



Centro de Investigación y
Desarrollo (CIDE)

LOS MODELOS LOGIT Y PROBIT EN LA INVESTIGACIÓN SOCIAL

**El caso de la Pobreza del Perú
en el año 2001**

Lima, Agosto 2002

DIRECCIÓN Y SUPERVISIÓN

Econ. Mirlena Villacorta Olazabal
Directora Técnica del CIDE

Documento Elaborado por:

Franck G. Pucutay Vásquez

Preparado	:	Centro de Investigación y Desarrollo del Instituto Nacional de Estadística e Informática (INEI)
Impreso	:	Talleres de la Oficina Técnica de Administración del INEI
Diagramación	:	Centro de Edición de la Oficina Técnica de Difusión del INEI
Tiraje	:	200 Ejemplares
Domicilio	:	Av. General Garzón 658, Jesús María. Lima - Perú
Orden de Impresión	:	Nº -OTA-INEI
Depósito Legal Nº	:	150113-2002-4014

Presentación

El INEI pone a disposición la investigación metodológica: **"LOS MODELOS LOGIT Y PROBIT EN LA INVESTIGACIÓN SOCIAL El caso de la pobreza del Perú en el año 2001"**, que por su nivel de especialización está dirigida principalmente a los miembros de la comunidad académica, profesionales de las oficinas de estadística y los investigadores interesados en mantener la actualidad de sus procedimientos estadísticos.

Esta investigación metodológica tiene por finalidad, generar instrumentos y procedimientos que permitirán validar, mejorar y actualizar los procesos estadísticos. Se caracteriza por ser innovadora en su campo de aplicación, por contener un rigor científico en su desarrollo integral, por la validez de sus procesos, por la vigencia y actualidad de sus metodologías aplicadas.

En esa misma dirección, la investigación presentada desarrolla la metodología asociada a los modelos de probabilidad con variable dependiente discreta dicotómica (modelo logit y probit), en función del fenómeno de pobreza en los jefes de hogar del Perú para el año 2001. Contribuyendo a la implementación de los enfoques asociados a estos modelos, según la viabilidad permitida por los factores de naturaleza cualitativa, cuantitativa y su aporte en la probabilidad de ser pobre del jefe de hogar.

Este estudio al igual que otros de carácter metodológico, ha sido elaborado por profesionales del Centro de Investigación y Desarrollo (CIDE), en el marco del desarrollo y promoción de investigaciones estadísticas y socioeconómicas que permitan elevar la calidad de la información del INEI y el SEN.

El INEI espera como resultado de esta investigación, incorporarse en el circuito de la producción del conocimiento y elevar los estándares de calidad de sus procesos, sentando con ello las bases de la investigación metodológica en la institución.

Lima, Agosto 2002



Gilberto Moncada Vigo
Jefe del INEI

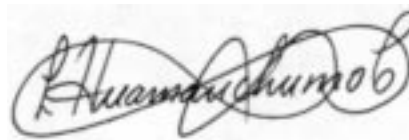
INDICE

Presentación	3
Prólogo	7
I. INTRODUCCIÓN	9
II. FUNDAMENTACION DEL PROBLEMA	11
2.1. Formulación del problema	13
III.OBJETIVOS DE LA INVESTIGACIÓN	15
3.1. Objetivo general	15
3.2. Objetivos específicos	15
IV. MARCO TEÓRICO	17
4.1. Breve formulación del caso de la Pobreza del Perú.	17
4.1.1. ¿Qué es la pobreza?	17
4.1.2. Enfoques y métodos para su medición	17
4.2. Los modelos logit y probit con variable dependiente dicotómica (VDD).	20
4.2.1. Modelos con variable dependiente dicotómica (vdd).	20
4.2.2. Formulación del modelo logit y el modelo probit con (vdd)	22
4.2.3. Caracterización de los modelos logit y probit aplicados al caso de la pobreza del Perú.	24
V. HIPÓTESIS	33
VI. METODOS	35
6.1. Tratamiento de errores en la adecuación de ambos modelos.	35
6.1.1. Métodos de estimación	35
VII. LA ENCUESTA NACIONAL DE HOGARES - 2001	
IV TRIMESTRE	39
7.1. Características de la muestra	41
7.2. Factores de relevancia para la explicación de la pobreza extraídos de la ENAHO IV trimestre 2001.	43
VIII. CRITERIOS DE FORMULACIÓN DE LOS MODELOS LOGIT Y PROBIT CON VDD APLICADOS AL CASO DE LA POBREZA DEL PERÚ.	49
8.1. Estudio a nivel descriptivo y exploratorio de algunas variables cualitativas y cuantitativas que inciden en la pobreza del Perú.	49
8.2. Formulación y adecuación de los modelos Logit y Probit con VDD, en función de las variables o factores explicativos más significativos.	58
IX. CONCLUSIONES	79
X. RECOMENDACIONES	81
XI. BIBLIOGRAFÍA	83
XII. ANEXOS	85
Anexo 1 Informe metodológico	87
Anexo 2. Indices de ecuaciones, cuadros y gráficos	95

Prólogo

Desde las aulas universitarias y siendo conocedor de la difícil tarea que significa la promoción y desarrollo de la labor científica en una realidad caracterizada por la escasez de recursos y las restricciones presupuestarias, saludo este esfuerzo del Centro de Investigación y Desarrollo (CIDE) orientado no sólo a elevar la calidad de la información generado por el Sistema Estadístico Nacional sino también a la producción científica y metodológica en nuestro medio. En este sentido, la presente publicación: "Los Modelos Logit y Probit en la Investigación Social: El Caso de la Pobreza del Perú en el Año 2001" satisface una necesidad no sólo para los consumidores potenciales de investigaciones sociales o eventuales investigadores sino también para el público interesado y universitario familiarizado con los elementos de la estadística.

Aunque este documento es metodológico por naturaleza, está orientado hacia las aplicaciones. A lo largo del estudio, se ha mantenido al mínimo las demostraciones teórico-matemáticas y se ha puesto énfasis en el desarrollo de un entendimiento claro de los resultados teóricos usuales en los estudios sociales con este tipo de modelos, el cual está plasmado en los objetivos planteados en el documento. De otro lado, se hace una breve pero precisa formulación de la pobreza en el Perú cuyo análisis se basa en las variables provenientes de La Encuesta Nacional de Hogares-2001 IV Trimestre (ENAH0). De ahí que, se incluye un capítulo que describe la definición de dichas variables para luego interactuar sistemáticamente en la formulación del problema y análisis con el enfoque de los Modelos Logit y Probit.



Luis Huamanchumo de la Cuba
Escuela Profesional de Ingeniería
Estadística-UNI

I. INTRODUCCION

La investigación metodológica juega un papel fundamental en el desarrollo de los estándares de calidad de las oficinas estadísticas más renombradas del mundo, con el objetivo de generar información relevante para la producción del conocimiento científico y por ende la toma de decisiones. Pero ello implica la utilización y exploración de técnicas vigentes e idóneas para el análisis de su información sin la cual no podrían mejorarse los procesos que implican su producción.

Muchos de los campos en los cuales se desarrollan estas investigaciones abordan diferentes aspectos tanto de la problemática social y económica. Aquí podemos decir que la Pobreza es uno de esos temas tan apasionantes y a la vez muy discutidos sobre el cual se desarrollan una gran diversidad de metodologías y explicaciones, sin ser estas concluyentes y menos aún aceptadas por todos. Desde el punto de vista estadístico, una explicación a este fenómeno es ampliamente beneficiada por la utilización de los modelos de elección discreta dicotómica-modelos logit y probit, para obtener la cuantificación del aporte de sus factores significativos en el análisis de la pobreza.

En esa misma línea se plantea en qué medida una explicación sobre la pobreza de los jefes de hogar del Perú en el año 2001, se vería beneficiada por la adopción de un enfoque de proporciones muestrales

o por un enfoque de observaciones individuales en la utilización de los modelos logit y probit.

Entonces diríamos que los factores de naturaleza cuantitativa como el ingreso per cápita mensual del hogar, expresado a través de sus deciles de ingresos, o los años de estudios, etc; generan un modelo correctamente ajustado a la probabilidad de ser pobre de los jefes de hogar con el enfoque de proporciones muestrales del modelo probit.

O que los factores explicativos de la pobreza del Perú en el año 2001 de naturaleza cualitativa y cuantitativa exclusivos del jefe de hogar como el nivel de educación, el tipo de colegio donde estudió, la categoría ocupacional, el tamaño de la firma donde labora, la tenencia de otro empleo, el estado civil, su edad, su indicador de experiencia laboral; en combinación con los factores de naturaleza cualitativa y cuantitativa exclusivos del hogar como la cantidad de miembros del hogar, la cantidad de miembros pertenecientes a la PET, el ingreso per cápita mensual, el acceso a activos públicos de agua y desagüe, si el hogar dedica un espacio físico de este a generación de ingresos, no permiten generar modelos correctamente ajustados a la probabilidad de ser pobre del jefe de hogar siguiendo el enfoque de observaciones individuales en los modelos logit y probit.

Para tal efecto, lo que se pretende es estudiar la relación entre los factores determinantes de naturaleza social, económica, demográfica de la pobreza del Perú y los enfoques de utilización de los modelos logit y los modelos probit.

Analizar el grado de afectación de los factores de naturaleza cuantitativa sobre el enfoque de proporciones muestrales. Comparar una estimación del modelo logit con una estimación del modelo probit en un enfoque de casos individuales, para cuantificar el aporte de los factores explicativos de la pobreza en los jefes de hogar.

Luego de esta sucinta introducción, el segundo capítulo de esta investigación desarrolla la fundamentación del problema y su sistematización; en el tercero se formulan los objetivos de esta, mientras que en el cuarto capítulo mostramos el marco teórico asociado a la pobreza, cómo esta viene definida y los métodos para su medición. Además en este se muestran la formulación teórica de los modelos logit y probit con variable dependiente discreta-dicotómica, su caracterización, similitudes y diferencias, enfoques de utilización y los efectos marginales asociados a cada uno de los modelos. En el quinto capítulo se muestran las hipótesis formuladas para esta investigación, mientras que en el sexto indicamos los métodos de estimación a partir de los enfoques utilizados.

En el séptimo capítulo se muestra de manera general la Encuesta Nacional de Hogares, sus objetivos y características específicas, así como los factores relevantes para la explicación de la pobreza de los jefes de hogar tomados en la ENAHO-IV trimestre 2001. El octavo capítulo muestra los criterios de aplicación

de los modelos logit y probit aplicados al caso de la pobreza en el Perú, partiendo de un análisis descriptivo-exploratorio de algunas variables significativas como la educación, el estado civil, región natural de residencia, los años de estudios, etc; para luego hacer una formulación y desarrollo metodológico de los modelos en función a sus enfoques y factores más significativos, terminando con una explicación de los resultados obtenidos. Los siguientes capítulos hacen referencia a las conclusiones, recomendaciones y anexos.

No quisiera terminar esta breve introducción sin mostrar mi agradecimiento al Instituto Nacional de Estadística e Informática por permitirme colaborar a través del desarrollo de esta investigación en avanzar más hacia el fortalecimiento de la cultura estadística en el país, y además, sentar las bases de la investigación metodológica en el Perú.

De la misma manera, las gracias infinitas a la señora Directora Técnica del Centro de Investigación y Desarrollo, Mirlena Villacorta, por sus valorables aportes en el desarrollo de esta investigación, su comprensión y constante apoyo, y porque me muestra que cada día es posible construir desde la inteligencia y la creatividad.

Así mismo, quiero expresar mi gratitud y reconocimiento a Luis Huamanchumo, auditor de esta investigación metodológica, por sus importantísimas apreciaciones y por su compromiso con la labor de investigación estadística, desde ya reciba las gracias infinitas. Para terminar las gracias a Dios, a mi Familia y a mis seres queridos.

II. FUNDAMENTACION DEL PROBLEMA

La pobreza es un fenómeno siempre presente, en mayor o menor medida, en todas las sociedades, razón por la cual ha sido objeto de estudio y se han buscado las herramientas de política para enfrentarla. Existen muchas definiciones respecto a lo que debe entenderse por pobreza y ninguna de ellas es precisa ni aceptada por todos, sin embargo, en general todas las definiciones apuntan a la situación en que se encuentran las personas que no disponen de los medios (de "producción", activos fijos, intelectuales, sociales, culturales, financieros y demás que permitan generar fuentes permanentes de ingresos) suficientes para satisfacer sus necesidades básicas definidas como tales para un grupo social específico y en un tiempo determinado, y que permitan su desarrollo personal y reflejen el estilo de vida de la formación social en su conjunto. Entre estas necesidades figuran la alimentación, salud, vivienda, educación básica, acceso a servicios esenciales de información, recreación, cultura, vestido, calzado, transporte y comunicaciones, participación e identidad en y con la comunidad, entre otras.

El Perú es uno de los países más pobres de América del Sur. Una breve mirada a cualquier listado que pretenda ordenar los países en función de su bienestar lo demuestra. Casi la cuarta parte de los peruanos carecen de recursos para alimentarse adecuadamente, es decir, viven en condiciones de pobreza extrema.

La pobreza se origina en la incapacidad de la economía peruana para generar suficientes empleos productivos. De los 140 mil jóvenes que se integran cada año a la fuerza laboral urbana, menos de un tercio obtiene un empleo adecuado y casi el 40% está desempleado o trabajando en algo que no implica capacitación alguna ni ofrece perspectivas de progreso futuro.

Dada la complejidad del fenómeno, existen distintas metodologías para medirlo, tales como el método de la línea de pobreza, el método de las NBIs, el método integrado, entre otros. No obstante, tan importante como tener una cuantificación rigurosa de los niveles de pobreza y las tendencias en su evolución temporal y espacial, es analizar las relaciones entre ellas y sus factores explicativos que podrían albergar opciones de política para solucionarla, tales como la educación.

Debido a lo expuesto líneas antes, se están realizando estudios e investigaciones avanzadas, pero aún incipientes en el aspecto metodológico, sobre el rol de factores como la posesión de activos privados, la distribución del ingreso, el acceso a activos públicos, en su explicación del fenómeno de la pobreza, o mejor dicho sobre la adquisición de tal condición. Investigaciones que estén orientadas y permitan un mejor diseño e implementación de las políticas públicas y

sociales. En tal sentido queremos afirmar que el desarrollo que se viene dando en el mundo en muchos aspectos del conocimiento, entre ellos en los campos de la economía y la estadística, permiten explicar fenómenos a través del desarrollo y análisis de enfoques cualitativos.

Con la selección del enfoque y la formulación de modelos donde la variable dependiente cualitativa es discreta dicotómica (2 niveles), para efectos de esta investigación, ser jefe de hogar pobre o jefe de hogar no pobre; y expresada a través de variables o características sociales y demográficas que pueden poseer los individuos (el estado civil, experiencia educacional, categoría ocupacional, acceso a activos públicos, etc) y el hogar (cantidad de miembros, ingreso per cápita mensual), podemos colaborar aún más con explicaciones razonables a este vasto fenómeno de LA POBREZA, constituido en el Perú actualmente por más de la mitad de su población.

Actualmente, en nuestro país pueden estarse aplicando inadecuadamente modelos con estas características, debido a la falta de un proceso de análisis adecuado que determine su correcta utilización y que puede responder a un desconocimiento y una falta de apoderamiento de las metodologías adecuadas para su implementación y como consecuencia directa una correcta explicación de los fenómenos bajo estudio, en donde la pobreza no se encontraría sola, sino también otras manifestaciones sociales como la situación laboral, la educación, o en el campo de la medicina para el caso de enfermedades como la diabetes, etc.

En muchas de las investigaciones de índole social y demográfica, la naturaleza del fenómeno a indagar, en este caso la condición ser jefe de hogar pobre o jefe de hogar no pobre y sus factores determinantes de naturaleza social, económica y demográfica, podrían determinar un tipo de enfoque (proporciones muestrales u observaciones individuales) a seguir dentro de lo que se constituyen los modelos con variable dependiente dicotómica y por lo tanto utilizar los modelos logit o probit en la explicación de este fenómeno.

Desde la perspectiva estadística con que se enfoca al fenómeno de pobreza, se puede evidenciar que no es aún clara la idea de optar por alguno de los dos modelos (el modelo Logit o el modelo Probit) siguiendo el enfoque de observaciones o casos individuales, pero que este último, permitiría saltar las limitaciones que posee el enfoque de proporciones muestrales o de clasificación de casos u observaciones en función a sus factores determinantes de índole social, demográfico, económico, etc.

En tal sentido nos vemos en la necesidad de proponer una metodología de análisis y utilización de los modelos logit y probit con variable dependiente dicotómica, valiéndonos para ello de una explicación de la pobreza en los jefes de hogar del Perú para el año 2001, cuantificando el aporte de sus factores determinantes entre ellos el nivel de educación, los años de estudios, la experiencia laboral, la condición ocupacional, el acceso a activos públicos, etc; y como esta posibilita la viabilidad de los enfoques de utilización de dichos modelos y un mejor acercamiento a su explicación.

Para esto se analizará la información recogida por La Encuesta Nacional de Hogares (ENAH) - IV TRIMESTRE Condiciones de Vida Y Pobreza, realizada por el Instituto Nacional de Estadística e Informática en el año de 2001.

2.1 Formulación del Problema

¿En qué medida una explicación de la pobreza en los jefes de hogar del Perú se vería beneficiada por la adopción de un enfoque de proporciones muestrales o por un enfoque de observaciones individuales?

Sistematización del Problema

- ¿Cómo la pobreza en los jefes de hogar del Perú a través de sus factores determinantes de naturaleza social, económica y demográfica hacen factible el enfoque de proporciones muestrales o el enfoque de observaciones individuales en los modelos logit y probit?.
- ¿En qué medida los factores determinantes de índole cuantitativo como el ingreso per cápita mensual, los años de estudios, la experiencia laboral, la cantidad de miembros en el hogar, etc, influyen sobre un enfoque de proporciones muestrales del modelo probit y la pobreza de los jefes de hogar del Perú?.
- ¿Resulta más eficaz una estimación del modelo logit que una estimación del modelo probit en un enfoque de observaciones individuales para cuantificar el aporte de los factores explicativos de la pobreza en los jefes de hogar?.
- ¿Qué similitudes y diferencias pueden ser encontradas en ambos modelos y cómo pueden verse estos reflejados y contrastados desde el caso de la pobreza en el Perú- año 2001?

III. OBJETIVOS DE LA INVESTIGACION

3.1 Objetivo General

Determinar en qué medida una explicación sobre la pobreza en los jefes de hogar del Perú para el año 2001 es beneficiada siguiendo un enfoque de proporciones muestrales o un enfoque de observaciones individuales.

3.2 Objetivos Específicos

- Estudiar la relación entre los factores determinantes de naturaleza social, económica, demográfica de la pobreza en los jefes de hogar y los enfoques de utilización de los modelos logit y los modelos probit.
- Analizar el grado de afectación de los factores de naturaleza cuantitativa

sobre el enfoque de proporciones muestrales del modelo probit y la pobreza en los jefes de hogar.

- Comparar una estimación del modelo logit con una estimación del modelo probit en un enfoque de casos individuales, para cuantificar el aporte de los factores explicativos de la pobreza en los jefes de hogar.
- Analizar las similitudes y diferencias que puedan ser encontradas en ambos modelos y como pueden verse estos reflejados y contrastados desde el caso de la pobreza en los jefes de hogar del Perú.

IV. MARCO TEORICO

4.1 BREVE FORMULACIÓN DEL CASO DE LA POBREZA EN EL PERÚ

4.1.1 ¿Qué es la pobreza?

La pobreza es una condición en la cual una o más personas tienen un nivel de bienestar inferior al mínimo socialmente aceptado. En una primera aproximación, la pobreza se asocia a la incapacidad de las personas para satisfacer sus necesidades básicas de alimentación. Luego se considera un concepto más amplio que incluye la salud, las condiciones de vivienda, educación, empleo, ingresos, gastos y aspectos más extensos como la identidad, los derechos humanos, la participación popular, entre otros¹.

En general, todas las definiciones apuntan a la situación en que se encuentran las personas que no disponen de los medios ("de producción", activos físicos, intelectuales, sociales, culturales, financieros y demás) que permitan su desarrollo personal y reflejen el estilo de vida de la formación social en su conjunto. El concepto de pobreza es evidentemente relativo y cambiante. Basta considerar las diferencias de aquello que define a un pobre en Suiza respecto de los satisfactores considerados relevantes en el Perú, así como las características de la pobreza del siglo XIX en plena revolución industrial versus la que presentan países como los

nuestros hoy en día en pleno tercer milenio.

4.1.2 Enfoques y métodos para la medición de la pobreza

Existen 3 grandes enfoques para medir la pobreza. El primero es el enfoque de la pobreza absoluta, que toma en cuenta el costo de una canasta mínima esencial de bienes y servicios y considera como pobres a todos aquellos cuyo consumo o ingreso está por debajo de este valor.

El enfoque de pobreza relativa considera al grupo de personas cuyo ingreso se encuentra por debajo de un determinado nivel. Por ejemplo, en algunos países se considera como pobres a todos aquellos que tienen remuneraciones inferiores a la mitad del ingreso promedio (Criterio aplicado en sociedades que han logrado erradicar la pobreza absoluta)

El enfoque de la exclusión social, de absoluta vigencia en Europa, presta atención a las personas que no pueden acceder a determinados servicios como por ejemplo el empleo, la educación superior, la vivienda propia, el empleo y otros.

Ahora dentro de lo correspondiente a los métodos de medición, solo nos centraremos en el método de línea de

1/ INEI. Metodologías Estadísticas, Año 1-Nº02 Metodología para la medición de la pobreza en el Perú.

pobreza. Una explicación al por qué de su elección se dará a continuación.

EL METODO DE LA LÍNEA DE POBREZA-LP

Este método centra su atención en la dimensión económica de la pobreza y utiliza el ingreso o el gasto como medidas del bienestar. Al determinar los niveles de pobreza, se compara el valor per cápita de ingreso o gasto en el hogar con el valor de una canasta mínima denominada línea de pobreza.

Cuando se utiliza el método de línea de pobreza por el consumo, se incorpora el valor de todos los bienes y servicios que consume el hogar, indistintamente de la forma de adquisición o consecución. La utilización del gasto de consumo tiene la ventaja de que es el mejor indicador para medir el bienestar porque se refiere a lo que realmente consume un hogar y no a lo que potencialmente puede consumir cuando se mide por el ingreso. Otro aspecto favorable es que el consumo es una variable más estable que el ingreso, lo que permite una mejor condición de la tendencia del nivel de pobreza.

Así como existen enfoques y métodos para medir la pobreza, existen definiciones que nos permitirán centrar aún mejor la idea de nivel de pobreza que puede tener una persona y/o un hogar en particular.

Pobreza Absoluta:

Comprende a las personas cuyos hogares tienen ingresos o consumo per cápita inferiores al costo de una canasta total de bienes y servicios mínimos esenciales. Línea de Pobreza Absoluta (LPA): Es el costo de una canasta mínima de bienes(incluido los alimentos) y servicios.

Pobreza Absoluta:

Consumo Hogar < Costo Canasta Básica

Consumo (LPA)

Pobreza Extrema:

Comprende a las personas cuyos hogares tienen ingresos o consumos per cápita inferiores al valor de una canasta mínima de alimentos.

Línea de Pobreza Extrema (LPE): Es el costo de una canasta mínima de alimentos.

Pobreza Extrema:

Consumo Hogar < Costo Canasta Básica

Alimenticia (LPE)

Después de mostrar, a modo general, algunos enfoques y a grosso modo métodos de medición de pobreza, que si bien es cierto no es uno de los objetivos explícitos de esta investigación medirla, pretendemos con ello guiar al lector hacia la concepción y formulación de una variable que permita clasificar a un individuo de acuerdo a su nivel de pobreza, en otras palabras, si este posee o no la condición de pobreza.

Según Sen (1992) la medición de pobreza requiere realizar dos ejercicios distintos pero interrelacionados: la identificación de los pobres por un lado y la agregación por otro. Este último es el usado para obtener indicadores resumen del nivel(incidencia) de la pobreza. Para el caso de esta investigación solo realizaremos en alguna medida el primer ejercicio.

Esto quiere decir que necesitaríamos de un método de identificación para construir lo que será la variable dependiente de las estimaciones y es allí donde entra a tallar el MÉTODO DE LÍNEA DE POBREZA(LP), definido anteriormente. Con lo cual se define específicamente los valores que

podría tomar la variable en estudio: igual a 1 si es un hogar pobre y 0 si es un hogar no pobre, y como vamos a trabajar a nivel de individuos, tomamos, 1 si fuese un Jefe de Hogar (JH) que pertenece a un hogar pobre y 0 si fuese un JH que no pertenece a un hogar pobre.

Obviamente, al trabajar con esta variable dependiente no se podrá conocer la intensidad de la pobreza². Esta limitación puede ser resuelta mediante el uso de las medidas FGT (Foster et al., 1984), que surgen de la siguiente expresión:

$$FGT_{\alpha} = \frac{1}{N} * \sum \left(\frac{z_i - y_i}{z_i} \right)^{\alpha} \quad \text{ECUACIÓN IV.1}$$

Donde: y_i es el ingreso de la i -ésima familia u hogar, z_i es la línea de pobreza de ese hogar y α el parámetro de aversión a la pobreza.

Entonces, si $\alpha = 0$ obtenemos, el porcentaje de pobres.

Si $\alpha = 1$, obtenemos lo concerniente a Brecha de Pobreza.

$$FGT_1 = \frac{1}{N} * \sum \left(\frac{z_i - y_i}{z_i} \right)^1$$

Este es el denominado poverty gap de profundidad o intensidad de la pobreza. A diferencia del anterior, esta medida muestra la brecha existente entre el ingreso de los hogares pobres y la línea de pobreza.

Si $\alpha = 2$, arroja un indicador de severidad de la pobreza y que puede ser interpretado como la suma de dos componentes: la brecha de pobreza y la desigualdad entre los pobres.

De acuerdo con lo anterior es posible asignar a cada hogar un valor de intensidad y de severidad de pobreza dado por los supuestos acerca del parámetro α de dichas medidas. Esta forma de proceder requiere de métodos de estimación diferentes³, que no es motivo de estudio en este trabajo.

LOS ACTIVOS DE LOS POBRES EN EL PERÚ⁴

Tanto la distribución del ingreso como los niveles de pobreza han registrado importantes modificaciones a lo largo de las últimas cuatro décadas en el Perú. Más allá de las diferencias metodológicas asociadas al cálculo de estos indicadores, la evidencia sugiere que en los últimos 40 años se habría reducido la dispersión en la distribución del ingreso. Asimismo, se habría producido una importante reducción en los niveles de pobreza.

De otro lado, un análisis de la distribución de activos durante los últimos 10 años revela en general una continuación de las tendencias de largo plazo. El nivel educativo medio sigue aumentando y la desigualdad en el acceso a educación es cada vez más baja. En cambio, en el caso de los servicios públicos, si bien el acceso

2/ En términos de Sen (Sen, 1976), ante la necesidad de obtener una medida agregada de la pobreza, H , ocurre que no se satisfacen los axiomas de monotonidad y transferencia. El primero establece que toda medida de pobreza bien conformada debe reflejar las variaciones del ingreso de los hogares situados por debajo de la LP. El segundo, que la medida de pobreza debe ser sensible a las transferencias de ingresos entre pobres y no pobres.

