuesto ésta modelación en ocasiones recibe dos nombres distintos, el modelo expresado como (2.8) conduce a "modelos Logit" debido al término Logit, mientras que el expresado como (2.9) conduce a "regresión logística" debido a la función de distribución logística acumulativa.

La distinción que se realiza entre estos dos nombres suele basarse en si las variables explicativas continuas se incluyen en el conjunto de variables x. Se suele llamar modelos Logit a aquellos que incluyen variables categóricas en el conjunto de las x y modelos de regresión logística a aquellos que incluyen variables categóricas y continuas para x.

2.6.2.1 Interpretación de los modelos Logit

Para interpretar los resultados de un modelo Logit de manera significativa, el modelo primeramente debe ajustarse a los datos. Dicho de otra manera, las variables explicativas incluidas en el modelo deben ser capaces de explicar la variable de respuesta significativamente mejor que los modelos de un solo interceptor, lo cual cierto para todos los modelos lineales generalizados. En un modelo de regresión clásico, se usa una prueba F; en un modelo de Logit (y otros modelos de probabilidad), la prueba más comúnmente utilizada es el test de la razón de verosimilitudes, que sigue aproximadamente la distribución de Chi-cuadrado (χ^2). Si el test de la razón de verosimilitud indica que el modelo se ajusta a los datos significativamente se procede a interpretar las estimaciones de los parámetros.

Como se mencionó antes existen cinco maneras en las cuales se podría interpretar los parámetros estimados para un modelo de probabilidad, la primera forma, la interpretación del signo de los parámetros estimados y su significancia estadística, será omitida ya que existen maneras más fáciles y útiles de interpretar los modelos Logit, la interpretación de los odds (la probabilidad de que suceda un evento dividido por la probabilidad de que no suceda) y odds ratios (la relación entre dos probabilidades).

1. *El efecto marginal de una variable explicativa sobre η o sobre una transformación de η*

   Debido a que los efectos lineales y aditivos de los parámetros en el Logit (logaritmo de la razón de las probabilidades) no son tan intuitivamente atractivos, se calcula el antilogaritmo de ambos lados de la función de enlace (Ecuación 3.3) y se obtiene:

$$\frac{\Pr(Y=1)}{1-\Pr(Y=1)} = e^{\sum_{k=1}^{K}\beta_k x_k} = \prod_{k=1}^{K}\beta_k x_k \qquad [2.10]$$

   El lado izquierdo de la expresión representa la probabilidad de que suceda un evento dividido por la probabilidad de que no suceda (llamado también odds), y el lado derecho da el efecto marginal de x_k en las probabilidades indicadas por e^{β_k}. Aquí el concepto de odds y de la razón de probabilidades (odds ratios), que es la relación entre dos probabilidades, se vuelven centrales.

La razón de probabilidades (odd ratios) tienen cuatro propiedades principales como medida de asociación (Morgan & Teachman, 1988).

a. Permite una interpretación clara de cómo el paso de un odds a otro provoca un cambio multiplicativo como el que se indica en la razón de probabilidades. Una razón de probabilidades mayor que 1 indica una mayor probabilidad de que ocurra un evento versus la probabilidad de que no ocurra, y una razón de probabilidades menor que 1 indica una disminución en la probabilidad de la ocurrencia de un evento versus la no ocurrencia del mismo.
b. Es invariable con respecto al orden de las variables, excepto por su "signo".
c. Es invariante con respecto al número de multiplicaciones en las variables, es decir, si se aumenta el tamaño de la muestra en un determinado estudio, se debería obtener aproximadamente los mismos odds y, por lo tanto, las mismas razones de probabilidades de como si no se lo hiciera.
d. Se puede usar en el estudio de variables y modelos de respuestas politómicas con múltiples variables explicativas.

2. *Los valores predichos o transformados de η dado un conjunto de valores en las variables explicativas.*

Debido a la facilidad para comprender la probabilidad de ocurrencia de un evento, en ocasiones también resulta interesante intentar calcular una predicción para la probabilidad dado un conjunto de valores para las variables x. Uno de los principales problemas es qué valores elegir para las variables x. Si se tiene muchas variables independientes, algunas de las cuales son categóricas y otras continuas, probablemente sea aconsejable centrarse en una o dos variables de interés a la vez y establecer los valores para las otras variables según sus medias muestrales. Si una de las dos variables seleccionadas es discreta y la otra es continua, se puede elegir medir los valores de probabilidad predichos con respecto a la variable continua x dentro de cada categoría de la variable discreta. La decisión de cuantos puntos de medición se debe tomar en la variable continua para realizar la predicción es principalmente una decisión basada en el juicio del investigador, al asunto es lograr un balance adecuado entre simplicidad y claridad por un lado y especificidad y precisión por el otro.

3. *El efecto marginal de una variable explicativa sobre la probabilidad de un evento.*

En lugar de examinar el efecto marginal de una variable x en los odds, también se puede observar el efecto marginal de la variable sobre la probabilidad del evento. El efecto viene dado por la siguiente ecuación:

$$\frac{\partial \Pr(Y=1)}{\partial x_k} = \frac{e^{\sum_{k=1}^{K}\beta_k x_k}}{\left(1+e^{\sum_{k=1}^{K}\beta_k x_k}\right)^2}\beta_k = P(1-P)\beta_k \qquad [2.11]$$

Donde, ∂ representa la derivada parcial de la probabilidad de ocurrencia de un evento respecto a x_k, P una abreviación para $\Pr(Y=1)$, y $1-P$ representa la $\Pr(Y=0)$. Sin embargo, a diferencia de la interpretación del efecto sobre las probabilidades, que es invariable a los valores de las variables independientes, el efecto marginal en la probabilidad cambia con los valores de x y, por lo tanto, con la probabilidad asociada con los valores de las x. Además, las derivadas parciales dan solo una estimación aproximada del efecto marginal de una variable ficticia sobre la probabilidad, aunque dan una aproximación cercana del efecto marginal de una variable continua sobre la probabilidad.

La derivada de una función se define solo para variables continuas. Por lo tanto, las derivadas no están definidos en principio para variables discretas, como una variable ficticia con valores de 0 y 1 solamente, aunque pueden dar una estimación aproximada del efecto marginal de una variable ficticia en la práctica. Una mejor estimación de dicho efecto marginal se calcula tomando la diferencia de la probabilidad condicional predicha en cada una de las dos categorías de la variable ficticia.

2.6.3 El modelo Probit

El modelo Probit es otro de los modelos estadísticos ampliamente utilizados para el estudio de datos con distribuciones del tipo binomial, estos modelos se encuentran dentro de la categoría de los modelos lineales generalizados y su función de enlace está dada por:

$$\eta = \Phi^{-1}(\mu)$$

donde ϕ^{-1} representa la inversa de la función de distribución acumulativa normal. El inverso de la función de distribución acumulativa normal es en efecto una variable

estandarizada, o una puntuación estándar (Z score). Al igual que con el modelo Logit, el modelo Probit se usa para estudiar una variable de resultado binaria. Se puede expresar modelos Probit en función de la probabilidad de ocurrencia de un evento.

$$\Pr(Y = 1) = 1 - F\left(-\sum_{k=1}^{K} \beta_k x_k\right) = F\left(\sum_{k=1}^{K} \beta_k x_k\right) = \Phi\left(\sum_{k=1}^{K} \beta_k x_k\right)$$

donde la forma más general de la función de distribución acumulativa, F, se reemplaza por la función de distribución acumulativa normal estándar Φ. A diferencia del modelo Logit, que podía expresarse en dos formas distintas (en términos de Logit y en términos de la probabilidad de un evento) el modelo Probit toma una única forma debido a que un modelo Probit expresado en función η es una regresión lineal de una puntuación estándar (Z-score) de la probabilidad de un evento. La probabilidad de no ocurrencia de un evento estará dada por $1 - \Pr(Y = 1)$, es decir,

$$\Pr(Y = 0) = 1 - \Phi\left(\sum_{k=1}^{K} \beta_k x_k\right) \qquad [2.13]$$

2.6.3.1 Interpretación de los modelos Probit

Dada la falta de formas alternativas para la representación de los modelos Probit, su interpretación se vuelve más limitada, sin embargo, se tratará de realizar una analogía respecto a los puntos tratados en el modelo Logit.

1. *El efecto marginal de una variable explicativa sobre η*

   Debido a que no existe una manera sencilla de transformar η, que en este caso toma la forma de la inversa de la función de distribución acumulativa normal estándar, interpretar el efecto marginal sobre η significa interpretar los efectos lineales y aditivos de las variables x en el inverso del función de distribución acumulativa normal estándar.

   Una alternativa para la interpretación resultaría de considerar la función de distribución acumulativa normal estándar como un Z-score y el efecto marginal sobre η como un efecto sobre el Z-score, sin embargo, esta opción no resulta tan simple

como parece, lo que resta su atractivo práctico a pesar de que genera buenos resultados.

Una segunda opción consistiría en transformar los Z-score en probabilidades e interpretar los efectos marginales sobre la probabilidad de un evento, esto lleva a las dos formas de interpretación que se mencionarán a continuación y que en la práctica resultan más útiles que trabajar con la interpretación de los parámetros estimados mediante la medición de sus efectos sobre un Z-score.

2. *Probabilidades estimadas dado un conjunto de valores en las variables explicativas.*

La interpretación de los resultados generados a través de un modelo Probit mediante el análisis de las probabilidades estimadas es idéntico al explicado para los modelos Logit, la única diferencia radica en el tipo de función de distribución acumulativa que se utiliza en cada modelo, otra diferencia a tomar en cuenta tiene que ver con el cálculo de las probabilidades estimadas; para los modelos Logit el cálculo resulta sencillo y puede ser realizado fácilmente con el uso de herramientas como una calculadora, sin embargo, para este cálculo en los modelos Probit será necesario recurrir al uso de un paquete estadístico o a la revisión de las tablas para el área bajo la curva de una curva de distribución normal.

Las probabilidades estimadas calculadas a partir del modelo Probit son casi idénticas a las del modelo Logit, y difieren en menos del 5 por mil. Por lo tanto, los resultados llevarán a conclusiones similares. Si el cálculo de la función Φ es fácilmente accesible, estas probabilidades pueden ser tan fáciles de calcular como las que se usan el modelo Logit.

3. *El efecto marginal de una variable explicativa sobre la probabilidad de un evento*

Al igual que en el modelo Logit el análisis estará basado en la aplicación de la derivada parcial de la probabilidad de ocurrencia de un evento respecto a x_k.

$$\frac{\partial \Pr(Y=1)}{\partial x_k} = \Phi\left(\sum_{k=1}^{K} \beta_k x_k\right)\beta_k \qquad [2.14]$$

> La derivada parcial de las variables ficticias sobrestima el efecto marginal, si se compara con las diferencias correspondientes con las probabilidades estimadas por las ecuaciones 2.12 y 2.13, como en el caso de Logit. Al comparar los efectos marginales basados en la ecuación 2.14 o la diferencia de probabilidad con los efectos correspondientes al modelo Logit, las diferencias están presentes, pero son pequeñas, lo que sugiere que cualquier método conducirá a conclusiones similares.

Dadas las similitudes entre los dos tipos de modelos, ambos modelos ofrecerán conclusiones sustantivas idénticas en la mayoría de las aplicaciones. De hecho, uno puede ir de un conjunto de estimaciones a otro. Si se multiplica una estimación Probit por un factor, se obtiene un valor aproximado de la estimación Logit correspondiente. Normalmente se cree que este factor es $\pi/\sqrt{3}$. Sin embargo, un valor de 1.6 es más recomendado por varios autores. El valor más preciso del factor se encuentra en algún lugar cercano a estos dos valores. Sin embargo, hay situaciones en las que las estimaciones de los modelos Logit y Probit pueden diferir sustancialmente y, en tales casos, se debe tener cuidado al elegir el modelo más apropiado. Estos son casos con un número extremadamente grande de observaciones y con una gran concentración de observaciones en las colas de la distribución.

En el siguiente capítulo se precederá a la utilización de estos dos últimos modelos para explicar el fenómeno social de la pobreza en el Ecuador.

CAPÍTULO 3

MODELACIÓN MATEMÁTICA PARA LA MEDICIÓN DE LA POBREZA EN LAS CIUDADES DE QUITO, GUAYAQUIL Y CUENCA

En el primer capítulo de este trabajo se abordó el fenómeno la pobreza como un fenómeno social detallando sus causas, perspectivas, definición y métodos de medición; posteriormente, en el capítulo dos se describió las técnicas estadísticas más utilizadas para la medición y descripción de los fenómenos sociales, en este punto se hizo énfasis en los modelos Logit y Probit y su aplicabilidad como una herramienta para trabajar en la descripción de fenómenos sociales como la pobreza; partiendo de esta base teórica en este capítulo se procederá a estudiar la pobreza y la relación entre los factores que la pueden llegar a determinar ya sean estos de naturaleza social, económica o demográfica y desde una perspectiva de los jefes de hogar utilizando como base los modelos Logit y los modelos Probit.

