



FACTORES CONCLUYENTES DE LA POBREZA EN BASE A UN MODELO LOGÍSTICO

Tesis para optar al Grado de Magíster en Bioestadística

Alumna: Berta Teitelboim

Profesor Guía: Gabriel Cavada

Abril 2006

INDICE

	Página
1.- Introducción	4
2.- Objetivos	7
2.1 Objetivos Generales	7
2.2 Objetivos Específicos	7
2.3 Hipótesis de Trabajo	8
3.- Marco Teórico Conceptual de Referencia	11
3.1 Marco Teórico	11
3.2 Método de Análisis	20
3.3 Modelo de Regresión logística	23
3.4 Modelo de Regresión Logística Múltiple	26
4.- Metodología	34
4.1 Información primaria a utilizar	34
4.2 Universo y unidades de muestreo	34
4.3 Variables del Estudio	36
4.4 Plan General de Análisis	38

5.- Resultados	39
5.1 Presentación general de la base de datos	39
5.2 Análisis Descriptivo de los datos	39
5.3 Análisis según objetivos específicos	49
6.- Conclusiones	68
7.- Revisión Bibliográfica	71
8.- Anexos	
Anexo N°1	Medición de la pobreza
Anexo N°2	Ficha técnica de la Encuesta CASEN 2003
Anexo N°3-A	Frecuencias Zona Urbana
Anexo N°3-B	Frecuencias Zona Rural
Anexo N°4-A	Resultados Computacionales Modelo de Regresión Logística Zona Urbana
Anexo N°4-B	Resultados Computacionales Modelo de Regresión Logística Zona Rural

1.- INTRODUCCIÓN:

En Chile, la mayoría de los programas sociales (educación, salud, subsidios, etc.) se asignan o se operan a nivel comunal y la asignación de recursos a éstos se efectúa de acuerdo a los niveles de pobreza que muestran las comunas. Uno de los principales objetivos de la política social en nuestro país es la focalización de los recursos en los sectores más pobres de la población.

La asignación de estos programas es estimada a través de diversas metodologías (Indicador de Desarrollo Humano Comunal del PNUD¹, SINIM², etc.), ya que no existe información de pobreza a nivel de todas las comunas del país (sólo para algunas comunas que están autorepresentadas en la Encuesta CASEN³) que permita ordenar o priorizar a los municipios según esta variable en forma precisa.

Poder determinar para todas las comunas de Chile y su desagregación territorial (zonas, distritos censales y manzana) su situación de pobreza, permitiría identificar a los sectores más pobres de la población, y así asignar o diseñar programas sociales como “Chile Solidario” dirigidos a éstos grupos que

¹ MIDEPLAN - PNUD, Desarrollo Humano en las Comunas de Chile; Temas de Desarrollo Humano Sustentable, N°3, Santiago, 2000.

² Sistema Nacional de Indicadores Municipales, Subsecretaría de Desarrollo Regional, Ministerio del Interior, Chile.

³ El número de comunas de las cuales se puede realizar inferencias estadísticas, varía en cada versión de la Encuesta, el diseño de la muestra es regional y por zona geográfica.

históricamente han quedado relativamente más marginados de la política social por problemas de falta información o acceso al sistema.

Este trabajo tiene gran importancia en la política social desarrollada por el Gobierno, que implica focalizar en los más pobres una parte importante de los recursos del Estado, debido a la escasez de éstos y poder diseñar adecuadamente políticas que tengan un fuerte impacto redistributivo.

El modelo estadístico desarrollado en esta investigación, permite determinar los niveles de pobreza a nivel regional, provincial, comunal, por zona, distrito y finalmente por manzana.

Para la clasificación de la situación que presenta un hogar, se buscó perfilar a ellos de acuerdo a un conjunto de características referidas al Jefe de Hogar y que tengan relación con los niveles de pobreza.

Este tesis desarrolló un modelo probabilístico utilizando “Regresión Logística” que permite determinar, en base a la información censal o encuestas, los niveles de pobreza a nivel regional, provincial, comunal, por zona, distrito y finalmente por manzana, lo que implicará poder focalizar los programas sociales territorialmente con gran precisión. Esta metodología no ha sido aplicada en nuestro país, por lo cual este trabajo es precursor en esta área. El modelo propuesto combina información social censal con georeferenciación y modelación probabilística.

En base al resultado obtenido, se puede calcular la probabilidad de ser pobre para cada una de las familias entrevistadas por el censo realizado en el año 2002 y así obtener, en base a variables estructurales y no coyunturales, una medición alternativa de pobreza e indigencia. Esta medición se realizó para un conjunto de variables que tienen estabilidad en el tiempo y que además pueden ser verificadas, lo que no sucede con los ingresos del hogar⁴, variable que es muy difícil de comprobar y que los estudios muestran que su subdeclaración llega al 47% en algunos casos⁵.

⁴ La metodología de medición de pobreza utilizada actualmente en Chile es exclusivamente en base a los ingresos per cápita del hogar.

⁵ MIDEPLAN; Distribución e Impacto Distributivo del Gasto Social en los Hogares, 1996. Santiago, mayo 1998. Mimeo.

2.- OBJETIVOS E HIPOTESIS

2.1 Objetivos Generales:

- Obtener un modelo que permita identificar las variables que son determinantes de la pobreza de los hogares.
- Tener un modelo de estimación de pobreza estructural, que sea usado por encuestas de muestras representativas de la población.
- Elaborar un modelo que permita focalizar social y territorialmente los programas sociales.

2.2 Objetivos Específicos:

- Obtener estimaciones de pobreza a nivel regional, comunal, por distrito y zona censal.
- Obtener estimaciones de pobreza e indigencia a nivel de manzanas y de hogares.
- Poder en el futuro, en base a la información censal y el modelo obtenido, georeferenciar cada una de las variables e identificar geográficamente los bolsones de pobreza existentes en el país, para poder diseñar políticas y programas sociales adecuados a realidades específicas de las familias.

2.3 Hipótesis de Trabajo:

La pobreza corresponde a múltiples factores, los que según estudios recientes, se ven reflejados por variables vinculadas al hogar y su jefatura (Larrañaga, 2002, Guerrero 2004, y otros autores mencionados en la bibliografía) .

El ser pobre o no, está determinado por un conjunto de características estructurales del hogar, vinculadas a las siguientes dimensiones:

- **Geográficas**
- **Demográficas**
- **Mercado Laboral**
- **Educación**
- **Vivienda**
- **Ingresos**
- **Patrimonio**

Las cuales se modelarán en base a un modelo de regresión logística, el cual necesita validar muy pocos supuestos acerca de cada una de las variables y además permite la regresión no lineal utilizando variables categóricas ordinales (Hosmer y Lemeshow (2000)).

La variable dependiente será el nivel de pobreza, pudiendo tomar dos valores: pobre y no pobre.

Las variables independientes consideran las siguientes dimensiones:

Geográficas:	Región y Zona
Demográficas:	Sexo del Jefe(a) de Hogar, Nº de personas en el hogar, Nº de menores de 14 en el hogar
Mercado Laboral:	Actividad del JH (Ocupado, Desocupado, Inactivo), Nº de ocupados en el hogar.
Educación:	Nivel Educativo del JH y Años de Escolaridad del JH
Vivienda⁶:	Calidad de la Vivienda, Acceso al Agua Potable, Acceso al alcantarillado, Acceso a la energía Eléctrica, Hacinamiento
Ingresos:	Nº de perceptores de ingresos en el hogar ⁷
Patrimonio	Tenencia de: Teléfono, Computador, Microondas, Videograbador, Lavadora,

⁶ Los Indicadores utilizados en vivienda corresponden a las definiciones realizadas por MIDEPLAN en el documento Situación habitacional en Chile, 1990-1998; Santiago, julio 1999 (anexo 3).

⁷ Se define como perceptores de ingresos en el hogar a las personas que tienen ocupación, perciben rentas, jubilaciones o subsidios monetarios.

Refrigerador,

Internet y

Celular

2.3.1 Hipótesis Específicas:

1. La inserción de los Jefes de Hogar en el mercado laboral tiene una fuerte incidencia en la probabilidad de que el hogar sea pobre o no.
2. La edad del Jefe de Hogar tiene incidencia en la probabilidad de que el hogar sea pobre o no, a menor edad mayor probabilidad de ser pobre.
3. La escolaridad del Jefe de Hogar tiene incidencia en la probabilidad de que el hogar sea pobre o no.
4. A mayor número de personas en el hogar mayor es la posibilidad de pobreza.
5. El genero del Jefe de Hogar tiene incidencia en la probabilidad de que el hogar sea pobre.
6. El acceso a bienes y servicios básicos incide en menores niveles de pobreza.
7. El ser miembro de un pueblo originario aumenta la posibilidad de ser pobre.

3. MARCO TEÓRICO CONCEPTUAL DE REFERENCIA:

3.1 Marco teórico

La pobreza tiene muchas caras pero se puede resumir simplemente como la falta de recursos, va más allá de la carencia de bienes económicos, también involucra la falta de oportunidades para el desarrollo de una vida decente para mantener y conservar la dignidad, la autoestima y el respeto de otros, trascendiendo de los bienes materiales⁸.

Se caracteriza por ser un fenómeno especialmente económico con dimensiones sociales, políticas y culturales (véase Moon 1991). La pobreza también está asociada a la escasa participación de las personas en los distintos ámbitos de la vida del país y se expresa en el subconsumo en los hogares. Las personas que se encuentran en esta situación se ven obligadas a elegir la satisfacción de unas necesidades sacrificando otras igualmente apremiantes para ellos.

Esta concepción de la pobreza sostiene que el foco tiene que ser la situación de bienestar de los pobres como tales, sin importar los factores que lo afecten. Según algunas corrientes de pensamiento esta medición se realiza simplemente contando el número de pobres; así la pobreza se expresa como la relación entre el número de pobres y la población total de la comunidad, tasa de incidencia de la pobreza (Sen 1992).