3/ Paz, Jorge. La pobreza en Argentina: una comparación entre regiones disímiles. Unas, 2001.

4/ Escobal, Javier; Saavedra, Jaime; Torero, Máximo. Los Activos de los pobres en el Perú. GRADE. 1998.

medio ha ido aumentando, los niveles de desigualdad siguen siendo muy elevados. Asimismo, el acceso al crédito es muy diferenciado según los quintiles de gasto, mientras que el ahorro y los bienes financieros durables, que son activos que pueden fungir como colaterales, están entre los activos peor distribuidos. Asimismo, reducciones en el tamaño de la familia tienen un impacto positivo y significativo sobre los activos privados. En ese sentido, la concepción que un mayor tamaño de la familia implica un aumento de los recursos productivos de la familia, y por lo tanto un mayor bienestar, no encuentra sustento teórico.

Debido a la multidimensionalidad de la pobreza, su análisis a través de los jefes de hogar en función a sus factores de índole cualitativo y cuantitativo, en el primer caso reflejado a través de posesión de activos privados como su nivel de educación, su condición ocupacional y tamaño de la empresa donde trabaja, si posee algún otro empleo, y la influencia de aquellos factores inseparables de su hogar y también de naturaleza cualitativa, como el acceso a activos públicos como agua y desagüe dentro de la vivienda, si algún lugar dentro de la vivienda es destinada para obtener ingresos juega un papel fundamental.

Dentro de factores de naturaleza cualitativa también resulta importante considerar la región y área natural de procedencia del jefe de hogar como factor preponderante para su condición de pobreza.

Con respecto a los factores de naturaleza cuantitativa propios del jefe de hogar son considerados los años de estudios, su

indicador proxy de experiencia laboral y su edad como aquellos que pueden determinar su condición. En la misma línea, factores correspondientes al hogar son la cantidad de miembros en el hogar, ingreso per cápita mensual, la cantidad de personas en edad de trabajar, serían aquellos que determinan una mayor probabilidad de ser pobre.

4.2. LOS MODELOS LOGIT Y PROBIT CON VARIABLE DEPENDIENTE DICOTÓMICA (VDD)

4.2.1. Modelos con Variable Dependiente Dicotómica (VDD).

En muchas situaciones, el fenómeno que queremos estudiar no es continuo, sino discreto. Por ejemplo, cuando queremos modelar la participación del mercado de trabajo, la condición de pobreza de un Jefe de Hogar, la decisión sobre si se hace una compra o no. En el caso de la condición de pobreza existen estudios que sugieren que factores como la educación, la edad, el número de hijos y ciertas características económicas, sociales, demográficas, etc; que podrían ser relevantes para explicar si un individuo Jefe de Hogar está más afecto a adquirir la condición de pobreza. Pero, obviamente, algo falta si se aplica en este caso el mismo tipo de modelo de regresión que utilizábamos para analizar el consumo o los costes de producción, o algún otro fenómeno de naturaleza cuantitativa.

Vamos a analizar algunos modelos conocidos como modelos de **respuesta cualitativa (RC)**, y lo que podemos mencionar es que tienen en común que

su variable dependiente es discreta, es decir puede tomar valores como "no" o "sí" que pueden ser codificados como "0" ó "1"; o quizás más valores "0", "1", "2" según sus categorías, para representar los resultados cualitativos respectivos.

Pero reflejemos esto en ejemplos concretos, el tipo de variable dependiente y el fenómeno en estudio:

- **Participación en el mercado de trabajo:** Donde 0 es "no" y 1 significa "sí", donde la participación es Var. Dependiente nominal dicotómica, y se representa con 0 y 1 por comodidad.
- **Opinión sobre cierto tipo de legislación:** Donde 0 sería "totalmente opuesto" 1 para "opuesto" 2 para "indiferente" 3 para "a favor" 4 para "totalmente a favor" Aquí vemos que se ordenan las respuestas no en función a su valor cuantitativo, sino por una cuestión de grado en la respuesta, en donde la OPINIÓN es Var. Dependiente Ordinal.
- **Área de trabajo escogida por un individuo:** Donde 0 es representar al vendedor, 1 para ingeniero, 2 para abogado, 3 para político, y así sucesivamente, y es de allí que podemos mencionar que AREA DE TRABAJO es Var. Dependiente nominal politómica⁵.

En ninguno de estos casos parece posible, en principio, utilizar el análisis de regresión clásico. Sin embargo, en todos ellos es posible construir modelos que enlacen el resultado o la decisión a tomar a través de su aporte en la probabilidad de la realización del fenómeno bajo estudio; con un conjunto de factores, con la misma filosofía que en regresión. Entonces lo que se hace es analizar cada uno de estos modelos dentro de un marco general de **LOS MODELOS DE PROBABILIDAD.**

$$\Pr(\text{Ocurre suceso } j) = \Pr(Y=j) = F(\text{efectos relevantes: parámetros})$$

Resulta conveniente agrupar estos modelos en dos grandes clases: aquellos que siguen un **enfoque binomial**, es decir, si el resultado o fenómeno depende de la elección o la situación en dos alternativas. Para ello podemos mencionar al respecto; que si tenemos a cada uno de N individuos, casos u objetos que pueden ser clasificados independientemente en 1 de 2 categorías complementarias, ejemplo de ello puede ser cara o sello de un lanzamiento de monedas, pacientes curados o no curados, personas por sobre o debajo de un nivel de ingreso, etc. En este caso se tiene que cada individuo tiene la misma probabilidad p de estar en una de las dos categorías, por ejemplo si el Jefe de Hogar es pobre ($0 \leq p \leq 1$); y la prob. 1-p de encontrarse en la otra categoría complementaria, es decir, si el Jefe de Hogar no es pobre.

Entonces la probabilidad de que X de los N individuos sean pobres es:

5/ Variable nominal politómica, es aquella que en sus categorías no denota ningún tipo de ordenamiento ni grado específico, otro ejemplo de ello lo constituye el estado civil (soltero, casado, viudo, conviviente, etc).

Donde $x=0,1,2,3...N$.

$$P[X=x]=\binom{N}{x}p^x(1-p)^{N-x}$$

ECUACIÓN IV.2

Así como existen en los modelos el **enfoque binomial**, existe asimismo el **enfoque multinomial**, que es aquel que depende de una elección o resultado de más de 2 alternativas y que pueden reflejarse en algunos de los ejemplos mencionados líneas antes y que resultan en algunos aspectos novedosos pero que son, en su mayor parte, extensiones de los casos binomiales.

Queremos dejar en claro aquí que en esta investigación el enfoque que seguirá el modelo será de tipo binomial, porque plantear uno de tipo multinomial implicaría, que los diversos niveles de condición de pobreza (no pobre, pobre y pobreza extrema), se encontrarían a un mismo nivel de selección, vale decir, que en el caso de pobreza absoluta y pobreza extrema tendrían que considerarse como grupos excluyentes uno del otro, lo cual no se da en este caso, pues la pobreza extrema es una condición mucho más precaria en todo aspecto que la pobreza absoluta. Es este entonces nuestro punto de partida para poder determinar y cuantificar el aporte de las variables o factores que inciden sobre la probabilidad que un jefe de hogar se encuentre en condición de pobreza. En tal sentido los modelos que más se adecuan en esta línea son el modelo logit con variable dependiente discreta dicotómica (a partir de ahora, VDD) y el modelo probit con VDD.

4.2.2. Formulación del modelo logit y el modelo probit con variable dependiente dicotómica (VDD)

Empezaremos esta formulación a partir de la suposición de un modelo de probabilidad de condición de pobreza para Jefes de Hogar (JH), donde:

- $Y=1$ El JH se encuentra en condición de pobreza
 $Y=0$ El JH no se encuentra en condición de pobreza.

Vamos a suponer que un vector de variables explicativas o conjunto de factores que expliquen este fenómeno venga dada por:

X = Nivel de Educación, Analfabetismo, Dominio Geográfico, Experiencia, Estado Civil, Categoría Ocupacional, etc.

Podemos imaginar que la primera idea intuitiva que gira alrededor del fenómeno de la condición de pobreza es que esta podría ser explicada a través del conjunto de factores mencionados, y cuánto es que estos contribuyen individual y en forma conjunta a dicho fenómeno.

Entonces, autores como Green plantean la siguiente idea:

Sea:

La probabilidad de que el Jefe de Hogar sea pobre:

$$P[Y=1]=F(X,\beta)$$

ECUACIÓN IV.3

Y la probabilidad de que el Jefe de Hogar no sea pobre:

$$P[Y = 0] = 1 - F(X, \beta)$$

Donde el vector de parámetros(β) refleja el aporte o impacto que X (vector de var. explicativas) tiene como parte de la función de distribución acumulada sobre la probabilidad. Por ejemplo uno de los factores que podría interesarnos sería ver el efecto que el nivel de educación tiene sobre la probabilidad de ser pobre.

A partir de este punto la incógnita cae sobre el lado derecho de dicha ecuación y sobre cómo plantear un modelo adecuado para este.

Una solución a ello puede darse en un modelo de regresión lineal.

$$F(X, \beta) = \beta'X$$

ECUACIÓN IV.4

Tenemos que:

$$E(Y/X) = 0 * \Pr(y=0) + 1 * \Pr(y=1)$$

$$E(Y/X) = F(X, \beta)$$

A partir de lo anterior podemos construir:

$$\begin{aligned} Y &= y = E[y/x] + [y - E[y/x]] \\ &= F(x, \beta) + e \end{aligned}$$

$$Y = \beta'X + e$$

Modelo de Probabilidad Lineal. ECUACIÓN IV.5

Pero este modelo de probabilidad lineal presenta algunos inconvenientes:

(i) e presenta heterocedasticidad que depende de β , esto es si:

$$Y = 0 \rightarrow \beta'X + e = 0 \rightarrow e = -\beta'X, \text{ donde } p(y=0) = 1 - F$$

$$Y = 1 \rightarrow \beta'X + e = 1 \rightarrow e = 1 - \beta'X, \text{ donde } p(y=1) = F$$

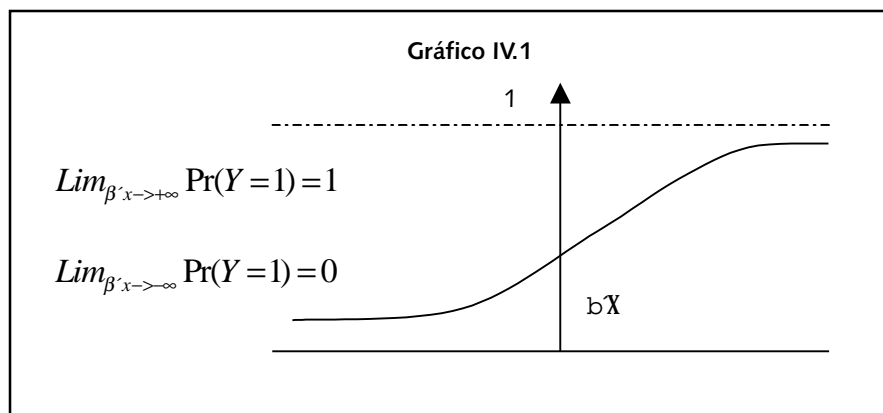
Entonces tendríamos que

$$V(e) = (-\beta'X)(1 - \beta'X)$$

ECUACIÓN IV.6

(ii) El inconveniente más serio es que no se puede asegurar que las predicciones parezcan verdaderas probabilidades. No se puede restringir $\beta'X$ al intervalo $[0,1]$, lo cual origina tanto varianzas negativas como probabilidades imposibles.

Entonces para un vector de regresores dado, esperaríamos que [Gráfico IV.1]:



Analizando el gráfico anterior, vemos que para los requisitos especificados anteriormente, en principio bastaría trabajar con una Función de Distribución Acumulada definida sobre la recta real, en tal sentido las funciones idóneas para tal efecto vienen a ser la Normal y la Logística.

Es natural ahora preguntarse ¿Cuál de las dos debe usarse?, la respuesta a esa pregunta y otras que se han empezado a formar, van a ir siendo resueltas en la medida de lo posible en la secciones siguientes.

Supongamos ahora que en vez de utilizar la función lineal de probabilidad de X para caracterizar esta condición, utilizamos una función monótona $[F(b'x)]$ creciente del producto $b'x$, es decir, variables explicativas y aportes. Esta formulación es más general que la del modelo lineal de probabilidad y, en consecuencia, el procedimiento de estimación de los parámetros así como la forma en que dichos valores deben interpretarse, es ahora diferente.

- Al usar esta transformación, podemos ver que basta tomar una función real F acotada entre 0 y 1 para que el problema que se originaba sobre el rango de las probabilidades desaparezca. En efecto, ahora vemos que $P = F(b'x)$ está siempre entre 0 y 1, con independencia de los valores que toman los factores explicativos y sus efectos marginales. Es por ello que las funciones de distribución de variables aleatorias son candidatos importantes a ser elegidas para estas transformaciones.
- Suponga que existe un indicador que depende de las características

individuales: $li = b'x$, que determina la decisión tomada para cada individuo o la condición que este posee frente a un fenómeno dado. Es decir, el individuo toma la decisión o se encuentra en el estado $Y_i = 1$ si el valor de su indicador es superior a un cierto valor crítico I^* , y la decisión contraria o se encuentre en el otro estado si $Y_i = 0$. Es decir, el indicador li refleja el sentimiento del decisor frente a la opción indicada $Y_i = 1$, de modo que si su predisposición, indicada por li es suficientemente grande (mayor que li^*), escoge dicha opción, y si no, elegirá la opción alternativa.

Por ser desconocido, consideramos el valor crítico I^* del indicador para cada individuo como una variable aleatoria. Entonces de acuerdo con esta interpretación, la probabilidad de que el individuo i -ésimo elija o posea la condición $Y_i = 1$ viene dada por:

$$P_i = P(Y_i = 1) = P(I_i^* \leq I_i) = F(\beta'x)$$

ECUACIÓN IV.7

Donde F es la distribución de probabilidad de la variable aleatoria I^* .

4.2.3. Caracterización de los modelos logit y probit aplicados al caso de la pobreza en el Perú.

4.2.3.1. El Modelo Logit

Supongamos el siguiente ejemplo, se tiene que Y (1 = jefe de hogar pobre, 0 = jefe de hogar no pobre) en función al ingreso familiar $X(S/.)$ para un conjunto de familias. Se tiene la siguiente representación de la condición de pobreza:

$$P_i = E(y=1/X_i) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}} \quad \text{ECUACIÓN IV.8}$$

IV.8

Para facilidad de la exposición, se escribe [ECUACIÓN IV.8] como:

$$P_i = E(y=1/X_i) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}} \quad \text{donde } Z_i = (\beta_1 + \beta_2 X_i) \quad \text{ECUACIÓN IV.9}$$

La [ECUACIÓN IV.9] representa lo que se conoce como función de distribución logística (acumulativa). Es fácil verificar que a medida que Z_i se encuentra dentro de un rango de $-\infty$ a $+\infty$, P_i se encuentra dentro de un rango 0 a 1 y que P_i no está linealmente relacionado con Z_i (es decir con X_i), satisfaciendo así requerimientos que son considerados. Pero parece que al satisfacer estos requerimientos se ha creado un problema de estimación porque P_i es no lineal no solamente con X sino también en los β , como puede verse claramente a partir de la [ECUACIÓN IV.8]. Esto significa que no se puede utilizar el procedimiento familiar MCO para estimar los parámetros. Pero este problema es más aparente que real porque la [ECUACIÓN IV.8] es intrínsecamente lineal, lo cual puede verse de la siguiente manera.

Si P_i , la probabilidad de ser jefe de hogar pobre, está dada por la [ECUACIÓN IV.9] entonces $(1 - P_i)$, la probabilidad de jefe de hogar no pobre.

$$1 - P_i = \frac{1}{1 + e^{Z_i}} \quad \text{ECUACIÓN IV.10}$$

Por consiguiente, se puede escribir

$$\frac{P_i}{1 - P_i} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} = e^{Z_i} \quad \text{ECUACIÓN IV.11}$$

Ahora $P_i / (1 - P_i)$ es sencillamente la razón de probabilidades ('Odds Ratio') a favor de ser jefe de hogar pobre- la razón de la probabilidad de que un jefe de hogar sea pobre a la probabilidad de que no sea pobre. Así, si $P_i = 0.8$, significa que las probabilidades son 4 a 1 a favor de que el jefe de hogar sea pobre.

Ahora, si se toma el logaritmo natural de [ECUACIÓN IV.11], se obtiene un resultado muy interesante, a saber,

$$L_i = \ln\left(\frac{P_i}{1 - P_i}\right) = Z_i = \beta_1 + \beta_2 X_i \quad \text{ECUACIÓN IV.12}$$

es decir, L_i , el logaritmo de la razón de probabilidades no es solamente lineal en X_i , sino también (desde el punto de vista de estimación) lineal en los parámetros, L es llamado Logit y de aquí el nombre modelo LOGIT para modelos como la [ECUACIÓN IV.12].

Obsérvense estas características del modelo Logit:

- A medida que P va de 0 a 1 (es decir, a medida que Z varía de $-\infty$ a $+\infty$, el Logit L va de $-\infty$ a $+\infty$). Es decir, aunque las probabilidades (por necesidad) se encuentran entre 0 y 1, los Logit no están limitados en esa forma.
- Aunque L es lineal en X , las probabilidades en sí mismas no lo son.

Esta propiedad hace contraste con el modelo de MPL, en donde las probabilidades aumentan linealmente con X.

- (c) La interpretación del modelo Logit es el siguiente: β_2 , la pendiente, mide el cambio en L ocasionado por un cambio unitario en X, es decir, **dice como el logaritmo de las probabilidades a favor de ser jefe de hogar pobre cambia a medida que el ingreso cambia en una unidad, por ejemplo de S/. 100.** El intercepto β_1 es el valor del logaritmo de las probabilidades a favor de ser jefe de hogar pobre si el ingreso es cero.

- (d) Dado un nivel determinado de ingresos, por ejemplo, X, si realmente se desea estimar la probabilidad misma de ser jefe de hogar pobre, y no las probabilidades a favor de ser jefe de hogar pobre, esto puede hacerse directamente a partir de la [ECUACIÓN IV.8], una vez que se disponga de las estimaciones de β_1 y β_2 .
- (e) Mientras que el MLP supone que Pi está linealmente relacionado con Xi, el modelo Logit supone que el logaritmo de la razón de probabilidades está relacionado linealmente con X.

4.2.3.2. EL MODELO PROBIT

Si se elige como función F la función de distribución Φ de una variable normal (0,1), se tiene:

$$P = E(y/x) = P(Y=1/x) = P[I^* \leq I] = \Phi(\beta'x)$$

ECUACIÓN IV.13

De modo que:

$$\beta'x = \Phi^{-1}(P)$$

La probabilidad correspondiente a un vector X de factores que contribuyen a explicar un fenómeno, como el de la pobreza es ahora:

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta'x} e^{-t^2/2} dt$$

ECUACIÓN IV.14

Como se ha podido apreciar anteriormente, para explicar el comportamiento de una variable dependiente dicotómica, es preciso utilizar una FDA seleccionada apropiadamente. El modelo Logit utiliza la función distribución logística acumulativa. Pero esta no es la única FDA que se puede utilizar. En algunas aplicaciones, la FDA normal se ha encontrado útil. El modelo de Estimación que surge de una FDA normal es comúnmente conocido como el modelo Probit, aunque algunas veces también es conocido como el modelo normit. En principio, se puede sustituir la FDA normal por la FDA logística y proceder de acuerdo al modelo Logit. Pero en lugar de seguir este camino se presentará el modelo probit basado en la teoría de utilidad o de la perspectiva de selección racional con base en el comportamiento, según el modelo desarrollado por McFadden.

Para motivar el modelo Probit supóngase el ejemplo de condición de pobreza, la posesión del i-ésimo jefe de hogar de la condición de pobreza o de no poseerla, medida a través de un índice imperfecto

de conveniencia I_i que está determinado por una o varias variables explicativas, por ejemplo, el ingreso X_i , de tal manera que entre mayor sea el valor del índice, mayor será la probabilidad de que el jefe de hogar sea pobre. Se expresa el índice I_i , como:

$$I_i = \beta_1 + \beta_2 X_i$$

ECUACIÓN IV.15

¿Cómo se relaciona el I_i , índice imperfecto de conveniencia con la condición específica de ser pobre? Sea $Y=1$ si es jefe de hogar pobre y $Y=0$ si no es. Ahora bien, es razonable suponer que para cada individuo hay un nivel crítico o umbral del índice, que se puede denominar I_i^* , tal que si I_i excede a I_i^* , el jefe de hogar es pobre, de lo contrario no lo es. El nivel crítico I_i^* , al igual que I_i , no es observable, y se supone que está distribuido normalmente con la misma media y varianza, y por lo tanto es posible no solamente estimar los parámetros del índice, sino también obtener alguna información sobre el índice imperfecto de

la cual depende nuestra variable observable.

Dado el supuesto de normalidad, la probabilidad de que I_i^* sea menor o igual que I_i , puede ser calculada a partir de la FDA normal estándar como:

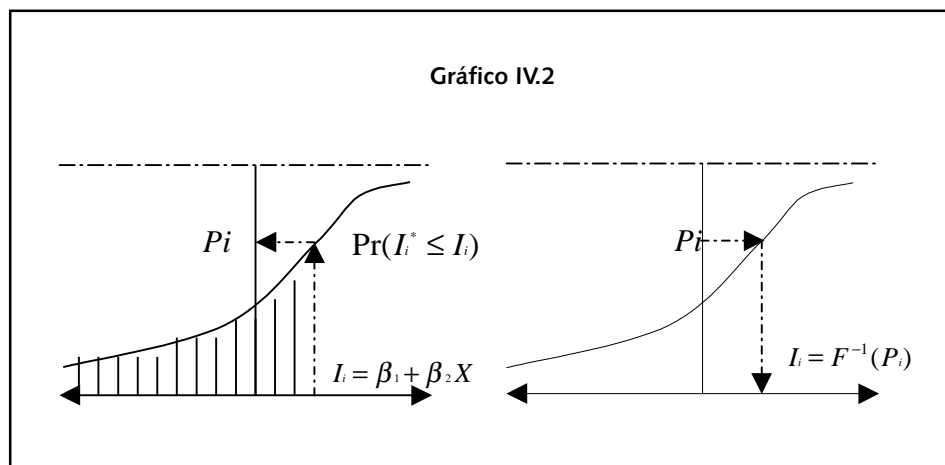
$$P_i = \Pr(Y=1) = \Pr(I_i^* \leq I_i) = F(I_i)$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1 + \beta_2 X_i} e^{-\frac{t^2}{2}} dt$$

ECUACION IV.16

donde t es una variable normal estandarizada, es decir, $t \sim N(0,1)$.

Puesto que P_i representa la probabilidad de que ocurra un evento, en este caso la probabilidad de poseer la condición de pobreza, ésta se mide por el área de la curva normal estándar de $-\infty$ a I_i , como se muestra en la figura siguiente [GRAFICO IV.2]:



Ahora, para obtener información sobre I_i , el índice de utilidad, lo mismo que para los coeficientes estimados se toma la inversa de [ECUACIÓN IV.7] para obtener:

$$I_i = F^{-1}(P_i) = \beta_1 + \beta_2 X_i$$

ECUACIÓN IV.17

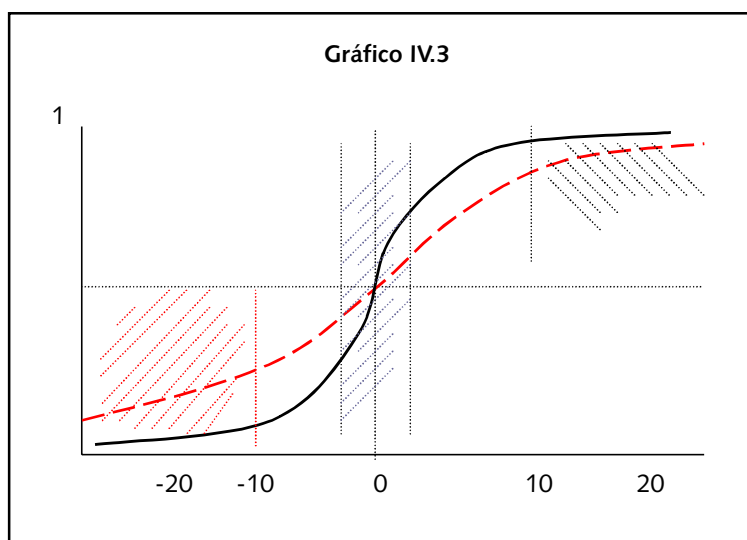
Donde F^{-1} es la inversa de la FDA normal. El significado de todo esto puede aclararse con la figura anterior [GRÁFICO IV.2], donde en la figura del lado izquierdo se obtiene (de la ordenada) la probabilidad (acumulada) de ser jefe de hogar pobre dado I_i^* menor o igual que I_i , mientras que en la parte derecha (de la abscisa) se

obtendría el valor de I_i , dado el valor de P_i , es decir, evaluar la probabilidad en la inversa de la FDA Normal.

4.2.3.3. Similitudes y diferencias en ambos modelos

Podemos ver que la distribución logística : es similar a la distribución normal, excepto por sus colas, (la distribución Logística se parece más a la distribución t con siete grados de libertad).

Analicemos la siguiente gráfica [GRÁFICO IV.3], que compara ambas funciones de distribución.



La FDA Normal es aquella gráfica alrededor(____) de la recta real más gruesa y la recta de segmentos (-----) viene a ser la FDA Logística y el eje vertical se desplaza entre 0 y 1.

De aquí podemos empezar a analizar cada una de las 3 regiones formadas:

- La PRIMERA REGIÓN, la inferior izquierda, muestra que para el fenómeno en estudio $Y=1$, por ejemplo: si el Jefe de Hogar es pobre, las estimaciones del modelo Logit producirían mayores contribuciones para la probabilidad de ser pobre en función a sus factores explicativos, que el modelo probit.

- La SEGUNDA REGIÓN, la central, muestra que las 2 distribuciones generan estimaciones similares, es más autores como Amemiya dan un intervalo de variación para dichas estimaciones similares $< -1.2, 1.2 >$.
- La TERCERA REGIÓN, la superior derecha, muestra que el modelo probit generaría mejores estimaciones de los aportes de las variables en la probabilidad de ser pobre que el modelo Logit.

Debemos hacer énfasis en que ambos modelos presentan muchas similitudes con respecto a su representación de probabilidad, visiblemente apreciable en la gráfica anterior, y con respecto a sus estimaciones en la mayoría de aplicaciones parece que se llega a los mismos resultados partiendo de una aplicación u otra.