El Instituto Nacional de Estadísticas y Censo (INEC) establece las dimensiones e indicadores que se aplican a nivel país considerando las Constitución de la República del Ecuador y la teorización respecto de la Pobreza. Como insumos para el desarrollo del presente capítulo se cuenta con la Encuesta Nacional de Empleo, Desempleo (ENEMDU), los modelos Logit, Probit y la Carta Magna a fin de poder determinar si a iguales variables de análisis para las tres principales ciudades del país, se presentan variaciones significativas en los parámetros estimado con objeto de demostrar si son necesarias políticas focalizadas para la disminución de los índices de pobreza en las distintas ciudades.

3.1 LA ENCUESTA NACIONAL DE EMPLEO, DESEMPLEO Y SUBEMPLEO

En el año de 1985 se crea en el Ecuador el Instituto Nacional de Empleo (INEM) cuyo objetivo era obtener información relacionada a la fuerza laboral del país, en la persecución de este objetivo se implementa en el año de 1987 la Encuesta Permanente de Empleo y Desempleo, la misma que se aplicaría a personas mayores de 12 años en el área urbana de las ciudades de Quito, Guayaquil y Cuenca con una periodicidad anual. Posteriormente para el año de 1988 se implementa la Encuesta Nacional Urbana sobre Empleo de

periodicidad anual ubicando el estudio en las ciudades de Portoviejo, Guayaquil, Machala, Tulcán, Quito, Ambato y Cuenca, para el levantamiento del año 1989 se amplió el número de ciudades objeto de la encuesta a 65 buscando una representatividad de las regiones costa, sierra y oriente, además se modificó el universo de estudio involucrando a todas aquellas personas mayores de 10 años, posteriormente para los años de 1990, 1991 y 1992 se mantuvieron los dominios y el universo de estudio.

Para el año 1993 se implementa la ya hoy conocida Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), la cual desde ese entonces ha estado a cargo del Instituto Nacional de Estadísticas y Censos, hasta el año 2003 la metodología, periodicidad y representatividad de la encuesta no cambiaron siendo en este año que se inicia una intervención de periodicidad trimestral donde además se incluyeron variables importantes relacionadas con el campo socioeconómico y demográfico. En el año 2007 se realiza una revisión metodológica y se incorporan mejoras a la misma con el propósito de institucionalizar y unificar la generación de estadística sobre empleo en el país absorbiendo el trabajo que hasta entonces realizaba el Banco Central del Ecuador con la encuesta de coyuntura del mercado laboral ecuatoriano.

En el año 2014 y 2015 se realizan cambios respecto al incremento del tamaño de la muestra, la actualización del marco conceptual de la Población Económicamente Activa y la actualización metodológica de la medición del empleo en el sector informal.

En la actualidad la ENEMDU capta información de las 24 provincias del país, es una encuesta por muestreo a hogares integrado por personas de 5 años y más, y se trabaja en las tres regiones: Costa, Sierra y Oriente considerando la división cantonal y la existencia de parroquias urbanas y rurales, también se considera la existencia de cinco ciudades auto representadas, Cuenca, Machala, Guayaquil, Quito y Ambato.

3.1.1 Antecedentes

El modelo neoclasista fundaba la medición de la pobreza en el Ingreso-Consumo, Amartya Sen rompe ese paradigma monetarista y considera la “capacidad” para contribuir a una mejor sociedad en donde la pobreza es un grado de privación que impide el desarrollo pleno de las capacidades, destrezas y derechos, asentados en Convenciones Internacionales y Constituciones de los Estados.

Esta evolución conceptual inicia en 1981 con Paúl Streeten y la CEPAL, introduciendo los principios de las Necesidades Básicas Insatisfechas (NBI); posteriormente, para 1989 Katman agrega al enfoque de las necesidades básicas insatisfechas la dimensión del ingreso y las definiciones de pobreza transitoria y crónica. En 1990 Mahbub ul Haq introduce el Índice de Desarrollo Humano basado en tres dimensiones: Esperanza de Vida, Educación y Riqueza; llegando a 2007 Alkire y Foster establecer el Índice de Pobreza Multidimensional.

En nuestro país, en el 2014 Ecuador plantea la Estrategia Nacional para la Igualdad y la Erradicación de la Pobreza dirigida por la Secretaría Nacional de Planificación y Desarrollo (SENPLADES) que se enmarca dentro de los 17 Objetivos Mundiales para el Desarrollo Sostenible de la Organización de las Naciones Unidas.

Dentro de esta Estrategia Nacional para la Igualdad y la Erradicación de la Pobreza, Ecuador emplea la metodología Alkire-Foster para determinar un índice de pobreza, esta metodología requiere información desagregada y simultanea de hogares y personas para lo que se emplea la encuesta Nacional de Empleo, Desempleo y Subempleo ENEMDU que es elaborada por un equipo multisectorial y aplicada por el Instituto Nacional de Estadísticas y Censos, INEC.

3.1.2 Definición

Según el Instituto Nacional de Estadísticas y Censo: “La Encuesta Nacional de Empleo, Desempleo y Subempleo es una encuesta por muestreo probabilístico, cuyo propósito principal es la medición y seguimiento del empleo, desempleo y la caracterización del mercado de trabajo, que permite conocer la actividad económica y las fuentes de ingresos de la población cuya cobertura es a nivel nacional, regional y provincial y a nivel urbano y rural”.

3.1.3 Objetivo

La finalidad principal de la Encuesta de Empleo, Desempleo y Subempleo, es conocer la actividad económica y las fuentes de ingresos de la población. La información recolectada está orientada a entregar datos de las principales categorías poblacionales en relación con el mercado de trabajo: activos (ocupados, desocupados) e inactivos y a obtener clasificaciones de estas categorías según diversas características. También posibilita

confeccionar series temporales homogéneas de resultados. Además, al ser las definiciones y criterios utilizados coherentes con los establecidos por los organismos internacionales que se ocupan de temas laborales, permite la comparación con datos de otros países.

3.1.4 Metodología

La ENEMDU se aplica en nuestro país como la conocemos hoy desde el 2009, y para su aplicación requiere de dos momentos: Identificación y Agregación. Para la Identificación se aplican dos etapas: primero definir los indicadores de medición y su umbral de corte, segundo determinar el número de privaciones para definir como pobre o no a una persona u hogar. La Agregación consiste en el cálculo de los índices de Incidencia, Intensidad y el Índice de Pobreza Multidimensional (tasa de Incidencia ajustada por la Intensidad de Pobreza).

3.1.4.1 Unidad de análisis e identificación

En la Encuesta la unidad de análisis e identificación es el hogar, donde las privaciones de uno de sus integrantes son trasladadas al hogar y viceversa, por lo que todos los miembros del hogar van a tener el mismo número de privaciones, se fomenta así la solidaridad y equidad intra-hogar. La encuesta está diseñada con seis indicadores de índice personal y seis indicadores de índole familiar.

3.1.4.2 Dimensiones

El armaje de la Encuesta Nacional de Empleo, Desempleo y Subempleo basa su accionar en el concepto del Buen Vivir donde prevalecen los derechos y capacidades sobre la acumulación y el consumo lo cual se ve reflejado en el segundo capítulo de nuestra Constitución. En primera instancia se definen cuatro dimensiones: Educación; Trabajo y Seguridad Social; Salud, agua y alimentación; Hábitat, vivienda y ambiente sano.

Para el Derecho de la Educación, los artículos 26 y 28 de la Marta Magna la señalan como deber ineludible e inexcusable del estado, área prioritaria de la inversión pública, su gratuidad hasta el nivel de bachillerato. Dentro de la Ley Orgánica de Educación Intercultural los artículos 28 y 42 sustentan la obligatoriedad de los niveles de básica (diez años) y bachillerato (tres años). La Ley Orgánica de Educación Superior en sus artículos

4 y 80 brinda el ejercicio efectivo de la igualdad de oportunidades y ratifica la educación superior pública gratuita vinculada a la responsabilidad académica del estudiante.

En Ecuador el trabajo para menores de 15 años está prohibido según el artículo 42 de la Ley Suprema de Montecristi y el artículo 82 del Código de la Niñez y Adolescencia. Para los adolescentes (15 a 17 años) la Constitución en su artículo 46, numeral 2, establece: "…que el trabajo será excepcional y no podrá conculcar su derecho a la educación ni realizarse en situaciones nocivas o peligrosas para su salud o su desarrollo personal". El artículo 33 de la Constitución establece que el trabajo es un derecho y un deber social, reconoce todas las formas de trabajo, entre ellos el no remunerado de auto-sustento y cuidado humano que se realiza en los hogares. En cuanto a Seguridad Social, el artículo 34 de la Constitución indica que "la seguridad social es un derecho irrenunciable de todas las personas, y será deber y responsabilidad primordial del Estado. La seguridad social se regirá por los principios de solidaridad, obligatoriedad, universalidad […]". Son sujetos a solicitar la protección del Seguro, todas las personas que perciben ingresos por su actividad económica (Ley de la Seguridad Social, art. 2) contemplando la obligatoriedad, solidaridad y subsidiariedad del sistema.

El Agua es un derecho fundamental e irrenunciable establecido en el artículo 314 en la Constitución y en el artículo 57 de la Ley Orgánica de Recursos Hídricos, Usos y Aprovechamiento del Agua (2014), establece el derecho de todas las personas a disponer de agua limpia, suficiente, salubre, […] para el uso personal y doméstico en cantidad, calidad, continuidad y cobertura. En el artículo 37 de la mencionada ley consideran la provisión de agua potable y saneamiento ambiental como servicios públicos básicos relacionados con el agua. El artículo 362 de la Carta Magna menciona que es una obligación del Estado garantizar la calidad, calidez, seguridad, información, transparencia y confidencialidad de la atención en salud con especial preferencia a los grupos vulnerables que establece la Constitución.

A la entrada en vigencia de la Constitución del 2008, todas las leyes infra debían adecuarse al enfoque de Derechos, estando a la fecha en debate en la Asamblea Nacional el Código de la Salud, sin embargo, de esto, el artículo 362 de la Carta Magna establece como una obligación del Estado garantizar la calidad, calidez, seguridad, información, transparencia y confidencialidad de la atención en salud. En la vigente Ley Orgánica de la Salud (2006) en su artículo 7 establece que toda la población tiene derecho sin

discriminación alguna al acceso universal, oportuno y de calidad de acciones y servicios de salud y que los servicios públicos estatales deben ser universales y gratuitos para todos los procedimientos y niveles de atención, con especial preferencia para los grupos vulnerables. En la actual metodología de la Encuesta hay limitaciones para evaluar la dimensión Salud por lo que se la evalúa transversalmente a través de agua, alimentación y ambiente sano.

Para el Derecho a la Alimentación, la Constitución en los artículos 13 y 281 plantea que todos los habitantes tienen derecho al acceso seguro y permanente a alimentos sanos, suficientes y nutritivos debiendo el Ecuador garantizar la soberanía alimentaria para una autosuficiencia de alimentos. En la ENEMDU no existen indicadores de consumo alimenticio, pero se realiza una aproximación mediante la línea pobreza extrema por ingresos que se construye a partir de la canasta alimenticia.

En la Dimensión de Hábitat, vivienda y ambiente nuestra Carta Magna en el artículo 30 reconoce el derecho de las personas “a un hábitat seguro y saludable, y a una vivienda adecuada y digna, con independencia de su situación social y económica”. En el artículo 66 se reconoce el derecho al agua potable, vivienda, saneamiento ambiental y el artículo 264 establece como competencias de los gobiernos municipales el prestar los servicios públicos de agua potable, alcantarillado, depuración de aguas residuales, manejo de desechos sólidos, actividades de saneamiento ambiental.

3.1.4.3 Indicadores

La metodología seguida es la del método deductivo, yendo de lo macro a lo micro, puesto que, partiendo de la Constitución de la República y Convenios Internacionales, se plantean las dimensiones que reflejen los derechos sustentados en los modelos de estudio de la pobreza, y es necesario desagregarlos para su operatividad de análisis e identificación.

Según la metodología planteada para la ENEMDU y en sustento a la teoría del IPM y la igualdad de jerarquía de los derechos constitucionales del Ecuador, las cuatro dimensiones tienen igual peso de ponderación, el mismo que se subdivide para sus indicadores, así:

Figura 4. Dimensiones, Ponderación, Indicadores y Población del IPM en Ecuador

Dimensión	Pesos	Indicador	Población aplicable
Educación (25%)	8.3%	1. Inasistencia a educación básica y bachillerato	5 a 17 años
	8.3%	2. No acceso a educación superior por razones económicas	18 a 29 años
	8.3%	3. Logro educativo incompleto	18 a 64 años
Trabajo y Seguridad social (25%)	8.3%	4. Empleo infantil y adolescente	5 a 17 años
	8.3%	5. Desempleo o empleo inadecuado	18 años y más
	8.3%	6. No contribución al sistema de pensiones	15 años y más
Salud, Agua y Alimentación (25%)	12.5%	7. Pobreza extrema por ingresos	Toda población
	12.5%	8. Sin servicio agua por red pública	Toda población
Hábitat, Vivienda y Ambiente sano (25%)	6.25%	9. Hacinamiento	Toda población
	6.25%	10. Déficit habitacional	Toda población
	6.25%	11. Sin saneamiento de excretas	Toda población
	6.25%	12. Sin servicio de recolección de basura	Toda población

Fuente: www.ecuadorencifras.gob.ec

En este contexto, según la ENEMDU que aplica la teoría de Alkin-Foster los valores de privación de cada indicador son variables dicotómicas 0-1 que no necesitan ser calibrados para diferentes niveles de privación en una sola variable por lo que se le asigna o no, el peso relativo. Así mismo, un hogar es definido como multidimensionalmente pobre extremo si tiene privaciones en al menos la mitad de los indicadores ponderados y pobre multidimensional si las privaciones alcanzan una tercera parte o más de los indicadores ponderados.