⁸ Altimir, Oscar; La dimensión de la pobreza en América Latina, Naciones Unidas, Cuadernos de la CEPAL, N°27, Santiago, 1979.

Los enfoques más utilizados para medir la pobreza son el de la desigualdad (Miller S.M. 1967) y el biológico.

El enfoque de la desigualdad consiste en medirla en base a la privación relativa de la población, en términos de su distribución de ingresos⁹ supone concebir la pobreza como un problema de desigualdad. Existen una serie de medidas propuestas para utilizar este método (Atkinson 1970).

El enfoque biológico es el que se utiliza en América Latina para medir pobreza y define a las familias en situación de “pobreza como aquellas” cuyos ingresos totales resultan insuficientes para cubrir las necesidades básicas relacionadas con el mantenimiento de la simple eficiencia física”, este enfoque ha sido intensamente atacado por varios autores (véase Townsend, y Rein 1974).

Según esta metodología, para delimitar la pobreza, se define un conjunto de necesidades básicas y, para cada una de éstas se especifican características mínimas de sus satisfactores, normas por debajo de las cuales se presenta una situación de

⁹ En este enfoque, se consideramos la naturaleza y la magnitud de las diferencias entre el 20 o el 10 por ciento más pobre de la población y el resto de ella. El interés se centra en cerrar las brechas entre los que están abajo y los que están mejor en cada dimensión de la estratificación social (Sen). La discusión que se ha planteado frente a este enfoque es que la desigualdad es fundamentalmente un problema distinto de la pobreza. La desigualdad y la pobreza están relacionadas pero ninguno de los conceptos subsume al otro.

insatisfacción. Las normas definidas tienen validez temporal y restringida, motivo por el cual deben revisarse periódicamente¹⁰.

De acuerdo con este método, a un individuo se le considera pobre si su nivel de ingreso se sitúa por debajo de un nivel mínimo que le permita satisfacer sus necesidades básicas; e indigente, si éste no le permite satisfacer sus necesidades alimentarias. Estos mínimos se denominan "línea de pobreza" y "línea de indigencia" respectivamente. Así, estas líneas constituyen el límite entre quiénes son pobres y quiénes no lo son y entre quiénes son indigentes y quiénes no lo son¹¹, es el método utilizado por CEPAL.

Entre las críticas que se realizan a este enfoque, está que los requerimientos, varían según el clima, actividad que se realiza, características demográficas tales como edad y sexo. Además los requerimientos nutricionales son difíciles de precisar, aún más los no alimentarios, por lo que generalmente su medición es arbitraria. Las canastas básicas resultantes tienen un costo excesivamente bajo (véase Stigler 1945), aún con éstos cuestionamientos este método presenta gran utilidad y algunos de los planteamientos se pueden ir superando, como puede ser a través del cálculo de los requerimientos nutricionales considerando las características climáticas de cada país, así como la estructura etaria y laboral de

¹⁰ En Chile, la última revisión se realizó en el año 1987.

¹¹ La línea de pobreza representa el ingreso mínimo necesario por persona para cubrir el costo de una canasta mínima individual para la satisfacción de las necesidades básicas, alimentarias y no alimentarias.

La línea de indigencia representa el ingreso mínimo necesario por persona para cubrir el costo de una canasta alimentaria.

su población, así como también se ha avanzado en construir canastas de consumo que consideren los hábitos de la población de referencia (véase CEPAL; Medición y análisis de la pobreza, 2002).

En América Latina, en los últimos años han predominado tres métodos de medición de la pobreza, los cuáles consideran solamente las necesidades económicas de las personas (enfoque biológico), que se expresan a través de los siguientes métodos:

- (i) El enfoque del ingreso o Línea de Pobreza (LP)
- (ii) El de las Necesidades Básicas Insatisfechas (NBI)
- (iii) El Método Integrado de medición de la pobreza

A nivel mundial existen otras mediciones que permiten su comparación, como la propuesta por el Banco Mundial, Esta línea de pobreza se basa en el consumo de bienes y servicios:

- En los países muy carentes de recursos una línea de pobreza fijada en 1 dólar por persona diarios (US\$ 30 mensuales)
- Se sugiere una línea de pobreza de 2 dólares diarios para América Latina y el Caribe.
- Para Europa oriental y las repúblicas de la ex Unión Soviética se ha utilizado una línea de pobreza de 4 dólares diarios.

- Para la comparación entre los países industrializados se ha utilizado una línea de pobreza que corresponde a la línea de pobreza de los Estados Unidos, de 14,40 dólares diarios por persona.
- La Unión Europea ha sugerido que para determinar la línea de pobreza de esos países se utilice la mitad de la mediana del ingreso personal disponible

3.1.1 La Pobreza en Chile:

El concepto de pobreza utilizado en Chile corresponde al definido por Naciones Unidas¹², que la precisa como: “Una situación que impide al individuo o a la familia satisfacer una o más necesidades básicas y participar plenamente en la vida social”. MIDEPLAN¹³ es la institución en Chile encargada de estimar la pobreza, y lo realiza a través del "método del ingreso" o "del costo de las necesidades básicas", que como se mencionó anteriormente es el método utilizado internacionalmente y propuesto por CEPAL¹⁴, el cual tiene como desventaja que es muy coyuntural al depender solamente de los ingresos corrientes del mes pasado y no de la acumulación de bienes. Mide la pobreza en base a los ingresos monetarios percibidos por la población¹⁵. Desde el año 1987 MIDEPLAN realiza la encuesta CASEN¹⁶, cuyo principal objetivo entre otros, es medir y analizar la evolución de la pobreza de los hogares.

¹² PNUD, Desarrollo sin Pobreza. II Conferencia Regional sobre la Pobreza en América Latina y el Caribe, Quito, Ecuador, 1990.

¹³ Institución en Chile encargada de medir los niveles de pobreza en el país.

¹⁴ CEPAL; Medición y análisis de la pobreza: Notas Técnicas, 26 de abril del 2002 (mimeo).

¹⁵ Se clasifican los hogares en 3 categorías : Indigentes, pobres no indigentes y no pobres

¹⁶ La Encuesta CASEN se aplica a hogares de todos los estratos socioeconómicos seleccionados por muestreo estratificado (región y zona), aleatorio dentro de los estratos y finalmente por

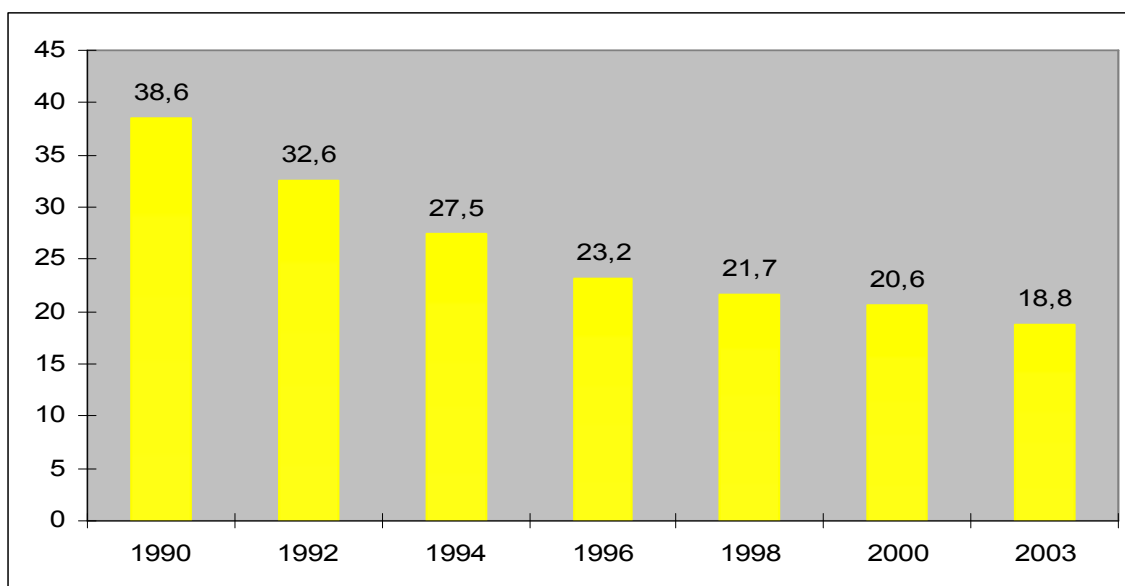
Se estiman dos líneas de pobreza e indigencia, una correspondiente a las zonas urbanas y la segunda a las zonas rurales¹⁷. Las diferencias entre ambas se derivan del supuesto de que en las zonas rurales hay un menor costo de los alimentos y del gasto no alimentario, según estudios previos realizados por CEPAL. En definitiva, la línea de indigencia rural representa un 77% del valor de la urbana. La “línea de pobreza” en las zonas urbanas se calcula como el doble del valor de la línea de indigencia urbana, en tanto que el de las zonas rurales aquélla asciende a 1.75 vez el valor de la línea de indigencia rural. Lo cual significa que la cuantía de recursos monetarios que necesitan los hogares para superar esta situación según área geográfica es distinta, siendo un 67,4% menor el valor de la canasta en el área urbana que en la rural.

Los datos provenientes de la Encuesta CASEN, muestran una importante disminución en los niveles de pobreza e indigencia en los últimos 13 años, tanto en el área rural como urbana, como se puede observar en los siguientes gráficos:

conglomerados, obteniéndose así una muestra representativa de todo el país. Además se obtiene representatividad regional, por zona y en algunas comunas. La Encuesta se ha realizado en noviembre de los años 1987, 1990, 1992, 1994, 1996, 1998 y 2000; la próxima se realizará en noviembre del 2003. (ver Anexo N°1).