Es de conocimiento que cabe esperar que los 2 modelos originen predicciones o aportes diferentes en los factores si la muestra contiene:

- Pocas respuestas afirmativas ($Y=1$), es decir, pocas observaciones para JH (Jefe de Hogar) en condición de pobreza; y del mismo modo, pocas respuestas para ($Y=0$), vale decir, pocas observaciones para JH en status de no pobreza
- Gran variación en una variable independiente de importancia, especialmente si se cumple lo mencionado en el párrafo anterior, por ejemplo: presentar una variación considerable con respecto a los niveles de educación de los JH de la población en estudio.

Por sencillez de cálculo pueden existir razones prácticas para preferir una u otra distribución; pero desde el punto de vista teórico resulta difícil justificar esta elección. Amemiya (1981) analiza varios aspectos relacionados con esta cuestión pero, en términos generales, puede decirse que este problema no se ha resuelto aún.

Los efectos marginales en ambos modelos

Hay que considerar que el modelo de probabilidad es un modelo de regresión:

$$E[y/x] = 0*[1-F(b'X)] + 1*[F(b'X)] \\ = [F(b'X)]$$

Y que sea cual fuere la distribución que se utilice, es importante observar que los parámetros ($b'X$) del modelo, como los de cualquier modelo de regresión no lineal, no son necesariamente los efectos marginales comunes analizados.

$$\frac{\partial E[y/x]}{\partial x} = \left\{ \frac{dF(\beta'x)}{d(\beta'x)} \right\} \beta = f(\beta'x)\beta$$

ECUACIÓN IV.18

Donde $f(\cdot)$ es la función de densidad asociada a la Función de distribución $F(\cdot)$.

Para la distribución normal:

$$\frac{\partial E[y/x]}{\partial x} = \phi(\beta'x)\beta$$

ECUACIÓN IV.19

Siendo $\mathcal{F}(b'x)$ la función densidad normal estándar.

Para el caso de la distribución logística:

$$\frac{d \Lambda[\beta'x]}{d(\beta'x)} = \frac{e^{\beta'x}}{(1 + e^{\beta'x})^2}$$

ECUACIÓN IV.20

Lo que es sumamente práctico. Así en el modelo logit.

$$\frac{\partial E[y/x]}{\partial x} = \Lambda(\beta'x)[1 - \Lambda(\beta'x)]\beta$$

ECUACIÓN IV.21

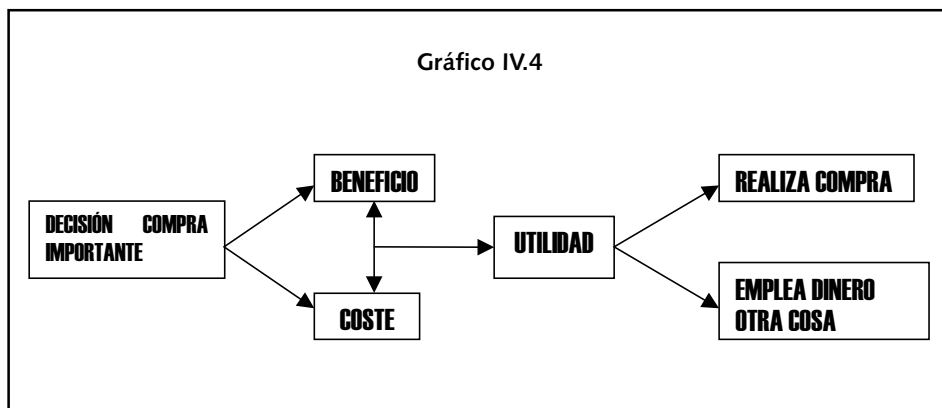
Para interpretar el modelo estimado, resulta útil calcular estos efectos marginales en varios valores de x: por ejemplo en las medias de los regresores o en otros puntos que puedan resultar de interés.

Los modelos de variable dependiente discreta aparecen con frecuencia como modelos con función índice, es decir, interpretamos el resultado de una elección discreta como un reflejo de una regresión subyacente. Pero el significado de esto lo entenderemos a través de un ejemplo:

Supongamos que tenemos que tomar una **decisión** para hacer una **compra importante**.

Teoría: Consumidor hace cálculo **Beneficio** marginal - **Coste** marginal : Obtener una **utilidad** y tomar una decisión..

Gráfico IV.4



Podemos apreciar que el beneficio marginal es evidentemente no observable (no tangible), modelizamos la diferencia entre beneficio y coste con una variable no observable y cumple.

$$y^* = \beta'x + \varepsilon$$

ECUACIÓN IV.22

Suponemos que la distribución de $\varepsilon \rightarrow N(0,1)$ ó Logística, entonces, no se observa el beneficio neto de la compra, sólo si esta se hace o no.

$$\begin{aligned} Y &= 1 && \text{si } Y^* > 0 \\ Y &= 0 && \text{si } Y^* \leq 0 \end{aligned} \quad (*)$$

Donde $b'X$, recibe el nombre de función índice.

Cabe recalcar que se deben considerar algunos aspectos en la construcción de (*):

- Primero: La hipótesis de varianza unitaria es una normalización que no juega ningún papel importante. Supongamos que la varianza de ε es en realidad σ^2 y multipliquemos entonces los coeficientes por σ . Nuestros datos observados no varían: y es 0 ó 1, dependiendo únicamente

del signo de y^* , no de la escala en que se midan los datos.

- Segundo: La hipótesis de que el umbral es 0, tampoco juega ningún papel si el modelo contiene término constante⁶.

La probabilidad del suceso $Y=1$, jefe de hogar pobre, se puede ver como sigue:

$$\Pr(y^* > 0) = \Pr(\beta'x + \varepsilon > 0) = \Pr(\varepsilon > -\beta'x) = 1 - F(-\beta'x)$$

Si la distribución es simétrica⁷, como lo son la normal y la logística.

$$\Pr(y^* > 0) = \Pr(\varepsilon > -\beta'x) = F(\beta'x)$$

De este modo se habría obtenido un modelo estructural para la probabilidad y como esta depende de la FDA Normal o Logística.

Análisis de datos a través de proporciones muestrales

Cuando se analizan respuestas binarias, los datos vendrán dados de una de las dos formas siguientes: o bien, tal y como se ha considerado hasta ahora, de **forma individual** (es decir, cada observación está formada por la respuesta del individuo y un vector de regresores asociados a él $[Y_i, X_i]$), O bien de **forma agrupada** (es decir, los datos consisten en proporciones o recuentos de observaciones). Los datos en forma agrupada se obtienen

observando la respuesta de n_i individuos, todos ellos con la misma x_i . La variable dependiente observada será la proporción (P_i) de los n_i individuos ij para los cuales $y_{ij} = 1$. Una observación es por tanto $[n_i, P_i, x_i]$, $i = 1, \dots, N$. Los datos electorales constituyen un ejemplo típico⁸. En el caso de datos dados en forma agrupada, pueden analizarse la relación entre P_i y x_i no sólo utilizando los estimadores de máxima verosimilitud, sino también métodos de regresión. La proporción observada P_i es un estimador de la cantidad poblacional $\pi_i = F(\beta'x_i)$. Si consideramos esta igualdad como un sencillo problema de muestreo en una población Bernoulli, utilizando los resultados básicos de estadística obtenemos que:

$$P_i = F(\beta'x_i) + \varepsilon_i = \pi_i + \varepsilon_i, \text{ siendo}$$

$$E(\varepsilon_i) = 0, \text{ Var}(\varepsilon_i) = \frac{\pi_i(1-\pi_i)}{n_i}$$

Este formato de regresión heterocedástica sugiere que los parámetros podrían estimarse utilizando una regresión de mínimos cuadrados ponderados no lineales. Para no detallar más podemos decir que este análisis y su estimación correspondiente será abordada en las secciones posteriores de esta investigación. Lo que podemos mencionar por ahora es que en la práctica surgen dos complicaciones con respecto a los errores en este enfoque. Primero, cuando la proporción se calcula utilizando una muestra de gran tamaño, la varianza del estimador puede llegar a ser sumamente pequeña. Ello originará que en la regresión

6/ Ha no ser que haya poderosas razones, los modelos binomiales deben incluir constante (Green)

7/ En distribuciones simétricas se tiene que: $1 - F(b'X) = F(-b'X)$

8/ Los estudios iniciales sobre modelos probit se desarrollaron en laboratorios. Cada observación consistía en n_i individuos que recibían una cierta dosis x_i , la proporción P_i respondían al tratamiento. Véase Finney (1971) y Cox (1970).

de mínimo chi-cuadrado los errores estándar sean inverosímilmente pequeños y los estadísticos t enormemente grandes. Por desgracia, todo esto es consecuencia de la estructura del modelo. A los mismos resultados se llega si se estima por máxima verosimilitud con datos de proporciones.

Segundo, es imposible obtener tanto el estimador de máxima verosimilitud como algún otro relacionado, si una de las dos proporciones es 0 ó 1. Se han sugerido varias posibles soluciones específicas para este caso, la que con más frecuencia se

utiliza consiste en sumar o restar un valor constante pequeño, por ejemplo, 0.001, al valor observado cuando éste sea 0 ó 1.

De lo anteriormente expresado, en esta investigación se tratará de desarrollar una metodología que permita cuantificar el aporte de los factores asociados a la pobreza con su correspondiente probabilidad de ser pobre para el jefe de hogar en los modelos Logit y Probit con variable dependiente dicotómica, utilizando criterios de enfoque a nivel teórico y práctico.

V. HIPOTESIS DE INVESTIGACIÓN

- Los factores de naturaleza cuantitativa como el ingreso per cápita mensual del hogar expresado a través de sus déciles de ingreso, los años de estudios del jefe hogar, etc, generan un modelo correctamente ajustado a la probabilidad de ser pobre de los jefes de hogar expresado a través de su estadístico de bondad de ajuste pearson χ^2 dejando de lado las limitaciones del tamaño de muestra, en el enfoque de proporciones muestrales del modelo probit.
- Los factores explicativos de la pobreza en los jefes de hogar de naturaleza cualitativa y cuantitativa exclusivos de estos como el nivel de educación, el tipo de colegio de estudio, la categoría ocupacional, el tamaño de la firma donde labora, la tenencia de otro empleo, el estado civil, su edad, su indicador de experiencia laboral; en combinación con los factores de naturaleza cualitativa y cuantitativa exclusivos del hogar como el indicador de si el hogar dedica un espacio del hogar a generación de ingresos, la cantidad de miembros en el hogar, la cantidad de miembros pertenecientes a la PET, el ingreso per cápita mensual, el acceso a activos públicos de agua y desagüe, no permiten generar modelos correctamente ajustados a la probabilidad de ser pobre del jefe de hogar siguiendo el enfoque de observaciones individuales o no clasificación en los modelos logit y probit.

VI. METODOS

6.1 Tratamiento de Errores en la adecuación de ambos modelos, análisis de factores explicativos observables y no observables

Casi todos los modelos de elección binaria, excepto el modelo de probabilidad lineal, se estiman habitualmente por el método de máxima verosimilitud. Cada observación se considera como realización individual de una variable aleatoria con distribución Bernoulli (es decir, binomial con $n=1$). La probabilidad conjunta o función de verosimilitud, de un modelo con probabilidad de éxito $F(b'X)$ y observaciones independientes es:

$$\Pr(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) =$$

$$\Pr(Y_1 = 1, Y_2 = 0, \dots, Y_n = 1, Y_n = 0) =$$

$$= \prod_{y_i=0} [1 - F(\beta'x_i)] \prod_{y_i=1} F(\beta'x_i)$$

ECUACIÓN VI.1

Podemos reescribir la fórmula anterior como:

$$L = \prod_{i=1}^n [F(\beta'x_i)]^{y_i} [1 - F(\beta'x_i)]^{1-y_i}$$

ECUACIÓN VI.2

Esta es la función de verosimilitud para una muestra de n observaciones.

$$\ln L = \sum_{i=1}^n [y_i \ln F(\beta'x_i) + (1 - y_i) \ln (1 - F(\beta'x_i))]$$

Las condiciones de primer orden del problema de maximización requieren que

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^n \left[\frac{y_i f_i}{F_i} + (1 - y_i) \frac{-f_i}{(1 - F_i)} \right] x_i = 0$$

ECUACIÓN VI.3

En la ecuación anterior y en lo que sigue, se utilizará el subíndice i para indicar que la función se evalúa en $b'X$, es decir, en el conjunto de factores explicativos al fenómeno. Al seleccionar una forma concreta para F_i se obtiene un modelo empírico.

A menos que se utilice el modelo de probabilidad lineal, las ecuaciones contenidas en la fórmula anterior serán no lineales y habrán de resolverse de modo iterativo.

6.1.1 Los métodos de estimación según enfoques

MODELO PROBIT

Estimación de mínimos cuadrados con enfoque de proporciones muestrales

El modelo original relaciona las frecuencias observadas p_i , con las probabilidades (P_i) que resultan de las clasificaciones de los factores explicativas, por ejemplo el ingreso per cápita mensual, por medio de:

9/ Como vimos hace un momento en distribuciones simétricas $1 - F(b'x) = F(-b'x)$. Definiendo $q = 2y - 1$, entonces $\ln L = \sum \ln F(qb'x)$.

$$p_i = (P_i + u_i)$$

por lo que

$$\Phi^{-1}(p_i) = \Phi^{-1}(P_i + u_i)$$

De aquí esta expresión puede aproximarse por:

$$\Phi^{-1}(p_i) \rightarrow X_i' \beta + \frac{1}{f(X_i' \beta)} u_i$$

El modelo probit puede por tanto estimarse de modo aproximado por una regresión de los llamados "probits" muestrales $\Phi^{-1}(p_i)$ sobre el vector X_i .

Se trata de calcular las frecuencias muestrales p_i , obtener los valores $\Phi^{-1}(p_i)$ a partir de la tablas de la distribución $N(0,1)$ y estimar la regresión descrita.

Ahora bien, los residuos tienen heterocedasticidad, puesto que:

$$\text{var}\left(\frac{u_i}{f(X_i' \beta)}\right) = \frac{P_i(1-P_i)}{n[f(X_i' \beta)]^2}$$

ECUACIÓN VI.4

Por lo que habrá que utilizar mínimos cuadrados generalizados.

$$\beta = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} \pi$$

Con una matriz Σ diagonal, con elementos genéricos dados por [Ecuación vi.4] donde π es el vector de probits muestrales. Como la matriz Σ es desconocida, hay que estimarla, para lo que se podría utilizar: a) las frecuencias observadas p_i , o bien b) las predicciones P_i obtenidas a partir de un modelo de probabilidad lineal previamente estimado.

Estimación de máxima verosimilitud para observaciones individuales

El procedimiento de estimación MV es preciso cuando no es posible agrupar las observaciones según los valores del vector X_i . En tal situación, carece de sentido hablar de proporciones muestrales. En dichos casos, la estimación por MV evita los problemas ya citados acerca de la estimación MCG del modelo lineal de probabilidad. Por otra parte el estimador de MV es eficiente, y se calcula sobre el modelo original, sin necesidad de ninguna aproximación.

En el caso del MODELO PROBIT, la función de verosimilitud sería:

Reemplazando [Ecuación iv.9] en [Ecuación vi.1]

$$L = \prod_1^N [\Phi(\beta' x_i)]^{y_i} [1 - \Phi(\beta' x_i)]^{1-y_i}$$

Nótese que para cada individuo i el término correspondiente en la función de verosimilitud es simplemente $\Phi(\beta' x_i)$ o $1 - \Phi(\beta' x_i)$, dependiendo de si $Y = 1$, jefe de hogar pobre, ó $Y = 0$, jefe de hogar no pobre.

Por tanto la función logaritmo de la verosimilitud se obtiene del logaritmo de la expresión anterior, y tomando sus derivadas con respecto al vector β se tienen las k condiciones necesarias de optimalidad:

$$\sum_1^N Y_i \frac{\phi_i}{\Phi_i} x_i + \sum_1^N (1 - Y_i) \frac{-\phi_i}{1 - \Phi_i} x_i = 0_k$$

ECUACIÓN VI.5

ó

$$S(\beta) = \sum_1^N \frac{Y_i - \Phi(\beta'x)}{\Phi(\beta'x)[1 - \Phi(\beta'x)]} \phi(\beta'x) x_i = 0$$

ECUACIÓN VI.6

donde $S(b)$ denota el vector gradiente de la función de verosimilitud. Si derivamos de nuevo en la expresión anterior con respecto al vector b , se obtiene la matriz Hessiana, y tomando esperanza en esta y cambiando de signo se obtiene finalmente la matriz de información, $I(b)$:

$$I(\beta) = \sum_1^N \frac{[\phi(\beta'x)]^2}{\Phi(\beta'x)[1 - \Phi(\beta'x)]} x_i x_i'$$

ECUACIÓN VI.7

Conviene hacer hincapié en que en las expresiones anteriores N denota el número total de observaciones, por lo que prescindiendo de clasificaciones, hay que considerar un sumando para cada observación muestral. En particular, en estos problemas es más sencillo utilizar el método del scoring¹⁰, razón por la que hemos calculado directamente la matriz de información a partir de la matriz de derivadas segundas de la función de verosimilitud con respecto al vector b . La inversa de la matriz de información será además la matriz de covarianzas del estimador de MV del vector b . El procedimiento de estimación de MV utilizaría:

$$I(\beta) = \sum_1^N \frac{[\phi(\beta'x)]^2}{\Phi(\beta'x)[1 - \Phi(\beta'x)]} x_i x_i'$$

ECUACIÓN VI.8

que proporciona la corrección que hay que introducir en el estimador del vector b en cada iteración. Al sustituir las expresiones

de $I(b)$ y $S(b)$ antes obtenidas puede verse fácilmente que si se hace el cambio de variables:

$$x_{ij}^* = \frac{x_{ij} \phi(\beta'x)}{\sqrt{\Phi(\beta'x)[1 - \Phi(\beta'x)]}},$$

$j=1,2,\dots,K$ ECUACIÓN VI.9

que forma, para cada observación i , un vector de dimensión k , e:

$$y_i^* = \frac{y_i - \Phi(\beta'x)}{\sqrt{\Phi(\beta'x)[1 - \Phi(\beta'x)]}}$$

ECUACIÓN VI.10

entonces la corrección a introducir en el estimador $\hat{\beta}_{n-1}$ coincide con los coeficientes estimados por mínimos cuadrados ordinarios en una regresión que utilizase y_i^* como variable a explicar, y x_i^* como vector de variables explicativas, utilizando los $\hat{\beta}_{n-1}$ para calcular x_{ij}^* y y_i^* .

MODELO LOGIT

Estimación de máxima verosimilitud para observaciones individuales.

La función de verosimilitud muestral es: Reemplazando [ECUACION IV.4] en [ECUACION VI.1]

$$L = \prod_{Y_i=1} F(\beta'x) \prod_{Y_i=0} [1 - F(\beta'x)] = \frac{e^{\sum_1^N Y_i(\beta'x)}}{\prod_1^N [1 + e^{\beta'x}]}$$

o, lo que es lo mismo:

$$\ln L = \sum_1^N Y_i(\beta'x) - \sum_1^N \ln(1 + e^{\beta'x}) =$$

^{10/} Especialmente diseñado para el caso en que se pretende obtener el EMV, este algoritmo se basa en la propiedad de que la esperanza matemática de la matriz hessiana de la función de verosimilitud (es decir, la matriz de información cambiada de signo). Así se ha sugerido como aproximación, sustituir la matriz de derivadas segundas por la matriz de información, teniéndose el llamado algoritmo de "scoring" $\hat{\theta}_s = \hat{\theta}_{s-1} + [I(\hat{\theta}_{s-1})]^{-1} \nabla \ln L(\hat{\theta}_{s-1})$

$$\sum_1^N (Y_i \gamma) \beta - \sum_1^N \ln(1 + e^{\beta' x})$$

ECUACIÓN VI.11

y denotando por $z' = \sum_1^N Y_i x_i'$ un vector fila $1 \times k$ se tiene:

$$\ln L = z' \beta - \sum_1^N \ln[1 + e^{\beta' x}]$$

y

$$S(\beta) = \frac{\partial \ln L}{\partial \beta} = z - \sum_1^N \frac{e^{\beta' x} x}{1 + e^{\beta' x}} = 0$$

ECUACIÓN VI.12

y este sistema de k ecuaciones no lineales debería, en principio, resolverse por procedimientos numéricos, para obtener el vector de estimaciones b . La matriz de información es:

$$I(\beta) = \sum_1^N \frac{e^{\beta' x} x x'}{1 + e^{\beta' x}} = \sum_1^N x P_i (1 - P_i) x'$$

ECUACIÓN VI.13

Para estimar el valor b por el algoritmo del "scoring" se comienza de un estimador b_0 y se actualiza por medio de:

$$\beta_1 = \beta_0 + [I(\beta_0)]^{-1} S(\beta_0)$$

En realidad, la matriz $S(b)$ puede escribirse también:

$$S(\beta) = \sum_1^N Y_i x_i - \sum_1^N \frac{x_i}{1 + e^{-\beta' x}} = \sum_1^N (Y_i - P_i) x_i$$

donde $P_i = \frac{1}{1 + e^{-\beta' x}}$, por lo que el algoritmo puede describirse como sigue:

1. A partir de un estimador inicial $\hat{\beta}_0$, calcular $\hat{P}_i(1 - \hat{P}_i)$.
2. Transformar las variables:

$$x_i^* = x_i \sqrt{\hat{P}_i(1 - \hat{P}_i)}$$

ECUACIÓN VI.14

$$Y_i^* = \frac{(Y_i - \hat{P}_i)}{\sqrt{\hat{P}_i(1 - \hat{P}_i)}}$$

ECUACIÓN VI.15

y el cambio a introducir en el vector $\hat{\beta}_0$ viene dado por los coeficientes estimados por mínimos cuadrados ordinarios en una regresión Y_i^* sobre el vector x_i^* .

El algoritmo se itera hasta conseguir su convergencia, y se utiliza la inversa de la matriz de información evaluada en el último estimador obtenido como estimación de la matriz de covarianzas de b . Por otra parte, los métodos de inferencia que consideran esta matriz de covarianzas son válidos, ya que el estimador máximo verosímil resultante tiene distribución normal asintótica. Las probabilidades de que un individuo con características X_i escoja la acción o se sitúe dentro del estado que hemos catalogado como $Y_i = 1$, (estar en condición de pobreza) se estiman mediante la expresión:

$$\hat{P}_i = \frac{e^{\beta' x}}{1 + e^{\beta' x}}$$

Luego de mostrar como la teoría estadística de ambos modelos propone su desarrollo, es aquí donde empezamos el trabajo de estudio empírico del fenómeno de pobreza en su conjunto. Las variables a considerar fueron recopiladas de un seguimiento de investigaciones las cuales existen en abundancia acerca del tema de la pobreza¹¹.

11/ La metodología desarrollada para la inclusión de variables se encuentra en el anexo metodológico al final de esta investigación.

VII. LA ENCUESTA NACIONAL DE HOGARES

La Encuesta Nacional de Hogares (ENAH), es un programa continuo de encuestas, que inició el Instituto Nacional de Estadística e Informática (INEI) a través de la Dirección Nacional de Censos y Encuestas el año 1995.

A partir del año 1997, el INEI ha puesto en ejecución el Programa de Mejoramiento de Encuestas y de la Medición de las Condiciones de Vida (MECOVI), bajo el auspicio financiero y técnico del Banco Interamericano de Desarrollo (BID), Banco Mundial (BM) y la Comisión Económica para América Latina y el Caribe (CEPAL), con el propósito central de fortalecer y mejorar el Sistema de Encuestas de Hogares, constituido por un conjunto de encuestas que se vienen realizando trimestralmente, y que representan una de las principales fuentes de información para el análisis, evaluación y seguimiento de la realidad demográfica, social y económica de la población peruana.

El Empleo y el Ingreso son módulos de seguimiento en todos los trimestres, pues son considerados los pilares para explicar los cambios en las condiciones de vida.

En el segundo y cuarto trimestre se efectúan las Encuestas Panel con el fin de estudiar los cambios en las características de la población en el tiempo.

En el marco de los nuevos lineamientos de política de gestión para identificar la demanda real de información y en coordinación con los usuarios el INEI ha identificado la necesidad de contar principalmente con indicadores de empleo y condiciones de vida que permitan cumplir con su principal misión.

OBJETIVOS

Objetivos Generales:

La encuesta del cuarto trimestre del 2001 tiene los objetivos generales siguientes:

- i. Generar indicadores anuales, que permitan conocer la evolución de la pobreza, el bienestar y las condiciones de vida de los hogares.
- ii. Efectuar diagnósticos (anuales) sobre las condiciones de vida y pobreza de la población.
- iii. Medir el alcance de los programas sociales en la mejora de las condiciones de vida de la población.
- iv. Servir de fuente de información a instituciones públicas y privadas, así como a investigadores.
- v. Permitir la comparabilidad con investigaciones similares en relación a las variables investigadas.

Objetivos Específicos

Los objetivos específicos de la ENAHO 2001 del cuarto trimestre son los siguientes:

- i. Determinar el grado de acceso a servicios básicos de la vivienda y del hogar y de los programas sociales orientados a la vivienda según diferentes estratos socio-económicos.
- ii. Obtener indicadores de riesgo de salud debido al hacinamiento y las condiciones sanitarias de los hogares según diferentes estratos socio-económicos.
- iii. Caracterizar las estructuras demográficas, según diferentes estratos socio-económicos, con el fin de medir la evolución y el impacto demográfico de los programas sociales.
- iv. Determinar el nivel educativo según diferentes estratos socio-económicos.
- v. Determinar el grado de acceso a la educación y el alcance de los programas sociales, según diferentes estratos socio-económicos.
- vi. Caracterizar los niveles de empleo, según diferentes estratos socio-económicos.
- vii. Determinar la estructura del ingreso de los hogares, teniendo en cuenta los ingresos provenientes de los propios hogares y el efecto redistributivo de los programas sociales.
- viii. Cuantificar el gasto de consumo de los hogares diferenciando el aporte de los

programas sociales según diferentes estratos socio-económicos.

- ix. Obtener información sobre morbilidad y acceso a los servicios de salud.
- x. Evaluar el grado de conocimiento y utilización de los Programas Sociales y Proyectos de Inversión Social.
- xi. Caracterizar a los hogares en pobreza extrema, pobres y no pobres en función a variables demográficas, educativas, otras sociales y económicas y el grado de acceso a los servicios

TEMAS A INVESTIGAR

- Módulo de Vivienda
- Características de los miembros del hogar
- Módulo básico de Educación
- Módulo básico de Salud
- Módulo básico de empleo
- Sistema de Pensiones
- Uso de Computadora e Internet en el Trabajo
- Ingresos del sector formal e informal
- Gastos
- Módulo básico de Programa Social
- Módulo Comunal para Informantes Calificados
- Módulo de Opinión

CARACTERISTICAS

La Investigación se desarrollará sobre la base de una muestra de hogares siendo los niveles de inferencia del diseño muestral: Nacional, Urbano Nacional, Rural Nacional, Resto Costa, Sierra, Selva y el Area Metropolitana de Lima y Callao.