3.2 ESTUDIO DE LAS VARIABLES QUE INCIDEN EN LA POBREZA A NIVEL DESCRIPTIVO

En esta sección se procederá a describir a aquellas variables que sirven para dar posibles explicaciones acerca del fenómeno de la pobreza desde un marco general y del jefe del hogar, la información base para realizar este análisis será la Encuesta Nacional de Empleo, Desempleo y Subempleo publicada para el mes de diciembre de 2016, de esta base de datos se procederá a describir las variables que podrían incidir para que un hogar se considere como pobre o no pobre tomando como primer indicador de pobreza el determinado a partir de la línea de pobreza por ingresos.

Se ha considerado que el estudio debe centrarse de forma general en los hogares debido a que la propensión a que un individuo sea juzgado como pobre o no pobre depende en gran medida de las condiciones de su núcleo familiar (ingresos por hogar, tipo de

vivienda, acceso a servicios, etc.). Entre las variables que se incluirán en este primer análisis se encuentran aquellas de carácter sociodemográfico como el género, la edad, el estado civil, el nivel de educación, medidas que en algunos casos son del tipo nominal - dicotómica como el sexo del jefe de hogar (Hombre-Mujer) y en otros casos nominal - politómica como el estado civil (Soltero-Casado-Conviviente, etc.); otras variables dentro de este análisis serán de carácter cuantitativo como lo son por ejemplo, el ingreso per cápita o la edad del jefe del hogar, lo que debe quedar claro en esta parte es que se está trabajando con un conjunto de variables que se miden en diferentes escalas y que dependiendo de estas escalas se procederá posteriormente a la aplicación de alguno de los modelos de medición de fenómenos sociales que ya se explicaron anteriormente.

3.2.1 Educación

Una de las principales características de la pobreza, que ha sido discutida por varios sociólogos e investigadores, es que esta va ligada directamente al nivel de educación; se afirma que individuos con niveles de educación más altos tienen una menor probabilidad de ser catalogados como pobres, versus aquellos con niveles de educación más bajos que tendrán una mayor probabilidad de ser clasificados como pobres. La ENEMDU clasifica el nivel de instrucción dentro de 9 categorías, (1) Ninguno, (2) Centro de alfabetización, (3) Jardín de infantes, (4) Primaria, (5) Educación Básica, (6) Secundaria, (7) Educación media, (8) Superior no universitario, (9) Superior Universitario y (10) Post-grado. Para efectos de análisis se ha recategorizado esta clasificación en 3 niveles de instrucción: (1) Bajo, que agrupa las categorías de la 1 a la 4, (2) Medio, que agrupa las categorías 5, 6 y 7, y finalmente (3) Alto, que agrupa las categorías 8, 9 y 10. En las tablas siguientes se muestra un resumen donde se puede relacionar, la condición de pobreza con el nivel de instrucción, esto primero a nivel nacional y posteriormente para cada una de las tres ciudades objeto del estudio que se propone en este trabajo.

Tabla 5. Perfil del jefe de hogar según estado civil y nivel de instrucción. Quito, Guayaquil y Cuenca

		Nivel de instrucción			
		Bajo	Medio	Alto	Total
Pobreza por Ingresos	NO POBRE	72.6%	84.7%	96.1%	81.1%
	POBRE	27.4%	15.3%	3.9%	18.9%
Total		100.0%	100.0%	100.0%	100.0%

Fuente: ENEMDU - Diciembre 2016

A nivel de las tres principales ciudades del país se puede observar que dentro de aquellos jefes de hogar con nivel educativo bajo y medio, el 27.4% y el 15.3% se encuentran en condiciones de pobreza respectivamente. Para el caso de los jefes de hogar con nivel de instrucción alto se puede observar que el 96.1% se encuentra en condiciones de pobreza en contraste con el 3.9% que no lo están.

En función del comportamiento entre la condición de pobreza y el nivel de instrucción del jefe de hogar se procede a analizar el grado de relación que existe entre estas variables. Analizando la Tabla 6 para el test Chi-Cuadrado se puede observar que la hipótesis de nulidad que supone la independencia entre el nivel de instrucción y la condición de pobreza del jefe de hogar es rechazada para un valor de $\chi^2 = 2983{,}725$ y un nivel de significancia ($\text{p} - \text{valor} = 0.000$), por lo que a medida que el jefe de hogar aumenta su nivel de instrucción, éste obtiene más capacidades para ir abandonando la condición de pobreza.

Tabla 6. Test Chi-cuadrado para la relación nivel de instrucción – condición de pobreza. Quito, Guayaquil y Cuenca.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	221005,443[a]	2	0.000
Likelihood Ratio	256626.866	2	0.000
Linear-by-Linear Association	220955.454	1	0.000
N of Valid Cases	4383268		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 144699,08.
Fuente: ENEMDU - Diciembre 2016

De la misma forma se procederá a mostrar los resultados obtenidos para las ciudades de Quito, Guayaquil y Cuenca.

Tabla 7. Perfil del jefe de hogar según estado civil y nivel de instrucción. Quito

		Nivel de instrucción			
		Bajo	Medio	Alto	Total
Pobreza por Ingresos	NO POBRE	85.9%	90.3%	98.1%	91.6%
	POBRE	14.1%	9.7%	1.9%	8.4%
Total		100.0%	100.0%	100.0%	100.0%

Fuente ENEMDU – Diciembre 2016

Para la ciudad de Quito se puede observar que dentro de aquellos jefes de hogar con nivel de instrucción bajo y medio la condición de pobreza se encuentra presente en el 14.1% y

el 9.7% de los individuos respectivamente, paralelamente, para los jefes de hogar con instrucción alta la condición de pobreza está presente en únicamente el 1.9% de los individuos.

Tabla 8. Test Chi-cuadrado para la relación nivel de instrucción – condición de pobreza. Quito

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	16427,318[a]	2	0.000
Likelihood Ratio	19545.115	2	0.000
Linear-by-Linear Association	15903.017	1	0.000
N of Valid Cases	541095		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 12433,96.
Fuente ENEMDU – Diciembre 2016

Para el análisis del comportamiento entre la condición de pobreza y el nivel de instrucción del jefe de hogar para la ciudad de Quito se procede de forma similar a lo realizado a nivel nacional. Analizando la tabla del test Chi-Cuadrado se puede observar que la hipótesis de nulidad que supone la independencia entre el nivel de instrucción y la condición de pobreza del jefe de hogar se rechaza para un valor de $\chi^2 = 17311{,}188$ y un nivel de significancia ($\mathrm{p} - \mathrm{valor} = 0.000$), por lo que a medida que el jefe de hogar aumenta su nivel de instrucción, éste irá abandonando la condición de pobreza.

Tabla 9. Perfil del jefe de hogar según estado civil y nivel de instrucción. Guayaquil

		Nivel de instrucción			
		Bajo	Medio	Alto	Total
Pobreza por Ingresos	NO POBRE	85.5%	90.7%	98.2%	90.4%
	POBRE	14.5%	9.3%	1.8%	9.6%
Total		100.0%	100.0%	100.0%	100.0%

Fuente ENEMDU – Diciembre 2016

Para la ciudad de Guayaquil, los jefes de hogar con nivel instrucción bajo y medio, el 14.5% y el 9.3% se encuentran en condiciones de pobreza respectivamente, mientras que el 1.8% de los jefes de hogar con instrucción alta se encuentran en condiciones de pobreza.

Tabla 10. Test Chi-cuadrado relación nivel de instrucción – condición de pobreza. Nacional

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	14829,501[a]	2	0.000
Likelihood Ratio	18037.760	2	0.000
Linear-by-Linear Association	14602.341	1	0.000
N of Valid Cases	668500		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 12016,88.
Fuente: ENEMDU – Diciembre 2016

La relación entre la condición de pobreza y el nivel de instrucción para el jefe de hogar en la ciudad de Guayaquil se analiza de igual manera a través de la tabla del test Chi-Cuadrado, se puede observar que la independencia entre el nivel de instrucción y la condición de pobreza del jefe de hogar es rechazada para un valor de $\chi^2 = 15406{,}668$ y un nivel de significancia ($p - valor = 0.000$), por lo que a medida que el jefe de hogar aumenta su nivel de instrucción, éste obtiene más capacidades para ir abandonando la condición de pobreza.

Tabla 11. Perfil del jefe de hogar según estado civil y nivel de instrucción. Cuenca

		Nivel de instrucción			
		Bajo	Medio	Alto	Total
Pobreza por Ingresos	NO POBRE	88.4%	96.5%	97.4%	94.1%
	POBRE	11.6%	3.5%	2.6%	5.9%
Total		100.0%	100.0%	100.0%	100.0%

Fuente ENEMDU – Diciembre 2016

En la ciudad de Cuenca de los jefes de hogar con nivel instrucción bajo y medio, el 11.6% y el 3.5% se encuentran en condiciones de pobreza respectivamente, mientras que de los jefes de hogar con nivel de instrucción alto el 2.6% se encuentra en condición de pobreza.

Tabla 12. Test Chi-cuadrado para la relación nivel de instrucción – condición de pobreza. Cuenca

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3032,455[a]	2	0.000
Likelihood Ratio	2828.326	2	0.000
Linear-by-Linear Association	2459.850	1	0.000
N of Valid Cases	103679		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 1895,63.
Fuente: ENEMDU - 2016

En cuanto a la relación entre la condición de pobreza y el nivel de instrucción para el jefe de hogar en la ciudad de Cuenca, la Tabla 12 muestra el test Chi-cuadrado que en este caso permite rechazar la hipótesis de independencia entre el nivel de instrucción y la condición de pobreza para un valor de $\chi^2 = 2983{,}725$ y un nivel de significancia ($p - valor = 0.000$), por lo que a medida que el jefe de hogar aumenta su instrucción, su capacidades para ir abandonando la condición de pobreza aumenta.

3.2.2 Estado Civil

A continuación, se analizará la relación que puede existir entre estado civil del jefe del hogar y la condición de pobreza, tanto a nivel de las tres principales ciudades del país en su conjunto como a nivel individual.

Tabla 13. Perfil del jefe de hogar según estado civil y condición pobreza, Quito, Guayaquil y Cuenca.

		6. Estado civil						
		Casado(a)	Separado(a)	Divorciado(a)	Viudo(a)	Unión libre	Soltero(a)	Total
Pobreza por Ingresos	NO POBRE	81.8%	82.5%	88.4%	81.3%	76.6%	85.5%	81.1%
	POBRE	18.2%	17.5%	11.6%	18.7%	23.4%	14.5%	18.9%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Fuente ENEMDU – Diciembre 2016

Se puede observar que dentro del estado civil casado(a) los jefes de hogar en condiciones de pobreza representan el 18.2%, de igual forma, de los individuos en la categoría separados(a) y divorciado(a) el 15.5% y el 11.6% se encuentran en condiciones de pobreza respectivamente, en lo que respecta a viudo(a) y unión libre el 18.7% y el 23.4% se encuentran en condiciones de pobreza correspondientemente, finalmente de los individuos en la categoría soltero(a) el 14.5% se encuentra en condiciones de pobreza.

Tabla 14. Test Chi-cuadrado para la relación estado civil – condición de pobreza, Quito, Guayaquil y Cuenca.

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	27017,069[a]	5	0.000
Likelihood Ratio	27380.453	5	0.000
Linear-by-Linear Association	2728.595	1	0.000
N of Valid Cases	4383269		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 34565,10.
Fuente: ENEMDU – Diciembre 2016

En cuanto a la relación que existe entre la condición de pobreza y el estado civil del jefe del hogar se puede ver en la Tabla 14 que según el test Chi-cuadrado se permite rechazar la hipótesis que formula la independencia de ambas variables con $\chi^2 = 27017{,}069$ y un nivel de significancia ($p - valor = 0.000$), es decir, que existe relación significativa entre el estado civil del jefe de hogar y la condición de pobreza a nivel de las tres principales ciudades del país en su conjunto.

A continuación, se procederá realizar un análisis similar pero independiente para las ciudades que se analizan en este trabajo.

Tabla 15. Perfil del jefe de hogar según estado civil y condición pobreza, Quito.

		6. Estado civil						
		Casado(a)	Separado(a)	Divorciado(a)	Viudo(a)	Unión libre	Soltero(a)	Total
Pobreza por Ingresos	NO POBRE	93.0%	86.3%	91.6%	92.1%	90.8%	90.1%	91.6%
	POBRE	7.0%	13.7%	8.4%	7.9%	9.2%	9.9%	8.4%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Fuente ENEMDU – Diciembre 2016

En la Tabla 15 se puede observar que para la ciudad de Quito dentro del estado civil casado(a) y separado(a) el 7% y el 13.7% se encuentra en condición de pobreza respectivamente, para las otras categorías, divorciado(a), viudo(a), unión libre, y soltero(a) se encuentra en condición de pobreza el 8.4%, el 7.9%, el 9.2% y el 9.9% para cada categoría y en ese orden respectivamente.