¹⁷ Ver Anexo N°1

Gráfico N°1: Evolución de la Pobreza en Chile

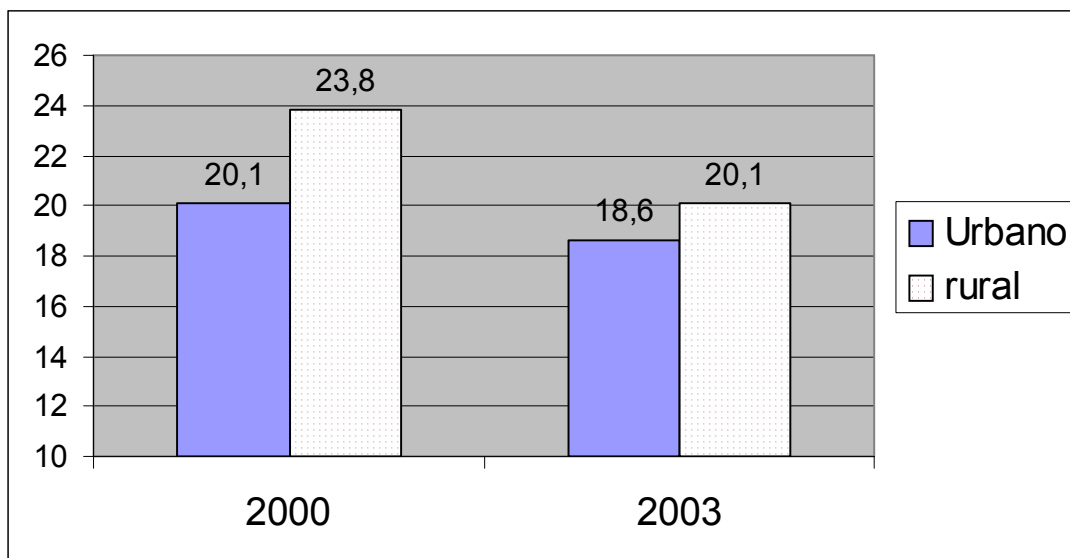


Fuente: MIDEPLAN; Volumen 1: "Pobreza, Distribución del Ingreso e Impacto Distributivo del Gasto Social"; Santiago; agosto 2004.

A fines del año 2003 la situación de pobreza total del país baja a un 18.8%, correspondiente a 2.907.700 personas aproximadamente, de las cuales un 4,7% son indigentes (728.100 personas) y 14,1% son pobres no indigentes (2.179.600)¹⁸.

¹⁸ MIDEPLAN; Volumen 1: "Pobreza, Distribución del Ingreso e Impacto Distributivo del Gasto Social"; Santiago; agosto 2004.

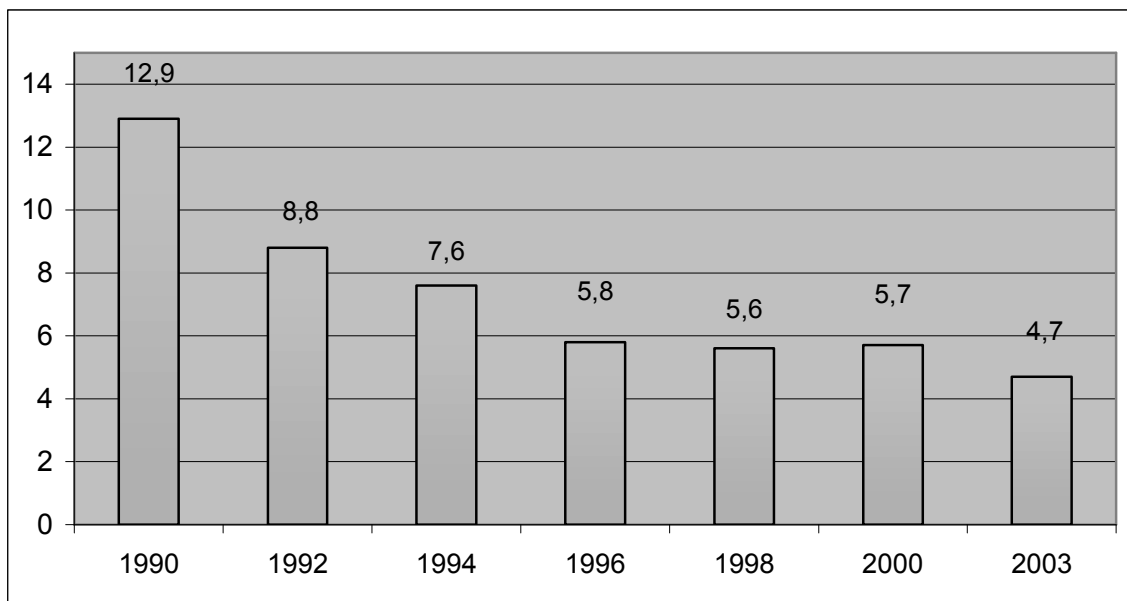
Gráfico N°2: Pobreza según área Geográfica: 2000-2003



Fuente: MIDEPLAN; Volumen 1: "Pobreza, Distribución del Ingreso e Impacto Distributivo del Gasto Social"; Santiago; agosto 2004.

En las zonas urbanas, la pobreza es menor a los niveles que se presentan en las zonas rurales y en el periodo 2000-2003 disminuyó en ambas áreas. La pobreza rural registró un gran avance, 86 mil 200 personas salieron de esta condición, lo que corresponde a una reducción en estos tres años del 23,8 al 20,1%. La misma senda siguió la indigencia, que bajó en 1,0 punto porcentual.

Gráfico N°3: Evolución de la Indigencia en Chile



Fuente: MIDEPLAN; Volumen 1: "Pobreza, Distribución del Ingreso e Impacto Distributivo del Gasto Social"; Santiago; agosto 2004.

Como se observa en los gráficos anteriores, la evolución de la pobreza e indigencia, muestra una importante disminución a partir de la década de los '90. Desde 1996, el ritmo de esta baja en las tasas disminuye, especialmente en la situación de indigencia, o también llamada "pobreza dura", donde se mantienen los mismos niveles entre el año 1996 y el 2000. Esta tendencia presenta un considerable cambio en el año 2003, pasando de 5,7% a 4,7%. Según los primeros análisis esta baja sería producto de las políticas sociales focalizadas en los sectores más pobres, viéndose estas representadas en estos grupos por el Sistema Chile Solidario, que es un programa de atención integral a las familias más pobres. Esta cifra muestra la importancia de poder producir información que permita focalizar cada vez mejor cada uno de los programas dirigidos a la superación de la pobreza.

3.2 Método de Análisis

Para encontrar los determinantes de la pobreza en los hogares chilenos, se utilizará un modelo que permita primero caracterizar y luego clasificar los hogares que requieren ayuda social en forma más prioritaria, para lo cual se utilizará un modelo de regresión.

En los casos en que el marco conceptual lleva a la especificación de un modelo en el cual existe una variable explicada (nivel de pobreza) por un conjunto de variables explicativas ($X_1, X_2, X_3, \dots, X_k$), la técnica estadística adecuada es el análisis de regresión. En este caso como el objetivo es determinar tanto las características de la pobreza como la probabilidad de ser pobre, se usará regresión logística que permite predecir probabilidades.

La proposición teórica que lleva a emplear esta herramienta supone que la variabilidad de la variable dependiente se puede explicar a través de las variables explicativas. El modelo estadístico correspondiente agrega que, además, opera como variable independiente un factor aleatorio no observable (x), llamado error aleatorio.

El ajuste del modelo estadístico especifica una serie de condiciones que debe satisfacer el error aleatorio y que permiten estimar, no sólo si el modelo ajusta adecuadamente (para lo cual se tienen pruebas de hipótesis e índices de bondad,

sino también la significación y magnitud de los efectos de las variables explicativas.

La medición de los impactos de las variables explicativas requiere que no haya relación entre ellas o que ésta sea baja, cuando esto no ocurre se presentan dificultades para estimar la magnitud real de los efectos debido a que el modelo es incapaz de separar cuál corresponde a cada variable. Los procedimientos de estimación se complican si se detecta que no se cumplen los supuestos relativos al término de error. Por lo cual se deben realizar las pruebas estadísticas para verificar los supuestos.

El planteamiento clásico del modelo de regresión supone que todas las variables son métricas (medidas en escala de intervalo o de razón). Esta condición se convierte en una fuerte barrera para aplicar ésta técnica estadística a los problemas típicos de las ciencias sociales. Sin embargo, esta limitación fue superada al introducir, *variables explicativas* dicotómicas (dummies) que dieran cuenta de la presencia o ausencia de un evento particular. Es así como la regresión se ha convertido en una herramienta de gran utilidad para las ciencias sociales. Este modelo aún mantiene la restricción a la variable dependiente, la cual es la continuidad de esta y que sus valores pueden variar entre menos infinito e infinito.

Estos problemas se superaron aplicando transformación logarítmicas a la variable dependiente (Theil H. 1971: 166 a 196; Aldrich é. y N. Forrest; 1984; Maddala G., 1983: 13 a 56; Agresti A., 1990)¹⁹ y utilizando un modelo denominado de regresión logística.

En las ciencias sociales se recurre cada vez más a utilizar técnicas de análisis de variables múltiples para analizar los distintos fenómenos. En el ajuste de este tipo de modelos tienden a predominar los métodos de estimación mínimo cuadrático y de máxima verosimilitud. Si bien hay razones técnicas para preferir uno u otro dependiendo del problema de que se trate.

En conclusión “los desarrollos estadísticos recientes, fuertemente condicionados por las demandas planteadas por las ciencias sociales, tienden a abandonar la naturaleza determinística de la experiencia y a reemplazarla por una concepción que reconoce el carácter aleatorio de la misma y a reconocer el papel activo del sujeto en la construcción del objeto”²⁰.

¹⁹ Córtes Fernando; Regresión Logística en la investigación social: potencialidades y limitaciones; CES, COLMEX, página web.

²⁰ Córtes Fernando; Regresión Logística en la investigación social: potencialidades y limitaciones; CES, COLMEX, página web.