Los cuestionarios a emplearse serán los mismos del Cuarto Trimestre de 2000, además de un módulo de opinión.

Los informantes serán todos los residentes habituales de 12 años y más de edad que conforman el hogar entrevistado.

La **Unidad de Investigación**: es el hogar, el cual está constituido por: 1) los integrantes del hogar familiar, 2) los trabajadores del hogar con cama adentro, reciban o no pago por sus servicios, 3) los integrantes de una pensión familiar que tienen como máximo 9 pensionistas, y 4) las personas que no son miembros del hogar familiar pero que estuvieron presentes en el hogar los últimos 30 días.

No serán investigados: 1) los integrantes de una pensión familiar que tiene de 10 a más pensionistas, y 2) los trabajadores del hogar con cama afuera.

7.1. CARACTERÍSTICAS DE LA MUESTRA

Población y Cobertura

El universo cubierto por la muestra de la ENAHO es todo el territorio nacional. Es decir, la población está definida como el conjunto de todas las viviendas particulares y sus ocupantes residentes del área urbana y rural del país.

Se excluye del estudio a la población residente en viviendas tipo colectivas como hospitales, cuarteles, comisarías, hoteles, centros de reclusión, etc.

Niveles de Inferencia de Resultados

A efectos de permitir el estudio de los cambios en las características de la

población en el tiempo, se ha considerado que en la ENAHO 2001 Cuarto Trimestre se trabaje con una muestra del tipo Panel y una muestra No Panel.

La muestra panel estará conformada por los hogares entrevistados en la ENAHO 2000 - Cuarto Trimestre. En el caso de la muestra no panel, esta será totalmente nueva.

Muestra panel: El principal objetivo del uso de una muestra panel en una encuesta, es realizar un seguimiento de las unidades de investigación, en este caso los hogares y los miembros que habitan en ella en un determinado período. Asimismo, esta muestra permite obtener estimaciones de las características socio-demográficas de la población para diferentes áreas, estratos o dominios de interés, para posteriormente realizar comparaciones de las unidades investigadas con referencia a la anterior investigación.

Muestra no panel: A través de la muestra No Panel, se puede obtener estimaciones de las características socio-demográficas de la población de estudio para diferentes áreas, estratos o dominios de interés. Además, esta muestra incluye las nuevas unidades estadísticas que se incrementan en el marco inicial de selección.

Metodología de Estimación

La metodología de estimación para procesar los datos de la ENAHO, involucra el uso de un peso o factor de expansión para cada registro que será multiplicado por todos los datos que conforman el registro correspondiente.

El factor final para cada registro tiene dos componentes:

El factor básico de muestreo y los factores de ajuste por la no entrevista.

El factor básico de expansión para cada hogar muestral es determinado por el diseño de la muestra. Equivale al inverso de su probabilidad final de selección, el mismo que es el producto de las probabilidades de selección en cada etapa.

El diseño de la muestra de la ENAHO, involucra hasta 3 etapas de muestreo donde las unidades son seleccionadas con probabilidades proporcionales al tamaño (ppt) excepto la última etapa. En la última etapa se seleccionará un número de viviendas para cada conglomerado teniendo en cuenta un intervalo de selección.

Errores de Muestreo

Trimestralmente, en la Encuesta Nacional de Hogares se calculan los errores de muestreo de las estimaciones de las principales variables investigadas en la encuesta.

El paquete estadístico utilizado en la ENAHO para el cálculo de las varianzas es el CENVAR (Sistema de Cálculo de Varianzas), el cual provee los estimadores de variabilidad muestral para parámetros poblacionales, como: totales, medias, razones y proporciones para los diferentes dominios de estimación.

Para cada parámetro especificado y dominio de estimación, CENVAR produce un cuadro de salida con los indicadores siguientes:

- El valor estimado del parámetro (estimación puntual)
- El error estándar

- El coeficiente de variación (CV)
- El intervalo con 95 por ciento de confianza
- El efecto del diseño (DEFT)
- El número de observaciones sobre el cual se basa la estimación

El algoritmo usado por el CENVAR se basa en el método de los estimadores de la varianza de los conglomerados últimos.

Cuestionarios

Se emplearán 6 tipos de cuestionarios:

ENAHO 01. Cuestionario individual para ser llenado con información del jefe del hogar y con entrevista directa a los informantes individuales. Comprende las características de la vivienda, del hogar y de los miembros del hogar, Gastos del Hogar, Programas Sociales y Otras Transacciones.

ENAHO 01A. Cuestionario individual para ser llenado con información del jefe del hogar y con entrevista directa a los informantes individuales, investiga las características de Educación, Salud, Empleo e Ingreso, Sistema de Pensiones y Uso de Computadora e Internet en el Centro de Trabajo.

ENAHO 01B. Cuestionario individual que es llenado por entrevista directa con información del Jefe del hogar, en este Módulo de Opinión se investiga Nivel de Vida/Situaciones Adversas, Participación ciudadana, Percepción sobre la comunidad, Seguridad y Violencia, ETNIA/RAZA y Educación de los Padres.

ENAHO 02. Cuestionario individual que es llenado por entrevista directa con cada productor agropecuario que conduce una

unidad agropecuaria. Se investiga los ingresos del productor agropecuario, el régimen de tenencia de las tierras, la posesión de títulos de propiedad, el destino de la producción y crianza, los gastos realizados en la actividad agropecuaria y el crédito agropecuario.

ENAH0 03. Cuestionario aplicado a informantes calificados del área rural o áreas periféricas de las ciudades, investiga acerca del acceso a servicios y programas sociales, así como a los organismos que financian dichos programas.

ENAH0 04. Ingreso del Trabajador Independiente (Sector Informal).

7.2. Factores de relevancia para la explicación de la pobreza extraídos de la ENAH0 IV trimestre 2001

Variable dependiente

[CONDICIÓN DE POBREZA]

pobreza Pobreza

Escala Nominal Categórica

Codificación: 1 Pobre Extremo, 2 Pobre No extremo, 3 No Pobre

Base de datos: Sumaria

Esta variable es recodificada en una nueva variable Epobre

Escala Nominal Dicotómica

Codificación: 0 No Pobre, 1 Pobre

Variables independientes

Características Sociodemográficas del Jefe de Hogar

[NIVEL DE EDUCACIÓN]

p301 Nivel educativo que aprobó

Conocer el grado de educación más alto aprobado por cada persona dentro del nivel educativo que alcanzó.

Escala Ordinal

Base de datos: Educación (CAP. 300)

Codificación: 1 Sin nivel, 2 Inicial, 3 Primaria incompleta, 4 Primaria completa, 5 Secundaria incompleta, 6 Secundaria completa, 7 Sup. No Univ. Incompleta, 8 Sup. No Univ. Completa, 9 Sup. Univ. Incompl., 10 Sup. Univ. Completa, 11 Post-grado Universitario, 99 Missing value

La variable nivel educativo que aprobó (p301) fue recodificada en la variable **adnivedu**-Nivel Educativo aprobado. Escala ordinal.

Codificación: 1 Sin nivel, 2 Primaria, 3 Secundaria, 4 Sup. No Univ., 5 Sup. Univ. 6 Post- Grado Univ.

Se creó la variable **aest**-Años de Estudios totales. Para su construcción se consideró las variables p301 y p301b(años de estudios que aprobó). Como es evidente, dependiendo del grado aprobado y los años que aprobó, se generan la cantidad total de años estudiados. Ejem: Si es un jefe de hogar que llegó solo a secundaria incompleta (3 años aprobados), entonces los años de estudios totales ($aest = 6$ (primaria completa) + 3 (secundaria incompleta)).

[CENTRO DE ESTUDIOS]

p301d Centro de Estudios

Determinar si el colegio de procedencia donde se estudio es de régimen estatal o particular.

Escala nominal

Base de datos: Educación (CAP. 300)

Codificación: 1 Estatal, 2 No estatal, 9 Missing value

[EDAD]

p208a Edad en años

Determinar el período transcurrido entre la fecha de nacimiento de la persona y el día de la entrevista.

Variable cuantitativa

Base de Datos: Características de los miembros del hogar (CAP. 200)

[GÉNERO]

p207 Sexo

Determinar el sexo de los entrevistados. Identificados como hombre o mujer.

Escala Nominal Dicotómica

Base de Datos: Características de los miembros del hogar (CAP. 200)

Codificación: 1 Hombre, 2 Mujer

[ESTADO CIVIL]

p209 Estado conyugal

Investigar sobre la naturaleza de la participación de la población en materia de comportamiento conyugal en especial durante el período de madurez.

Escala Nominal Categórica

Base de Datos: Características de los miembros del hogar (CAP. 200)

Codificación: 1 Conviviente, 2 Casado(a), 3 Viudo(a), 4 Divorciado(a), 5 Separado(a), 6 Soltero(a)

[DOMINIO GEOGRÁFICO]

dominio Dominio Geográfico

Escala Nominal Categórica

Base de Datos: Identificador presente en todas las bases de datos

Codificación: 1 Costa Norte, 2 Costa Centro, 3 Costa Sur, 4 Sierra Norte, 5 Sierra Centro, 6 Sierra Sur, 7 Selva, 8 Lima Metropolitana

[GRANDES DOMINIOS]

gdomini Grandes Dominios Geog.

Escala Nominal Categórica

Codificación: 1 Costa, 2 Sierra, 3 Selva, 4 Lima Metropolitana

[ESTRATO]

estrato Estrato Geográfico

Determinar la cantidad de viviendas albergadas en cada estrato geográfico.

Escala Nominal Categórica

Base de Datos: Identificador presente en todas las bases de datos

Codificación: 1 Mayor de 100,000 viviendas, 2 De 20,001 a 100,000 viviendas, 3 De 10,001 a 20,000 viviendas, 4 De 4,001 a 10,000 viviendas, 5 De 401 a 4,000 viviendas, 6 Menos de 400 viviendas, 7 AER compuestos, 8 AER simples.

[AREA]

urb_ru Área urbano-rural

Esta variable resulta de recodificar la variable estrato cuyas 5 primeras categorías proceden a conformar el área urbana y las categorías de 6 a la 8 conformarían el área rural.

Escala Nominal Dicotómica

Codificación: 1 Área Urbana, 2 Área Rural

[DOMINIO URBANO RURAL]

domur_ru Grandes dominios por área urb o ru.

Esta variable resulta del cruce de las variables gdomini (Costa, Sierra, etc) y área (urbano, rural).

Escala Nominal categórica

Codificación: 11 Costa Urbana, 12 Costa Rural, 21 Sierra Urbana, 22 Sierra Rural, 31 Selva Urbana, 32 Selva Rural, 41 Lima Urbana

Características de la inserción ocupacional del JH

[EXPERIENCIA LABORAL]

exper1 Indicador proxy de exper. laboral

Este indicador resulta de restar a la edad del jefe de hogar los años de estudios,

aunque no considera si el jefe de hogar ha trabajado y estudiado a la vez, se aproxima en buena medida a la experiencia laboral del jefe de hogar. Se resta "6" al final del indicador, pues es a la edad en que debería empezar el colegio. Variable cuantitativa

Por ejemplo, un jefe de hogar con 45 años y secundaria completa, tendrá:

$$\text{Exper1} = \text{p208a} - \text{aest} - 6 = 45 - (6+5) - 6 = 28 \text{ años de experiencia laboral}$$

[CATEGORÍA OCUPACIONAL]

p507 Cargo en Centro Ocupacional. Conocer la relación del trabajador con su empleo, es decir, la forma que tienen los trabajadores de insertarse en el mercado laboral. Una primera distinción básica es si se trabaja en forma dependiente o independiente, ya que en ambos casos son muy diferentes las relaciones económicas y laborales involucradas.

Escala Nominal Categórica

Base de datos: Empleo e Ingresos (Cap. 500)

Codificación: 1 Empleador o patrono, 2 Trabajador independiente, 3 Empleado, 4 Obrero, 5 Trabajador Familiar no remunerado, 6 Trabajador del Hogar, 7 Otro.

[TAMAÑO DE LA FIRMA]

p512a Tamaño de la empresa. Disponer de información sobre el volumen de la fuerza de trabajo según el tamaño del establecimiento, el mismo que conjuntamente con otras características, son útiles para determinar el volumen de la población del sector informal.

Escala Ordinal

Base de datos: Empleo e Ingresos (Cap. 500)

Codificación: 1 menos de 100 personas, 2 De 100 a 499 personas, 3 De 500 y más personas.

[TRABAJO ADICIONAL]

p514 Ocupación secundaria. Determinar el número de ocupados que tienen actividad secundaria, que desarrollan simultáneamente con la actividad principal en la semana de referencia, ya sea en forma dependiente o independiente.

Escala Nominal Dicotómica

Base de datos: Empleo e Ingresos (Cap. 500)

Codificación: 1 Si, 2 No

Características del hogar

[NUCLEOS EN EL HOGAR]

nuchoga Cant. de núcleos en hogar. Determinar el número de núcleos presentes en cada hogar para intentar captar a cuántos hogares se alberga fuera del propio hogar del jefe de hogar, dentro de la misma vivienda.

Variable cuantitativa

Base de datos: Generada a partir de características de los miembros del hogar (cap.200)

[HIJOS EN EL HOGAR]

hijxhog Cant. de hijos del jh. Determinar la cantidad de hijos presentes en el hogar.

Variable cuantitativa

Base de datos: Generada a partir de características de los miembros del hogar (cap.200)

[MIEMBROS EN EL HOGAR]

mieperhog Cant. de miembros en hogar. Determinar la cantidad de miembros pertenecientes al hogar, se excluyen a las

trabajadoras del hogar que no permanecen más de 30 días en el hogar.

Variable Cuantitativa

Base de Datos: Sumaria

[PERCEPTORES EN HOGAR]

percep hog Cant. de preceptores en hogar.

Determinar la cantidad de preceptores dentro de cada hogar.

Variable Cuantitativa

Base de Datos: Sumaria

[INGRESO NETO TRIMESTRAL]

inghog2d Ingreso neto trimestral
Trimestralizar la información de ingresos, llevándola a un mismo período de investigación.

Variable Cuantitativa

Base de Datos: Sumaria

[INGRESO PER CÁPITA M]

ingper hog Ingreso per cápita mens. del hogar

Determinar el ingreso per cápita mensual de los hogares.

Variable Cuantitativa

Base de datos: Generada a partir de Sumaria.

[ACCESO LUZ]

p112 Tipo de alumbrado en su hogar
Determinar si el hogar dispone o no de alumbrado eléctrico. Asimismo se conocerá la cantidad de hogares que carecen de servicio y permitirá establecer la relación que existe entre padecimientos respiratorios y algunas formas de alumbrado como el uso de kerosene y vela. Asimismo, las deficiencias en la agudeza visual, también pueden estar en relación al uso de alumbrado no eléctrico.

Escala Nominal Categórica

Base de datos: Características de la vivienda y el hogar. (cap. 100)

Codificación: 1 Electricidad, 2 kerosene (mechero/lamparin), 3 Petróleo/gas (lampara), 4 Vela, 5 Generador, 6 Otro.

[ACCESO AGUA]

p110 Abastecimiento agua en hogar
Conocer la cantidad de personas y lugares donde se carece de este líquido vital, lo que será de utilidad para la ejecución de programas de saneamiento.

Escala Nominal Categórica

Base de datos: Características de la vivienda y el hogar. (cap. 100)

Codificación: 1 Red pública, dentro de la vivienda, 2 Red pública, fuera de la vivienda, 3 Pílon de uso público, 4 Camión -cisterna u otro similar, 5 Pozo, 6 Río, acequia, manantial o similar, 7 otro.

[ACCESO A RED SANITARIA]

p111 El servicio higiénico está conectado

Conocer si el hogar dispone o no de servicio higiénico, su ubicación y forma de eliminación de los residuos humanos. Permitirá tener una apreciación más general respecto a la calidad de vida de los habitantes; y es un componente elemental en la determinación del bienestar social. Además permitirá conocer las posibilidades de contaminación en la comunidad en general y las causas que producen la hepatitis y la tifoidea.

Escala Nominal Categórica

Base de datos: Características de la vivienda y el hogar. (cap. 100)

Codificación: 1 Red pública dentro de la vivienda, 2 red pública fuera de la vivienda, 3 pozo séptico, 4 pozo ciego o negro/letrina, 5 río, acequia o canal, 6 no tiene

Si se desea obtener más información sobre otras variables de medición de condiciones de vida y pobreza, se puede consultar "Variables investigadas en la ENAHO-2001 IV TRIMESTRE .INEI - MECOVI"

VARIABLES INDICADORAS

Se generaron variables indicadoras para trabajar sobre los modelos de probabilidad del jefe de hogar y poder captar el efecto de las categorías de las variables en escalas nominales sobre sus categorías base. El mismo proceder se efectuó para las variables en escalas ordinales.

En tal sentido el nivel educativo agrupado fue dividido en 5 indicadoras, donde la categoría que está fuera de riesgo es el Post-Grado Univ.

- DSINNIV = 1
Si JH no tiene nivel educativo y 0 en otro caso
- DPRIM = 1
Si JH tiene primaria y 0 en otro caso
- DSECUND = 1
Si JH tiene secundaria y 0 en otro caso
- DSUPNU = 1
Si JH tiene educ. sup no univ. y 0 en otro caso
- DSUPUN = 1
Si JH tiene educ. sup. Univer. y 0 en otro caso

Variable indicadora de colegio estatal.

- DCOLEGIO = 1
Si JH a estudiado en colegio estatal y 0 en otro caso

La variable categoría ocupacional generaría 4 variables categóricas, donde se considera

que la categoría empleador o patrono es la categoría base.

- DINDEP = 1
Si JH es trabajador independiente y 0 en o.c.
- DEMPLEADO = 1
Si JH es empleado y 0 en o.c.
- DOBRERO = 1
Si JH es obrero y 0 en otro caso
- DOTRO = 1
Si JH se encuentra en otra situación ocupacional

Se generan variables indicadoras para dominio en área urbana y rural.

- DCORU = 1
Si JH habita en Costa urbana y 0 en otro caso
- DSIEUR = 1
Si JH habita en Sierra urbana y 0 en otro caso
- DSELUR = 1
Si JH habita en Selva urbana y 0 en otro caso

La variable en escala ordinal Tamaño de la firma genera dos variables indicadoras, tomando como categoría base el tamaño de la firma de 500 y más personas.

- DME100 = 1
Si JH trabaja empresa con menos de 100 personas y 0 o.c.
- DME499 = 1
Si JH trabaja empresa con 100 a 499 personas y 0 o.c.

Variable indicadora de si el JH tiene ejerce una actividad secundaria

- DOEMP = 1
Si JH tiene otro trabajo y 0 en otro caso

De igual manera el estado civil generaría 5 variables indicadoras, donde tomaríamos como categoría base cuando el jefe de hogar es soltero.

- DCONVI = 1
Si JH es conviviente y 0 en otro caso
- DCASAD = 1
Si JH es casado y 0 en o.c.
- DVIUDO = 1
Si JH es viudo y 0 en o.c.
- DDIVOR = 1
Si JH es divorciado y 0 en o.c.
- DSEPAR = 1
Si JH es separado y 0 en o.c.

Generamos variables indicadoras cuando el hogar tiene conexión de servicios higiénicos a red pública dentro de vivienda.

- DSHDV = 1
Si SS.HH. hogar - conectados a red pública dentro de vivienda y 0 en o.c.

De la misma manera, se generan una variable indicadora cuando el hogar no tiene ss.hh. dentro del hogar.

- DSHNO = 1
Si hogar no tiene SS.HH. dentro de hogar y 0 en o.c.

Para el caso de abastecimiento de agua dentro de la vivienda a través de red pública, generamos la variable indicadora.

- DAGUDV = 1
Si abastecimiento de agua es través de red pública dentro de vivienda

Variable indicadora que defina si el alumbrado público en la vivienda es través de kerosene.

- P1122 = 1
Si el tipo de alumbrado en la vivienda es a través de kerosene

Variable indicadora que define si se utiliza espacio de la vivienda que destine ingresos para el hogar.

- P115 = 1
Si se utiliza espacio en la vivienda que destine ingresos al hogar.

VIII. CRITERIOS DE APLICACIÓN DE LOS MODELOS LOGIT Y PROBIT CON VDD APLICADOS AL CASO DE LA POBREZA EN EL PERÚ

8.1 Estudio a nivel descriptivo y exploratorio de variables cualitativas y cuantitativas que incidan en la pobreza en el Perú¹²

Empezaremos esta parte de nuestra investigación describiendo aquellas variables que nos brindan posibles explicaciones de la pobreza en un marco general y de los individuos Jefe de Hogar en este caso en particular. Entre ellas se encuentran aquellas de carácter sociodemográfico como el género, la edad, el estado civil, el nivel de educación, el alfabetismo (variables de naturaleza cualitativa), medidas en algunos casos en escala 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).

En otro tipo de escalas se encuentran aquellas que denotan un orden ascendente en sus categorías como lo es el nivel de educación (Sin Nivel- Educación Primaria-etc.) alcanzado por el Jefe de Hogar.

Además encontramos variables de naturaleza cuantitativa como lo constituyen

la edad y el indicador proxy de experiencia laboral , que no considera aquel caso en el que el Jefe del Hogar, ha estudiado y trabajado, sino solo aquel tiempo en el cual el Jefe de Hogar solo trabaja.

Uno de los objetivos que perseguimos y que esta implícito dentro de esta investigación es dejar en claro que estamos trabajando con variables a diferentes escalas, mencionadas como ejemplos líneas antes, y que dependiendo de estas escalas determinarán la implementación y ejecución de las metodologías que sirvan de herramientas para una explicación de la pobreza en el Perú en particular.

EDUCACIÓN

Una de las características de la pobreza es que esta se encuentra asociada a niveles bajos de educación alcanzados por el Jefe de Hogar. Podemos ver [Cuadro VIII.1] como dentro de aquellos Jefes de Hogar sin nivel educativo y nivel educativo primario, el 72.5% y el 61.3% se encuentran en condición de pobreza, respectivamente.

Cuadro VIII.1 Perfil del Jefe de Hogar según Nivel Educativo y Estado pobreza

		Nivel Educativo Agrupado ^a						Total
		sin nivel	primaria	secundaria	sup. no univ.	universitaria	post univers.	
estado pobreza	pobre	72.5%	61.3%	40.0%	19.3%	8.2%	.2%	46.8%
	no pobre	27.5%	38.7%	60.0%	80.7%	91.8%	99.8%	53.2%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

a. Percentage de columnas

12/ Debemos de recordar que si bien el objetivo de nuestra investigación gira en torno a los modelos logit y probit, vemos que no podemos ser ajenos a aquellos previos fundamentales de análisis descriptivo para poder seguir avanzando sobre niveles de análisis más complejos, como lo podrían constituir el análisis multivariado y los modelos de probabilidad con enfoques binomial.

En el caso de los que tienen nivel educativo secundaria, se aprecia que el estado de pobreza se encuentra repartido para sus dos categorías, con 40% para el caso de jefes de hogar pobres y con 60% para el caso de no pobreza. Un panorama diferente y en sentido opuesto se da en los niveles superiores donde en el caso de jefes de hogar con nivel educativo superior no universitario el 80.7% se encuentran en estado de no pobreza. En una tendencia creciente mucho mayor y estado de no pobreza se encuentran aquellos jefes de hogar que tiene niveles como el universitario y el post-grado, con 91.8% y 99.8% para la primera y segunda respectivamente.

En función del comportamiento entre el estado de pobreza y el nivel educativo del jefe de hogar nos toca ahora analizar el grado de relación que estas tienen, y definir más aún este comportamiento a nivel poblacional. Analizando el cuadro que muestra el Test Chi-Square [CUADRO VIII.2] podemos ver que la hipótesis de nulidad que supone independiente el nivel educativo y la condición de pobreza del jefe de hogar es rechazada para un valor de $\chi^2 = 859730.5$ y un nivel de significancia (p-valor = 0.000), por lo tanto a medida que el jefe de hogar va adquiriendo un mayor nivel de educación, éste obtiene mas capacidades para ir abandonando la condición de pobreza, más aún si tiene niveles de educación superiores.

Cuadro VIII.2 Chi-Square Test

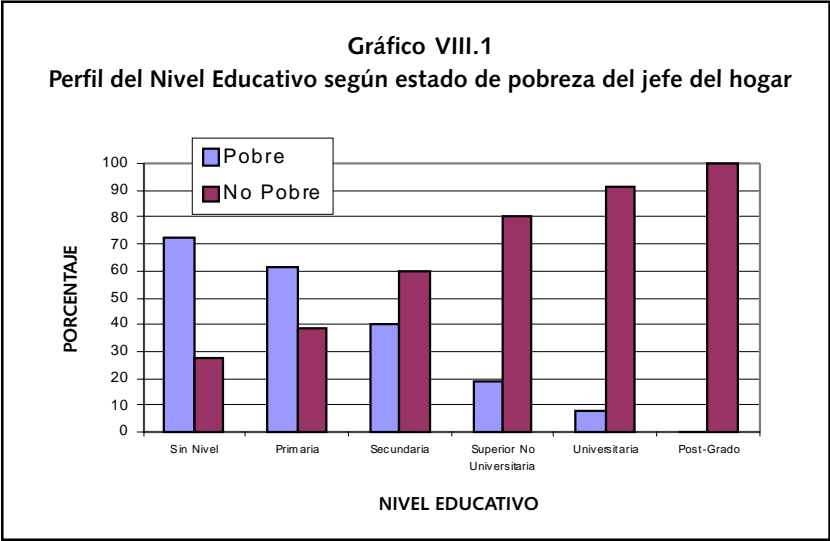
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	859730.5	5	.000
Likelihood Ratio	954693.3	5	.000
Linear-by-Linear Association	842469.1	1	.000
N of Valid Cases	5834837		

Una respuesta a la pregunta de ¿cual es el grado de asociación? Se puede apreciar en el cuadro [CUADRO VIII.3] Symmetric Measures donde los Coeficientes de correlación de Spearman (0.375), Gamma (0.567) y Kendall´s tau-c (0.409) denotan un fuerte grado de asociación, pues nos

dan la idea de la asociación que debe darse entre el nivel socioeconómico y el nivel educativo, hablando de la población en su conjunto. Debemos de tomar en cuenta que estamos aislando en alguna medida el fenómeno de pobreza y poniéndolo solo en términos del nivel educativo.

Cuadro VIII.3 Symmetric Measures

		Value	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	.346	.000
	Kendall's tau-c	.409	.000
	Gamma	.567	.000
	Spearman Correlation	.375	.000
N of Valid Cases		5834837	



ESTADO CIVIL - SEXO DEL JEFE DE HOGAR

A continuación analizaremos el comportamiento y la relación existente entre el sexo del jefe de hogar y el estado civil que posee. Se puede apreciar [CUADRO VIII.4] que dentro de la condición de pobreza, los jefes de hogar hombres constituyen el 96.4% del estado

civil conviviente, en igual sentido se encuentra los jefe de hogar casados que vienen a estar constituidos en un 97.7% por los hombres. La figura cambia para las otras categorías de estado civil, donde para el estado civil viuda, divorciada, separada y soltera, las mujeres constituyen el 76.36%, el 70.86%, el 84.13% y el 50.48, para cada categoría y en ese orden respectivamente.