Tabla 16. Test Chi-cuadrado para la relación estado civil – condición de pobreza. Quito

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2525,491[a]	5	0.000
Likelihood Ratio	2324.900	5	0.000
Linear-by-Linear Association	571.099	1	.000
N of Valid Cases	541095		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 3027,67.
Fuente: ENEMDU – Diciembre 2016

El test Chi-cuadrado para la relación estado civil – condición de pobreza para la ciudad Quito permite rechazar la hipótesis que formula la independencia de ambas variables con $\chi^2 = 2525{,}491$ y un nivel de significancia ($p - valor = 0.00$), es decir, que existe relación significativa entre el estado civil del jefe de hogar y la condición de pobreza en la ciudad de Quito.

Para la ciudad de Guayaquil la Tabla 17 muestra el perfil del jefe de hogar según estado civil y condición pobreza. En la categoría casado(a) el 7.9% se encuentra en condición de pobreza, para la categoría separado(a) el 11.5% se califica como pobre, en las categorías divorciado(a) y viudo(a) el 1.0% y el 9.2% se encuentra en condición de pobreza respectivamente, y para las condiciones unión libre y soltero(a) el 13.3% y el 1% se encuentra en condición de pobreza respectivamente.

Tabla 17. Perfil del jefe de hogar según estado civil y condición pobreza, Guayaquil.

		6. Estado civil						
		Casado(a)	Separado(a)	Divorciado(a)	Viudo(a)	Unión libre	Soltero(a)	Total
Pobreza por Ingresos	NO POBRE	92.1%	88.5%	99.0%	90.8%	86.7%	99.0%	90.4%
	POBRE	7.9%	11.5%	1.0%	9.2%	13.3%	1.0%	9.6%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Fuente ENEMDU – Diciembre 2016

En cuanto al test Chi-cuadrado para la relación estado civil - condición de pobreza que se muestra a continuación se pude decir que se rechaza la hipótesis de independencia con $\chi^2 = 10013{,}855$ y un nivel de significancia ($\text{p} - \text{valor} = 0.000$) lo que significa que existe relación entre la condición de pobreza y el estado civil del jefe de hogar para la ciudad de Guayaquil.

Tabla 18. Test Chi-cuadrado para la relación estado civil - condición de pobreza. Guayaquil

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	10013,855[a]	5	0.000
Likelihood Ratio	13013.291	5	0.000
Linear-by-Linear Association	233.328	1	.000
N of Valid Cases	668500		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 1777,99.
Fuente: ENEMDU – Diciembre 2016

Para la ciudad de Cuenca el perfil del jefe de hogar según estado civil y condición pobreza, y el test Chi-cuadrado para la relación entre estas variables se muestran en las Tablas 19 y 20 respectivamente.

Tabla 19. Perfil del jefe de hogar según estado civil y condición pobreza, Guayaquil.

		6. Estado civil						
		Casado(a)	Separado(a)	Divorciado(a)	Viudo(a)	Unión libre	Soltero(a)	Total
Pobreza por Ingresos	NO POBRE	92.0%	92.1%	97.6%	96.9%	97.1%	99.5%	94.1%
	POBRE	8.0%	7.9%	2.4%	3.1%	2.9%	.5%	5.9%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Fuente ENEMDU – Diciembre 2016

Tabla 20. Test Chi-cuadrado para la relación estado civil - condición de pobreza. Cuenca

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1534,374[a]	5	0.000
Likelihood Ratio	1943.944	5	0.000
Linear-by-Linear Association	1416.309	1	0.000
N of Valid Cases	103679		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 280,42.
Fuente: ENEMDU – Diciembre 2016

Dentro de las categorías casado(a), separado(a), divorciado(a), viudo(a), unión libre y soltero(a) el porcentaje de individuos en condición de pobreza para la ciudad de Cuenca es del 8.0%, el 7.9%, el 2.4%, el 3.1%, el 2.9% y el 0.5% respectivamente para cada categoría.

Según los resultados del Test-Chi-cuadrado, $\chi^2 = 1534{,}374$ y nivel de significancia ($p - valor = 0.000$) se pude negar la hipótesis de independencia entre las variables lo que determina que existe relación entre la ciudad del jefe del hogar y la condición de pobreza para la ciudad de Cuenca.

3.2.3 Ciudad de residencia

Si bien el estudio realizado en este trabajo pretende diferenciar los factores que determinan la pobreza en las tres principales ciudades del Ecuador es importante echar una mirada a cuál es la proporción de jefes de hogar en condición de pobreza en las ciudades de Quito, Guayaquil y Cuenca.

Tabla 21. Perfil del jefe de hogar según ciudad y condición pobreza.

		Ciudades			
		Cuenca	Guayaquil	Quito	Total
Pobreza por Ingresos	NO POBRE	94.1%	90.4%	91.6%	91.2%
	POBRE	5.9%	9.6%	8.4%	8.8%
Total		100.0%	100.0%	100.0%	100.0%

Fuente: ENEMDU – Diciembre 2016

La Tabla 21 muestra que el porcentaje de jefes de hogar en condición de pobreza en las ciudades de Quito, Guayaquil y Cuenca es del 8.4%, el 9.6% y el 5.9% respectivamente.

Tabla 22. Test Chi-cuadrado para la relación estado civil - condición de pobreza. Cuenca

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1676,423[a]	2	0.000
Likelihood Ratio	1796.379	2	0.000
Linear-by-Linear Association	219.387	1	.000
N of Valid Cases	1313274		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 9110,87.
Fuente: ENEMDU – Diciembre 2016

Según el test Chi-cuadrado se pude rechazar la condición de independencia entre las variables iniciadas con valores $\chi^2 = 1676{,}423$ y nivel de significancia $(\mathrm{p} - \mathrm{valor} = 0)$, por lo tanto se pude decir que existe una relación entre la ciudad donde reside el jefe de hogar y la condición de pobreza.

3.2.4 Número de miembros del hogar

Varios investigadores que abordan el fenómeno de la pobreza desde el punto de vista social, como por ejemplo Musgrove (1977), sostienen que la condición de la pobreza va ligada al número de miembros que comparten o dependen de un hogar; en las siguientes tablas se realiza un análisis de la condición de pobreza según el número de miembros que componen un hogar agrupados en categorías de, (1) 1 a 2 miembros, (2) de 3 a 5 miembros y (3) de 5 miembros en adelante.

Tabla 23. Perfil del jefe de hogar según estado el número de miembros del hogar, Quito, Guayaquil y Cuenca.

		Miembros del hogar			
		de 1 a 2 miembros	de 3 a 5 miembros	más de 5 miembros	Total
Pobreza por Ingresos	NO POBRE	88.5%	81.8%	65.4%	81.1%
	POBRE	11.5%	18.2%	34.6%	18.9%
Total		100.0%	100.0%	100.0%	100.0%

Fuente: ENEMDU – Diciembre 2016

Como se puede observar en la Tabla 23 de los jefes de hogar con hogares de 1 a 2 miembros el 11.5% se considera en situación de pobreza, de los jefes con hogares de entre 3 y 5 miembros el 18.2% se clasifica en condición de pobreza y de los jefes de hogar con hogares de más de 5 miembros el 34.6% es clasificado en condición de pobreza.

Tabla 24. Test Chi-cuadrado para la relación número de miembros del hogar - condición de pobreza. Quito, Guayaquil y Cuenca.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	150320,296[a]	2	0.000
Likelihood Ratio	140057.675	2	0.000
Linear-by-Linear Association	135112.419	1	0.000
N of Valid Cases	4383269		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 127147,00.
Fuente: ENEMDU – Diciembre 2016

En la Tabla 24 del test Chi-cuadrado se puede rechazar la hipótesis de independencia para un valor $\chi^2 = 150320{,}296$ y nivel de significancia $(\mathrm{p} - \mathrm{valor} = 0)$, por lo que se concluye que existe relación entre el número de miembros del hogar y la condición de pobreza para las tres ciudades en su conjunto.

Tabla 25. Perfil del jefe de hogar según estado el número de miembros del hogar, Quito.

		Miembros del hogar			
		de 1 a 2 miembros	de 3 a 5 miembros	más de 5 miembros	Total
Pobreza por Ingresos	NO POBRE	93.4%	92.2%	80.7%	91.6%
	POBRE	6.6%	7.8%	19.3%	8.4%
Total		100.0%	100.0%	100.0%	100.0%

Fuente ENEMDU – Diciembre 2016

Tabla 26. Test Chi-cuadrado para la relación número de miembros del hogar - condición de pobreza. Quito

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	7959,495[a]	2	0.000
Likelihood Ratio	6272.718	2	0.000
Linear-by-Linear Association	4479.199	1	0.000
N of Valid Cases	541096		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 3840,64.
Fuente: ENEMDU – Diciembre 2016

Las tablas 25 y 26 muestra un análisis similar al anterior para la ciudad de Quito, para las categorías, 1 a 2 miembros, de 3 a 5 miembro y más 5 miembros, la condición de pobreza se encuentra presente en el 6.6%, el 7.8% y el 19.3% respectivamente. La condición de independencia se puede negar según los valores de $\chi^2 = 7959{,}495$ y nivel de significancia $(\mathrm{p} - \mathrm{valor} = 0)$ lo que indica que existe relación entre el número de miembros del hogar y la condición de pobreza para la ciudad de Quito.

En la tabla 27 y 28 se muestra análisis similar para la ciudad de Guayaquil.

Tabla 27. Perfil del jefe de hogar según estado el número de miembros del hogar, Guayaquil.

		Miembros de hogar			
		de 1 a 2 miembros	de 3 a 5 miembros	más de 5 miembros	Total
Pobreza por Ingresos	NO POBRE	96.6%	91.0%	78.4%	90.4%
	POBRE	3.4%	9.0%	21.6%	9.6%
Total		100.0%	100.0%	100.0%	100.0%

Fuente: ENEMDU – Diciembre 2016

Tabla 28. Test Chi-cuadrado para la relación número de miembros del hogar - condición de pobreza. Guayaquil.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	25643,415[a]	2	0.000
Likelihood Ratio	23683.624	2	0.000
Linear-by-Linear Association	23419.072	1	0.000
N of Valid Cases	668501		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 10242,38.
Fuente: ENEMDU – Diciembre 2016

Para este caso la condición de pobreza está presente en el 3.4% de hogares de 1 a 2 miembros, en el 9.0% de hogares de 3 a 5 miembros y en el 21.6% de hogares con más

de 5 miembros. Según el test Chi-cuadrado, donde se observan valores de $\chi^2 = 25643{,}415$ y nivel de significancia $(\text{p} - \text{valor} = 0)$, se puede concluir que existe relación entre el número de miembros del hogar y la condición de pobreza para la ciudad de Guayaquil.

Finalmente, para la ciudad de Cuenca, se puede observar en la Tabla 29 que, dentro de los hogares de 1 a 2 miembros, el 1.2% presenta la condición de pobreza, para hogares de 3 a 5 miembros el 6.2% presenta de igual manera condiciones de pobreza y finalmente en hogares con más de cinco miembros la pobreza se encuentra presente en el 14.1%.

Tabla 29. Perfil del jefe de hogar según estado el número de miembros del hogar, Cuenca.

		Miembros del hogar			
		de 0 a 2 miembros	de 3 a 5 miembros	más de 5 miembros	Total
Pobreza por Ingresos	NO POBRE	98.8%	93.8%	85.9%	94.1%
	POBRE	1.2%	6.2%	14.1%	5.9%
Total		100.0%	100.0%	100.0%	100.0%

Fuente: ENEMDU – Diciembre 2016

Tabla 30. Test Chi-cuadrado para la relación número de miembros del hogar - condición de pobreza. Cuenca.

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2546,668[a]	2	0.000
Likelihood Ratio	2632.192	2	0.000
Linear-by-Linear Association	2465.534	1	0.000
N of Valid Cases	103677		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 717,75.
Fuente: ENEMDU – Diciembre - 2016

La Tabla 30 permite ver que existe una relación entre la condición de pobreza y el número de miembros del hogar en la ciudad de Cuenca, los valores del test Chi-cuadrado que llevan a esta conclusión son, $\chi^2 = 2546{,}668$ y nivel de significancia $(\text{p} - \text{valor} = 0)$.

3.2.5 Género

Dado la importancia que ha tomado hoy en día las ideologías de equidad de género es importante realizar una comparación acerca de este factor y la condición de pobreza que puede afectar a los jefes de hogar.

Tabla 31. Perfil del jefe de hogar según sexo y ciudad de residencia.

		Ciudad			Total
		Cuenca	Guayaquil	Quito	
SEXO	Hombre	72.4%	65.0%	72.7%	68.8%
	Mujer	27.6%	35.0%	27.3%	31.2%
Total		100.0%	100.0%	100.0%	100.0%

Fuente: ENEMDU – Diciembre 2016

Como se puede observar en la Tabla 31 existe una marcada tendencia a que los jefes de hogar sean hombres, siendo así que en la ciudad de Cuenca el 72.4% de los jefes de hogar corresponden a esta categoría, en las ciudades de Guayaquil y Quito, el 65.0% y el 72.2% de los jefes de hogar son hombres respectivamente.

Tabla 32. Perfil del jefe de hogar según sexo a nivel de Quito, Guayaquil y Cuenca.