3.3 Modelo de Regresión Logística ²¹:

Es un modelo no lineal, los datos no se ajustan a una línea recta, a las variables explicativas no se les exige una distribución determinada. Permite construir modelos donde la variable dependiente es cualitativa, esta puede ser dicotómica o politómica y dentro de éstas las variables pueden ser ordinales o nominales. Las variables explicativas pueden ser cualitativas y/o cuantitativas.

La función logística es una curva sigmoidea en forma de letra S, utilizando los logaritmos es posible linealizar el modelo.

Si la variable dependiente es dicotómica se utiliza para predecir la probabilidad estimada de que la variable dependiente "Y" presente uno de los valores posibles (1=sí o 0=no) en función de los diferentes valores que adoptan el conjunto de variables independientes "X_i"²².

Las variables predictoras "X_i" pueden ser dicotómicas, medidas a nivel nominal, ordinal, intervalo o razón.

²¹ La formulación matemática de este modelo está basada en el libro de Hosmer D. Y Lemeshow S.; Applied Logistic Regression; Wiley Series in Probability and Statistics; 2ª edición, Canadá, 2000.

²² Jovel; Albert J.; "Análisis de Regresión Logística" Centro de Investigaciones Sociológicas; Madrid, 1995.

Jovell plantea que los objetivos del modelo logístico son “determinar la existencia o ausencia de relación entre una o más variables independientes y la variable dependiente; medir la magnitud de dicha relación y estimar o predecir la probabilidad de que se produzca un suceso $Y=1$ en función de los valores que adopten las variables independientes X_i ”²³.

En la mayoría de los casos las regresiones categóricas además pueden utilizarse para medir el rendimiento de los parámetros alternativos de determinación de beneficiarios de programas y transferencias sociales. Se trata de un tema importante, ya que las autoridades suelen utilizar variables representativas para evaluar el nivel de renta o de consumo de un hogar o individuo con el objeto de determinar si les corresponde o no ser incluidos en dichos programas. Las regresiones categóricas pueden ayudar a seleccionar las mejores variables representativas para identificar a pobres y no pobres, o bien más ampliamente para seleccionar beneficiarios²⁴.

3.3.1 Definición:

Sea Y una variable dependiente binaria que toma dos valores posibles (0 y 1).

Sean X_1, \dots, X_k un conjunto de variables independientes observadas con el fin de explicar y/o predecir el valor de Y .

El objetivo es determinar:

²³ Vivanco, M., Análisis Estadístico Multivariado, Teoría y práctica. Ed. Universitaria, Santiago, Chile, 1999.

²⁴ Boltvinik Julio.

$P[Y=1/ X_1, \dots, X_k]$, donde P indica probabilidad

Así $P[Y=0/ X_1, \dots, X_k] = 1 - P[Y=1/ X_1, \dots, X_k]$.

Se construye un modelo :

$$P[Y=1/ X_1, \dots, X_k] = p(X_1, \dots, X_k; \beta)$$

Donde $p(X_1, \dots, X_k; \beta)$ es una función que recibe el nombre de función de enlace (función de probabilidad) cuyo valor depende de un vector de parámetros

$$\beta = (\beta_1, \dots, \beta_k)'$$

3.3.2 Función de Verosimilitud

Con el fin de estimar β y analizar el comportamiento modelo, observamos una muestra aleatoria simple de tamaño n dada por $\{(X_i', Y_i; i=1, \dots, n)\}$

donde $X_i = (X_{i1}, \dots, X_{ik})$, es el valor de las variables independientes e $Y_i = \{0, 1\}$ es el valor observado de Y en el i -ésimo elemento de la muestra.

$Y/(X_1, \dots, X_k)$ se distribuye Binomial $(1, p(Y=1/ X_1, \dots, X_k; \beta))$.

La función de verosimilitud es:

$$L(\beta/(x_1, y_1), \dots, (x_n, Y_n)) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} \quad (1)$$

Donde $p_i = p(x_i', \beta) = p(x_{i1}, \dots, x_{ik}; \beta)$ con $i=1, \dots, n$

3.4 Modelo de regresión logística múltiple

El modelo matemático se construye en base a probabilidades, las cuales se obtienen considerando la probabilidad de que ocurra un suceso determinado $P(Y)$ en relación con la dependencia de que dicha probabilidad no ocurra $1 - P(Y)$.

La probabilidad proporciona predicciones consistentes y fáciles de los resultados en términos del "Odds" del evento $Y=1$.

$$\text{Odds}(Y=1) = \frac{P(y)}{1 - p(y)} \quad (2)$$

El modelo de regresión logística múltiple, relaciona la probabilidad de que ocurra un determinado suceso denotado por el vector $X' = (X_1, \dots, X_k)$ con probabilidad condicional $P(Y=1 | X)$ en función de k variables independientes que pueden ser cuantitativas, cualitativas o ambas según sea el tipo de diseño de estudio.

El modelo logístico múltiple es:

$$L = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (3)$$

$$\text{O también: } p_i = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k X_k}} = \frac{1}{1 + e^{-z}} \quad (4)$$

Donde:

- L es el logaritmo natural de "Odds"²⁵ también denominado "Logit" o "L"

²⁵ Corresponde a la razón de las probabilidades $P(Y)$ y $1 - p(y)$.

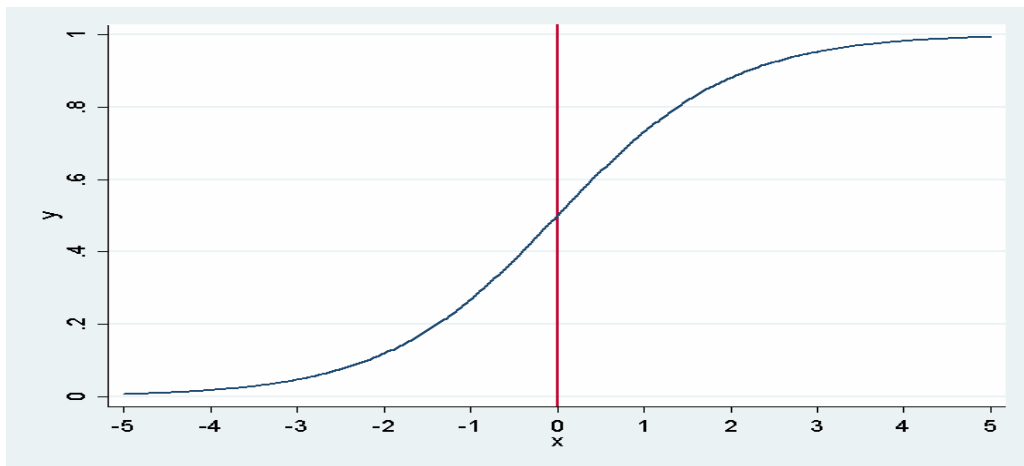
- $\beta_0, \beta_1, \dots, \beta_k$ son constantes
- X una variable explicativa que puede ser continua o discreta.

Como los coeficientes del modelo logístico no tienen restricciones éstos son fácilmente interpretables en términos independencia o asociación entre las variables.

Gráficamente la función es simétrica

- Continua y creciente sobre el intervalo 0 y 1
- Sigue una curva sigmoidea

Gráfico n° 4 Gráfico del Modelo de Regresión Logística



3.4.1 Estimación de los Parámetros

Sea una muestra de n observaciones independientes definido por $(X_{i1}, X_{i2}, \dots, X_{ik}, Y_i)$, $i=1, \dots, n$; se elige el vector $\beta' = (\beta_0, \beta_1, \dots, \beta_k)$ el método más usado es el

de verosimilitud, con $k+1$ ecuaciones de verosimilitud que se obtienen derivando el ln de la función de verosimilitud respecto a $k+1$ coeficientes.

Las ecuaciones de verosimilitud son:

$$\sum_{i=1}^n (y_i - p_i) = 0$$

$$\sum_{i=1}^n X_{ij}(y_i - p_i) = 0, j = 1, \dots, k; i = 1, \dots, n$$

Mediante calculo diferencial se encuentran las soluciones a este conjunto de ecuaciones, actualmente existen software estadísticos para estimar los parámetros.

3.4.2 Pruebas de Significación

Una vez estimado los coeficientes del modelo, se tiene que verificar si el modelo predice de manera adecuada a la variable dependiente. Para evaluar la bondad del modelo se utiliza el logaritmo del cociente de verosimilitud y la prueba de Hosmer-Lemeshow.

3.4.2.1 Modelo

Se procede de la siguiente manera

- 1.- Se prueba la hipótesis de que el modelo encontrado es el que mejor se ajusta a través de la razón de verosimilitud²⁶.

²⁶ Hosmer D. Y Lemeshow S.; Applied Logistic Regression; Wiley Series in Probability and Statistics; 2ª edición, Canadá, 2000.

$$D = -2 \ln \frac{\text{verosimilitud del modelo sin la variable (modelo analizado)}}{\text{verosimilitud del modelo con la variable (modelo saturado)}}$$

La diferencia entre estos dos valores de $-2nL$ se llama **Devianza** y prueba si la o las variables X_i son significativas.

$$H_0 : \beta_0 = \beta_1 = \dots = \beta_k = 0$$

H_1 : Algún β_i distinto de 0.

Estadístico de prueba $D \sim \chi^2$ con $n-k-1$ grados de libertad.

Región de Rechazo: si $D > \chi^2_{\alpha, (n-k-1)}$ si se rechaza H_0 , significa que al menos uno de los coeficientes es diferente de cero y la variable correspondiente es significativa en la probabilidad de que ocurra o no el suceso en estudio²⁷.

- 2.- Cuando el número de variables del modelo es grande y/o tiene variables independientes cuantitativas, se debe utilizar la prueba de Hosmer-Lemeshow, como consecuencia de la gran cantidad de variables, el número de pautas de variación existente entre ellas es tan elevado que puede invalidar la utilidad de los estadísticos de bondad de ajuste clásicos.