Cuadro VIII.4
Perfil de jefe de hogar según género y estado civil o conyugal en estado de pobreza

estado pobreza			Qual essu estado civil o conyugal ^a						Total
			Conviviente	Casado	Viudo	Divorciado	Separado	Soltero	
pobre	Sexo	Hombre	96.4%	97.7%	23.6%	29.1%	15.9%	49.5%	81.2%
		Mujer	3.6%	2.3%	76.4%	70.9%	84.1%	50.5%	18.8%
	Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

a. Percentage en columnas

Resultado de la incorporación de los efectos de diseño vemos [CUADRO VIII.5] que según el Chi-Square Tets, obtenemos el coeficiente Pearson χ^2 (5) = 1685365 y un p-valor=0.000, nos

permite rechazar la hipótesis que formula la independencia de ambas variables, es decir, que existe relación significativa entre el estado civil y el sexo del jefe de hogar en condición de pobreza.

Cuadro VIII.5 Chi-Square Test

estado pobreza		Value	df	Asymp. Sig. (2-sided)
pobre	Pearson Chi-Square	1685365	5	.000
	Likelihood Ratio	1478553	5	.000
	N of Valid Cases	2764186		

El grado de asociación del cual estamos hablando lo podemos observar en el cuadro siguiente Symetric Measures [CUADRO VIII.6], donde coeficientes como el de contingencia (CC) con un valor de 0.615 , el Phi = 0.781 y el Cramer´s V=.781 denotan un alto grado de asociación¹³, lo cual nos muestra que la pobreza afecta en todos los estados civiles sean estos en

condición de pareja (casado-conviviente), como sería en el caso de los jefes de hogar hombres. En aquellas situaciones en las que la mujer tiene que responder como jefe de hogar y "no tendría pareja" (Viuda-Divorciada-Soltera-etc.), podría motivar un mayor riesgo para la adquisición de la condición de pobreza.

Cuadro VIII.6 Symmetric Measures

estado pobreza			Value	Approx. Sig.
pobre	Nom	Phi	.781	.000
	by	Cramer's V	.781	.000
	Nom	Contingency Coefficient	.615	.000
	N of Valid Cases		2764186	

Al analizar las mismas variables para los jefes de hogar no pobres, podemos observar [CUADRO VIII.7] un comportamiento parecido al de los jefes de hogar en estado de pobreza. Es decir, para los jefes de hogar con estado civil de convivencia y casado, los hombres

constituyen el 94.1% y el 97.2% dentro de cada categoría respectivamente. Para los jefes de hogar en los estados civiles viudo, divorciado y separado, las mujeres constituyen el 67.8%, el 60.2% y el 67.8% de cada categoría y en ese orden respectivamente.

Cuadro VIII.7

Perfil de jefe de hogar según género y estado civil o conyugal en estado de no pobreza

estado pobreza		Cual es su estado civil o conyugal ^a						Total
		Conviviente	Casado	Viudo	Divorciado	Separado	Soltero	
no pobre	Sexo							
	Hombre	94.1%	97.2%	32.2%	39.8%	32.2%	62.1%	78.2%
	Mujer	5.9%	2.8%	67.8%	60.2%	67.8%	37.9%	21.8%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

a. Porcentage de Columnas

13/ Debemos de considerar que si bien estos coeficientes nos confirman los grados de asociación de variables estos poseen limitaciones como el caso del coeficiente de contingencia, que a pesar que nos puede expresar que un valor de cero confirma la no asociación de variables, no posee la propiedad en el caso extremo de un grado de asociación total, es decir, igual a 1. Para más detalle consultar Siegel , Sidney.

Se confirma la relación significativa y diferente de cero existente entre el sexo y el estado civil del jefe de hogar, esto en el Chi-Square Test [CUADRO VIII.8], a través de los coeficientes de Pearson $\chi^2(5)=1356731$, el χ^2 de la razón de verosimilitud (likelihood ratio)= 1327980

con resultados en el mismo sentido debido al tamaño de la población. A través de su significancia (0.000) que permite rechazar la hipótesis que sostiene la independencia del sexo y el estado civil del jefe de hogar dentro de cada categoría de pobreza.

Cuadro VIII.8 Chi-Square Test

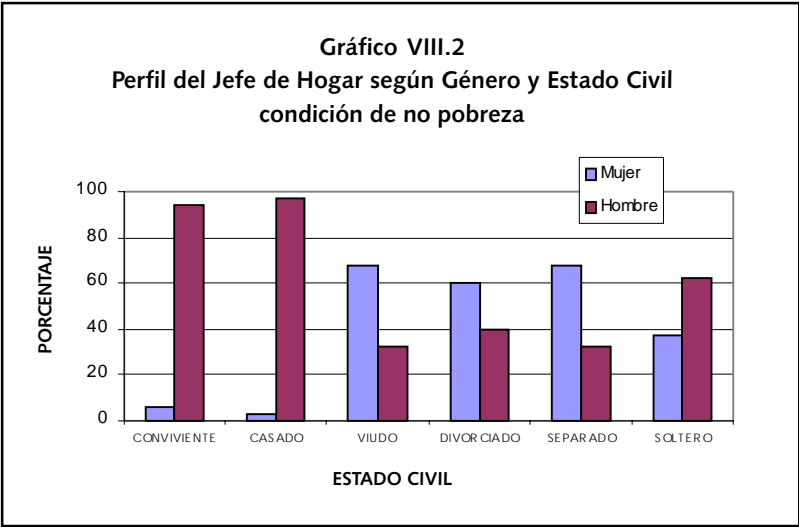
estado pobreza		Value	df	Asymp. Sig. (2-sided)
no pobre	Pearson Chi-Square	1356731	5	.000
	Likelihood Ratio	1327980	5	.000
	N of Valid Cases	3144673		

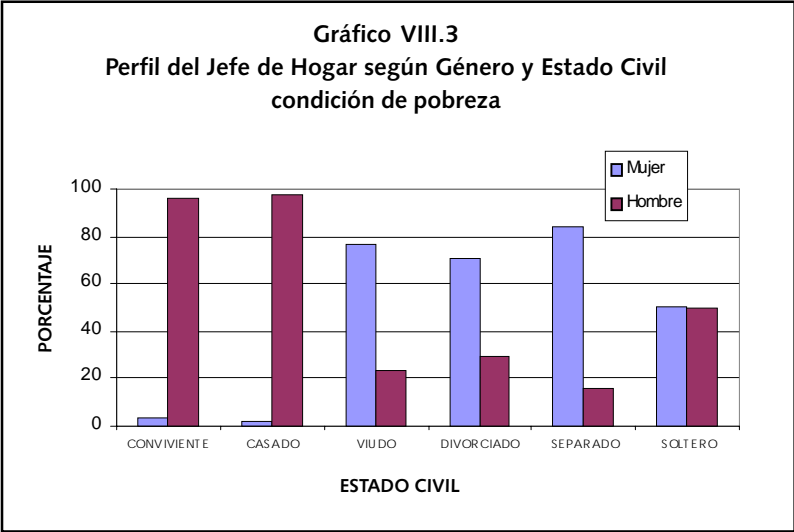
De lo explicado anteriormente, resulta interesante ver que a pesar que exista una relación entre el sexo del jefe de hogar y su correspondiente estado civil, la pobreza de los jefes de hogar del Perú ya no solo

capta a los hogares con jefes de hogar mujeres sin pareja, ahora en el año 2001 ya ha captado a los jefes de hogar con pareja, ya sea formal o informal.

Cuadro VIII.9 Symmetric Measures

estado pobreza			Value	Appro x. Sig.
no pobre	Nom by Phi		.657	.000
	Nom Cramer's V		.657	.000
	Contingency Coefficient		.549	.000
	N of Valid Cases		31 446 73	





AREAS DE RESIDENCIA - REGIÓN NATURAL

La pobreza en nuestros días ha conquistado todos los ámbitos del vasto territorio peruano, lo cual sumado a condiciones de focalización mal llevadas en la década pasada permitió que aquellos que se encontraban en estado de pobreza no fueran quienes recibieran la ayuda y los medios adecuados paleativos para subsistir.

En el año 2001, podemos ver [CUADRO VIII.10] que en condición de pobreza los

jefes de hogar del área urbana constituye el 84.3% para la región de la costa. En un nivel porcentual menor pero igual de considerable, la región de la sierra está constituida en un 75.9% por jefes de hogar del área rural. Si bien para el área rural en la región selvática los jefes de hogar pobres son alrededor del 58% , no es tan evidente la diferencia como en el resto de regiones, debido a que en el área urbana se encuentra un 42.1% en estado de pobreza.

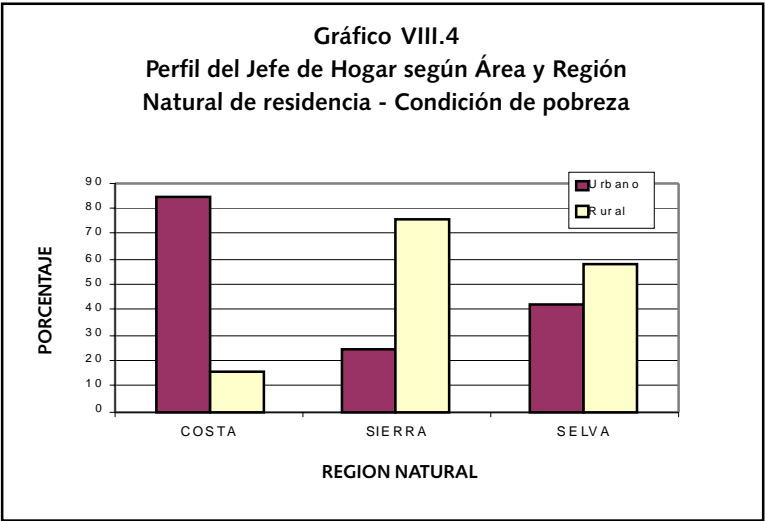
Cuadro VIII.10
Perfil de jefe de hogar según Area y Región Natural en estado de pobreza

estado pobreza			regiones naturales ^a			Total
			costa	sierra	selva	
pobre	Urbano_Rural	urbano	84.3%	24.1%	42.1%	47.4%
		rural	15.7%	75.9%	57.9%	52.6%
	Total		100.0%	100.0%	100.0%	100.0%

a. Porcentage de Columnas

Resulta dramático evidenciar cómo la condición de pobreza afecta en mayor medida a aquellos que menos acceso tienen en cuanto a activos públicos de infraestructura, educación, salud, etc; que son justamente aquellos jefes de hogar

que se encuentran en la sierra rural del Perú, a pesar de que se sostenga en algunas investigaciones que la brecha de acceso a estos activos se ha ido reduciendo en los últimos años.



La relación entre estas variables es significativa a un nivel de confianza del 95%. Esto como sabemos podemos deducirlo del rechazo de la hipótesis que supone la independencia del área (urbano o rural) y la región natural de residencia

del jefe de hogar [CUADRO VIII.11], ambos en escala nominal para la condición de pobreza, al haber obtenido valores del coeficiente de Pearson $\chi^2(2) = 819652.7$ y un p-valor = 0.000, esto con una significancia (α) de 5%.

Cuadro VIII.11 Chi-Square Test

estado pobreza01		Value	df	Asymp. Sig. (2-sided)
pobre	Pearson Chi-Square	819652.7	2	.000
	Likelihood Ratio	879899.6	2	.000
	N of Valid Cases	2764184		

El grado de asociación que se encuentra en estas variables se puede apreciar en el cuadro [CUADRO VIII.12] Symmetric Measures, el cual nos muestra los coeficientes de asociación $\Phi = 0.545$,

Cramer's $V = 0.545$, y el ya conocido coeficiente de contingencia (CC)=0.478, que a pesar estos últimos de mostrar un alto grado de relación entre 2 variables medidas en una escala simple como es la

nominal, puede mostrarnos cómo va determinando en alguna medida su influencia en la condición de pobreza que posee el jefe de hogar y cómo esta entrará a tallar dentro la probabilidad que el individuo obtenga dicha condición.

Cuadro VIII.12 Symmetric Measures

estado pobreza			Value	Approx. Sig.
pobre	Nom by	Phi	.545	.000
	Nom	Cramer's V	.545	.000
		Contingency Coefficient	.478	.000
N of Valid Cases			2764184	

ANÁLISIS CUANTITATIVO

Si bien hemos mostrado relaciones con variables importantes que en principio nos muestran panoramas independientes de la pobreza y el efecto que sobre la condición del jefe de hogar ejercen estas variables de naturaleza cualitativa, extrayéndolas del fenómeno multidimensional en su conjunto, podemos también detenernos por instantes dado que no es objetivo de esta investigación analizar como cierto tipo de activo de capital humano se comportan dentro de las distintas condiciones de pobreza, entre ellos, años de estudios del jefe de hogar.

AÑOS DE ESTUDIOS DEL JEFE DE HOGAR

Uno de los indicadores que muchos consideran de suma importancia al analizar la pobreza, es el de los años de estudios del jefe de hogar. Para el año 2001 podemos observar [CUADRO VIII.13] que el activo de capital humano referente a los años de estudios del jefe de hogar es bajo, basta solo con observar que para el

estado de pobreza el promedio de años de estudios se encuentra alrededor de 5, pequeño en comparación a que los jefes de hogar en condición de no pobreza tienen en promedio alrededor de 10 años de estudios.

Dada la heterogeneidad y dispersión de las poblaciones en estudio, el tomar como valor del promedio de años de estudios los valores obtenidos, resulta erróneo. En tal sentido deberemos optar por la mediana como mejor medida de ajuste a esta variable. Se puede apreciar que en promedio, el jefe de hogar pobre estaría alcanzando el grado de primaria (6 años), diferencia sustantiva con respecto a los años alcanzados por el jefe de hogar no pobre que podría estar alcanzando al completar el nivel de secundaria con los 11 años en promedio de estudio que este tendría. No debemos además olvidar que estamos considerando su conducta desde la idea de un corte transversal hecho en el tiempo y deben ser consideradas sus implicancias respectivas, en caso se decidiese, probarla con alguna otra variable y analizar su comportamiento.

Cuadro VIII.13
Cuadro de años de estudio del Jefe de Hogar según Condición de Pobreza

Estado Pobreza			Statistic	Std. Error
Años de estudios JH	pobre	Promedio	5.36	2.60E-03
		Promedio sin 5% VE	5.17	
		Mediana	6.00	
		Desviacion Std.	4.30	
		Asimetría	.345	.001
		Kurtosis	-.870	.003
	no pobre	Promedio	9.29	2.80E-03
		Promedio sin 5% VE	9.38	
		Mediana	11.00	
		Desviacion Std.	4.93	
		Asimetría	-.298	.001
		Kurtosis	-.842	.003

Más allá de que la desigualdad existente entre los dos grupos sea evidente, es interesante confirmar cómo a través de la prueba de diferencia de medias confirmamos tal. El [CUADRO VIII.14] nos muestra en primer lugar el estadístico de Levene¹⁴, el cual nos permite rechazar la idea de igualdad de poblaciones pobre y no pobre. Como ya probamos la diferencia existente entre los años de estudios de jefes de hogar pobres y no pobres, con el test de equivalencia de medias rechazamos tal hipótesis, primero tomando la fila de varianzas diferentes, que ya ha sido probada y luego podemos ver que la

diferencia de medias es distinta de cero (3.92), debido a un $t = -1026.487$ y una significancia de 0.000, lo cual estadísticamente hablando nos permite rechazar la idea de igualdad de medias al nivel poblacional.

Probada esta diferencia de medias nos toca ahora mostrar cómo la diferencia en años de estudios se hace cada vez más notoria a medida que el jefe de hogar se encuentra en pobreza extrema, pobreza no extrema y no pobreza, la cual será mostrada por única vez, dado que no es fin primordial de esta investigación.

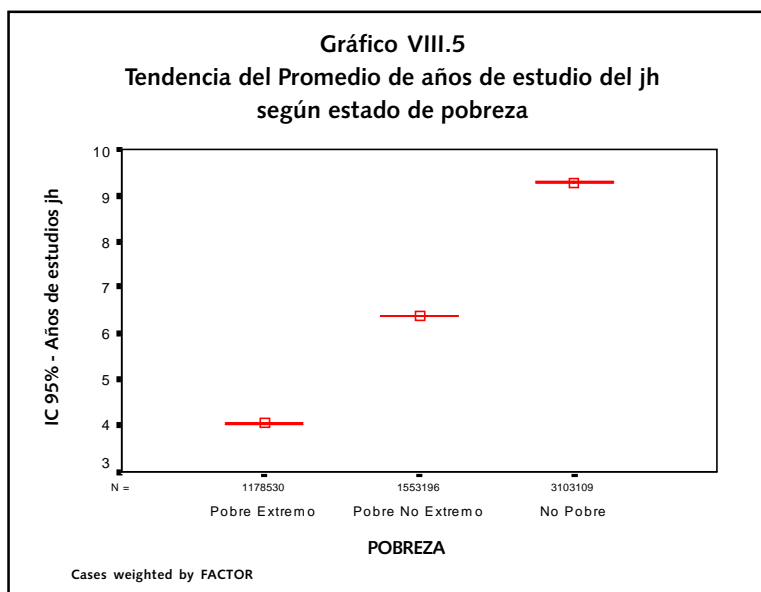
Cuadro VIII.14
Prueba de Levene para igualdad de varianzas y diferencia de medias en los años de estudios del JH

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Años de estudios JH	Asumiendo varianzas iguales	57766.511	.000	-1017.677	5834834	.000	-3.92	3.86E-03
	Sin asumir varianzas iguales			-1026.487	5834386	.000	-3.92	3.82E-03

14/ Sabemos que el estadístico de Levene es resistente a la ausencia de normalidad de una distribución analizada, en tal sentido se propone como más idóneo para medir la homogeneidad de las varianzas.

Una prueba a la afirmación anterior, la podemos confirmar con la tendencia creciente de la diferencia en los años de estudios del jefe de hogar a los diversos niveles de pobreza que este posee. Además de confirmada la diferencia en los promedios de años de estudios, la

brecha diferenciadora y por ende el acceso al capital humano importantísimo en la época actual se hace cada vez más grande porque resulta evidente que haber perdido un año de estudios en la década pasada no es igual que perder un año en la actualidad.



8.2 Formulación y adecuación de los modelos Logit y Probit con VDD, en función de sus enfoques y las variables o factores explicativos más significativos

Como ya vimos en capítulos anteriores, la diferenciación teórica de ambos modelos aún no ha sido descubierta en su total cabalidad, se mencionó que Amemiya (1981) intenta hacer aproximaciones hacia la resolución de este problema pero en general no se obtuvieron resultados concretos.

Si bien tenemos la necesidad de mostrar esta diferenciación, partimos ahora de un

criterio eminentemente práctico. Se sostiene que el ANÁLISIS PROBIT¹⁵ está estrechamente ligado a la regresión logística. Es que si usamos la transformación logit estaríamos calculando esencialmente la regresión antes mencionada. En general, el análisis probit con enfoque de proporciones muestrales sería más apropiado para diseño de experimentos, en función a que su procedimiento mide la relación entre la potencia de un estímulo y la proporción de casos que presentan una respuesta a ese estímulo.

Se materializa esta idea a través de un ejemplo, querer saber qué tan efectivo

15/ SPSS para Windows, Realease 10.0.7 (2000)

es un nuevo pesticida para liquidar insectos, ¿cual sería la concentración letal efectiva a usar? Uno podría diseñar un experimento en el cual se exponen muestras de insectos a diferentes concentraciones del insecticida, es decir, tener el número de insectos expuestos en dicha concentración y entonces determinar el número de insectos muertos bajo tal efecto. Aplicando un análisis probit a estos datos, uno puede determinar la potencia de la relación entre la muerte de los insectos y la dosis de pesticida adecuado y determinar cuál es el grado de concentración apropiado del insecticida que me permitiría estar seguro de matar al 95% de los insectos expuestos.

Como resulta evidente, se está realizando un enfoque cuantitativo, confirmado esto además en la línea seguida por autores como Green o Gujarati al formular sus modelos de probabilidad. Pero como puede apreciarse se limita al hecho de inclusión de pocas variables, más aún siendo estas de índole cuantitativo y porque necesariamente tendría que obtener la frecuencia de observaciones que implican en primer lugar a las variables independientes, vale decir por ejemplo cantidad de jefes de hogar dentro de los deciles de ingreso, luego determinar la cantidad de jefes de hogar pobres dentro de cada décil, para poder saber la tasa de respuesta a ese nivel de ingreso dentro de la variable dependiente (la condición de pobreza). Si se decidiera incluir más variables independientes se tiene que tomar en cuenta que debemos obtener las frecuencias que resultan del cruce de estas. Imaginarse tan solo el cruce los diversos valores del ingreso per cápita con el estado civil, la cantidad de miembros de la familia y con los tipos de accesos a activos públicos como la luz, agua, entre otros, empiezan a figurar la inviabilidad de esta aplicación.

Aquí es donde se producirán limitaciones en la aplicación de este enfoque debido a

que si bien se pueden obtener el cruce de ellas, no esta regido a parámetros de medición exacta y continua, pues la pobreza es un fenómeno de escala multidimensional. Sería interesante poder saber si la mezcla de escalas determinan un nivel de medición que me permitiese acceder a la probabilidad exacta de ser pobre.

A parte de la condición de que las observaciones deben ser independientes. Si tenemos un gran cantidad de estas para las variables independientes, la Chi-cuadrado (χ^2) y los estadísticos de bondad de ajuste pueden no ser válidos.

Retomando la diferenciación de ambos modelos, podemos mencionar que la regresión logística con enfoque de casos individuales es más apropiada para estudios observacionales. Usada también para situaciones en las cuales uno quiere ser capaz de predecir la presencia o ausencia de una característica o resultado basado en valores de un conjunto de variables predictoras. Los coeficientes del modelo logit pueden ser usados para estimar Odds Ratios (‘Razón de probabilidades’) para cada una de las variables independientes en el modelo. La regresión logística y por ende el modelo logit, es aplicable a un rango mayor de situaciones de investigación. Adicionalmente, como en otras formas de regresión, la multicolinealidad de las variables independientes si no es manejable puede generar estimadores sesgados o inflar el error estándar.

8.2.1 Metodologías de estimación de los modelos de probabilidad

MODELO DE PROBABILIDAD DE LA POBREZA EN EL JEFE DE HOGAR

A continuación mostramos la metodología de estimación del modelo de probabilidad del jefe de hogar pobre del Perú, para el

año 2001, tomada una muestra aleatoria de alrededor del 50% (8233) de los jefes de hogar de la encuesta nacional de hogares según los factores considerados de importancia en la condición de pobreza de éste y obtenidos según la información de la ENAHO 2001 -IV TRIMESTRE, Condiciones de Vida y Pobreza.

Enfoque de Observaciones Individuales

Coefficientes para el modelo logit

El [CUADRO VIII.15] contiene los coeficientes estimados (bajo la columna encabezada por B) y los estadísticos asociados al modelo que predice la probabilidad de ser pobre del jefe de hogar del Perú en el año 2001 en función a características individuales del jefe de hogar como el no tener nivel educativo,

edad, su indicador de experiencia laboral, indicador de condición ocupacional, tamaño de la firma, otro empleo; y características asociadas al hogar como el ingreso per cápita mensual, la cantidad de miembros, y el indicador PET del hogar, además si la conexión de agua y servicios es a través de red pública dentro de la vivienda. Las variables (sinnivel, costa urbana, costa rural, sierra urbana, sierra rural, selva urbana, selva rural, obrero, menos100pe, otroemple, conviviente, sehigdviv, y las cuatro últimas variables) son indicadoras, codificadas con 0 y 1. El valor de 1 para OBRERO indica que el jefe de hogar tiene condición ocupacional de obrero, el valor de 1 para MENOS100PE indica que el tamaño de la firma en la que trabaja es de menos de 100 personas, el valor de 1 para OTROEMPLE indica que el jefe de hogar posee actividad secundaria por la que percibe ingresos.

Cuadro VIII.15 Variables en la ecuación

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
							Lower	Upper
SINNIVEL	.911	.231	15.548	1	.000	2.488	1.582	3.913
EDAD	-.072	.011	41.221	1	.000	.931	.910	.951
COSTA URBANA	-1.189	.167	50.439	1	.000	.304	.219	.423
SIERRA URBAN	.570	.098	33.642	1	.000	1.768	1.459	2.144
SELVA URBANA	.899	.121	55.483	1	.000	2.458	1.940	3.114
SELVA RURAL	-.807	.125	41.844	1	.000	.446	.349	.570
INDEXPLABO	.070	.010	49.495	1	.000	1.072	1.051	1.093
COLEGESTA	.584	.189	9.540	1	.002	1.793	1.238	2.598
OBRERO	.451	.092	23.940	1	.000	1.570	1.311	1.881
MENOS100PE	.324	.095	11.606	1	.001	1.383	1.148	1.666
OTROEMPLE	-.226	.085	7.090	1	.008	.798	.675	.942
INPERCAM	-.011	.000	876.865	1	.000	.989	.988	.990
MIEPERHO	.517	.031	283.825	1	.000	1.677	1.579	1.781
PETHOGAR	-.313	.040	60.764	1	.000	.731	.676	.791
CONVIVIENTE	.269	.080	11.274	1	.001	1.308	1.118	1.530
SEHIGDVIV	-.388	.093	17.414	1	.000	.678	.565	.814
NOSERVHIGIE	.270	.089	9.281	1	.002	1.310	1.101	1.558
AGUAVIV	-.168	.082	4.137	1	.042	.846	.719	.994
ALUMKEROSE	.627	.092	46.847	1	.000	1.873	1.565	2.241
EMPVIVINGRE	.288	.083	11.916	1	.001	1.334	1.133	1.571
Constant	.761	.379	4.035	1	.045	2.141		

Dados estos coeficientes, la ecuación para la probabilidad de ser pobre del jefe de hogar se puede escribir como sigue:

Pr(Jefe de Hogar Pobre) =

$$Pi = \frac{1}{1 + e^{-Zi}}$$

ECUACIÓN VIII.1

Donde:

$$\begin{aligned} Zi = & 0.761 + 0.911(\text{sinnivel}) - \\ & 0.072(\text{edad}) - 1.189(\text{costa urbana}) + \\ & 0.570(\text{sierra urbana}) + 0.899(\text{selva urbana}) - \\ & 0.807(\text{selva rural}) + 0.07(\text{indexplabo}) + 0.584(\text{colegesta}) + \\ & 0.451(\text{obrero}) + 0.324(\text{menos100pe}) - \\ & 0.226(\text{otroemple}) - 0.011(\text{inpercam}) + \\ & 0.517(\text{mieperhog}) - 0.313(\text{pethogar}) + \\ & 0.269(\text{conviviente}) - 0.388(\text{sehigdviv}) + \\ & 0.270(\text{noservhigie}) - 0.168(\text{aguaviv}) + \\ & 0.627(\text{alumkerose}) + 0.288(\text{empvivingre}). \end{aligned}$$

ECUACIÓN VIII.2

Aplicando esto a un jefe de hogar sin nivel de educación con 60 años de edad de Lima con un indicador de experiencia laboral de 54 años, desocupado, con un ingreso per cápita mensual de s/.143.00, con 4 miembros en su hogar, con 3 personas en la PET, que no es conviviente, que no use el kerosene como tipo de alumbrado en su hogar, que posee conexión a red pública dentro de la vivienda tanto de agua como de alcantarillado y no emplea parte de la vivienda para ingresos dentro del hogar.

$$\begin{aligned} Zi = & 0.761 + 0.911(1) - \\ & 0.072(60) - 0.388(1) + 0.270(0) - \\ & 0.168(1) + 0.627(0) + 0.288(0) \end{aligned}$$

ECUACIÓN VIII.3

Entonces la probabilidad de ser pobre del jefe de hogar es:

$$\text{Pr(Jefe de hogar pobre)} = 0.65847$$

ECUACIÓN VIII.4

Basados en este estimado, podemos predecir que el jefe de hogar con estas características es pobre. En general, si la probabilidad estimada del evento es menor a 0.5, podemos decir que el evento no va a ocurrir. Si la probabilidad es mejor que 0.5, podemos decir que el evento va a ocurrir y por lo tanto como en el ejemplo, que el jefe de hogar es pobre.