		Sexo		
		Hombre	Mujer	Total
Pobreza por Ingresos	NO POBRE	92.0%	89.4%	91.2%
	POBRE	8.0%	10.6%	8.8%
Total		100.0%	100.0%	100.0%

Fuente: ENEMDU – Diciembre 2016

Como se observa en la Tabla 32 entre el total de jefes de hogar hombres el 8.0% se encuentra en condición de pobreza, de igual manera ente las mujeres jefas de hogar el 10.6% se encuentra en condición de pobreza.

Tabla 33. Test Chi-cuadrado para la relación sexo - condición de pobreza. Quito, Guayaquil y Cuenca.

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2364,187[a]	1	0.000		
Continuity Correction[b]	2363.864	1	0.000		
Likelihood Ratio	2295.973	1	0.000		
Fisher's Exact Test				0.000	0.000
Linear-by-Linear Association	2364.185	1	0.000		
N of Valid Cases	1313273				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 36182,40.

b. Computed only for a 2x2 table

Fuente: ENEMDU – Diciembre 2016

Del test Chi-cuadrado se puede decir que existe dependencia entre la condición de la pobreza y el sexo del jefe del hogar para las ciudades de Quito, Guayaquil y Cuenca en su conjunto, los valores $\chi^2 = 2364{,}187$ y nivel de significancia $(\text{p} - \text{valor} = 0)$ respaldan esta conclusión.

Las tres ciudades analizadas de forma independiente presentan características similares y ratifican la condición de no independencia entre las variables. Los resultados que ratifican estas conclusiones se muestran en las tablas siguientes, tablas 34 y 35 para Quito, tablas 36 y 37 para Guayaquil y tablas 38 y 39 para Cuenca.

Tabla 34. Perfil del jefe de hogar según sexo para la ciudad de Quito

		Sexo		
		Hombre	Mujer	Total
Pobreza por Ingresos	NO POBRE	92.9%	88.1%	91.6%
	POBRE	7.1%	11.9%	8.4%
Total		100.0%	100.0%	100.0%

Fuente: ENEMDU – Diciembre 2016

Tabla 35. Test Chi-cuadrado para la relación sexo - condición de pobreza, Quito.

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3215,433[a]	1	0.000		
Continuity Correction[b]	3214.812	1	0.000		
Likelihood Ratio	3019.552	1	0.000		
Fisher's Exact Test				0.000	0.000
Linear-by-Linear Association	3215.427	1	0.000		
N of Valid Cases	541096				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 12547,58.

b. Computed only for a 2x2 table

Fuente: ENEMDU – Diciembre 2016

Tabla 36. Perfil del jefe de hogar según sexo para la ciudad de Guayaquil

		Sexo		
		Hombre	Mujer	Total
Pobreza por Ingresos	NO POBRE	91.0%	89.4%	90.4%
	POBRE	9.0%	10.6%	9.6%
Total		100.0%	100.0%	100.0%

Fuente: ENEMDU – Diciembre 2016

Tabla 37. Test Chi-cuadrado para la relación sexo - condición de pobreza, Guayaquil.

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	424,481[a]	1	.000		
Continuity Correction[b]	424.301	1	.000		
Likelihood Ratio	418.804	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	424.480	1	.000		
N of Valid Cases	668500				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 22323,02.

b. Computed only for a 2x2 table
Fuente: ENEMDU – Diciembre 2016

Tabla 38. Perfil del jefe de hogar según sexo para la ciudad de Quito

		Sexo		
		Hombre	Mujer	Total
Pobreza por Ingresos	NO POBRE	93.3%	96.2%	94.1%
	POBRE	6.7%	3.8%	5.9%
Total		100.0%	100.0%	100.0%

Fuente: ENEMDU – Diciembre 2016

Tabla 39. Test Chi-cuadrado para la relación sexo - condición de pobreza, Cuenca.

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	302,645[a]	1	.000		
Continuity Correction[b]	302.132	1	.000		
Likelihood Ratio	328.654	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	302.642	1	.000		
N of Valid Cases	103679				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 1686,62.

b. Computed only for a 2x2 table
Fuente: ENEMDU – Diciembre 2016

3.2.6 Otras variables a considerar

Se considera que antes de entrar al desarrollo del modelo que permita determinar la probabilidad de que un jefe de hogar sea considerado como pobre o no pobre también es importante mirar ciertas variables determinantes de la pobreza desde el punto de vista descriptivo-cuantitativo.

En la Tabla 40 se puede observar el número de miembros promedio en un hogar tanto en condición como en condición de no pobreza, se puede ver que para las ciudades de Guayaquil y Cuenca el número promedio de miembros en el hogar en condiciones de no pobreza es de 3.66, mientras que para la ciudad de Quito es de 3.34 miembros.

Para la condición de pobreza el número de miembros promedio por hogar en las tres ciudades es: Cuenca, 4.74 miembros; Guayaquil, 3.94 miembros y Quito 4.13 miembros. Esta situación respalda el análisis realizado anteriormente donde se había confirmado la relación que existe entre el número de miembros de un hogar y la condición de pobreza.

Tabla 40. Número de miembros del hogar en las tres principales ciudades del Ecuador

				Número de miembros del hogar		
				Promedio	Desviación Estándar	Varianza
Pobreza por Ingresos	NO POBRE	Ciudades	Cuenca	3.66	1.70	2.91
			Guayaquil	3.66	1.86	3.47
			Quito	3.34	1.57	2.46
	POBRE	Ciudades	Cuenca	4.74	1.99	3.97
			Guayaquil	4.94	1.88	3.55
			Quito	4.13	2.16	4.65

Fuente: ENEMDU – Diciembre 2016

Otra variable objeto de este análisis es la edad del jefe del hogar, como se puede observar en la Tabla 41 las edades promedio tanto en las tres ciudades como en las dos distintas condiciones son casi similares, así para la ciudad de Cuenca la edad promedio del jefe de hogar pobre es de 49.61 años versus los 49.66 años del no pobre; en la ciudad de Guayaquil la edad promedio del jefe de hogar pobre es de 44.54 años y de 48.64 años para el no pobre. En la ciudad de Quito las edades promedio del jefe de un hogar pobre y no pobre son 47.60 años y 48.10 años respectivamente.

A continuación, se analizará el ingreso per cápita promedio de cada una de las tres principales ciudades del Ecuador, León y Vos indican que el ingreso per cápita de cada hogar se obtiene sumando el ingreso de todos los perceptores del hogar y dividiéndolo para el número de miembros del hogar incluido los que realizan tareas domésticas (León & Vos, 2000)

Tabla 41. Edad del jefe de hogar en las tres principales ciudades del Ecuador

				Edad		
				Promedio	Desviación Estándar	Varianza
Pobreza por Ingresos	NO POBRE	Ciudades	Cuenca	49.66	16.35	267.42
			Guayaquil	48.64	15.07	227.14
			Quito	48.10	15.34	235.43
	POBRE	Ciudades	Cuenca	49.61	14.76	217.86
			Guayaquil	44.54	12.07	145.79
			Quito	47.60	14.54	211.55

Fuente: ENEMDU – Diciembre 2016

Tabla 42. Ingreso per cápita en los hogares de las tres principales ciudades del Ecuador

				Ingreso Per cápita		
				Promedio	Desviación Estándar	Varianza
Pobreza por Ingresos	NO POBRE	Ciudades	Cuenca	388.35	409.18	167430.28
			Guayaquil	299.02	314.67	99016.94
			Quito	421.76	462.10	213536.30
	POBRE	Ciudades	Cuenca	59.83	16.86	284.15
			Guayaquil	60.79	17.98	323.16
			Quito	53.59	21.14	447.03

Fuente: ENEMDU – Diciembre 2016

Como se puede observar en la Tabla 42 el ingreso promedio para los hogares considerados no pobres es de $388.35 en Cuenca, $299.02 en Guayaquil y $421,76 en Quito. Para los hogares considerados como pobres el ingreso per cápita promedio es de $59.83 en la ciudad de Cuenca, $60.79 en Guayaquil y $53,59 en Quito, lo que deja ver que efectivamente existe una diferencia en cómo se determina la pobreza en las tres principales ciudades del Ecuador.

Finalmente, se mostrará una visión general de cuál es la condición de pobreza en las cinco ciudades que se encuentran autorepresentadas en la Encuesta Nacional de Empleo y Subempleo, la tabla siguiente muestra la composición de la pobreza en estas cinco ciudades.

Tabla 43. Composición de la pobreza en las cinco ciudades autorepresentadas en la ENEMDU

		Pobreza por Ingresos			
		NO POBRE		POBRE	
		Count	Row N %	Count	Row N %
Ciudades	Cuenca	97575	94,1%	6104	5,9%
	Machala	64651	90,3%	6917	9,7%
	Guayaquil	604648	90,4%	63852	9,6%
	Quito	495646	91,6%	45449	8,4%
	Ambato	49849	89,9%	5608	10,1%

Fuente: ENEMDU – Diciembre 2016

Algunas variables que serán consideradas en los modelos Logit y Probit no se han mencionado en esta sección ya que el objetivo de la misma era dar una idea de cómo se ciertas variables consideradas básicas por ciertos autores inciden en el fenómeno de la pobreza.

3.3 FORMULACIÓN DE LOS MODELOS LOGIT Y PROBIT EN FUNCIÓN DE LAS VARIABLES EXPLICATIVAS MÁS RELEVANTES

Como se mencionó previamente, las variables ya mencionadas no son las únicas que ayudan a describir el fenómeno de la pobreza, en este apartado se incluirán ciertas variables que según varios autores son determinantes para asegurar condiciones de vida mínimas y dignas las cuales tienen influencia directa en la probabilidad de que un hogar sea considerado como pobre o no pobre, estas variables tienen que ver especialmente con el acceso a vivienda, servicios básicos y la condición de empleo del jefe del hogar.

El análisis estadístico se realizará con la ayuda de los paquetes estadísticos SPSS y EViews, y como ya se hizo anteriormente la información necesaria para el análisis será obtenida de la Encuesta Nacional de Empleo y Subempleo (ENEMDU) de diciembre del año 2016, de esta base las variables independientes escogidas para ser incluidas en el modelo fueron las siguientes:

- Como variable dependiente se encuentra la pobreza [POBREZA] donde 1 indica un jefe de hogar pobre y 0 un jefe de hogar no pobre.
- Alfabetización [ALFABET], toma el valor de 1 si la persona jefa del hogar sabe leer y escribir, caso contrario 0.

- Beneficiario del bono de desarrollo humano [BONO], esta variable toma el valor de 1 si el jefe de hogar el beneficiario del bono desarrollo humano, caso contrario 0.
- Vivienda adecuada [CONDVIV], toma en cuenta si la condición de la vivienda donde habita el jefe de hogar y sus miembros se puede considerar de calidad, esta clasificación se ha tomado en base a una de las recomendaciones de la CEPAL donde se toma en consideración la materialidad del piso, para el caso se ha considerado una vivienda adecuada y de calidad a aquella que tiene el piso de madera tratada, cerámica o mármol, de ser así la variable tomará un valor de 1 caso contrario 0.
- Edad [EDAD], muestra la edad del jefe del hogar.
- Empleo formal [EMPFORM], toma el valor de 1 si el jefe de hogar realiza una actividad laboral formal, caso contrario 0.
- Experiencia laboral [EXLABORAL], es de carácter cuantitativo y mide la experiencia laboral del jefe del hogar en años.
- Formación profesional [FORMPR], toma el valor de 1 si el jefe del hogar tiene formación profesional, caso contrario 0, se entiende por formación profesional la equivalente a educación, superior no universitaria, superior universitaria y post-grado.
- Horas totales de trabajo del jefe de hogar [HTRABAJOT], variable cuantitativa que contiene el número de horas que un jefe de hogar trabaja a la semana, incluye las horas del trabajo principal y de existir las del trabajo secundario y otros trabajos.
- Ingreso total [INGTOTAL], representa el ingreso total del grupo familiar.
- Número de empleos [MASDUT], toma el valor de uno si el jefe del hogar tiene más de un trabajo, caso contrario 0.
- Número de dependientes [NDEPT], número de miembros del hogar que no se encuentran económicamente activos y dependen del ingreso del hogar.
- Pareja [PAREJ], variable que toma el valor de 1 si el jefe de hogar tiene esposa o conviviente, caso contrario 0.
- Servicios básicos [SBASICOS], variable que indica si el hogar tiene acceso a servicios básicos adecuados, toma el valor de 1 si el hogar tiene acceso a: (1) luz eléctrica mediante red pública o planta privada, (2) acceso a agua dentro de la

vivienda y a través de la red pública y (3) servicio higiénico dentro de la vivienda y conectado a la red de alcantarillado; de no poseer alguno de estos servicios la variable toma el valor de 0.

- Tenencia de vivienda [TENENCIAVIV] esta variable toma el valor de 1 si la vivienda del jefe del hogar es propia y se encuentra totalmente pagada o si es propia y se encuentra pagando por la misma, caso contario se toma el valor de 0.
- Trabajo fijo[TRFIJO], si el jefe del hogar está empleado mediante la modalidad de nombramiento o contrato indefinido, la variable toma el valor de 1 de lo contrario 0.
- Vivienda adecuada [VIVADECUADA], toma el valor de 1 si la vivienda donde habita el jefe de hogar se encuentra en la clasificación de casa, villa o departamento, caso contrario 0.