²⁷ Jovel; Albert J.; "Análisis de Regresión Logística" Centro de Investigaciones Sociológicas; Madrid, 1995.

Las hipótesis que se plantean son:

$$H_0 : P_{\text{observada}} = P_{\text{estimada}} \text{ (el modelo es significativo)}$$

$$H_1 : P_{\text{observada}} \neq P_{\text{estimada}}$$

Donde $P_{\text{observada}} = \text{Valor observado}$ y

$P_{\text{estimada}} = \text{Probabilidad estimada por el modelo.}$

Evalúa la bondad del modelo construyendo una tabla de contingencia, divide la muestra en 10 grupos a partir de los deciles de las probabilidades estimadas (a los 10 grupos resultantes se les denomina deciles de riesgo). En cada decil de riesgo se calcula el número de casos que pertenecen a cada categoría de la variable dependiente (número observado) y el número de casos que el modelo pronostica que pertenecen a cada categoría de la variable dependiente (número esperado). A continuación se compara el número de casos observado con el número de casos esperado (esta comparación se realiza en cada una de las 20 casillas definidas por la combinación de las 2 categorías de la variable dependiente con los 10 deciles de riesgo).

Este estadístico permite contrastar la hipótesis nula de igualdad de distribuciones, es decir, la hipótesis de que la variable dependiente se distribuye de la misma manera en los 10 deciles de riesgo o, que no existen diferencias entre las

frecuencias observadas y las esperadas. El estadístico de contraste tiene siempre k grados de libertad²⁸.

Las hipótesis que se contrastan son:

H_0 : El modelo es adecuado

H_1 : El modelo ajustado no es adecuado

Decisión si el estadístico de prueba es mayor o igual que $X^2_{\alpha, (n^\circ \text{ de grupos}-2)}$, se rechaza H_0 y se concluye que el modelo no es adecuado con un nivel de significación α .

3.4.2.2 Pruebas de coeficientes

Una vez encontrado el mejor conjunto de variables explicativas que predicen la variable dependiente Y , se debe evaluar cada coeficiente para determinar cuál o cuáles ingresan al modelo, este proceso se realiza mediante el estadístico Wald.

La hipótesis que se plantea es la siguiente:

H_0 : $\beta_i = 0$

H_1 : β_i no es igual a 0

Para contrastar la hipótesis señalada se usa el estadístico de Wald:

²⁸ Hosmer D. Y Lemeshow S.; Applied Logistic Regression; Wiley Series in Probability and Statistics; 2ª edición, Canadá, 2000.

$$WALD = \frac{\beta^2_i}{S^2_{\beta_i}} \sim \chi^2_{\alpha,1} \quad (5)$$

en el caso de la regresión logística multivariada corresponde a un vector donde cada celda es la división entre el coeficiente β_i y el error estándar de éste. Donde S_{β_i} , es el error estándar del coeficiente de regresión logística muestral y k es el número de variables independientes.

Se formula la hipótesis

H_0 : La variable independiente no influye sobre p_i .

H_1 : La variable independiente influye sobre p_i .

3.4.3 Interpretación de los Resultados

La interpretación de los resultados obtenidos se realiza a partir de los coeficientes del modelo. Para ello basta tener en cuenta que si el modelo ajustado es adecuado, entonces se dice que el modelo es significativo, pero además se debe analizar el grado de asociación estadística que existen en sus parámetros. Si:

$\beta_1 > 0$ el factor de riesgo será mayor que 1 y $p(X_1, X_2, \dots, X_k; \beta)$ aumentará

$\beta_1 < 0$ el factor de riesgo será menor que 1 y $p(X_1, X_2, \dots, X_k; \beta)$ disminuirá.

$\beta_1 = 0$ la variable X_1 , no ejerce ningún efecto sobre la probabilidad de riesgo.

Si el modelo de Regresión Logística es significativo y una de variables

independientes es dicotómica con valores de 0 y 1, el número e^{β_i} es el OR, denominado factor de riesgo o protección que implica un aumento unitario de la variable independiente.

En el caso de una variable cuantitativa, e^{β_i} es el número de veces que aumenta la “chance” de que ocurra el suceso, en este caso de ser pobre, por cada unidad de aumento de la variable independiente.

4.- METODOLOGÍA

4.1 Información Primaria a Utilizar

La Encuesta CASEN tiene una gran cantidad de información, que permitiría construir un modelo para determinar la probabilidad de ser pobre, y tiene como variable dependiente los niveles de pobreza, por lo cual se propone utilizar un modelo de regresión logística que determine esta probabilidad.

Las variables utilizadas en el modelo deben también existir en el Censo de Población, para poder determinar con las variables que contiene, niveles de pobreza en las comunas, distritos, zonas censales y manzanas. La construcción de los niveles de pobreza de estos últimos, sólo será propuesta metodológicamente, para que pueda ser utilizada por instituciones como MIDEPLAN, Min. de Salud y Educación, Secretarías Regionales, Gobernaciones y Municipios.

4.2 Universo, Muestra y Unidades de muestreo

Se seleccionaron dos muestras (1 para cada área) urbana y rural, estratificadas proporcionales y aleatorias (cada región un estrato), lo que significó construir 13 muestras aleatorias regionales urbanas y 13 rurales basadas en la encuesta

CASEN 2003, con un número total de 2000 casos para la zona urbana y 1000 para la zona rural.

Estas fueron proporcionales a la representación de cada área y región de el país, ya que el tamaño muestral de la Encuesta CASEN 2003 es muy grande (68.150 hogares), lo que podría significar que la mayoría de las variables del modelo podrían ser estadísticamente significativas, sólo por el tamaño de la muestra, además hay regiones que están sobrerrepresentadas en la muestra por razones administrativas²⁹.

Tabla 1 Número de hogares del país

Región	Zona Urbana			Zona Rural		
	Urbana	%	Muestra	Rural	%	Muestra
I	95.822	2,7	54	5.156	0,9	9
II	110.688	3,1	62	1.678	0,3	3
III	66.428	1,9	37	5.775	1,1	10
IV	117.172	3,3	66	39.984	7,3	72
V	400.522	11,3	225	31.600	5,7	57
VI	155.168	4,4	87	66.823	12,1	121
VII	158.785	4,5	89	83.688	15,2	152
VIII	426.433	12,0	239	96.674	17,5	175
IX	159.795	4,5	90	71.399	12,9	129
X	187.781	5,3	105	95.442	17,3	173
XI	19.810	0,6	11	6.464	1,2	12
XII	43.107	1,2	24	3.578	0,7	6
R.M.	1.619.601	45,5	910	43.465	7,9	79
Total	3.561.112	100,0	2.000	551.726	100,0	1.000

²⁹ El objetivo de la muestra es realizar inferencias estadísticas a nivel regional y en 302 comunas del país, por lo cual no es proporcional en cada región.

Los resultados obtenidos con esta muestra se validaron con el total de la información proveniente de la Encuesta CASEN 2003.

4.3 Variables del Estudio

La variable dependiente será el nivel de pobreza, pudiendo tomar dos valores: pobre y no pobre.

Las variables independientes consideran las siguientes dimensiones y variables independientes en los modelos explicativos:

Tabla 2 Variables

Dimensión	Variable	Etiqueta	Categoría
Geográfica	z	Zona	Cualitativa
Demográficas	sexojh	Sexo del jefe de hogar	Cualitativa
	edad	Edad del jefe de Hogar	Cuantitativa
	disca	¿Tiene alguna Discapacidad:?	Cualitativa
	etnia	Pertenece a alguna etnia	Cualitativa
	numper	Número de Personas en el hogar	Cuantitativa
Mercado Laboral	ocupado	¿Es ocupado el jefe de Hogar:?	Cualitativa
	inactivo	¿Es inactivo:?	Cualitativa
Educación	esc	Escolaridad del JH	Cuantitativa
Patrimonio	lavadora	Lavadora Automática	Cualitativa
	refri	Refrigerador	Cualitativa
	fono	Teléfono fijo	Cualitativa
	video	Videograbador	Cualitativa
	micro	Microondas	Cualitativa
	computa	Computador	Cualitativa
	calefont	Cálefont	Cualitativa
	celular	Teléfono Móvil	Cualitativa
	internet	internet	Cualitativa
	tv cable	TV Cable	Cualitativa
Vivienda	agua	Agua Potable	Cualitativa
	alcanta	Alcantarillado	Cualitativa
	vivienda	calidad vivienda	Cualitativa
	nhogar	hogares en la vivienda	Cualitativa
	hacina	Hacinamiento	Cualitativa

4.4 Plan General de Análisis

a) Preparación de la base de datos

Objetivo: Verificación y recodificación de los datos originales.

b) Análisis Exploratorio

Objetivo: Obtención de la información acerca del comportamiento de cada variable. Verificación de supuestos principales.

c) Análisis de Regresión Logística con Respuesta Binaria

Objetivo: A partir del primer modelo predictivo completo, se seleccionaron las variables con mejor capacidad predictiva, se analizaron los efectos de la interacción de las variables.

d) Validación del modelo obtenido

Objetivo: Se validó el modelo obtenido en las dos submuestras, tomando la muestra de la Encuesta CASEN completa separada por área urbana y rural.

5. RESULTADOS

Se construyeron 2 matrices de datos, la matriz con información de la zona urbana contiene 1999 unidades muestrales y la rural a 1000, cada unidad muestral corresponde a un hogar³⁰; cada unidad de respuesta está representada por un vector de dimensión igual al número de variables, éstas corresponden a 23³¹.

Las variables son de tipo numéricas y nominales y corresponden a la información del Jefe de hogar y al hogar que habitan. La definición de medición de pobreza es por hogar y no por individuo, si un hogar es pobre todos los individuos de ese hogar enfrentan la misma condición, lo que implica que la unidad de análisis es el hogar.