Prueba de hipótesis sobre los coeficientes

Para tamaños de muestra grande, la prueba de que un coeficiente es 0 es basada en el **estadístico de Wald**, el cual tiene una distribución chi-cuadrado. Cuando una variable tiene un grado de libertad, el estadístico de Wald es el cuadrado del ratio del coeficiente entre su error estándar. Para variables categóricas, el estadístico de Wald tiene sus grados de libertad equivalentes a uno menos el número de categorías.

Por ejemplo, en el [CUADRO VIII.15] el coeficiente de sin nivel de educación es 0.911 y su error estándar es 0.231 (en el cuadro la columna llamada S.E.). El estadístico de Wald es $(0.911/0.231)^2$ ó 15.548. El nivel de significancia para el estadístico de Wald es mostrada en la columna Sig. En nuestro caso todas las variables obtenidas en el modelo resultan ser significativamente diferentes de cero, para un nivel de significancia de 0.05.

Otra manera ver de este análisis, es a través de la prueba asociada a su efecto:

$H_0 : \beta(\text{SINNIV}) = 0$, lo cual quiere decir que el efecto de no tener educación es irrelevante, pero obteniendo un valor para el estadístico Wald de 15.548 y una significancia de 0.000, rechazábamos esa hipótesis nula afirmando que tal efecto es significativamente distinto de cero y por tal razón relevante para la probabilidad de ser pobre del jefe de hogar.

De la misma manera podemos afirmar, por ejemplo, con respecto a la significancia de la cantidad de miembros en el hogar, obteniendo un valor de 283.825 para el estadístico de Wald, que se obtiene de elevar al cuadrado la división del coeficiente entre su respectivo Error Estándar de estimación, demás está mencionar la relevancia de esta variable y confirmar el efecto positivo (0.517) que juega dentro de la probabilidad de ser pobre del jefe de hogar.

Desafortunadamente, el estadístico de Wald posee una propiedad indeseable. Cuando el valor absoluto del coeficiente de regresión llega a ser demasiado grande, el error estándar también lo es. Esto produce que el estadístico de Wald, sea muy pequeño, por tal motivo no estaríamos rechazando la hipótesis nula de que el coeficiente es 0, cuando en realidad sí deberíamos. De allí que cuando tenemos un coeficiente grande, uno no debería de confiar en el estadístico de Wald para prueba de hipótesis. En vez de ello, se debería construir un modelo con la variable y otro sin variable y basar la prueba de hipótesis en el cambio del log de la función verosimilitud. (Hauck & Donner, 1977).

Interpretación de los coeficientes

Para entender la interpretación de los coeficientes de esta regresión, consideramos un reordenamiento de la ecuación para el modelo logístico. Este puede ser escrito como un reordenamiento en términos de las razones ('Odds') de ocurrencia de un evento. (Las **Odds** ('razón de probabilidades') de ocurrencia de un evento es definido como el ratio de la probabilidad de que ocurra un evento sobre la probabilidad de que no ocurra un evento. Por ejemplo, la odds de obtener cara en un lanzamiento de una moneda sería $0.5/0.5 = 1$. Similarmente, la odds de obtener un corazón en una repartición de cartas será $0.25/0.75 = 1/3$. No debe confundirse este significado técnico de la odds con su uso informal de un simple promedio de probabilidad)

Escribimos el modelo logístico en términos del logaritmo de las odds, llamada como sabemos **logit**:

$$\log\left(\frac{\Pr(jhpobre)}{\Pr(jhnopobre)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

ECUACIÓN VIII.5

De la ecuación anterior, el coeficiente logístico puede ser interpretado como un cambio en el **log odds** asociado con un cambio unitario en la variable independiente. Por ejemplo, del [CUADRO VIII.15] podemos ver que el coeficiente para sinnivel es 0.911. Esto nos indica que cuando el jefe de hogar no posee nivel educativo y los valores de las otras variables independientes se mantienen constantes, el log odds (razón de probabilidades) se incrementa en un 0.911.

La ecuación de regresión puede ser escrita en términos de los odds como:

$$\frac{Pr ob(jhpobre)}{Pr ob(jhnopobre)} =$$

$$e^{\beta_0+\beta_1X_1+.....+\beta_pX_p} = e^{\beta_0} e^{\beta_1X_1} ...e^{\beta_pX_p}$$

ECUACIÓN VIII.6

En *e* radica el poder de *B_i*, que es el factor por el cual la razón de probabilidades (‘Odds’) cambia cuando la *i*-ésima variable independiente se incrementa en una unidad. Si *B_i* positivo, este factor va a ser mayor que 1, lo cual significa que la Odds se incrementa; si *B_i* es negativo, el factor va a ser menor que uno, lo cual significa que la Odds decrece. Cuando *B_i* es 0, el factor equivale a 1, lo cual significa que la odds no cambia.

En ese sentido resultó interesante ver como al ser obrero es 1.6 veces más probable ser pobre que no serlo, es decir, la Odds a favor de ser jefe de hogar pobre cambia positivamente en 100 (1.570 - 1)% =57% al tener condición ocupacional de obrero. De la misma manera pudimos dar una lectura parecida, pero con los criterios adecuados del caso, al afirmar que el tener abastecimiento de agua dentro de la vivienda a través de red pública le reduce la probabilidad de ser pobre al jefe de

hogar, es decir, que su Odds se reduce o cambia negativamente en 100(0.846-1)% = 15.4%, y como mencionamos anteriormente la probabilidad de ser jefe de hogar pobre sería menor con respecto a la probabilidad de ser no pobre.

Otra de las variables que resulta de interés, es la del número de miembros del hogar, donde se puede apreciar que dicha variable aumenta positivamente el logit de la probabilidad de ser pobre con respecto a no serlo en 0.517. Es decir, que la Odds (‘razón de probabilidades’) a favor de ser pobre cambia positivamente en 100(1.677-1)% = 67.7% al producirse el aumento de 1 persona en la cantidad de miembros del hogar.

Determinación de la bondad de ajuste del modelo

Existen varios caminos para determinar de un modo u otro la calidad de ajuste del modelo a los datos.

Tabla de clasificación

Un camino para determinar que tan bien nuestro modelo ajusta los datos es comparar nuestras predicciones con los resultados observados. El [CUADRO VIII.16] muestra la tabla de clasificación para el modelo obtenido.

Cuadro VIII.16 Tabla de Clasificación ^a

Observed		Predicted		
		Estado pobreza		Percentage Correct
		no pobre	pobre	
Estado pobreza	no pobre	3400	701	82.9
	pobre	519	3550	87.2
Overall Percentage				85.1

a. El punto de corte es .50

Del cuadro anterior vemos que 3400 jefes de hogar no pobres fueron correctamente clasificados por el modelo como jefes de hogar en estado de no pobreza. Similarmente, 3550 jefes de hogar pobres fueron correctamente clasificados en estado de pobreza. Los elementos fuera de la diagonal de la tabla nos dicen cuántos jefes de hogar fueron incorrectamente clasificados. Un total de 1220 jefes de hogar fueron mal clasificados- 701 jefes de hogar no pobres y 519 jefes de hogar pobres. De los jefes de hogar no pobres el 82.9% fueron correctamente clasificados. De los jefes de hogar pobres el 87.2% fueron correctamente clasificados. Del total, el 85.1% de los 8170 jefes de hogar tomados para ser analizados fueron correctamente clasificados.

Bondad de ajuste del modelo

Observando que tan bien clasifica el modelo los casos observados es uno de los caminos para determinar la potencia del modelo logit y la regresión logística.

Otro camino de acceso a la bondad de ajuste del modelo es examinar que tan bien los resultados de la muestra actual nos dan los parámetros estimados. La probabilidad de los resultados observados, dados los parámetros estimados, es conocida como la verosimilitud (‘*likelihood*’). Como la verosimilitud es un número pequeño menor que uno, se usa generalmente -2 veces el logaritmo de la verosimilitud (-2LL) como una medida para verificar que tan bien el modelo estimado ajusta los datos. Un buen modelo es aquel que tiene una alta verosimilitud obtenida de los resultados observados. Lo cual se traduciría en un pequeño valor para -2LL. (Si un modelo ajusta perfectamente, la verosimilitud es 1, y -2 veces el log likelihood es 0).

Ahora, para el modelo logit y de regresión logística de los jefes de hogar, un modelo con solo la constante nos da un -2LL igual a 11325.9 como se muestra en el [CUADRO VIII.17].

Cuadro VIII.17 Historia de Iteración ^{a,b,c}

Iteración		-2 Log likelihood	Coefficients
Step 0			Constant
1		11325.900	-.008

- a. Constante es incluida en el modelo.
- b. Inicial -2 Log Likelihood(-2LL): 11325.900
- c. Estimación terminada en iteración N° 1 porque log-likelihood decrece en menos de .01%.

Bondad de ajuste con todas las variables

El [CUADRO VIII.18] nos muestra la bondad de ajuste para el modelo con todas las variables independientes [CUADRO VIII.15]. Para este modelo el valor de -2LL

es 5683.192, el cual es menor que el -2LL para el modelo solo con la constante. La bondad de ajuste puede apreciarse además en los estadísticos R2 de Cox & Snell (0.499) y de Nagelkerke (0.665), que nos explican el porcentaje de variación explicada por el modelo.

Cuadro VIII.18 Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
20	5683.192	.499	.665

Existen otros estadísticos [CUADRO VIII.19] que nos permiten abordar el ajuste del modelo. Ellos son llamados X2 para el modelo, el bloque y el paso de iteración. Para el caso de los jefes de hogar, la X2 del modelo es la diferencia entre -2LL para el modelo solo con la constante y -2LL para el modelo con todas las variables independientes. Es decir, se prueba la hipótesis nula que sostiene que el modelo con solo la constante es mejor que el modelo con todos los factores incluidos, es decir:

$\chi^2(20 \text{ g.l.}) = -2 (\text{LnL}(\text{solo con cte.}) - \text{LnL}(\text{con factores covariantes})) = 11325.9 - 5683.192 = 5642.707$, la cual frente a un $\chi^2(20 \text{ g.l., } 5\%) = 31.41$, rechaza la hipótesis nula de que el efecto de todas las variables explicativas incluidas, excepto

la constante, es 0. De allí que podemos afirmar que este modelo es mejor y mantener $L = b'X$. La misma manera de proceder se realiza para una comparación entre diversos modelos que incluyen una cantidad diversa de factores explicativos. Este estadístico es comparable a la prueba F para la regresión clásica. Los grados de libertad para χ^2 modelo son la diferencia entre el número de parámetros de los dos modelos.

Con respecto a la χ^2 del paso (Step), se refiere al cambio en -2LL para este último paso en la construcción del modelo. Probaría la hipótesis nula que el efecto de este paso no es significativamente diferente de 0, lo cual es rechazado de acuerdo a su significancia de 0.042

Cuadro VIII.19 Omnibus Test of Model Coefficients

		Chi-square	df	Sig.
Step 20	Step	4.126	1	.042
	Block	5642.707	20	.000
	Model	5642.707	20	.000

Con respecto a la χ^2 del bloque sería el cambio en -2LL entre los sucesivos bloques en la construcción del modelo. Prueba la hipótesis nula que los coeficientes para el conjunto de variables adicionadas en el último paso son 0. En el caso de los jefes de hogar, consideramos solo dos modelos: el modelo solo con la constante y el modelo con todas las variables

independientes [CUADRO VIII.15], por tal motivo la χ^2 del modelo y del bloque tienen los mismos valores. Si se consideran secuencialmente otros modelos a parte de estos dos, usando los métodos de selección de variables Forward o Backward, las χ^2 para el modelo y para el bloque van a ser diferentes.

Métodos de diagnóstico

Cuando se construye un modelo estadístico, es importante examinar la calidad de los resultados obtenidos. En regresión lineal, observamos una variedad de residuales, medidas de influencia e indicadores de colinealidad. Existen herramientas valiosas, para identificar puntos donde el modelo no puede ajustar bien, puntos que ejercen fuerte influencia sobre los coeficientes estimados, y variables que son altamente relacionadas unas con otras. En regresión logística y el modelo logit existen diagnósticos comparables que deberían ser usados para detectar problemas.

El **residual** es la diferencia entre la probabilidad observada del evento y la probabilidad predicha del evento basado en el modelo. Por ejemplo, si predecimos que la probabilidad de ser pobre es 0.75 para un jefe de hogar pobre, su residual es $1 - 0.75 = 0.25$.

El residual estandarizado (**standardized residual**) es el residual dividido por un estimado de su desviación estándar. En este caso será:

$$Z_i = \frac{\text{residual}_i}{\sqrt{P_i(1 - P_i)}}$$

ECUACIÓN VIII.7

Para cada observación, el residual estandarizado puede además ser considerado una componente del estadístico de bondad de ajuste χ^2 . Si el tamaño de muestra es grande, el residual estandarizado debería aproximarse a una distribución normal, con media 0 y desviación estándar de 1.

Para cada observación, la **deviance** es calculada como:

$$-2 \cdot \log(\text{probabilidad calculada para el grupo observado})$$

ECUACIÓN VIII.8

La deviance es calculada tomando la raíz cuadrada del estadístico anterior y adicionándole un signo negativo si el evento no ocurre para dicha observación. Por ejemplo, la deviance para un jefe de hogar no pobre y una probabilidad calculada de 0.45 de ser no pobre es

$$\text{Deviance} = -\sqrt{-2 \log(0.45)} = -0.833$$

ECUACIÓN VIII.9

Valores grandes para la deviance indican que el modelo no ajusta bien la observación. Para tamaños de muestra grande, la deviance es aproximadamente una distribución normal.

El **Studentized residual** para cada observación es el cambio en la deviance del modelo si el caso es excluido. Discrepancias entre la deviance y el studentized residual pueden identificar casos inusuales. Una gráfica de probabilidad normal de los studentized residuals puede ser muy útil.

El **logit residual** es el residual para el modelo si la predicción esta en la escala logit.

$$\text{Logitresid}_i = \frac{\text{residual}_i}{P_i(1 - P_i)}$$

ECUACIÓN VIII.10

El **leverage** en la regresión logística es en muchos aspectos análogo al leverage en

la regresión de mínimos cuadrados. Los valores de leverage son siempre usados para detectar observaciones que tienen gran impacto en los valores predichos. A diferencia de la regresión lineal, los valores **leverage** en esta regresión dependen de las puntuaciones de la variable dependiente y del diseño de la matriz. Los valores de los leverage están restringidos entre 0 y 1. Su valor promedio es p/n , donde p es el número de parámetros estimados en el modelo, incluyendo la constante, y n es el tamaño de la muestra.

La **Cook's distance** es la medida de la influencia de una observación. Nos dice cuánto afecta no solo en el residual para esta observación, sino también sobre el residual del resto de observaciones restantes cuando se elimina dicho caso. La Cook's distance (D) depende de su residual estandarizado, así como de su leverage. Esta definido como:

$$D_i = \frac{Z_i * h_i}{(1 - h_i)^2}$$

ECUACIÓN VIII.11

Donde Z_i es su residual estandarizado y h_i es su leverage.

Otra medida de diagnóstico útil es el cambio en los coeficientes del modelo cuando una observación es retirada del modelo, o **DfBeta**. Se puede calcular el cambio en cada coeficiente, incluyendo la constante. Un ejemplo sería el cambio en el primer coeficiente cuando el caso i es borrado

$$DfBeta(B_1^{(i)}) = B_1 - B_1^{(i)}$$

ECUACIÓN VIII.12

Donde B_i es el valor del coeficiente cuando todas las observaciones son incluidas en el modelo y $B_1^{(i)}$ es el valor del coeficiente cuando la i -ésima observación es retirada. Valores grandes de cambio identifican observaciones que deben ser examinadas.

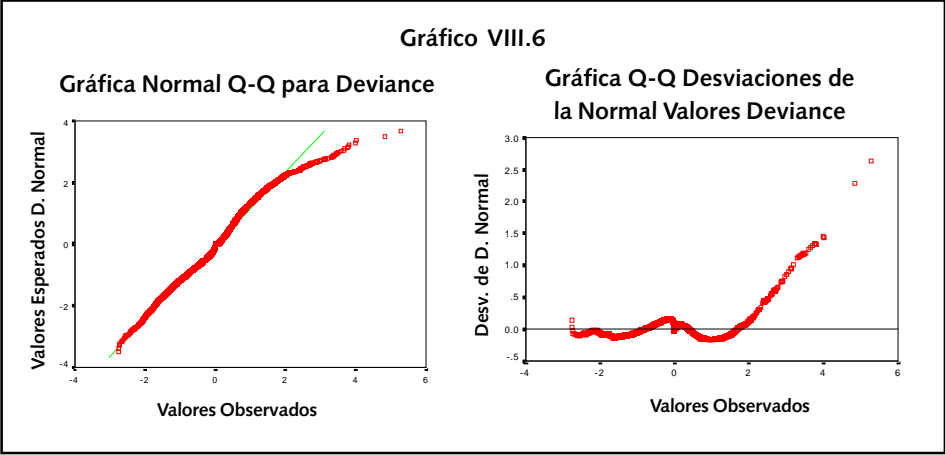
Diagnósticos de gráficas

Todos los estadísticos descritos hasta ahora pueden ser grabados para este análisis en la matriz de casos/variables. Cuando se considere conveniente, se pueden obtener gráficas de probabilidad normal usando el procedimiento de exploración y gráficas de diagnóstico usando procedimientos gráficos.

El [GRÁFICO VIII.6] muestra la gráfica de probabilidad normal Q-Q y la gráfica Q-Q de desviaciones respecto a la normal. En el caso de la gráfica izquierda Q-Q (Quantiles reales y teóricos de una distribución normal) de probabilidad normal, los valores correspondientes a una distribución normal teórica vienen representados por la recta y los puntos son los valores de la deviance de los jefes de hogar. Como vemos estos puntos, en su mayoría, están próximos a la recta, lo cual indica que el ajuste es aceptable. Confirma esto la suposición hecha con respecto a los tamaños de muestra grande. Para el caso de aquellos valores de deviance muy superiores a 2, el modelo no ajusta muy bien dichas observaciones, pero se debe considerar que estos corresponden probabilidades de mala clasificación de los jefes de hogar, en tal sentido, se puede optar por su eliminación o por el análisis de los estadísticos de Cook's y DfBetas para ver la influencia de dichas observaciones.

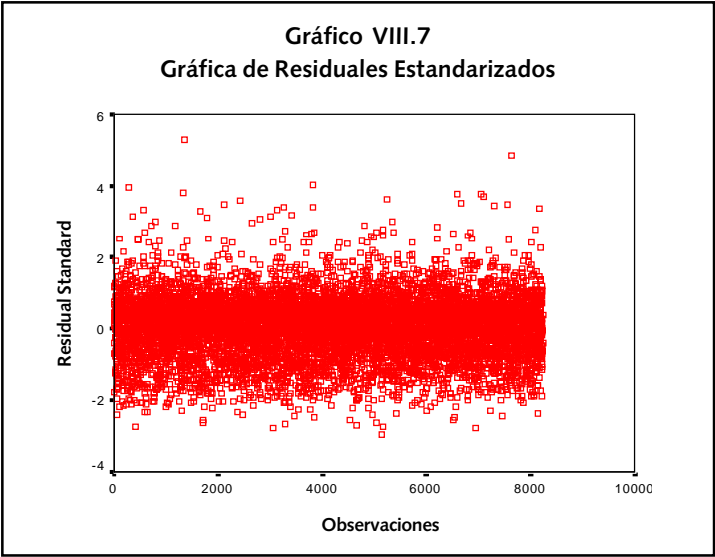
Con respecto a la gráfica Q-Q del lado derecho de desviaciones de la normal, recoge las deviances de los jefes de hogar respecto de la recta. Si la muestra procede de una población normal, los puntos deben fluctuar alrededor de "0" y no seguir un patrón determinado. Como se puede

apreciar con respecto a la mayoría de observaciones sus desviaciones fluctúan alrededor de "0", lo que confirma la idea de distribución normal para tamaños de muestra grandes por parte de las deviances.



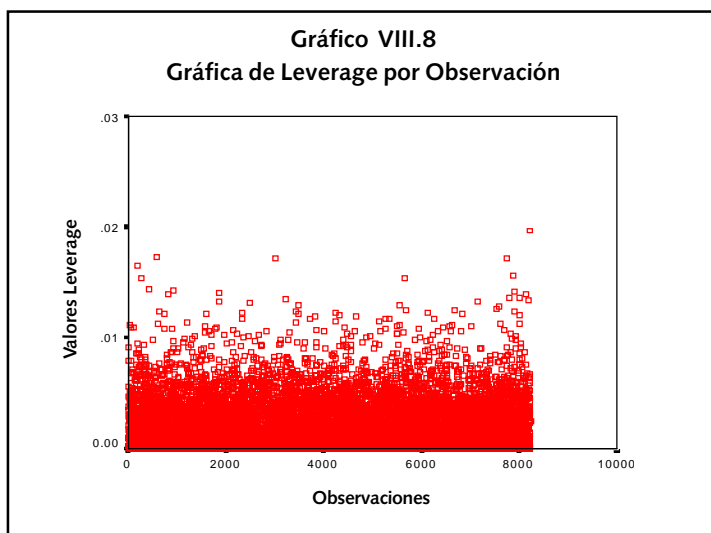
Una gráfica de los residuales estandarizados para cada observación se muestra en la [GRÁFICA VIII.7], allí podemos ver como los residuales toman valores grandes fuera

del rango de variación permitido. Hablamos específicamente para aquellos casos donde sus residuales toman valores superiores a 3.



En la [GRÁFICA VIII.8] se muestra que no existen en general casos que tengan un valor de leverage muy diferenciador unos de otros, cumpliéndose en primer lugar el rango de variación sobre el cual varía, 0 y 1. En segundo lugar, si se quiere

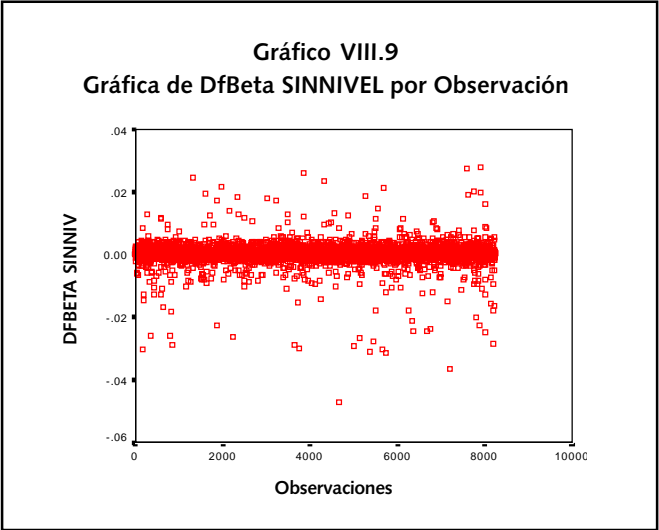
ser estricto, el mayor valor diferenciador que se encontró para el leverage se encuentra alrededor de 0.02, lo que no nos llevaría a afirmar que dicho caso tiene un gran impacto sobre los valores predichos.



En la [GRÁFICA VIII.9] se muestra que existen algunos casos que tienen sustancial impacto en la estimación del coeficiente de la variable indicadora sin nivel educativo (casos 4656 y 7198), los valores para Dfbeta sinnivel de estas observaciones se encuentran alrededor de -0.04 - valores extremos. Examinando la data se revela que el primer caso (4656), más próximo a -0.06, es un jefe de hogar sin nivel de educación, de la selva rural, hombre de 41 años, conviviente pero en estado de no pobreza, que trabaja solo en su actividad principal. En el caso de la segunda observación (7198), valor más próximo a -0.04, es un jefe de hogar con primaria,

de la costa urbana, mujer de 50 años, en estado de pobreza no extrema, trabajadora independiente con solo actividad principal. Podemos ver que estos dos casos son inusuales de acuerdo a las relaciones obtenidas en el [CUADRO VIII.15].

Si retirásemos el caso 4656 del análisis, el coeficiente de la variable sinnivel de educación cambiaría de 0.911 a -0.04731, con lo cual se vería perjudicado, y se convertiría en un mal predictor o variable explicativa. Lo mismo sucede si retirásemos el caso 7198, con lo cual el coeficiente variaría de 0.911 a -0.03660, con los efectos del caso ya conocidos.



Enfoque de proporciones muestrales

Modelo probit

En primer lugar intentaremos ejemplificar a qué nos referimos con la idea de tasas de respuesta o de proporciones

muestrales dentro de este análisis. Supongamos que se desea estimar la condición de pobreza del individuo en función a sus deciles de ingreso, en tal sentido correspondería obtener la siguiente información [CUADRO VIII.20]

Cuadro VIII.20

Deciles de Ingreso	Jefes de hogar en deciles (Ni)	Jefes de hogar pobres en deciles (ni)
Decil I	1789	1737
Decil II	1811	1711
Decil III	1782	1554
Decil IV	1705	1211
Decil V	1702	855
Decil VI	1730	531
Decil VII	1623	264
Decil VIII	1582	154
Decil IX	1498	89
Decil X	1293	21

Como aquí se tiene las probabilidades (Pi) o proporciones muestrales de ser jefe de hogar pobre según decil de ingreso (ninperca), obviando las limitaciones que la técnica posee en cuanto al tamaño de

observaciones que deben entrar en el análisis, estimamos un modelo de probabilidad de ser pobre del jefe de hogar en función al décil de ingreso.