Como es de esperarse la pobreza no puede ser descrita de la misma forma en distintas ciudades del país, en especial cuando existen marcadas diferencias en cuanto a aspectos sociales, económicos y demográficos, por tal motivo no todas las variables mencionadas sirven para describir la pobreza en cada una de las tres ciudades que se analizan en este trabajo, como se verá a continuación ciertas variables estarán presentes en los modelos utilizados para describir la pobreza en estas tres ciudades pero otras estarán únicamente presentes en ciertas ciudades ya que tienen un mayor impacto en esa localidad, incluso algunas de las variables propuestas podrían no aparecer en ninguno de los modelos si no resultan ser relevantes.

3.3.1 Modelo de probabilidad para la pobreza del jefe de hogar basada en el modelo Logit

En esta sección se buscará describir el fenómeno de la pobreza en las ciudades de Quito Guayaquil y Cuenca desde el punto de vista del modelo Logit, para este propósito se hará referencia a los conceptos de desarrollados en la sección 2.6.2 y a la aplicación de la ecuación 2.9 la cual permite presentar el modelo de una forma adecuada.

Para el caso de Quito se realizará una descripción a detalle de los estadísticos y los criterios que se deben tomar en cuenta para evaluar la capacidad predictora del modelo según los lineamientos Logit, para las ciudades de Guayaquil y Cuenca esta descripción

no entrará en mayor detalle ya que será la misma previamente explicada y el análisis se centrará netamente en la evaluación de los estadísticos.

3.3.1.1 Coeficientes para el modelo Logit, caso Quito

Una vez que se ha descrito de manera general las variables que podrían intervenir en el modelo, se ha procedido a realizar una selección de aquella que podrían ser más relevantes para la ciudad de Quito y se ha obtenido la tabla siguiente donde se observan los coeficientes estimados (bajo la columna encabezada por B) y los estadísticos asociados al modelo que predice la probabilidad de ser pobre para el jefe de hogar en esta ciudad.

Tabla 44. Coeficientes del modelo Logit, caso Quito

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	FORMPR	-.7588	.115	43.837	1	.000	.468
	HTRABAJOT	.0876	.004	576.109	1	.000	1.092
	INGTOTAL	-.0638	.001	2558.246	1	0.000	.938
	NMDEP	5.2989	.094	3207.238	1	0.000	200.114
	PAREJ	.9704	.151	41.357	1	.000	2.639
	SBASICOS	-1.4259	.149	92.130	1	.000	.240
	TENENCIAVIV	-.6816	.143	22.701	1	.000	.506
	TRFIJO	-.2376	.112	4.475	1	.034	.789
	VIVADECUADA	-1.5261	.117	170.334	1	.000	.217
	Constant	5.4636	.192	807.368	1	.000	235.943

a. Variable(s) entered on step 1: FORMPR, HTRABAJOT, INGTOTAL, NMDEP, PAREJ, SBASICOS, TENENCIAVIV, TRFIJO, VIVADECUADA.
Fuente: Autores

La ecuación 2.9 en el capítulo anterior mostraba la forma general de un modelo Logit de la manera que se presenta a continuación:

$$\Pr(Y = 1) = \frac{e^{\sum_{k=1}^{K} \beta_k x_k}}{1 + e^{\sum_{k=1}^{K} \beta_k x_k}}$$

Dicha ecuación puede ser reescrita como:

$$\Pr(Y = 1) = \frac{1}{1 + e^{-\sum_{k=1}^{K} \beta_k x_k}} = \frac{1}{1 + e^{-Z_i}} \qquad [3.1]$$

donde $Z_i = \sum_{k=1}^{K} \beta_k x_k$. Dada esta nueva ecuación y los coeficientes calculados en la Tabla 43 se puede escribir la ecuación para la probabilidad de ser pobre para el jefe del hogar, para lo cual tendríamos primeramente que calcular Z_i.

$$\begin{aligned} Z_i = &-0.7588(\text{FORMPR}) + 0.0876(\text{HTRABAJOT}) - 0.0638(\text{INGTOT}) \\ &+ 5.2928(\text{NMDEPT}) + 0.9704(\text{PAREJ}) - 1.4529(\text{SBASICOS}) \\ &- 0.6816(\text{TENENCIAVIV}) - 0.2376(\text{TRFIJO}) \\ &- 1.5261(\text{VIVADECUADA}) + 5.4536 \end{aligned}$$

Como se mencionó en el capítulo anterior, los modelos lineales generalizados no necesariamente cumplen con todas las pruebas de bondad de ajuste (pruebas para los coeficientes estimados) que son necesarias en los modelos de regresión lineal, por lo que las pruebas recomendadas para este tipo de modelos tienen que ver más con el poder predictivo del modelo. A más de lo ya mencionado, cabe recalcar que en este tipo de modelos las pruebas respecto a los errores, como los son la distribución de los errores y la homoscedasticidad, no son condiciones necesarias para validar el modelo por lo que el análisis se centrará en describir que tan bien el modelo pude predecir la condición de pobreza.

Primeramente, el investigador debe cerciorarse si el modelo es capaz de clasificar a la población en dos grupos (para este caso pobres y no pobres). El estimador mínimo Chi-cuadrado utiliza mínimos cuadrados ponderados para estimar un modelo lineal en el cual la variable dependiente es el Logit de estas proporciones, esta comprobación se realiza mediante la prueba Omnibus para los coeficientes del modelo. El criterio de decisión es verificar si la significancia del estadístico (columna Sig.) es menor al nivel de significancia, de ser así se acepta la capacidad del modelo de clasificar a la población. La tabla a continuación muestra los resultados de la prueba.

Tabla 45. Prueba Omnibus para los coeficientes del modelo, caso Quito

		Chi-square	df	Sig.
Step 1	Step	57902.623	9	0.000
	Block	57902.623	9	0.000
	Model	57902.623	9	0.000

Fuente: Autores

Para este caso, el nivel de significancia del estadístico Chi-cuadrado es menor a 0.00 < 0.05 por lo que se puede decir que el modelo es capaz de clasificar a los jefes de hogar de la muestra en dos grupos diferentes, lo que significa que el grupo de jefes de hogar pobres es estadísticamente diferente del grupo no pobre.

Segundo, se debe verificar la capacidad explicativa del modelo (bondad de ajuste), en los modelos de variable dependiente dicotómica este análisis se realiza mediante estadísticos similares al R^2 (coeficiente de determinación) usados en los modelos de regresión lineal y que resultan ser aproximaciones del mismo. El coeficiente que se analizará en este caso será el R^2 de Nagelkerke, si este arroja un valor de 0.5 o superior se puede considerar al modelo como bueno.

Tabla 46. Resumen para bondad de ajuste, caso Quito

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	6952,458[a]	.198	.904

a. Estimation terminated at iteration number 15 because parameter estimates changed by less than ,001.
Fuente: Autores

Para este caso, la $\mathbf{R}^2$ de Nagelkerke arroja un valor de 0.904, por lo que se puede decir que se tiene un modelo bastante aceptable y significativo.

El siguiente paso consiste en comprobar si el modelo configura o no subgrupos en la población, existen dos maneras de realizar esta comprobación, la primera consiste en verificar la clasificación final de los jefes de hogar al terminar el ajuste del modelo de regresión, y la segunda es verificar el estadístico de la prueba de Hosmer y Lemeshow.

Tabla 47. Tabla de clasificación del modelo, caso Quito

Observed			Predicted		
			Pobreza por Ingresos		Percentage Correct
			NO POBRE	POBRE	
Step 1	Pobreza por Ingresos	NO POBRE	254773	409	99.8
		POBRE	724	6320	89.7
	Overall Percentage				99.6

Fuente: Autores

Tabla 48. Prueba de Hosmer y Lemeshow, caso Quito

Step	Chi-square	df	Sig.
1	.134	8	1.000

Fuente: Autores

En la Tabla 47 se puede observar la relación que existe entre los casos realmente observados y el pronóstico del comportamiento del modelo ajustado. Se puede observar que dentro de la categoría "NO POBRE" se han clasificado correctamente el 99.8% de los casos y que en la categoría "POBRE" el 89.7% de los casos fueron bien clasificados, en total un 99.6% ha sido clasificado correctamente den función de las variables analizadas, por lo que se puede decir que el modelo es bastante bueno.

La tabla 48 muestra el estadístico Chi-cuadrado para la prueba de Hosmer y Lemeshow, si el nivel de significancia es superior a 0.05 se puede decir que se está trabajando con un buen modelo, lo cual resulta ser cierto para este caso.

Finalmente, se debe verificar el efecto década variable explicativa en el modelo, en la Tabla 45 se puede observar el coeficiente de Wald que sirva para contrastar las hipótesis de que los coeficientes son significativamente distintos de cero, si la significancia de estos estadísticos es inferior a 0.05 entonces se puede decir que los coeficientes son distintos de cero y aportan información al modelo. Observando el nivel de significancia para cada uno de los coeficientes estimados en la Tabla 44, los cuales son igual a cero, se puede decir que el efecto de cada una de las variables analizadas es significativamente distinto de cero y por tal razón relevantes en la probabilidad de ser pobre para un jefe de hogar.

De la misma forma en la que se procedió para el análisis del modelo de pobreza para la ciudad de Quito, se procederá para el análisis de las ciudades de Guayaquil y Cuenca. En los apartados siguientes no se volverá a hacer hincapié en los tipos de análisis que se realizarán ni sus criterios de decisión ya que son los mismos explicados en esta sección.

3.3.1.2 Coeficientes para el modelo Logit, caso Guayaquil

De manera similar a como se procedió en el caso de Quito se procederá a realizar el análisis de las variables que podrían intervenir en el modelo para la ciudad de Guayaquil, como se mencionó anteriormente las diferencias sociales, demográficas y económicas de

esta ciudad podrán causar que el número y el tipo de variables difieran con respecto al caso anterior, lo cual respaldaría uno de los objetivos de este trabajo.

Tal y como sucedió en la ciudad de Quito se puede notar que fue necesario realizar una reclasificación de las variables que se incluyeron en el caso de Guayaquil con el propósito de explicar el fenómeno de la pobreza más apropiadamente. Las variables para el nuevo modelo y sus estadísticos se presentan en la tabla siguiente.

Tabla 49. Coeficientes del modelo Logit ajustados, caso Guayaquil

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	ALFABET	-4.056	.068	3513.960	1	0.000	.017
	HTRABAJOT	-.023	.001	348.816	1	.000	.978
	INGTOTAL	-.043	.000	10572.948	1	0.000	.958
	MASDUT	2.188	.067	1080.417	1	.000	8.918
	NMDEP	3.498	.034	10409.912	1	0.000	33.035
	PAREJ	3.805	.058	4302.757	1	0.000	44.904
	SBASICOS	-.108	.032	11.619	1	.001	.897
	VIVADECUADA	-.398	.041	94.083	1	.000	.672
	Constant	8.306	.110	5707.673	1	0.000	4047.840

a. Variable(s) entered on step 1: ALFABET, HTRABAJOT, INGTOTAL, MASDUT, NMDEP, PAREJ, SBASICOS, VIVADECUADA.
Fuente: Autores

De la Tabla 49 se puede observar que los coeficientes calculados se ajustan bien al modelo debido a que el nivel de significancia para los estadísticos de Wald son menores a 0.05, por lo cual se puede concluir que los mismos aportan información de valor al modelo. Los estadísticos para las pruebas restantes se muestran en las tablas que siguen.

Tabla 50. Prueba Omnibus para los coeficientes del modelo, caso Guayaquil

		Chi-square	df	Sig.
Step 1	Step	114960.238	8	0.000
	Block	114960.238	8	0.000
	Model	114960.238	8	0.000

Fuente: Autores

Tabla 51. Resumen para bondad de ajuste, caso Guayaquil

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	29189,058[a]	.509	.863

a. Estimation terminated at iteration number 11.
Fuente: Autores

Tabla 52. Tabla de clasificación del modelo, caso Guayaquil

Observed			Predicted		
			Pobreza por Ingresos		Percentage Correct
			NO POBRE	POBRE	
Step 1	Pobreza por Ingresos	NO POBRE	132820	2362	98.3
		POBRE	4417	22030	83.3
	Overall Percentage				95.8

Fuente: Autores

Tabla 53. Prueba de Hosmer y Lemeshow, caso Guayaquil

Step	Chi-square	df	Sig.
1	446.049	8	2.2141E-39

Fuente: Autores

Resumiendo lo obtenido de las tablas 50, 51, 52 y 53 tenemos:

- La Chi-cuadrado de la tabla Omnibus tiene especificación 0.000 < 0.005.
- La R^2 de Negelkerke es bastante elevada, 0.863.
- El modelo clasifica bien un 95.8% de los casos.
- La prueba de Hosmer y Lemeshow tiene significancia 0.00<0.05

Primeramente, se puede ver que una de las pruebas no es significativa, la de Homer y Lemeshow, esto si bien afecta el poder de predicción del modelo, no se pude considerar como un factor determinante para rechazar el mismo ya que el resto de pruebas se ajustan bastante bien a las condiciones que plantea cada una.