5.1 Análisis de las variables

Una vez obtenida las muestras urbana y rural, se recodificaron los datos, para facilitar el análisis e interpretación de los coeficientes del modelo de regresión logística, las variables discretas quedaron como “0” cuando había ausencia del evento y 1 en caso de presencia. La variable “Participación en la Fuerza de Trabajo”, que tenía 3 categorías (Ocupado, Desocupado e Inactivo), se transformó en dos variables ficticias, ocupados e Inactivos).

³⁰ Hogar: aquel constituido por una persona o un grupo de personas, con o sin vínculos de parentesco, que comparten vivienda y tienen presupuesto de alimentación común.

³¹ Se excluye la variable zona.

5.1.1 Zona Urbana

5.1.1.1 Variables Discretas

En la tabla 3 se presentan la información de las variables utilizadas en el modelo completo para el área urbana cruzadas por situación de pobreza.

Se observa que las principales diferencias entre los Jefes de Hogar “pobres” y “no pobres” en la zona urbana en el año 2003 en el país son:

- Los jefes de hogares pobres de la zona urbana en mayor proporción pertenecen a una etnia.
- Las viviendas de los hogares pobres urbanos tienen menor acceso al agua potable, alcantarillado, una proporción mayor de hogares hacinados y viviendas de mala calidad.
- Los hogares pobres en su gran mayoría carecen de patrimonio.

Tabla 3 Características de los hogares urbanos

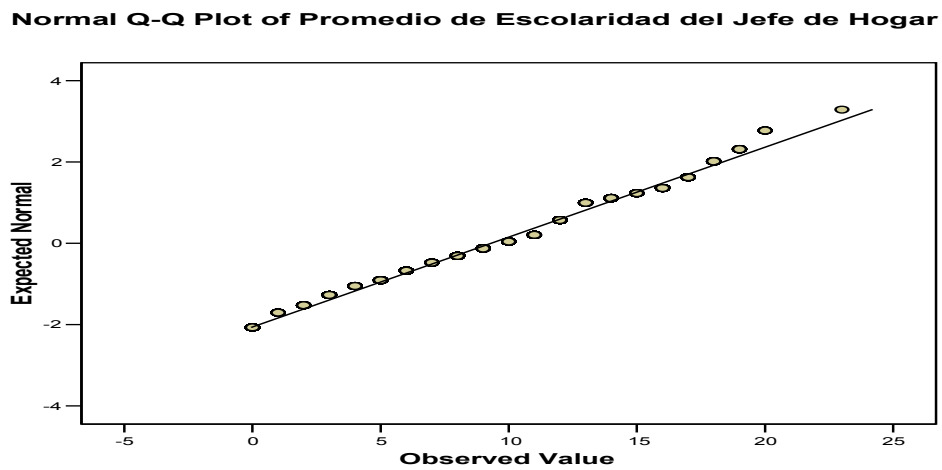
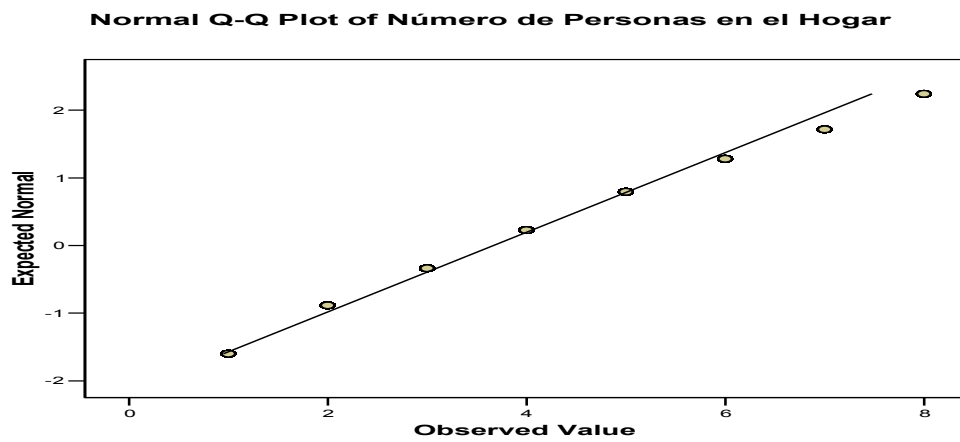
Zona Urbana		EL HOGAR ES POBRE		
		NO	SI	TOTAL
El Jefe de hogar esta Ocupado:	NO	27,9	44,3	31,0
	SI	72,1	55,7	69,0
El Jefe de hogar es Inactivo:	NO	74,4	72,7	74,1
	SI	25,6	27,3	25,9
Sexo del Jefe de Hogar	MUJER	26,3	30,5	27,1
	HOMBRE	73,7	69,5	72,9
El hogar tiene Lavadora Automática	NO	38,2	68,4	43,8
	SI	61,8	31,6	56,2
Refrigerador	NO	13,1	29,5	16,1
	SI	86,9	70,5	83,9
Teléfono fijo	NO	43,8	81,4	50,8
	SI	56,2	18,6	49,2
Videograbador	NO	61,8	88,4	66,7
	SI	38,2	11,6	33,3
Microondas	NO	61,4	88,4	66,4
	SI	38,6	11,6	33,6
Computador	NO	74,0	93,8	77,6
	SI	26,0	6,2	22,4
Cálefont	NO	31,8	71,6	39,2
	SI	68,2	28,4	60,8
Teléfono Móvil	NO	55,1	66,8	57,3
	SI	44,9	33,2	42,7
Internet	NO	88,9	99,2	90,8
	SI	11,1	0,8	9,2
Tvcable	NO	75,3	94,1	78,8
	SI	24,7	5,9	21,2
Hogar pertenece a alguna etnia	NO	97,1	92,2	96,1
	SI	2,9	7,8	3,9
La vivienda tiene agua potable:	NO	3,7	11,4	5,2
	SI	96,3	88,6	94,8
La vivienda tiene Alcantarillado	NO	5,8	16,2	7,7
	SI	94,2	83,8	92,3
La vivienda es de buena calidad	NO	1,9	6,2	2,7
	SI	98,1	93,8	97,3
Más de 1 hogar en la vivienda	NO	96,4	94,3	96,0
	SI	3,6	5,7	4,0
Hay Hacinamiento en el hogar	NO	98,0	88,8	96,3
	SI	2,0	11,2	3,7

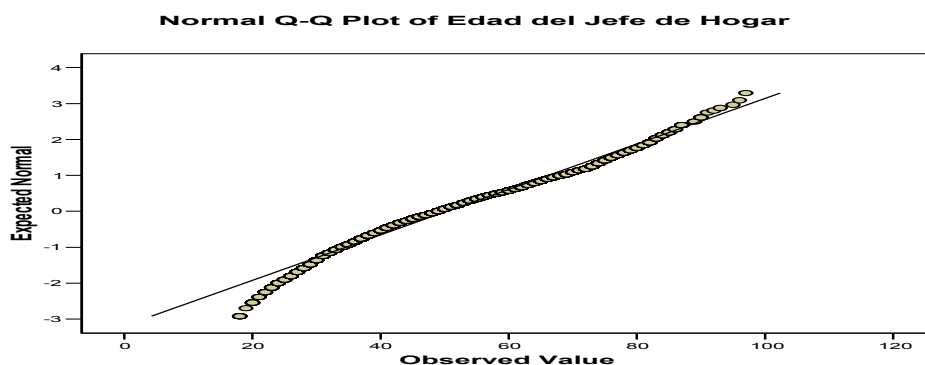
5.1.1.2 Variables Continúas

Tabla 4 Análisis Descriptivo de las variables

Variable	Promedio	Mediana	Desviación estándar
Número de personas en el hogar	3,7	4	0,038
Años de escolaridad del jefe de hogar	9,3	10	0,102
Edad del jefe de hogar	50,4	49	0,353

Gráficos:





En los tres casos se puede observar que la media y la mediana están bastante cercanas, además en base al gráfico "Plot-Plot" se puede asumir normalidad.

Una vez asumida la normalidad de las variables se realizaron las pruebas para verificar la diferencia de medias de las tres variables analizadas, según situación de pobreza.

Tabla 5 Promedio de personas por hogar, escolaridad y edad promedio del Jefe de Hogar según situación de pobreza

	EL HOGAR ES POBRE		TOTAL
	NO	SI	Promedio
	Promedio	Promedio	
Número de Personas en el Hogar	3,5	4,4	3,7
Promedio de Escolaridad del Jefe de Hogar (años)	9,6	7,9	9,3
Edad del Jefe de Hogar (años)	51,4	45,7	50,4

Los jefes de hogares pobres de la zona urbana, tienen un promedio de edad inferior a los jefes de hogares no pobres, son más jóvenes y el número de personas por hogar es mayor, como lo muestran las pruebas estadísticas que se utilizaron.

Tabla 6 Test de significación estadística de promedio de personas por hogar, escolaridad y edad promedio del Jefe de Hogar según situación de pobreza

Independent Samples Test										
		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	95% Confidence Interval of the Difference	
									Lower	Upper
Promedio de Escolaridad del Jefe de Hogar	Equal variance assumed	26,692	,000	6,743	1987	,000	1,742	,258	1,235	2,248
	Equal variance not assumed			7,804	664,869	,000	1,742	,223	1,303	2,180
Número de Personas en el Hogar	Equal variance assumed	1,273	,259	-9,289	1997	,000	-,924	,099	-1,119	-,729
	Equal variance not assumed			-9,318	551,161	,000	-,924	,099	-1,118	-,729
Edad del Jefe de Hogar	Equal variance assumed	9,798	,002	6,377	1997	,000	5,745	,901	3,978	7,512
	Equal variance not assumed			6,794	592,132	,000	5,745	,846	4,084	7,406

5.1.2 Zona Rural

Para la zona rural se realiza el mismo análisis, en la tabla nº7 se presenta la información de las variables utilizadas en el modelo completo para el área rural cruzadas por situación de pobreza.