PROBIT ANALYSIS			
MODEL Information			
ONLY Normal Sigmoid is requested.			
Parameter estimates converged after 11 iterations.			
Optimal solution found.			
Parameter Estimates (PROBIT model: (PROBIT(p)) = Intercept + BX):			
	Regression Coeff.	Standard Error	Coeff./S.E.
NINPERCA	-.00678	.00013	-53.42949
Intercept	S.Error	Intercept/S.E.	EPOBRE
.72278	.00094	767.70752	pobre
.75554	.00057	1322.62124	no pobre
Pearson Goodness-of-Fit Chi Square = 9392340.38 DF = 16512 P = .000			
Parallelism Test Chi Square = 9300657.30 DF = 1 P = .000			

Dado que el nivel de significancia de la bondad de ajuste del estadístico Pearson- χ^2 es pequeño, empiezan a evidenciarse razones para dudar del modelo.

Se puede apreciar [CUADRO VIII.21] que el modelo parece no ajustar correctamente las probabilidades observadas de los jefes de hogar en función a su décil de ingreso.

Cuadro VIII.21 Frecuencias Observadas y Esperadas

EPOBRE	NINPERCA	Number of Subjects	Observed Responses	Expected Responses	Residual	Prob
0	6.00	1730.0	531.0	1319.386	-788.386	.76265
0	8.00	1582.0	154.0	1199.852	-1045.852	.75844
0	3.00	1782.0	1554.0	1370.164	183.836	.76889
0	6.00	1730.0	531.0	1319.386	-788.386	.76265
0	5.00	1702.0	855.0	1301.590	-446.590	.76474
0	9.00	1498.0	89.0	1132.967	-1043.967	.75632
0	3.00	1782.0	1554.0	1370.164	183.836	.76889
0	6.00	1730.0	531.0	1319.386	-788.386	.76265
0	5.00	1702.0	855.0	1301.590	-446.590	.76474
0	4.00	1705.0	1211.0	1307.430	-96.430	.76682

En ese sentido tendríamos que la ecuación con respecto a la condición de pobreza:

(PROBIT(p)) =

Intercept + BX

(PROBIT(p)) =

.72278 -.00678*

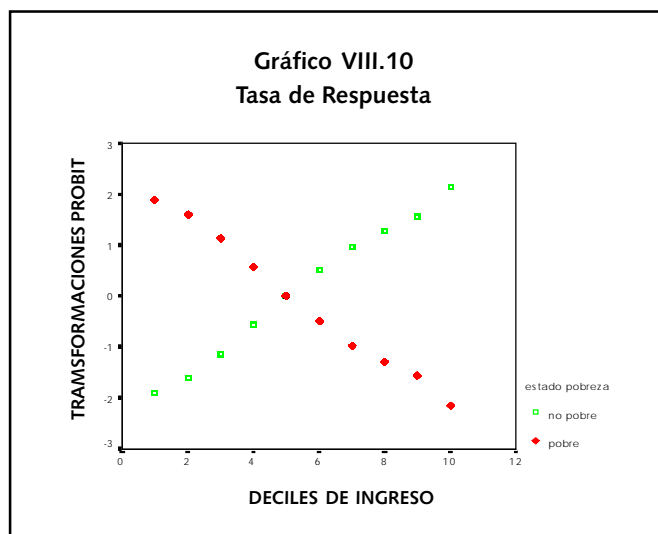
NINPERCA

El resultado obtenido debe ser tomado con cuidado en el sentido del cambio que se produciría sobre el probit de la probabilidad de ser pobre. Ésta debe ser transformada para llegar a la cantidad sobre la cual variaría en la distribución normal (Un aumento hacia los deciles más ricos reduce el valor de la transformación probit alrededor de un 0.6%).

Resulta interesante mostrar como existe una relación en algún sentido lineal, presente entre las transformaciones probit (Z) en la distribución normal y las probabilidades de la variable respuesta, es decir, la frecuencia relativa con respecto del jefe de hogar pobre en el decil correspondiente, podemos ver [GRÁFICO VIII.10] la relación inversa en el estado de pobreza de las transformaciones probit y el decil de ingreso, es decir, como a medida que se da un aumento de los deciles, por ende del nivel de ingreso per

cápita, la probabilidades de ser pobre asociadas al jefe de hogar y su transformación probit disminuyen.

No debemos olvidar algunas de las consideraciones que hicimos al empezar este análisis donde se sostuvo que algunos de los estimadores pueden resultar no útiles cuando manejamos tamaños de muestra bastante grandes, en este sentido resultaría importante el modelo logit con observaciones individuales para este tipo de condicionamientos, además si no se hubiese procedido a agrupar la muestra por niveles o valores del estímulo (deciles de ingreso), se podría haber aplicado este último, y haber obtenido las cuantificaciones de aportes en la probabilidad de que un jefe de hogar sea clasificado como pobre o no, en función a su nivel de educación, la cantidad de hijos que este tiene, si este accede a luz eléctrica dentro de su hogar, etc.



ANÁLISIS DEL MODELO SOBRE LA PROBABILIDAD DE POBREZA EN EL JEFE DE HOGAR DEL PERÚ PARA EL AÑO 2001

EL MODELO LOGIT (Observaciones individuales)

Con respecto al cuadro siguiente, se van a analizar 3 modelos, el modelo n° 1, que no incorpora los efectos de diseño de la muestra, con el objetivo de mostrar en que medida varían la inferencia a nivel poblacional comparada con los modelos que sí incorporan los factores de expansión, en este caso los modelos n° 2 y n° 3, en el caso del modelo n° 2, se expande la muestra tomada para construir la metodología (muestra aleatoria de 8233 hogares) y en el caso del modelo n°3 se toma el resto de la muestra (8282 hogares)

para medir la validez y confiabilidad de la estimación realizada en la metodología.

La probabilidad de ser un jefe de hogar pobre para el año 2001 [CUADRO VIII.22] vendría a estar fuertemente ligada a la ausencia de activos de capital humano como el de no tener nivel de educación alguno, contribuyendo este a elevar el logit de su probabilidad de ser pobre en un 94.7% (MODELO N° 1), creciendo hasta un 114.8% (MODELO N° 2) y reduciéndose hasta un 78.5% (MODELO N° 3). Resulta interesantísimo y no puede ser dejado de lado por su significancia, ver que ser un jefe de hogar que ha recibido su educación en colegio estatal aumenta el logit de su probabilidad alrededor de un 60%, manteniendo el efecto de las demás variables constantes

Cuadro VIII.22
Incidencia en la probabilidad de ser pobre a nivel nacional

EPOBRE	MODELO N° 1		MODELO N° 2		MODELO N° 3	
	Coef.	P> z	Coef.	P> t	Coef.	P> t
SINNIVEL	0.947	0.000	1.148	0.000	0.785	0.001
EDAD	-0.07	0.000	-0.083	0.000	-0.068	0.000
A.RURAL	-0.996	0.000				
SIERRA	0.852	0.000				
SELVA	0.573	0.000				
COSTA URB			-0.358	0.004	-0.392	0.001
COSTA RUR			-1.388	0.000	-1.456	0.000
SIERRA URB			0.232	0.094		
SIERRA RUR			-0.172	0.277	-0.571	0.000
SELVA RUR			-0.991	0.000	-1.113	0.000
INDEXPLAB	0.069	0.000	0.079	0.000	0.070	0.000
COLEG ESTA	0.601	0.000	0.802	0.000	0.502	0.009
OBRAERO	0.444	0.000	0.386	0.001	0.485	0.000
MENOS100PE	0.351	0.000	0.333	0.011	0.238	0.026
OTROEMPL	-0.223	0.01	-0.279	0.019		
INPERCAM	-0.011	0.000	-0.009	0.000	-0.011	0.000
MIERPERHO	0.515	0.000	0.539	0.000	0.482	0.000
PETHOGAR	-0.318	0.000	-0.322	0.000	-0.250	0.000
CONVIVIE	0.267	0.000	0.320	0.002	0.116	0.243
SEHGDVIV	-0.416	0.000	-0.369	0.003	-0.515	0.000
NOSEHIG	0.286	0.000	0.317	0.010	0.223	0.048
AGUAVIV	-0.139	0.000	-0.235	0.023		
ALUMKERO	0.593	0.000	0.666	0.000	0.639	0.000
EMPVIVI	-0.284	0.000	-0.143	0.189		
CONST	1.236	0.000	1.127	0.008	1.394	0.000

Observaciones	8170	Observaciones	8170	Observaciones	8202
LR chi-square	5604.5	Tam. Poblac	2924940	Tam. Poblac	2909895
Prob > chi2	0.000	F(21, 2593)	38.400	F(17, 2607)	49.530
Seudo R ajust	0.490	Prob > F	0.000	Prob > F	0.000
CC	84.010	CC	84.1	CC	84.6

Significancia: 5%

Fuente: ENAHO IV TRIM. 2001

Elaboración: CIDE

para el primer modelo; en el caso de los modelos 2 y 3, aumentan alrededor de un 80% y un 50% respectivamente. Como ya se habrá podido percibir, el poseer coeficientes positivos sobre el modelo logit de probabilidad beneficia en términos absolutos a la probabilidad de ser pobre comparada con su complementaria en una razón de probabilidades.

En el aspecto de la situación ocupacional en que se encuentra el jefe de hogar, podemos mencionar que estar en condición de obrero aumenta el logit de su probabilidad de ser pobre entre un 39% y un 49% en los 2 últimos modelos respectivamente. Una situación muy distinta poseería el jefe de hogar que tendría otra ocupación, donde el efecto en el logit de su probabilidad de ser pobre descende entre un 22% y un 27% en los 2 primeros modelos, manteniendo constante el efecto del resto de variables, mientras que en el tercero resultaría no significativo.

La idea de una reducción del tamaño de la familia para reducir su probabilidad de ser pobre es sustentada aquí, en función del hecho de que un aumento de miembros en el hogar estaría contribuyendo entre un 52% y un 54% en los 2 primeros modelos y en un 48% en el tercer modelo, a elevar el logit de su probabilidad, siempre que se mantenga el efecto del resto de variables constantes.

Con respecto a activos públicos a los que podría acceder el hogar, la no tenencia de acceso a servicios higiénicos en su vivienda, aparte de detectar una condición precaria, provoca un incremento en su logit de probabilidad de alrededor de un 29% y 32% en el primer y segundo modelo respectivamente, en el caso del

tercero, el incremento es de alrededor de un 22%; distinto el panorama y más aún de seguro en su condición, cuando este posee conexión a red pública de alcantarillado dentro de su vivienda, reduciendo así su logit en un 41.6%. (MODELO N° 1), hasta un 36.9% (MODELO N° 2) y hasta un 51.5% (MODELO N° 3), con lo cual se sigue ratificando a la condición de inaccesibilidad a servicios públicos como un factor muy importante dentro de la condición de pobreza que puede poseer el jefe de hogar.

Observando la parte estadística el primer modelo presenta estimaciones pseudo - R^2 del 50% y un porcentaje de clasificación de 84.01%. En el caso de los modelos 2 y 3, al incluir los factores de expansión tratamos con tamaños de población de alrededor de 2'924,940 y 2'909,895 hogares, para el primero y segundo respectivamente; el porcentaje de correcta clasificación para estos modelos fluctúa entre un 84.1 (modelo n°2) y un 84.6 (modelo n°3).

Debemos hacer hincapié en un aspecto que puede generar discusión en el sentido teórico estadístico estricto. En el primer modelo existen influencias negativas en la probabilidad, es decir, que si el jefe de hogar pertenece al área rural su probabilidad de ser pobre decrece en un 99.6%, mientras que si este pertenece a las regiones naturales de la sierra y de la selva sus probabilidades aumentan alrededor de un 85% y un 57%. Este resultado es producido por la multicolinealidad existente en la relación entre región natural y área de residencia del jefe de hogar, cuya relación fue probada en la parte de análisis descriptivo [CUADRO VIII.10], Una de las medidas

tomadas para contrarrestar este efecto fue el de retirar esas variables y colocar otras indicadoras que representan a las regiones naturales y el área de residencia del jefe de hogar, vale decir costa urbana, sierra rural, etc, por mencionar algunos ejemplos; lo que sucede allí con las probabilidades de ser pobre es que estas se reducirían, es decir, que las otras regiones tienen comportamientos diferenciadores con respecto a ingresos y gastos, sueldos, salarios, condiciones de acceso muy por debajo de los que se presentan en Lima. En tal sentido, las características diferenciadoras más específicas de cada región determinan aplicaciones de modelos inherentes a cada una.

Dadas las características de esta investigación metodológica, no se ha profundizado aún más en la generación de tales modelos, pues los objetivos buscados están más orientados hacia la parte de diferenciación de enfoques y métodos de estimación de los modelos logit y probit, y más específicamente hacia el enfoque de proporciones individuales en estos dos últimos modelos estimados.

EL MODELO PROBIT

De la misma manera que en el modelo logit, realizamos la estimación de 3 modelos, donde en el primer caso no se emplean los efectos de diseño provenientes de la encuesta compleja que es la ENAHO. En el caso de los modelos n° 2 y n° 3, como se mencionó líneas antes, en el primero se trabaja con la muestra aleatoria con la que se construyó la metodología (8233 hogares) y que constituye el 50% del total de observaciones, para que luego dicha estimación sea evaluada en cuanto a

validez y confiabilidad con el resto de casos (8282 hogares), y aplicando en cada uno de ellos los factores de ponderación correspondientes.

Siguiendo el enfoque de estimación máximo-verosímil, procedimos a estimar el modelo de probabilidad de un jefe de hogar pobre apoyados en la suposición que la distribución de los errores sigue una distribución normal debido a la gran cantidad de observaciones manejadas. Podemos observar [CUADRO VIII.23], que no tener nivel de educación alguno tiene un efecto positivo sobre la probabilidad de ser un jefe de hogar pobre, aumentando el probit alrededor de un 53% y un 68% en los 2 primeros modelos y en un 42.3% en el tercero.

Planteamos aquí la misma idea de discusión del modelo logit, acerca de la multicolinealidad con respecto al área y regiones naturales de residencia, sobre qué medida asumir, para el caso del modelo n°1 sin aplicación de factores de expansión.

La dimensión regional es importante, es decir, el hecho de vivir en zonas urbanas como rurales de la costa sierra y selva (modelos n° 2 y n° 3), controlando por el resto de variables (algunas de las cuales tratan de enfocarse en la capacidad de consumo y de generación de ingresos del hogar), estarían reduciendo la probabilidad de ser pobre, pues los ingresos por trabajo serían superiores a los de las otras regiones y el monto de la canasta básica en zonas rurales es tremendamente menor que en zonas urbanas y sobre todo comparadas con la categoría base que sería Lima. Ello estaría explicando por qué el solo hecho de vivir en la costa rural, luego de controlar el efecto por el resto de variables indicadas, reduce el probit de la

Cuadro VIII.23
Incidencia en la probabilidad de ser pobre a nivel nacional

EPOBRE	MODELO N° 1		MODELO N° 2		MODELO N° 3			
	Coef.	P > z	Coef.	P > t	Coef.	P > t		
SINNIVEL	0.534	0.000	0.683	0.000	0.423	0.002		
EDAD	-0.045	0.000	-0.050	0.000	-0.048	0.000		
ARURAL	-0.540	0.000						
SIERRA	0.500	0.000						
SELVA	0.307	0.000						
COSTA URB			-0.232	0.001	-0.233	0.000		
COSTA RUR			-0.830	0.000	-0.805	0.000		
SIERRA URB								
SIERRA RUR			-0.111	0.175	-0.224	0.013		
SELVA RUR			-0.602	0.000	-0.616	0.000		
INDEXPLAB	0.043	0.000	0.048	0.000	0.048	0.000		
COLEGESTA	0.333	0.001	0.469	0.000	0.284	0.010		
OBREO	0.206	0.000	0.178	0.008	0.250	0.000		
MENOS100PE	0.243	0.000	0.222	0.003	0.168	0.004		
OTROEMPL	-0.122	0.010	-0.156	0.026				
INPERCAM	-0.005	0.000	-0.004	0.000	-0.005	0.000		
MIEPERHO	0.304	0.000	0.309	0.000	0.304	0.000		
PETHOGAR	-0.200	0.000	-0.195	0.000	-0.180	0.000		
CONVIVIE	0.151	0.001	0.187	0.002	0.091	0.121		
SEHIGDIVI	-0.284	0.000	-0.282	0.000	-0.328	0.000		
NOSEHIG	0.168	0.001	0.184	0.009	0.139	0.031		
AGUAVIV	-0.073	0.117	-0.107	0.080				
ALUMKERO	0.357	0.000	0.382	0.000	0.389	0.000		
EMPVIVI	-0.181	0.000	-0.079	0.218				
CONST	0.592	0.002	0.595	0.012	0.590	0.008		
<hr/>								
Observaciones	8170		Observaciones	8170		Observaciones	8202	
LR chi-square	5470.3		Tam. Poblac	2924940		Tam. Poblac	2909895	
Prob > chi2	0.000		F(20, 2594)	45.380		F(17, 2607)	63.940	
Seudo R ajust	0.483		Prob > F	0.000		Prob > F	0.000	
CC	84.0		CC	83.4		CC	84.3	
Significancia: 5%								
Fuente: ENAHO IV TRIM. 2001				Elaboración: CIDE				

probabilidad en un 83% en el segundo modelo y en un 80.5% en el caso del tercero.

En el caso de la sierra rural se reduciría el probit de la probabilidad en un 11% y 22.4% en el segundo y tercer modelo respectivamente, comparadas siempre con la categoría base que es Lima. Donde se están apreciando brechas bastante notorias es en la selva rural, que comparadas con Lima se estaría reduciendo el probit de su probabilidad en 60.2% (modelo n° 2) y 61.6% (modelo n°3).

Algunos resultados obtenidos en el mismo sentido que en el modelo Logit anteriormente estimado recibirían igual interpretación, como el hecho de destinar

dentro de la vivienda un espacio para obtener ingresos, reduce el probit de su probabilidad entre un 18.1% (modelo n°1, sin considerar factor de expansión) y 11.7% (modelo n° 2, considerando factores de expansión), pero que en el caso del tercer modelo resulta siendo no significativo. Además se mantiene la idea de que otro empleo para obtener ingresos, reduce su probit de probabilidad en una pequeña medida, alrededor de un 12.2% (modelo n° 1) y 15.6% (modelo n° 2). Nuevamente en el caso del tercer modelo resultaría no significativa.

Queda definitivamente claro que el acceso a activos públicos beneficia al jefe de hogar y al hogar en su conjunto en una reducción de su probabilidad de ser pobre,

reduciéndola en 28.2% (MODELO N° 2) y hasta 32.8% (MODELO N° 3) cuando se tiene acceso a servicios de alcantarillado dentro de vivienda, y un 10.7% (MODELO N° 2) cuando se tiene abastecimiento de agua dentro de la vivienda del hogar, siendo no significativa esta variable en el caso del modelo n° 3.

Según lo anteriormente mostrado, la elección por alguno de los dos modelos conforme al enfoque de estimación de observaciones individuales no es diferenciable. Sobre todo si recalamos la idea que la gran cantidad de observaciones tratadas sesgan la idea de la distribución

normal de los errores, podría primar como criterio para la elección del modelo probit según este enfoque.

Desde el punto de vista práctico, la elección del modelo logit y de la correspondiente regresión logística habilitaría una mayor cantidad de herramientas de control de la bondad de ajuste del modelo estimado y por ende una mejor validación de éste, en tal sentido se propondría como el más idóneo, para el cálculo del modelo de probabilidad del jefe de hogar pobre del Perú, en el año 2001.

IX. CONCLUSIONES

El presente trabajo de investigación aborda los modelos logit y probit aplicados en la investigación social para el caso de la pobreza en el Perú, durante el año 2001, desde la formulación de la condición de la pobreza en el Perú y algunos activos pertenecientes a los jefes de hogar y al hogar que pertenece bajo esta condición. Para luego sentar las bases (enfoque teórico) sobre los cuales se construyó los modelos y se adaptaron a este caso en particular; de allí que se puede concluir, desde el punto de vista práctico, que aspectos como el acceso a activos públicos benefician al desarrollo de activos privados de los jefes de hogar a través de un decrecimiento en su probabilidad de ser pobre. Demás esta confirmar el aumento en una gran medida sobre la probabilidad de ser pobre cuando el jefe de hogar no tiene nivel de educación alguno, etc.

1. Con respecto a la hipótesis que sostiene que los factores de naturaleza cuantitativa como el ingreso per cápita, a través de sus deciles de ingreso, etc; ajustarían correctamente un modelo para la probabilidad de pobreza en el jefe de hogar del Perú se rechazaría, en primer lugar, porque expresado a través del estadístico de bondad de ajuste Pearson X^2 mostraría dudas acerca de las probabilidades generadas, y por ende no se convertiría en el más idóneo sobre el cual trabajar. En segundo lugar, dada la naturaleza del fenómeno de la pobreza en estudio, excluirla de sus factores de naturaleza

cualitativa, los cuales no hacen posible la viabilidad de este enfoque según los indicadores existentes, resultaría contraproducente y por tal motivo el modelo probit con variable dependiente dicotómica con enfoque de proporciones muestrales no es el más adecuado para la estimación de la probabilidad de ser pobre del jefe de hogar en función a sus factores determinantes.

2. Con respecto a la hipótesis que sostiene que los factores explicativos de la pobreza de los jefes de hogar del Perú de naturaleza cualitativa y cuantitativa exclusivos del jefe de hogar como el nivel de educación, el tipo de colegio de estudio, la categoría ocupacional, el tamaño de la firma donde labora, la tenencia de otro empleo, el estado civil, su edad, su indicador de experiencia laboral; en combinación con los factores de naturaleza cualitativa y cuantitativa exclusivos del hogar como el indicador de si el hogar dedica un espacio del hogar a generación de ingresos, la cantidad de miembros en el hogar, la cantidad de miembros pertenecientes a la PET, el ingreso per cápita mensual, el acceso a activos públicos de agua y desagüe, etc, no generan modelos correctamente ajustados sería rechazada porque, en primer lugar, las probabilidades de correcta clasificación para los modelos estimados se encuentran alrededor del 84% en el

modelo logit y en el modelo probit, ambos con enfoque de observaciones individuales. En segundo lugar, la obtención de efectos significativos en los factores de naturaleza cualitativa y cuantitativa mencionados líneas antes y exclusivos al jefe y al hogar, a través de las estimaciones de sus respectivos estadísticos de Wald y significancias

obtenidas, en el caso del modelo logit, y de sus respectivos efectos en la transformación probit, en el caso del modelo del mismo nombre, viabilizan una explicación del fenómeno de la pobreza en función a los determinantes antes mencionados.

X. RECOMENDACIONES

- Si se estuviese interesado en el análisis probit con variable dependiente dicotómica desde el enfoque de proporciones para el estudio de la pobreza bajo ciertos factores como el nivel de educación, el área de procedencia, acceso a ciertos tipos de activos públicos deberían en primer lugar reducirse las cantidades de observaciones a estudiar para no afectar a los estimadores, y en segundo lugar, tratar de generar variables-estímulos (cruce de variables explicativas) que puedan seguir un enfoque cuasi-cuantitativo, de tal manera de que se pueda determinar la medida de cambio exacto en la obtención de la condición de pobreza. En tal sentido y dada la limitación de este enfoque, se puede pasar al estudio de niveles más avanzados, como los modelos probit con variable dependiente ordinal o modelos con enfoques multinomiales.
- Si además de intentar obtener aportes de variables explicativas, está intentando predecir un modelo de clasificación idóneo a sus criterios, el modelo logit ampliamente difundido aporta una mayor variedad de herramientas de validación de la bondad de ajuste del modelo y no dejaría de lado su función discriminadora, beneficiosa en este caso en el que la variable dependiente posee solo dos categorías.
- Si bien no ha sido empíricamente mostrado en esta investigación, sugiero la inclusión de una mayor cantidad de variables de naturaleza cuantitativa en modelos probit con enfoque de observaciones individuales, pues son variables más idóneas a ser sometidas a pruebas de normalidad y análisis más complejos; además porque su participación en bloques nos permitiría obtener cambios exactos a los cuales responderían los jefes de hogar pobres. No deben dejar de considerarse las variables de naturaleza cualitativa, pues son fundamentales para la explicación de fenómenos sociales y podrían plantearse indicadores más idóneos para captar los cambios de naturaleza cualitativa de los individuos, y no solo expresarse en presencia o ausencia de factores de riesgo.

XI. BIBLIOGRAFÍA

- "CATEGORICAL DATA ANALYSIS",
Agresti, Alan.
Florida, Wiley, 1990.
- "ANÁLISIS ECONOMÉTRICO".
Green, William H.
Prentice Hall. 1998.
- "ESTADÍSTICA MULTIVARIANTE EN LAS CIENCIAS DE LA VIDA".
Carrasco, José Luis; Hernán, Miguel Angel.
CIBEST. España, Editorial Ciencia, 1993.
- "ECONOMETRÍA",
Gujarati, Damodár N.
Colombia, McGraw-HILL, 1997
- "MÁS ALLÁ DE LA FOCALIZACIÓN, RIESGOS DE LA LUCHA CONTRA LA POBREZA EN EL PERÚ"
Chacaltana J., Juan.
PERÚ, Consorcio de Investigación Económica y Social, 2001.
- "METODOLOGÍA DE LA INVESTIGACIÓN".
Hernández, Roberto; Fernández, Carlos; Baptista, Pilar.
México. McGraw Hill, 1994.
- "IMPACTO DE LOS SERVICIOS PÚBLICOS DE SALUD SOBRE LA PRODUCTIVIDAD Y LA POBREZA",
Cortez Valdivia, Rafael.
Lima, INEI, 2000.
- "MÉTODOS MULTIVARIADOS APLICADOS AL ANÁLISIS DE DATOS".
Johnson,
México, ITP, 2000.
- "BIOESTADÍSTICA".
Díaz, Gabriela; Gunther, Bruno.
Chile, Mediterráneo, 1994.
- "ESTUDIO SOBRE DETERMINANTES DEL ACCESO A LOS SERVICIOS DE SALUD EN EL PERÚ",
Lama More, Antonio.
Lima, INEI, 2000.
- "LOS ACTIVOS DE LOS POBRES EN EL PERÚ".
Escobal, Javier; Saavedra, Jaime; Torero, Máximo.
Lima, GRADE. 1998.
- "LIMITED-DEPENDENT AND QUALITATIVE VARIABLES IN ECONOMETRICS".
Maddala G. S.
New York, Cambridge University Press, 1996.
- "EL ANÁLISIS DE DATOS EN MÉTODO DE SELECCIÓN DICOTÓMICA DE LA VARIABLE CONTINGENTE",
Fasciolo, Graciela.
Mendoza, 1997.
- "SPSS ADVANCE STATISTICS",
Majira J. Norusis.
Chicago, SPSS Inc, 1994

- "METODOLOGÍA, GUÍA PARA ELABORAR DISEÑOS DE INVESTIGACIÓN EN CIENCIAS ECONÓMICAS, CONTABLES Y ADMINISTRATIVAS",
Méndez Alvarez, Carlos.
Colombia, Editorial McGraw - HILL., 1995.
- "MULTIVARIATE ANALYSIS: SOCIOLOGY 203",
Department of Sociology, University of California, Riverside.
- "EVALUACIÓN DE LOS PROGRAMAS DE COMPLEMENTACIÓN ALIMENTARIA, GASTO EN ALIMENTOS Y CONDICIONES DE VIDA EN EL PERÚ EN EL PERÍODO 1998-1999",
Navarro Levano, José Carlos.
Lima, INEI, 2001.
- "EL RETO 2001, COMPETIR Y CREAR EMPLEO",
Ortiz de Zevallos, Felipe y Kuczynski, Pedro Pablo.
Lima, El Comercio, 2001.
- "LA POBREZA EN ARGENTINA: UNA COMPARACIÓN ENTRE REGIONES DISÍMILES, BUENOS AIRES, 2DA REUNIÓN ANUAL SOBRE POBREZA Y DISTRIBUCIÓN DEL INGRESO- LACEA/BID/BM/"
Paz, Jorge A.
Universidad Torcuato Ditella, 2001.
- "POBREZA Y ECONOMÍA SOCIAL - ANÁLISIS DE UNA ENCUESTA ENNIV- 1997.LA EDUCACIÓN Y LA PROBABILIDAD DE SER POBRE EN EL PERÚ DE HOY, LA APLICACIÓN DE UN MODELO PROBIT DE MÁXIMA VEROSIMILITUD",
Shack Yalta, Nelson Eduardo.
Perú, DESA, 1999.
- "ANÁLISIS ESTADÍSTICO CON SPSS PARA WINDOWS".
Visauta Vinacua, B.
España, Mc Graw Hill, 1998.