Observando que variable fueron incluidas y cuales excluidas lo más destacable hace relación a la ausencia de variables relacionadas con la vivienda, lo que se podría ser una consecuencia de que la gente valora más servicios básicos de calidad que la propiedad de una vivienda.

En cuanto a lo relacionado a la educación tuvo más peso la variable que mide el alfabetismo [ALFABET] en lugar de la que mide la formación profesional [FORMPR], esto puede verse complementado con el hecho de que resulto mejor clasificar el trabajo en formal e informal antes que en fijo y ocasional, además se tuvo que incluir una variable relaciona al número de trabajos que realiza el jefe del hogar [MASDUT].

Según los coeficientes mostrados en la Tabla 49 y el modelo dado por la ecuación 2.9, el elemento Z_i del modelo para medir la probabilidad de pobreza de un hogar en la ciudad de Guayaquil es:

$$Z_i = -4.0561(\text{ALFABET}) - 0.0227(\text{HTRABAJOT}) - 0.0426(\text{INGTOT}) + 2.1881(\text{MASDUT}) + 3.4976(\text{NMDEPT}) + 3.8045(\text{PAREJ}) - 0.1084(\text{SBASICOS}) - 0.3979(\text{VIVADECUADA}) + 8.3059$$

3.3.1.3 Coeficientes para el modelo Logit, caso Cuenca

Para el caso de la ciudad de Cuenca se procederá una vez más de forma similar a como se lo hizo para las ciudades de Quito y Guayaquil. Las variables seleccionadas luego del análisis de los respectivos estadísticos se muestran en la tabla siguiente.

Tabla 54. Coeficientes del modelo Logit ajustados, caso Cuenca

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	FORMPR	-13.4081	.850	248.722	1	.000	.000
	HTRABAJOT	0.5255	.037	203.076	1	.000	1.691
	INGTOTAL	-0.3084	.019	274.235	1	.000	.735
	NMDEP	21.9478	1.295	287.383	1	.000	3.40E+09
	PAREJ	-3.3960	.441	59.324	1	.000	.034
	TENENCIAVIV	-12.5509	.682	338.578	1	.000	3.54E-06
	TRFIJO	-17.4095	1.103	249.206	1	.000	2.75E-08
	Constant	40.0035	2.418	273.762	1	.000	2.36E+17

a. Variable(s) entered on step 1: FORMPR, HTRABAJOT, INGTOTAL, NMDEP, PAREJ, TENENCIAVIV, TRFIJO.
Fuente: Autores

Como sucedió en las dos ciudades anteriores se puede notar que fue necesario realizar una reclasificación de las variables que se incluyeron en el análisis del modelo para así poder describir el fenómeno de la pobreza en la ciudad de Cuenca más adecuadamente.

De la Tabla 54 se puede observar que los coeficientes elegidos se ajustan de buena forma al modelo debido a que el nivel de significancia para los estadísticos de Wald son menores a 0.05, por lo cual se puede concluir que los mismos aportan información de valor al modelo.

Tabla 55. Prueba Omnibus para los coeficientes del modelo, caso Cuenca

		Chi-square	df	Sig.
Step 1	Step	7946.195	7	0.000
	Block	7946.195	7	0.000
	Model	7946.195	7	0.000

Fuente: Autores

Tabla 56. Resumen para bondad de ajuste, caso Cuenca

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1324,167[a]	.169	.870

a. Estimation terminated at iteration number 18 because parameter estimates changed by less than ,001.
Fuente: Autores

Tabla 57. Tabla de clasificación del modelo, caso Cuenca

			Predicted		
			Pobreza por Ingresos		Percentage Correct
Observed			NO POBRE	POBRE	
Step 1	Pobreza por Ingresos	NO POBRE	41780	155	99.6
		POBRE	0	970	100.0
	Overall Percentage				99.6

Fuente: Autores

Tabla 58. Prueba de Hosmer y Lemeshow, caso Cuenca

Step	Chi-square	df	Sig.
1	.000	8	1.000

Fuente: Autores

Acotando a lo ya mencionado se resume a continuación la información obtenida del resto de estadísticos los cuales se pueden observar en las tablas 55, 56, 57 y 58.

- La Chi-cuadrado de la tabla Omnibus tiene especificación 0.000 < 0.005.
- La R^2 de Negelkerke es bastante elevada, 0.870.
- El modelo clasifica bien un 99.6% de los casos.
- La prueba de Hosmer y Lemeshow tiene significancia 1.00>0.05.

De lo expuesto se puede decir que el modelo es bastante significativo pues las pruebas así lo indican. Según los coeficientes mostrados en la Tabla 55 y el modelo dado por la

ecuación 2.9, el elemento Z_i del modelo para medir la probabilidad de pobreza de un hogar en la ciudad de Cuenca está dado por:

$$Z_i = -13.4081(\text{FORMPR}) + 0.5255(\text{HTRABAJOT}) - 0.3084(\text{INGTOT}) + 21.9478(\text{NMDEPT}) - 3.3960(\text{PAREJ}) - 12.5509(\text{TENENCIAVIV}) - 17.4095(\text{TRFIJO}) + 40.0035$$

En el caso de Cuenca, la ciudad con menor pobreza de las analizadas, se debieron excluir variables como [SBASICOS] y [VIVADECUADA] lo que da una idea de que las condiciones de vida en cuanto a infraestructura y acceso a servicios básicos son mucho más altas que en resto del país, incluso llegando al punto de no ser determinantes en la pobreza, lo que se podría interpretar que casi la totalidad de los cuencanos sin importar su condición de "POBRE" o "NO POBRE" tienen acceso a servicios básicos adecuadas y a una vivienda de calidad.

Como se pudo observar cada una de las ciudades estudiadas tiene sus propias características razón por la cual el modelo para medir la probabilidad de que un jefe de hogar sea considerado "POBRE" o "NO POBRE" difiere de una ciudad a otra.

3.3.2 Modelo de probabilidad para la pobreza del jefe de hogar basada en el modelo Probit

Tal y como se procedió para los modelos de probabilidad basados en modelos tipos Logit en esta sección se procederá a desarrollar modelos para describir la pobreza, pero basados en el modelo Probit, para este objetivo se hará referencia a la teoría tratada en la sección 2.6.3 y la función de probabilidad mostrada en la ecuación 2.12.

Las variables a considerar en el análisis serán las mismas ya descritas anteriormente y dada la naturaleza de estos modelos y su complejidad las pruebas de bondad de ajuste no serán tan amplias como antes, además se debe recordar que los resultados obtenido con modelos Probit son muy similares a los obtenidos con modelos Logit según la teoría. El análisis estadístico en este caso se realizará con el paquete EViews y se centrará únicamente en verificar si los coeficientes de las variables escogidas con o significativos, cabe recordar que las pruebas relacionadas a los residuos (homoscedasticidad y normalidad en los residuos) serán omitidas dado que no son condiciones indispensables en este tipo de modelos.

3.3.2.1 Coeficientes para el modelo Probit, caso Quito

Una consideración a tomar en cuenta es que el ingreso total del hogar no podrá ser tomado en cuenta para este análisis debido a que es una variable que de cierta manera está explicando en sí misma la condición de pobreza; otro factor importante y que se debe mencionar es que los datos que se encuentran recopilados en la ENEMDU necesitan ser expandidos para un mejor análisis lo que causa que no se pueda realizar un análisis Probit común en el software descrito sino median un modelo lineal generalizado tipo Probit lo que provoca que se pierda un estadístico importante como lo es el R^2 deMcFadden. Las variables escogidas y sus coeficientes se muestran a continuación.

Tabla 59. Coeficientes del modelo Probit ajustados, caso Quito

Variable	Coefficient	Std. Error	z-Statistic	Prob.
FORMPR	-0.328951	0.016937	-19.42178	0.0000
HTRABAJOT	-0.013588	0.000540	-25.18437	0.0000
NMDEP	0.447551	0.005337	83.85313	0.0000
PAREJ	-0.730782	0.016432	-44.47221	0.0000
SBASICOS	-0.563092	0.019849	-28.36887	0.0000
TENENCIAVIV	-0.032363	0.015485	-2.089975	0.0366
TRFIJO	-1.169901	0.015885	-73.64648	0.0000
VIVADECUADA	-0.512610	0.016587	-30.90363	0.0000
C	-0.715465	0.026292	-27.21210	0.0000

Fuente: Autores

Una primera manera de constatar si los coeficientes son significativamente distintos de cero y aportan información relevante al modelo consiste en observar el p-valor (columna Prob) asociado al estadístico Z de cada coeficiente, si este valor es menor al nivel de significancia (0.05) se puede decir que los coeficientes aportan información al modelo. Como se puede notar de la Tabla 59 todos los coeficientes son relevantes y las variables seleccionadas son prácticamente las mismas del modelo Logit a excepción de la relacionada con el ingreso.

Una segunda manera de comprobar si los coeficientes son representativos, es determinar el intervalo de confianza. El intervalo de confianza proporciona un rango de posibles valores para el parámetro, por lo que si el valor estimado no pertenece a dicho intervalo se puede rechazar la hipótesis de que el parámetro es igual a cero. A continuación, se presentan los intervalos de confianza para los coeficientes y variables seleccionadas.

Tabla 60. Intervalos de confianza para los coeficientes estimados del modelo Probit, Quito

		90% CI		95% CI		99% CI	
Variable	Coefficient	Low	High	Low	High	Low	High
FORMPR	-0.328951	-0.356854	-0.301048	-0.362215	-0.295687	-0.372720	-0.285183
HTRABAJOT	-0.013588	-0.014477	-0.012699	-0.014647	-0.012528	-0.014982	-0.012194
NMDEP	0.447551	0.438759	0.456344	0.437069	0.458034	0.433759	0.461344
PAREJ	-0.730782	-0.757853	-0.703711	-0.763055	-0.698509	-0.773246	-0.688318
SBASICOS	-0.563092	-0.595792	-0.530393	-0.602075	-0.524109	-0.614385	-0.511800
TENENCIAVIV	-0.032363	-0.057874	-0.006853	-0.062775	-0.001951	-0.072379	0.007652
TRFIJO	-1.169901	-1.196071	-1.143731	-1.201099	-1.138702	-1.210951	-1.128850
VIVADECUADA	-0.512610	-0.539937	-0.485284	-0.545187	-0.480033	-0.555474	-0.469746
C	-0.715465	-0.758779	-0.672150	-0.767102	-0.663828	-0.783408	-0.647522

Fuente: Autores

Como se puede observar, todos los coeficientes están dentro de sus respectivos intervalos de confianza por lo cual se ratifica que los coeficientes son representativos.

3.3.2.2 Coeficientes para el modelo Probit, caso Guayaquil

Tal y como se realizó para el caso de Quito, se puede observar a continuación la tabla de coeficientes del modelo Probit para la ciudad de Guayaquil, de igual manera las variables incluidas resultan ser las mismas del modelo Logit y se puede observar que todas resultan significativas ya que el p-valor asociado al estadístico Z de cada coeficiente es menor a 0.05 lo cual significa que las variables pueden aportar información importante al modelo.

Tabla 61. Coeficientes del modelo Probit ajustados, caso Guayaquil

Variable	Coefficient	Std. Error	z-Statistic	Prob.
ALFABET	-0.182905	0.015850	-11.53994	0.0000
HTRABAJOT	-0.022449	0.000271	-82.79492	0.0000
MASDUT	0.417277	0.015756	26.48335	0.0000
NMDEP	0.250936	0.002387	105.1474	0.0000
PAREJ	0.202048	0.009993	20.21813	0.0000
SBASICOS	-0.242894	0.008453	-28.73364	0.0000
VIVADECUADA	-0.154345	0.011270	-13.69564	0.0000
C	-0.437141	0.021413	-20.41489	0.0000

Fuente: Autores

La Tabla 62, indica los intervalos de confianza para los coeficientes calculaos, es así pues que se puede notar que todo estos cumplen la condición de significancia lo cual respalda la afirmación realizada anteriormente respecto a estos.

Tabla 62. Intervalos de confianza para los coeficientes estimados del modelo Probit, Guayaquil

		90% CI		95% CI		99% CI	
Variable	Coefficient	Low	High	Low	High	Low	High
ALFABET	-0.182905	-0.209052	-0.156758	-0.214090	-0.151721	-0.223979	-0.141831
HTRABAJOT	-0.022449	-0.022897	-0.022002	-0.022983	-0.021916	-0.023152	-0.021747
MASDUT	0.417277	0.391284	0.443270	0.386276	0.448277	0.376445	0.458109
NMDEP	0.250936	0.246999	0.254873	0.246241	0.255632	0.244752	0.257121
PAREJ	0.202048	0.185562	0.218534	0.182386	0.221710	0.176150	0.227946
SBASICOS	-0.242894	-0.256839	-0.228949	-0.259526	-0.226262	-0.264800	-0.220987
VIVADECUADA	-0.154345	-0.172936	-0.135753	-0.176518	-0.132172	-0.183550	-0.125140
C	-0.437141	-0.472466	-0.401816	-0.479271	-0.395011	-0.492632	-0.381650

Fuente: Autores

3.3.2.3 Coeficientes para el modelo Probit, caso Cuenca

Como se puede observar en la tabla siguiente las variables y los coeficientes que escogidos para la ciudad de Cuenca nuevamente resultan ser muy similares a los que se plantearon para el modelo Logit, sin embargo, los resultados indican que la variable relaciona a la pareja o conviviente [PAREJ] podría ser retirada del modelo ya que no aporta significancia al modelo debido a que el p-valor asociado a su estadístico es mayor a 0.05.