5.1.2.1 Variables Discretas

Tabla 7 Características de los hogares urbanos

Muestra Urbana		EL HOGAR ES POBRE		
		NO	SI	TOTAL
El Jefe de hogar esta Ocupado:	NO	30,9	40,5	32,5
	SI	69,1	59,5	67,5
El Jefe de hogar es Inactivo:	NO	70,4	70,6	70,4
	SI	29,6	29,4	29,6
Sexo del Jefe de Hogar	MUJER	17,1	18,4	17,3
	HOMBRE	82,9	81,6	82,7
El hogar tiene Lavadora Automática	NO	72,1	87,7	74,6
	SI	27,9	12,3	25,4
Refrigerador	NO	32,5	54,0	36,0
	SI	67,5	46,0	64,0
Teléfono fijo	NO	93,3	97,5	94,0
	SI	6,7	2,5	6,0
Videograbador	NO	89,1	96,3	90,3
	SI	10,9	3,7	9,7
Microondas	NO	92,0	98,8	93,1
	SI	8,0	1,2	6,9
Computador	NO	95,7	98,8	96,2
	SI	4,3	1,2	3,8
Cálefont	NO	78,4	92,0	80,6
	SI	21,6	8,0	19,4
Teléfono Móvil	NO	63,7	77,3	65,9
	SI	36,3	22,7	34,1
Internet	NO	98,9	100,0	99,1
	SI	1,1	0,0	0,9
Tvcable	NO	97,2	100,0	97,7
	SI	2,8	0,0	2,3
Hogar pertenece a alguna etnia	NO	88,0	81,0	86,9
	SI	12,0	19,0	13,1
La vivienda tiene agua potable:	NO	61,9	75,5	64,1
	SI	38,1	24,5	35,9
La vivienda tiene Alcantarillado	NO	56,4	79,8	60,2
	SI	43,6	20,2	39,8
La vivienda es de buena calidad	NO	3,7	9,2	4,6
	SI	96,3	90,8	95,4
Más de 1 hogar en la vivienda	NO	98,7	99,4	98,8
	SI	1,3	0,6	1,2
Hay Hacinamiento en el hogar	NO	96,5	82,4	94,2
	SI	3,5	17,6	5,8

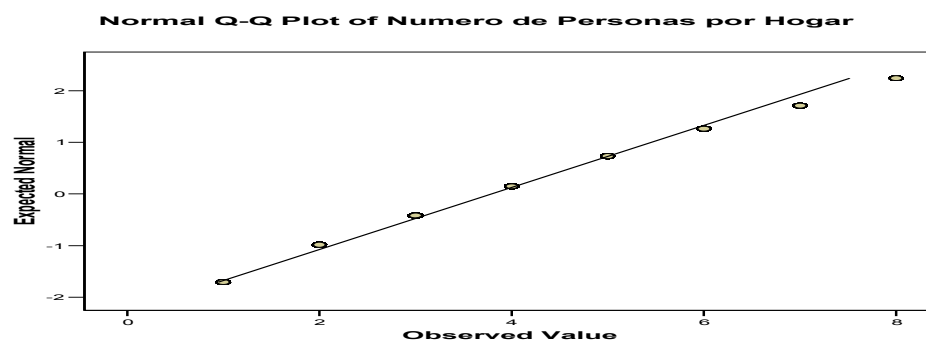
En la tabla anterior se observa que las principales diferencias entre los Jefes de Hogar “pobres” y “no pobres” en la zona rural en el año 2003 en el país:

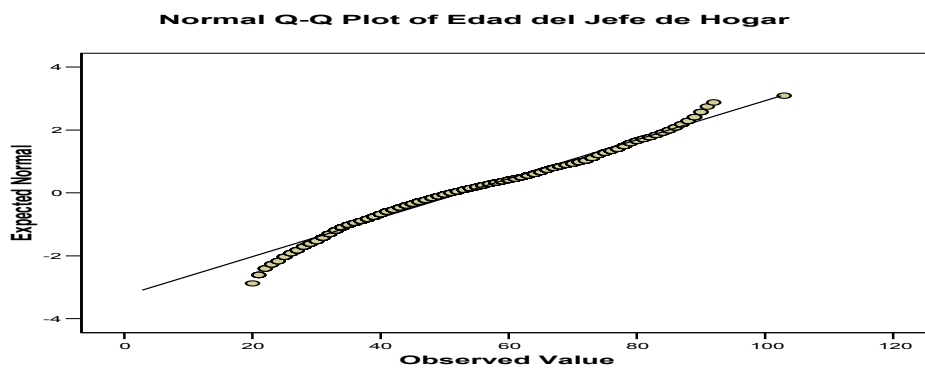
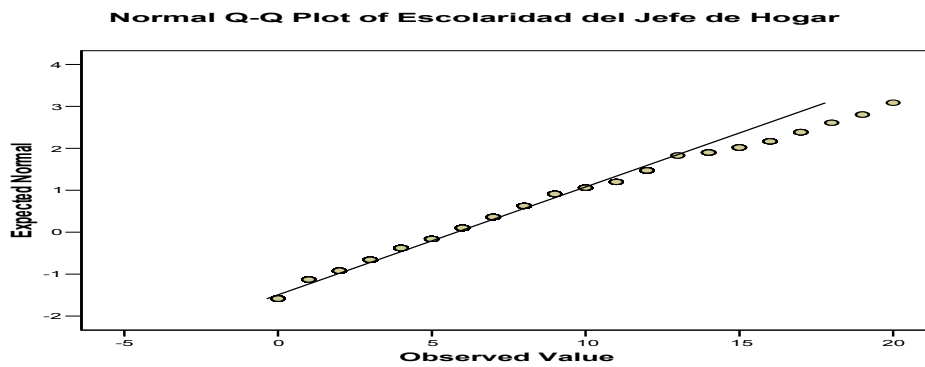
- Los jefes de hogares pobres de la zona rural al igual que en la zona urbana en mayor proporción pertenecen a una etnia.
- Las viviendas de los hogares pobres rurales si bien tienen menor acceso a los servicios como agua potable y alcantarillado esta es una carencia que se observa en ambos grupos.
- También hay una proporción mayor de hogares hacinados y viviendas de mala calidad en los hogares pobres.
- Los hogares de la zona rural en su gran mayoría carecen de bienes y en algunos casos, ningún grupo los posee (Tv cable), lo que fundamenta aún más la decisión de construir dos modelos separados.

5.1.2.2 Variables Continúas

Tabla 8 Análisis Descriptivo de las variables zona rural

Variable	Promedio	Mediana	Desviación estándar
Número de personas en el hogar	3,79	4	1,691
Años de escolaridad del jefe de hogar	5,79	6	3,885
Edad del jefe de hogar	52,67	51	16,138





En base a los gráficos “Plot-Plot” se puede asumir normalidad para las tres variables.

Al igual que en la zona urbana, una vez asumida la normalidad de las variables se realizaron las pruebas para verificar la diferencia de medias de las tres variables analizadas, según situación de pobreza.

Tabla 9 Promedio de personas por hogar, escolaridad y edad promedio del Jefe de Hogar según situación de pobreza – Zona Rural

	NIVEL DE POBREZA	N	Mean	Std. Deviation	Std. Error Mean
Numero de Personas por Hogar	NO	835	3,60	1,601	,055
	SI	163	4,74	1,661	,130
Escolaridad del Jefe de Hogar	NO	834	5,88	3,997	,138
	SI	163	5,34	3,229	,253
Edad del Jefe de Hogar	NO	835	53,89	16,309	,564
	SI	163	46,28	13,636	1,068

Tabla 10 Test de significación estadística de promedio de personas por hogar, escolaridad y edad promedio del Jefe de Hogar según situación de pobreza – zona rural

Independent Samples Test										
		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	95% Confidence Interval of the Difference	
									Lower	Upper
Numero de Person por Hogar	Equal varianc assumed	,238	,626	-8,305	996	,000	-1,146	,138	-1,417	-,875
	Equal varianc not assumed			-8,102	224,659	,000	-1,146	,141	-1,425	-,867
Escolaridad del Jef Hogar	Equal varianc assumed	10,880	,001	1,607	995	,108	,534	,332	-,118	1,187
	Equal varianc not assumed			1,853	268,868	,065	,534	,288	-,033	1,102
Edad del Jefe de H	Equal varianc assumed	12,767	,000	5,593	996	,000	7,617	1,362	4,945	10,290
	Equal varianc not assumed			6,306	261,154	,000	7,617	1,208	5,239	9,996

Los jefes de hogares pobres de la zona rural, al igual que en la zona urbana tienen un promedio de edad más bajo que los no pobres, son más jóvenes y el número de personas por hogar es mayor, como lo muestran las pruebas estadísticas que se utilizaron.

5.2 Modelo de Regresión Logística

5.2.1 Modelo Completo Inicial

$$\text{Logit } P(\text{pobreza}) = X' \hat{\beta}$$

con $X' =$ (1, NUMPER, ESC, OCUPADO, INACTIVO , SEXOJH , EDADJH, DISCA, LAVADORA, REFRI, FONON, VIDEO, MICRO, COMPUTA, CALEFONT, CELULAR, INTERNET, TVCABLE, ETNIA, AGUA, ALCANTA, VIVIENDA, NHOGAR, HACINA)

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{23})$$

5.2.2 Proceso de Selección de variables

A partir del modelo completo, la selección de variables se realizó en base a la probabilidad del coeficiente de regresión estimado por cada variable con un nivel de significación $\alpha = 0.09$. Este ajuste se efectuó mediante el software estadístico SPSS versión 13.