XII. Anexos

ANEXO 1. Informe metodológico

ANEXO 2. Indices de ecuaciones, cuadros y gráficos

ANEXO 1

INFORME METODOLÓGICO

Fuentes de información utilizada.

Para cumplir con los objetivos de la investigación se tomó en consideración la información de la ENCUESTA NACIONAL DE HOGARES 2001 - IV TRIMESTRE, aquí mencionamos algunas de sus características, que son de importancia relevante para nuestra investigación y que nos permitieron comprobar algunas ideas referentes a los factores determinantes de la pobreza en el Perú, y como estos beneficiarían o afectarían, los criterios de utilización de modelos logit y probit. Para dicho propósito se analizarán variables relacionadas con el Jefe de Hogar, en sus aspectos sociodemográficos y de inserción ocupacional.

Como sabemos, el fenómeno de la pobreza no afecta solo a los jefes de hogar, esta trae consigo una afectación al conjunto familiar, por lo tanto usar solo como unidad de análisis al individuo y no considerar al hogar, puede llevar a significativos sesgos sobrestimados al respecto, pero que constituye un punto de partida fundamental en razón al comportamiento de la sociedad y específicamente de la familia peruana, y del rol que este juega dentro y sobre el desarrollo de su hogar, además de tener en cuenta que estamos desarrollando una investigación que devela una metodología sobre los modelos mencionados anteriormente.

POBLACIÓN OBJETIVO: Fueron las viviendas particulares y sus residentes habituales (miembros permanentes del hogar), excluyéndose a los residentes en

viviendas colectivas (hoteles, cárceles, asilos, etc).

MARCO DE LA MUESTRA: La muestra es probabilística, de áreas, estratificada, multietápica e independiente en cada departamento.

La muestra es probabilística porque las unidades han sido seleccionadas mediante métodos al azar, lo cual permite efectuar inferencias a la población en base a la teoría de probabilidades.

La muestra es de áreas, porque la probabilidad de la población de ser seleccionada, está asociada a áreas geográficas.

La muestra es estratificada, porque previamente a la selección, la población se ha dividido en estratos, con el objeto de mejorar su representatividad.

En la primera y segunda etapa se utiliza la selección sistemática con probabilidad proporcional al tamaño (PPT) de viviendas.

En la última etapa (selección de las viviendas) la selección es sistemática simple con arranque aleatorio.

COBERTURA GEOGRAFICA: Nacional, Urbana y Rural (24 departamentos y la Provincia Constitucional del Callao).

TAMAÑO DE MUESTRA: El tamaño total de la muestra en el ámbito nacional es de 18,863 viviendas particulares, dentro de los cuales tomamos a 16515 Jefes de Hogar y sus respectivos hogares.

METODOLOGÍA PARA LA INCLUSIÓN DE VARIABLES EN LA ESTIMACIÓN DE MODELO LOGIT DE POBREZA

En primer lugar se emplea un modelo semiautomático de selección de variables (Forward Stepwise Wald).

Block 0: Beginning Block

En el paso 0 se cuenta con un modelo configurado solo por el término

independiente. Sobre el se estudiará la adición o no de las demás variables, con un nivel de significación tanto de entrada del 0.05 y de salida 0.10. En [CUADRO A.1] de historia de iteración obtenemos un -2 Log likelihood = 11325.900 resultante de su función de verosimilitud y un valor para la constante de -0.008.

Cuadro A.1 Iteration History a,b,c

Iteration	-2 Log likelihood	Coefficients
		Constant
Step 0 1	11325.900	-.008

- a. Constant is included in the model.
- b. Initial -2 Log Likelihood: 11325.900
- c. Estimation terminated at iteration number 1 because log-likelihood decreased by less than .010 percent.

A partir de aquí se debe comparar el valor que se obtenga en -2 Log likelihood en los modelos con las nuevas variables incluídas, de tal manera, que contrastemos la hipótesis nula de que el modelo sin covariables (factores de riesgo de pobreza), es tan bueno como el modelo que las contiene. La importancia del modelo y del conjunto de variables significativas es validado posteriormente con el test de coeficientes del modelo (Ómnibus Test for Model Coefficients), [CUADRO A.6] en

este caso, que rechaza o no dicha hipótesis.

El modelo con la constante posee una tabla de clasificación del 50.2%, esto quiere decirnos que solo la constante, ya esta clasificando como pobres a 5 de cada 10, apreciable en el [CUADRO A.2]. Pero el valor asignado al coeficiente de la constante es de -0.008, el cual no posee efecto significativo alguno, pues su significancia es de 0.723 [CUADRO A.3].

Cuadro A.2 Classification Table a,b

			Predicted		
			estado pobreza01		Percentage Correct
			no pobre	pobre	
Step 0	estado pobreza01	no pobre	4101	0	100.0
		pobre	4069	0	.0
Overall Percentage					50.2

- a. Constant is included in the model.
- b. The cut value is .500

También viene acompañado de su error estándar y el χ^2 de Wald, que prueba en este caso la significancia individual que en esta caso particular no se da; además de mostrar el Odds Ratio (‘Razón de

probabilidades’), que será fundamental en la determinación de la cantidad de veces que un jefe de hogar puede ser más pobre con respecto a no serlo respecto de la presencia o no de un factor.

Cuadro A.3 Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-.008	.022	.125	1	.723	.992

Antes de terminar este paso y proceder al siguiente, observemos [CUADRO A.4], se realiza un análisis de las variables todavía fuera de la ecuación a través de su p-valor o significancia, que indica la importancia relativa que cada una de ellas tendría en caso de entrar al modelo descrito. Se seleccionará, entre las variables aquella cuya significancia se encuentre bajo el nivel 0.05, tomando también en

consideración su score, en este nuestro caso, sería DSHDV CON $P = 0.000$, es decir la variable indicadora sobre abastecimiento de red de alcantarillado dentro de la vivienda.

Así finaliza el paso 0 del Forward Stepwise de Wald que se ha limitado a estudiar el modelo de partida.

Cuadro A.4 Variables not in the Equation ^a

Step	Variables	Score	df	Sig.
0	DSINNIV	188.492	1	.000
	P208A	55.014	1	.000
	DCOURB	223.726	1	.000
	DCORU	.152	1	.697
	DSIEUR	15.617	1	.000
	DSIERU	836.127	1	.000
	DSELUR	.371	1	.542
	DSELRU	50.785	1	.000
	DLIUR	446.997	1	.000
	EXPER1	8.833	1	.003
	DCOLEGIO	.075	1	.785
	DOBRERO	31.997	1	.000
	DME100	528.833	1	.000
	DOEMP	21.192	1	.000
	INPERCAM	860.534	1	.000
	MIEPERHO	819.967	1	.000
	PETHOGAR	33.780	1	.000
	DCONVI	180.384	1	.000
	DSHDV	1307.200	1	.000
	DSHNO	674.480	1	.000
	DAGUDV	669.772	1	.000
	P1122	1114.157	1	.000
	P115	57.575	1	.000

a. Residual Chi-Squares are not computed because of redundancies.

Block 1: Method = Forward Stepwise (Wald)

PASO 1: Variable ingresada DSHDV

Al ingresar la variable indicadora sobre abastecimiento de alcantarillado dentro de la vivienda, vemos que obtenemos un -2Log likelihood = 9973.258, ver [CUADRO A.5], obtenido este de la

función de máxima versosimilitud, menor con respecto al modelo solo con constante (11325.900), además tenemos un Cox & Snell-R² de 0.153 y un valor de Nagelker 0.203 que nos indica el grado de variabilidad explicada por el modelo, en esta caso demasiado baja, y que por ese lado podemos seguir intentando buscar un modelo más ideoneo.

Cuadro A.5 Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	9973.258	.153	.203
20	5683.192	.499	.665

Se observa en el [CUADRO A.6] siguiente (Test Omnibus) o prueba de los coeficientes del modelo, que arroja efectos significativos sobre el ingreso de dicha variable y por lo tanto rechazamos la hipótesis nula de que el modelo sólo con la constante es igual de bueno que el

modelo con la constante y DSHDV (tipo de acceso a alcantarillado), esto en función de la chi-square obtenida que es de 1352.641 es mayor comparada con aquella chi-square (1 g.l., 5%)=3.84, en tal sentido el modelo obtenido hasta ese momento es el más adecuado.

Cuadro A.6 Omnibus Test of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	1352.641	1	.000
	Block	1352.641	1	.000
	Model	1352.641	1	.000
Step 20	Step	4.126	1	.042
	Block	5642.707	20	.000
	Model	5642.707	20	.000

El modelo estimado sería: $\ln(p/1-p) = \beta_0 + \beta_1 \cdot \text{DSHDV} = 0.668 - 1.753 \cdot \text{DSHDV}$, apreciable en el [CUADRO A.9], donde ya apreciamos que el tener conexión a red

pública de alcantarillado dentro de la vivienda produce un efecto negativo sobre el logit de la probabilidad de ser pobre.

Cuadro A.7 Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
2	21.938	8	.005
20	525.158	8	.000

Si bien el Test de Hosmer y Lemershow [CUADRO A.7] deberían ser atendidos en el siguiente paso que incluye una nueva variable, este no será tratado en virtud de que el resto de procedimientos son similares en cada iteración, por eso solo nos detendremos aquí a afirmar que frente a un valor de chi-square de 21.938 y una significación 0.005 esta última se va reduciendo a medida que se van aumentando más variables en el modelo. La comparación se hace respecto a una chi-square (8 g.l., 5%)=15.51 obtenido de tabla de aquí el test rechaza la hipótesis de nulidad.

Volviendo al paso 1 encontramos que este modelo ya tiene un porcentaje de clasificación general del 69.5% obtenido de su tabla [CUADRO A.8], de aquí mismo la subclasificación hecha para la condición de pobreza de los jefes de hogar pobres obtiene un 79.8% de subclasificación.

Con respecto a los estimadores de los coeficientes tenemos que el término

independiente 0.688 y el coeficiente del acceso a red pública de alcantarillado dentro de la vivienda -1.753 formulan el modelo de probabilidad de pobreza [CUADRO A.9], el cual por alguna de las razones de índole estadística y empírica se queda corto al intentar brindar explicaciones del fenómeno en su conjunto y proceder al análisis de los siguientes pasos de iteración.

De la misma manera que en el paso 0, se analizan las variables que están fuera del modelo en este paso (1), antes de proceder al siguiente [CUADRO A.10], viendo su significancia (p-valor < 0.005), para su ingreso al modelo y además su score, en tal sentido algunas de las candidatas a ingresar serán MIEPERHO (0.000, 962.249), INPERCAM (0.000, 561.907), P1122 (0.000, 393.466), pero como es evidente de la ingresante en el siguiente paso será el total de miembros en el hogar (MIEPERHO).

Cuadro A.8 Classification Table ^a

Observed			Predicted		
			estado pobreza01		Percentage Correct
			no pobre	pobre	
Step 1	estado pobreza01	no pobre	2436	1665	59.4
		pobre	823	3246	79.8
Overall Percentage					69.5
Step 20	estado pobreza01	no pobre	3400	701	82.9
		pobre	519	3550	87.2
Overall Percentage					85.1

a. The cut value is .500

El mismo análisis se realiza en cada uno de los siguientes pasos de iteración, en tal sentido, para ser una lectura más dinámica se presenta el primer paso de iteración y el último donde se detiene la estimación del modelo.

PASO 20: Variable ingresada DAGUDV La variable ingresante en este paso, ver [CUADRO A.9], es si el tipo de abastecimiento de agua en la vivienda es a través de red pública. Las hipótesis acerca de la validación de un adecuado modelo

frente a uno que no contiene factores significativos es fácil de verificar en el Test Ómnibus, ver [CUADRO A.7], y que permite aceptar este último como mejor modelo frente a aquel que solo contiene el término constante.

Confirmada la idea anterior se puede redondear esta, es decir, confirmar el hecho de que nos encontramos frente a un modelo con un desajuste menor, además de obtener un grado de clasificación general de alrededor del 85%, ver [CUADRO A.8] .

Este será el último paso ha realizarse en la estimación del modelo de probabilidad de ser jefe de hogar pobre. Primero dejando en claro que no es el único y definitivo sobre el cual trabajar para estructurar

explicaciones de este fenómeno multidimensional, esto desde el punto de vista empírico y porque además debemos de dejar en claro que si bien nos aproximamos en una buena medida a la realidad con la herramienta estadística esta no resulta ser determinante y menos reemplazante de esta.

Segundo, que desde el punto de vista estadístico, ninguna de las variables restantes tiene una significación adecuada para ingresar al modelo [CUADRO A.10], en tal sentido, la estimación es detenida, y se empiezan a realizar los análisis mostrados en capítulos anteriores en el mismo sentido de haber ingresado la primera variable, pues este ha sido un proceso continuo durante toda la iteración.

MODELOS EN CADA PASO DE ITERACIÓN

Cuadro A.9 Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
							Lower	Upper
Step 1	DSHDV	-1.753	.050	1212.252	1	.000	.173	.157
	Constant	.668	.030	490.482	1	.000	1.950	
Step 20	DSINNIV	.911	.231	15.548	1	.000	2.488	1.582
	P208A	-.072	.011	41.221	1	.000	.931	.910
	DCORU	-1.189	.167	50.439	1	.000	.304	.219
	DSIEUR	.570	.098	33.642	1	.000	1.768	1.459
	DSELUR	.899	.121	55.483	1	.000	2.458	1.940
	DSELRU	-.807	.125	41.844	1	.000	.446	.349
	EXPER1	.070	.010	49.495	1	.000	1.072	1.051
	DCOLEGIO	.584	.189	9.540	1	.002	1.793	1.238
	DOBRERO	.451	.092	23.940	1	.000	1.570	1.311
	DME100	.324	.095	11.606	1	.001	1.383	1.148
	DOEMP	-.226	.085	7.090	1	.008	.798	.675
	INPERCAM	-.011	.000	876.865	1	.000	.989	.988
	MIEPERHO	.517	.031	283.825	1	.000	1.677	1.579
	PETHOGAR	-.313	.040	60.764	1	.000	.731	.676
	DCONVI	.269	.080	11.274	1	.001	1.308	1.118
	DSHDV	-.388	.093	17.414	1	.000	.678	.565
	DSHNO	.270	.089	9.281	1	.002	1.310	1.101
	DAGUDV	-.168	.082	4.137	1	.042	.846	.719
	P1122	.627	.092	46.847	1	.000	1.873	1.565
	P115	.288	.083	11.916	1	.001	1.334	1.133
	Constant	.761	.379	4.035	1	.045	2.141	

a. Variable(s) entered on step 1: DSHDV.

t. Variable(s) entered on step 20: DAGUDV.

VARIABLES EXCLUIDAS EN CADA PASO DE ITERACIÓN

Cuadro A.10 Variables not in the Equation

			Score	df	Sig.
Step 1	Variables	DSINNIV	69.683	1	.000
		P208A	13.868	1	.000
		DCOURB	29.125	1	.000
		DCORU	39.325	1	.000
		DSIEUR	.985	1	.321
		DSIERU	259.136	1	.000
		DSELUR	.699	1	.403
		DSELRU	1.023	1	.312
		DLIUR	125.821	1	.000
		EXPER1	8.835	1	.003
		DCOLEGIO	.090	1	.764
		DOBRERO	13.894	1	.000
		DME100	206.545	1	.000
		DOEMP	3.052	1	.081
		INPERCAM	561.907	1	.000
		MIEPERHO	962.249	1	.000
		PETHOGAR	158.062	1	.000
		DCONVI	79.704	1	.000
		DSHNO	107.207	1	.000
		DAGUDV	33.928	1	.000
Step 20	Variables	P1122	393.466	1	.000
		P115	16.562	1	.000
		DCOURB	1.150	1	.283
		DSIERU	.005	1	.942
		DLIUR	1.496	1	.221

a. Residual Chi-Squares are not computed because of redundancies.

ANEXO 2

INDICES DE ECUACIONES, CUADROS Y GRAFICOS

INDICE DE ECUACIONES

CAPÍTULO IV

ECUACIÓN IV.1	Indicadores de Pobreza (Foster-Greer-Thordecke)
ECUACIÓN IV.2	Función de Distribución Binomial
ECUACIÓN IV.3	Probabilidad de ser pobre del jefe de hogar en función a una FDA.
ECUACIÓN IV.4	Modelo de regresión lineal asociado a una FDA
ECUACIÓN IV.5	Modelo de probabilidad lineal
ECUACIÓN IV.6	Heterocedasticidad del modelo lineal de probabilidad.
ECUACIÓN IV.7	Probabilidad de poseer una condición o presencia de fenómeno asociada a una FDA
ECUACIÓN IV.8	Probabilidad asociada a la Función de Distribución Logística
ECUACIÓN IV.9	Transformación de la probabilidad de FDA Logística
ECUACIÓN IV.10	Probabilidad del evento complementario, $Y=0$, o ausencia de la condición o fenómeno.
ECUACIÓN IV.11	Razón de Probabilidades ('Odds ratio')
ECUACIÓN IV.12	Modelo Logit. Logaritmo de la razón de probabilidades en función a factores explicativos y sus aportes.
ECUACIÓN IV.13	Probabilidad asociada a la FDA Normal.
ECUACIÓN IV.14	Probabilidad del evento en función a una FDA normal y su representación matemática.
ECUACIÓN IV.15	Índice imperfecto de conveniencia
ECUACIÓN IV.16	Probabilidad asociada al índice imperfecto de conveniencia y una FDA Normal.
ECUACIÓN IV.17	Linealidad del modelo probit
ECUACIÓN IV.18	Efectos marginales asociados a una FDA.
ECUACIÓN IV.19	Efectos marginales para la Función de Distribución Normal
ECUACIÓN IV.20	Derivadas parciales respecto de los coeficientes de los factores en la FDA Logística.
ECUACIÓN IV.21	Efecto marginal para la FDA Logística
ECUACIÓN IV.22	Diferencia de beneficio-coste con una var. Observable

CAPÍTULO VI

ECUACIÓN VI.1	Probabilidad conjunta de un modelo de probabilidad .
ECUACIÓN VI.2	Reformulación de la función de verosimilitud.
ECUACIÓN VI.3	Condiciones de primer orden del problema de maximización.

ECUACIÓN VI.4	Heterocedasticidad de los errores en el enfoque de proporciones muestrales
ECUACIÓN VI.5	Condiciones necesarias de optimalidad de los coeficientes b
ECUACIÓN VI.6	Vector gradiente de la función de verosimilitud
ECUACIÓN VI.7	Matriz de información en función de la matriz Hessiana.
ECUACIÓN VI.8	Método Scoring para estimación de Máxima Verosimilitud.
ECUACIÓN VI.9	Transformación de la variable dependiente Xij en el modelo probit
ECUACIÓN VI.10	Transformación de la variable dependiente Yi en el modelo probit.
ECUACIÓN VI.11	Logaritmo de la verosimilitud de "n" observaciones en el modelo logit
ECUACIÓN VI.12	Vector gradiente de la verosimilitud en el modelo logit
ECUACIÓN VI.13	Matriz de información del modelo logit.
ECUACIÓN VI.14	Transformación de la variable dependiente Xi en el modelo logit.
ECUACIÓN VI.15	Transformación de la variable dependiente Yi en el modelo logit

CAPÍTULO VIII

ECUACIÓN VIII.1	Probabilidad del jefe de hogar pobre con FDA Logística.
ECUACIÓN VIII.2	Estimación del modelo Logit de probabilidad en función de sus factores explicativos.
ECUACIÓN VIII.3	Análisis de un caso particular de jefe de hogar para el modelo logit.
ECUACIÓN VIII.4	Probabilidad asociada al jefe de hogar con un caso específico.
ECUACIÓN VIII.5	Logit de la probabilidad de ser pobre frente a no ser pobre de acuerdo a sus factores explicativos.
ECUACIÓN VIII.6	Ecuación de regresión en términos de Odds Ratios (‘Razones de probabilidad’)
ECUACIÓN VIII.7	Residual estandarizado
ECUACIÓN VIII.8	Deviance de observaciones en presencia de pobreza.
ECUACIÓN VIII.9	Deviance de observaciones en ausencia de pobreza.
ECUACIÓN VIII.10	Logit residual del modelo de prob. de los jefes de hogar.
ECUACIÓN VIII.11	Medida de influencia de cada observación.
ECUACIÓN VIII.12	Cambio en el primer coeficiente al retirar la observación "i".

INDICE DE CUADROS

CAPÍTULO VIII

CUADRO VIII.1	Perfil del jefe de hogar según nivel educativo y estado de pobreza
CUADRO VIII.2	Test Chi-cuadrado (X ²) de independencia entre el nivel educativo y el estado de pobreza.
CUADRO VIII.3	Cuadro de medidas simétricas sobre el grado de relación entre el nivel educativo y el estado de pobreza.

CUADRO VIII.4	Perfil del jefe de hogar según género y estado civil en condición de pobreza.
CUADRO VIII.5	Test Chi-cuadrado (X ²) de independencia entre el género y el estado civil en condición de pobreza del jefe de hogar.
CUADRO VIII.6	Cuadro de medidas simétricas sobre el grado de relación entre el género y el estado civil del jefe de hogar pobre.
CUADRO VIII.7	Perfil del jefe de hogar según género y estado civil en condición de no pobreza.
CUADRO VIII.8	Test Chi-cuadrado (X ²) de independencia entre el género y el estado civil en condición de no pobreza del jefe de hogar.
CUADRO VIII.9	Cuadro de medidas simétricas sobre el grado de relación entre el género y el estado civil del jefe de hogar no pobre.
CUADRO VIII.10	Perfil del jefe de hogar según área y región natural en estado de pobreza
CUADRO VIII.11	Test Chi-cuadrado (X ²) de independencia entre el área y la región natural en condición de no pobreza del jefe de hogar.
CUADRO VIII.12	Cuadro de medidas simétricas sobre el grado de relación entre el área y la región natural en estado de pobreza del jefe de hogar.
CUADRO VIII.13	Cuadro de los años de estudios del jefe de hogar según condición de pobreza
CUADRO VIII.14	Prueba de Levene para la igualdad de varianzas y diferencia de medias en los años de estudios por condición de pobreza.
CUADRO VIII.15	Cuadro de factores explicativos de la pobreza de los jefes de hogar en el modelo logit con enfoque de observaciones individuales.
CUADRO VIII.16	Tabla de correcta clasificación de Jefes de hogar.
CUADRO VIII.17	Historia de Iteraciones
CUADRO VIII.18	Cuadro de estadísticos de bondad de ajuste del modelo
CUADRO VIII.19	Cuadro de evaluación de los coeficientes del modelo.
CUADRO VIII.20	Frecuencia de jefes de hogar según deciles de ingreso y jefes de hogar pobres en cada decil de ingreso para el modelo probit con enfoque de proporciones muestrales.
CUADRO VIII.21	Frecuencias observadas y esperadas de los jefes de hogar en base a modelo probit estimado.
CUADRO VIII.22	Incidencia en la probabilidad de ser pobre a nivel nacional del jefe de hogar según el modelo logit con enfoque de observaciones individuales.
CUADRO VIII.23	Incidencia en la probabilidad de ser pobre a nivel nacional del jefe de hogar según el modelo probit con enfoque de observaciones individuales.

ANEXO

CUADRO A.1	Historia de iteración en el paso 0.
CUADRO A.2	Tabla de clasificación en el paso 0.
CUADRO A.3	Cuadro de variables incluidas en el paso 0 (solo cte.)

CUADRO A.4	Cuadro de variables excluidas en el paso 0.
CUADRO A.5	Cuadro de resumen de modelos en los 1 y 20.
CUADRO A.6	Cuadro de evaluación de los coeficientes de los modelos generados en el paso 1 y el paso 20.
CUADRO A.7	Cuadro de prueba de Hosmer y Lemershow en las iteraciones o pasos 1 y 20.
CUADRO A.8	Tablas de correcta clasificación de los modelos generados en los pasos 1 y 20.
CUADRO A.9	Cuadro de variables incluidas en cada paso de iteración.
CUADRO A.10	Cuadro de variables excluidas en cada paso de iteración.

INDICE DE GRÁFICOS

CAPÍTULO IV

GRÁFICO IV.1	Gráfica de los límites de probabilidad asociado a la presencia del evento.
GRÁFICO IV.2	Gráfica de región asociada a su probabilidad y en función al índice imperfecto o de utilidad.
GRÁFICO IV.3	Gráfica de comparación de la FDA Logística y Normal.
GRÁFICO IV.4	Decisión de compra en base al consumidor

CAPÍTULO VIII

GRÁFICO VIII.1	Gráfica del perfil del nivel educativo según estado de pobreza del jefe de hogar
GRÁFICO VIII.2	Gráfica del perfil del jefe de hogar según género y estado civil en condición de no pobreza.
GRÁFICO VIII.3	Gráfica del perfil del jefe de hogar según género y estado civil en condición de pobreza.
GRÁFICO VIII.4	Gráfica del perfil del jefe de hogar según área y región natural en condición de pobreza.
GRÁFICO VIII.5	Gráfica de la tendencia del promedio de años de estudios del jefe de hogar según estado de pobreza.
GRÁFICO VIII.6	Gráfica de distribución normal y desviaciones respecto de la distribución normal de las Deviances obtenidas para cada JH.
GRÁFICO VIII.7	Gráfica de residuales estandarizados según observaciones.
GRÁFICO VIII.8	Gráfica de Leverages obtenidos según observaciones.
GRÁFICO VIII.9	Gráfica de la influencia de observaciones sobre el coeficiente del indicador sinnivel (sin nivel de educación).
GRÁFICO VIII.10	Gráfica de transformaciones probit según deciles de ingreso