Tabla 63. Coeficientes del modelo Probit ajustados, caso Cuenca

Variable	Coefficient	Std. Error	z-Statistic	Prob.
FORMPR	-0.950370	0.056742	-16.74908	0.0000
HTRABAJOT	-0.042851	0.002096	-20.44290	0.0000
NMDEP	0.386453	0.013387	28.86728	0.0000
PAREJ	-0.066203	0.045708	-1.448396	0.1475
TENENCIAVIV	-0.302205	0.038353	-7.879488	0.0000
TRFIJO	-1.417347	0.050954	-27.81596	0.0000
C	-0.308620	0.066801	-4.620009	0.0000

Fuente: Autores

En la tabla siguiente se puede observar que los intervalos de confianza respaldas las conclusiones realizadas a cerca de la significancia de los coeficientes del modelo.

Tabla 64. Intervalos de confianza para los coeficientes estimados del modelo Probit, Cuenca

		90% CI		95% CI		99% CI	
Variable	Coefficient	Low	High	Low	High	Low	High
FORMPR	-0.963397	-1.056143	-0.870652	-1.074003	-0.852791	-1.109060	-0.817734
HTRABAJOT	-0.043985	-0.047208	-0.040762	-0.047829	-0.040142	-0.049047	-0.038924
NMDEP	0.383530	0.361845	0.405214	0.357669	0.409390	0.349473	0.417586
TENENCIAVIV	-0.310568	-0.373046	-0.248090	-0.385078	-0.236058	-0.408694	-0.212442
TRFIJO	-1.423379	-1.506879	-1.339879	-1.522959	-1.323799	-1.554521	-1.292236
C	-0.298034	-0.407793	-0.188275	-0.428931	-0.167137	-0.470419	-0.125650

Fuente: Autores

Como se ha podido observar a lo largo de este capítulo no resulta fácil definir las variables descriptivas de la pobreza y más aún cuando cada ciudad bajo análisis es una realidad diferente, el problema radica principalmente en que el proceso de selección de las variables es sumamente subjetivo y está sujeto prácticamente a un proceso de prueba y error.

La tabla que se presenta a continuación resume la realidad de cada ciudad en cuanto a los factores que inciden en la determinación de la pobreza, las variables consideradas son aquellas obtenidas del uso del modelo Logit ya que las obtenidas con el uso del modelo Probit resultan ser muy similares.

Tabla 65. Variables determinantes de la pobreza en las ciudades de Quito, Guayaquil y Cuenca

Variable	**Quito**	**Guayaquil**	**Cuenca**
Alfabetización		X	
Formación profesional	X		X
Horas totales de trabajo del jefe de hogar	X	X	X
Ingreso total	X	X	X
Más de un empleo		X	
Número de dependientes	X	X	X
Pareja	X	X	X
Servicios básicos	X	X	
Tenencia de vivienda	X		X
Trabajo fijo	X		X
Vivienda adecuada	X	X	

Fuente: Autores

Como se puede ver en la Tabla 65, cada ciudad presenta ciertas diferencias, lo que lleva a pensar que cada una tiene necesidades y realidades diferentes, por lo cual intentar

mitigar de la pobreza de la misma manera en estas ciudades no produciría los mismos resultados.

Matemática y estadísticamente se puede ahondar mucho más en el análisis del problema pero para tal efecto es necesario una base de datos más robusta que ayude a realizar un trabajo más sólido; de la complejidad de este estudio se puede entender por qué muchos organismos gubernamentales optan por métodos mucho más sencillos para la medición de la pobreza como son el establecimiento de la línea base de pobreza a partir del ingreso per cápita o mediante el método de las necesidades básicas insatisfechas

CONCLUSIONES Y RECOMENDACIONES

Estudiar la pobreza es un tema complejo, para comprender este fenómeno social es necesario analizar sus conceptos, causas, perspectivas y sus métodos de medición, concluyéndose que la pobreza es de carácter multidimensional y por lo tanto debe ser medida multidimensionalmente. Esta investigación no se ha centrado en estudiar las causas de la pobreza, sino en tratar de establecer las variables que causan privaciones en los hogares no permitiendo se logren alcanzar las condiciones mínimas de vida y convirtiendo a los hogares en pobres.

Las metodologías de medición de la pobreza no son estáticas y evolucionan a través del tiempo-espacio, partiendo desde los métodos indirectos basados en el consumo energético y/o de alimentos, bienes y servicios que las personas usan para realizar sus actividades, hasta los métodos directos basados en las privaciones que pueden sufrir los individuos en su contexto, llegando a los métodos combinados que integran los dos mencionados, es decir, ingresos y capacidades reflejando la multidimensionalidad del fenómeno. .

La metodología Alkin-Foster sustenta su aplicación en la teoría de Amartya Sen quien evoluciona al no relacionar exclusivamente la pobreza con los ingresos, sino considera la "capacidad" para contribuir a una mejor sociedad en donde la pobreza es un grado de privación que impide el desarrollo pleno de las destrezas y derechos amparados en convenios internacionales y constituciones de los estados.

Existe un gran debate sobre las dimensiones que se deben analizar en un estudio de la pobreza, en esta investigación se concluye que deben ser dimensiones que engloben indicadores que representen las necesidades de la población, fruto de debates a nivel macro como los 17 Objetivos del Milenio para el Desarrollo Sostenible y los Derechos del Buen Vivir de la Constitución de la República se debe entender que cada territorio posee un nivel de desagregación y especificidad acorde a su realidad lo que debe llevar a considerar factores que faciliten la toma de decisiones sobre políticas públicas más puntuales, estas afirmaciones se ven respaldadas en el hecho de que no se pudo medir la pobreza de igual manera en las tres ciudades analizadas en este trabajo.

Matemáticamente, el hecho de definir a una persona como "POBRE" o "NO POBRE" conlleva el uso de una variable dependiente binaria o dicotómica que está en función de

variables independientes cualitativas y cuantitativas y son los modelos Logit y Probit los que permiten una investigación cuantitativa a través de un enfoque multivariable.

Para interpretar los resultados de un modelo de manera significativa, el modelo debe ajustarse a los datos para lo cual la prueba más utilizada es el test de la razón de verosimilitudes, que sigue aproximadamente la distribución de Chi-cuadrado (χ^2), una vez validado el modelo las probabilidades estimadas calculadas a partir del modelo Probit son casi idénticas a las del modelo Logit, difiriendo en menos del cinco por mil. Por lo tanto, los resultados llevarán a conclusiones similares.

En el presente estudio, las variables independientes determinadas, tanto cualitativa como cuantitativamente, son el resultado del aporte de los autores de la investigación, en base a la teoría de la medición de la pobreza, marco de derechos del Buen Vivir, diálogo con expertos y la experiencia propia, con propósito de definir una metodología propia para el análisis de la pobreza en Quito, Guayaquil y Cuenca. El primer problema se presenta aquí ya que la elección de las variables es netamente subjetiva y no existe un método estandarizado o matemático que permita seleccionar las variables más apropiadas, pudiendo llevar esto a que distintos investigadores obtengan distintos resultados según el enfoque y los objetivos de su estudio, esto conlleva a que en algunas ocasiones se tenga que recurrir a un proceso de prueba y error hasta encontrar el mejor conjunto de variables para describir el fenómeno.

A nivel país, deben reformularse los indicadores de la dimensión salud pues estos en la ENEMDU no son explícitos y se los está midiendo transversalmente a través de agua, alimentación y ambiente sano, justificándose por la aún no vigencia del Código de la Salud enmarcada dentro de la Constitución de Montecristi, a más de esta falencia la Encuesta Nacional de Empleo y Subempleo presenta muchos vacíos de información lo cual dificulta el análisis de los datos ya que los casos que pueden utilizarse efectivamente para el análisis estadístico se ven drásticamente reducidos.

En cuanto a los resultados obtenidos, se pudo demostrar que alrededor del fenómeno de la pobreza las ciudades de Quito, Guayaquil y Cuenca presentan realidades diferentes lo cual ya se podía deducir en cierto grado por el conocimiento que se tiene de la situación social de cada ciudad y región del país. Anteriormente ya se detallaron ciertos aspectos, pero cabe destacar aspecto como los referentes a la vivienda, los servicios básicos y la

educación en cada una de estas ciudades. Tanto en Quito como en Cuenca se pudo notar que la formación profesional tiene su relevancia en contraste con la ciudad de Guayaquil donde la educación se ve más representada por el nivel de alfabetismo. Tanto en Quito como en Guayaquil el acceso a servicios básicos resulta importante en contraposición con Cuenca donde estos no juegan un gran papel ya que debido a la infraestructura de la ciudad prácticamente todos sus habitantes tienen acceso a viviendas y servicios de calidad, esto se puede corroborar al verificar que en la ciudad de Cuenca no existen suburbios ni zonas marginales.

En cuanto a los ingresos, el trabajo y el número de miembros que dependen del jefe del hogar se pudo notar que estos tienen su nivel de significancia en las tres ciudades estudiadas, lo cual tiene bastante lógica ya que son indicativos directos para poder medir la pobreza o vulnerabilidad económica de un hogar.

La presente investigación, deja la puerta abierta para la incorporación de nuevas variables de estudio en el fenómeno social de la pobreza y la aplicación de modelos matemáticos como el Tobit que resulta ser una variación de los modelos Logit y Probit. Además, la simulación es otro de los caminos que se pude tomar para tratar de entender este fenómeno sin embargo queda demostrado que un solo enfoque no representa necesariamente la realidad y depende mucho de la subjetividad del investigador.

BIBLIOGRAFÍA

Aldrich, J., & Nelson, F. (1984). *Linear Probability, Logit and Probit Models.* Londres: Sage.

Alkire, S., & Foster, J. (2007). Recuento y medición multidimensional de la pobreza. (O. W. Series, Ed.) *Oxford Poverty & Human Development Initiative*.

Atkinson, A., & Bourguignon, F. (2000). *Handbook of Income Distribution.* North Holland.

Bolvinik, J. (2003). Tipología de los métodos de medición de la pobreza. Los métodos combinados. *Revista Comercio Exterior*(53), 453-465.

Bujosa, M. (2007). *Especificación y Estimación del Modelo Lineal General.* Madrid: Creative Commons.

Catell, R. (1966). *Handbook of Multivariate Experimental Psychology*. Chicago: Rand McNally.

Gujarati, D., & Porter, D. (2010). *Econometría* (Quinta ed.). México D.F.: McGraw-Hill.

Instituto nacional de estadísticas y Cencos. (Diciembre de 2016). Encuesta Nacionl de Empleo, Desempleo y Subempleo. Quito, Ecuador.

Instituto Nacional de Estadísticas y Censos. (5 de Agosto de 2013). *Ecuador - Encuesta Nacional de Empleo, Desempleo y Subempleo - Diciembre 2011, RONDA XXXIV-12-2011*. Recuperado el 2 de Octubre de 2017, de http://anda.inec.gob.ec/anda/index.php/catalog/269

León, M., & Vos, R. (2000). *La pobreza urbana en el Ecuador, 1988-1998: mitos y realidades.* Quito: Abya-Yala.

Liao, T. (1994). *Interpreting Probability Models: Logit,*. Londres: Sage.

López Gonzáles, E., & Ruiz Soler, M. (2011). Análisis de datos con el Modelo Lineal Generalizado. Una aplicación con R. *Revista española de pedagogía*(248), 59-80.

López Roldan, P., & Fachelli, S. (2015). *Metodología de la Investigación Social Cuantitativa* (Primera ed., Vol. III). Barcelona: Creative Commons.

López-Aranguren, E. (2005). *Problemas sociales: desigualdad, pobreza, exclusión social.* Madrid, España: Biblioteca Nueva.

Morales, E. (2001). *Introducción a la Econometría* (Vol. I). Quito: Abya-Yala.

Morgan, S., & Teachman, J. (1988). Logistic regression: Description, examples, and comparisons. *Journal of Marriage and the Family*(50), 929-936.

Musgrove, P. (Diciembre de 1977). Tamaño y y composición del hogar, ocupación y pobreza en América Latina. *Estudios de economía, 4*(2), 40-57.

NU. CEPAL. Oficina de Bogotá. (2007). La medida de Necesidades Básicas Insatisfechas (NBI) como instrumento de medición de la pobreza y focalización de programas. *CEPAL - Serie Estudio y Perspectivas*(18).

Phélan, M. (2006). *La Pobreza en Venezuela.* Caracas, Venezuela: undación Escuela de Gerencia Social.

Putucay, F. G. (2002). *Los modelos Logit y Probit en la investigación social. El caso de la pobreza del Perú 2001*. Lima: Talleres de la Oficina Técnica de Administración del INEI.

Rao, P., & Miller, R. L. (1971). *Applied Econometrics* . Wadsworth: Belmont Calif.

Spicker, P., Álvarez, L., & David, G. (2009). *Pobreza: Un Glosario Internacional.* Buenos Aires: Consejo Latinoamericano de Ciencias Sociales.

Universidad Internacional de La Rioja. (2016). *Regresión logística simple.* Logroño.