Como se mencionó en la sección anterior se construyeron dos modelos uno para la zona urbana y otro para la zona rural, como se explicó en el capítulo 3, la metodología de medición de pobreza indica “Canastas Alimentarias” distintas para cada una de las zonas, la canasta de la zona rural es un 66% más económica que

la de la zona urbana, por lo cual mantener como variable ficticia a “Zona” implica incluir una variable “confusora” en el modelo, ya que al construir el modelo lo más probable es que el signo del coeficiente de la variable indique que si la persona vive en la zona urbana la probabilidad de ser pobre aumenta, y esta situación sólo se explica por la metodología utilizada en la medición de pobreza y no por que efectivamente ocurra³² así.

5.2.3 Modelo Área Urbana

5.2.3.1 Parámetros Estimados del Modelo Completo

En la tabla 11 se presentan los parámetros del modelo completo, considerando 7 iteraciones que arrojó el software SPSS.

Se utilizó el estadístico de Wald para la prueba de hipótesis de cada uno de los coeficientes.

Las hipótesis son las siguientes: $H_0 : \beta_i = 0$

$H_1 : \beta_i$ no es igual a 0

³² Fue analizado computacionalmente y efectivamente Zona actúa como variable confusora.

Tabla 11 Variables del Modelo Completo

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	NUMPER	,373	,042	79,594	1	,000	1,452
	ESC	-,039	,023	2,882	1	,090	,962
	OCUPADO	-2,411	,282	73,054	1	,000	,090
	INACTIVO	-1,204	,312	14,846	1	,000	,300
	SEXO	,118	,173	,464	1	,496	1,125
	EDAD	-,045	,007	45,322	1	,000	,956
	DISCA	-,582	,359	2,624	1	,105	,559
	LAVADOR A	-,359	,164	4,794	1	,029	,698
	REFRI	-,092	,182	,255	1	,613	,912
	TELEFONO	-,739	,184	16,134	1	,000	,477
	VIDEO	-,653	,213	9,367	1	,002	,520
	MICROON DAS	-,394	,221	3,184	1	,074	,674
	COMPUTA DOR	-,192	,287	,446	1	,504	,825
	CALEFONT	-,639	,165	14,971	1	,000	,528
	CELULAR	-,089	,158	,317	1	,573	,915
	INTERNET	-,896	,686	1,708	1	,191	,408
	TVCABLE	-,640	,261	5,984	1	,014	,527
	ETNIA	,555	,303	3,359	1	,067	1,741
	AGUA	-,021	,364	,003	1	,954	,979
	ALCANTA	-,319	,302	1,117	1	,291	,727
VIVIENDA	,461	,436	1,119	1	,290	1,586	
HACINA	,466	,306	2,311	1	,128	1,593	
Constant	2,349	,668	12,379	1	,000	10,478	

a Variable(s) entered on step 1: NUMPER, ESC, OCUPADO, INACTIVO, SEXO, EDAD, DISCA, LAVADORA, REFRI, TELEFONO, VIDEO, MICROONDAS, COMPUTADOR, CALEFONT, CELULAR, INTERNET, TVCABLE, ETNIA, AGUA, ALCANTA, VIVIENDA, HACINA.

Como se mencionó en el capítulo 3, este estadístico se distribuye $X^2_{(1)}$. El valor crítico utilizado es de 9%, por lo cual se rechaza la hipótesis nula si el valor de la estadística de WALD es mayor que el valor crítico.

5.2.3.2 Bondad de Ajuste del modelo con 12 variables

Se eligió el mejor modelo logístico a través del método denominado “Adelante Condicional”³³.

Para el Modelo que contiene sólo el término constante, $-2\log$ Likelihood, es igual a 1908,095 (ver Anexo 4A) que se distribuye $\chi^2_{(n-1)}$.

Cuando el modelo tiene 12 variables explicativas, $-2\log$ Likelihood, es igual a 1350,949, que se distribuye como $\chi^2_{(n-1)}$.

Las Hipótesis son:

$$H_0 : \beta_0 = \beta_1 = \dots = \beta_k = 0$$

$$H_1 : \text{Algún } \beta_i \text{ distinto de } 0.$$

Esta estadística se distribuye como $\chi^2_{(n-k-1)}$, y se rechaza la hipótesis nula, si el valor de $-2\log$ Likelihood es mayor que $\chi^2_{\alpha(n-k-1)}$, donde α es el nivel de significación, en este caso es 0,05. Se rechaza H_0 , por lo tanto el modelo con 12 variables contribuye igual que el modelo saturado.

³³ Adelante Condicional. Método de selección por pasos *hacia adelante* que, partiendo del modelo *nulo*, va incorporando aquellas variables cuyo *estadístico de puntuación*, siendo significativo, posee la probabilidad asociada más pequeña. Tras incorporar al modelo una nueva variable, todas las variables incluidas hasta ese momento son revisadas para determinar si existe alguna que debe ser excluida (es decir, para determinar si, como consecuencia de la nueva incorporación, el *estadístico de puntuación* de alguna variable ha dejado de ser significativo). El proceso se detiene cuando entre las variables no incluidas el modelo no queda ninguna cuyo estadístico de puntuación sea significativo.

Tabla 12

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1350,949 ^a	,244	,396

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Si se utiliza el Test de Hosmer y Lemeshow³⁴, sugerido para un número de variables significativo, se plantea la siguiente hipótesis:

H₀ : El modelo ajustado es significativo , se ajusta correctamente

H₁: El modelo ajustado no es significativo, no se ajusta correctamente

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8,844	8	,356

De la tabla se observa que no hay antecedentes para rechazar H₀, por lo cual se concluye que el modelo ajustado es significativo.

³⁴ Esta prueba es utilizada cuando el modelo contiene un gran número de variables independientes. Evalúa la bondad del modelo construyendo una tabla de contingencia, divide la muestra en 10 grupos a partir de los deciles de las probabilidades estimadas (a los 10 grupos resultantes se les denomina deciles de riesgo). En cada decil de riesgo se calcula el número de casos que pertenecen a cada categoría de la variable dependiente (número observado) y el número de casos que el modelo pronostica que pertenecen a cada categoría de la variable dependiente (número esperado). Se compara el número de casos observado con el número de casos esperado. Este estadístico permite contrastar la hipótesis nula de igualdad de distribuciones, es decir, la hipótesis de que la variable dependiente se distribuye de la misma manera en los 10 deciles de riesgo o, que no existen diferencias entre las frecuencias observadas y las esperadas.

El mejor modelo para el área urbana, se obtuvo con 12 variables, que son las siguientes:

- Número de personas en el hogar
- Escolaridad del Jefe de Hogar
- Jefe de Hogar Ocupado
- Jefe de Hogar Inactivo
- Edad del Jefe de Hogar
- Hogar con lavadora
- Hogar con teléfono
- Hogar con video
- Hogar con microondas
- Hogar con calefont
- Hogar con TV cable
- Jefe de hogar pertenece a un pueblo originario

La mejor ecuación de regresión logística para predecir la situación de pobreza de los hogares en la zona urbana es:

$$P_i = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k X_k}} = \frac{1}{1 + e^{-z}}$$

$$\text{Logit P(pobreza)} = Z = X' \beta$$

con X' = (1, NUMPER, ESC, OCUPADO, INACTIVO, EDADJH, LAVADORA, TELEFONO, VIDEO, MICROONDAS, CALEFONT, TVCABLE, ETNIA)

y β = (2,705; 0,379; -0,048; -2,394; -1,201; -0,046; -0,376; -0,785; -0,727; -0,476; -0,703; -0,742; 0,509)

Como ya se mencionó se utilizó el método denominado “Adelante Condicional” para determinar las variables que se incorporan en el modelo.

5.2.3.3 Interpretación de los parámetros Estimados del Modelo – ODDS Ratio

El ODDS Ratio cambia cuando la i -ésima variable explicativa regresora se incrementa en una unidad, si:

$\beta_i > 0$ significa que el ODDS RATIO se incrementa.

$\beta_i < 0$ significa que el ODDS RATIO decrece.

$\beta_i = 0$ significa que el factor es igual a uno, lo cual hace que ODDS RATIO no varía.

Tabla 13

Variables in the Equation		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	NUMPER	,379	,040	91,429	1	,000	1,460
	ESC	-,048	,022	4,740	1	,029	,953
	OCUPADO	-2,394	,271	78,131	1	,000	,091
	INACTIVO	-1,201	,297	16,296	1	,000	,301
	EDAD	-,046	,006	52,371	1	,000	,955
	LAVADORA	-,376	,159	5,612	1	,018	,687
	TELEFONO	-,785	,179	19,329	1	,000	,456
	VIDEO	-,727	,209	12,106	1	,001	,483
	MICROONDAS	-,476	,215	4,883	1	,027	,621
	CALEFONT	-,703	,158	19,783	1	,000	,495
	TVCABLE	-,742	,258	8,312	1	,004	,476
	ETNIA	,509	,295	2,978	1	,084	1,664
	Constant	2,705	,496	29,765	1	,000	14,948

a. Variable(s) entered on step 1: NUMPER, ESC, OCUPADO, INACTIVO, EDAD, LAVADORA, TELEFONO, VIDEO, MICROONDAS, CALEFONT, TVCABLE, ETNIA.

En la columna $\exp(\beta)$ de la tabla 13, se observa que:

β_1 y $\beta_{12} > 0$, por lo tanto, por cada 1 unidad que aumente el número de personas por hogar, la probabilidad de ser pobre aumenta en 46%, manteniendo el resto de las variables constantes. De la misma manera si que el Jefe de hogar pertenece a un pueblo originario la probabilidad de que el hogar sea pobre aumenta en 66,4%.

$\beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$ y $\beta_{11} < 0$, por lo tanto, las variables, escolaridad, edad y actividad del Jefe de hogar, tenencia de lavadora, teléfono, video, microondas, calefón y conexión TV cable son factores que disminuyen la probabilidad de que el hogar sea pobre. Así, el hogar que posee lavadora, disminuye la probabilidad de ser pobre en 31,3% si mantiene constante el resto de las variables.

5.2.3.4 Curva de ROC – Sensibilidad y Especificidad

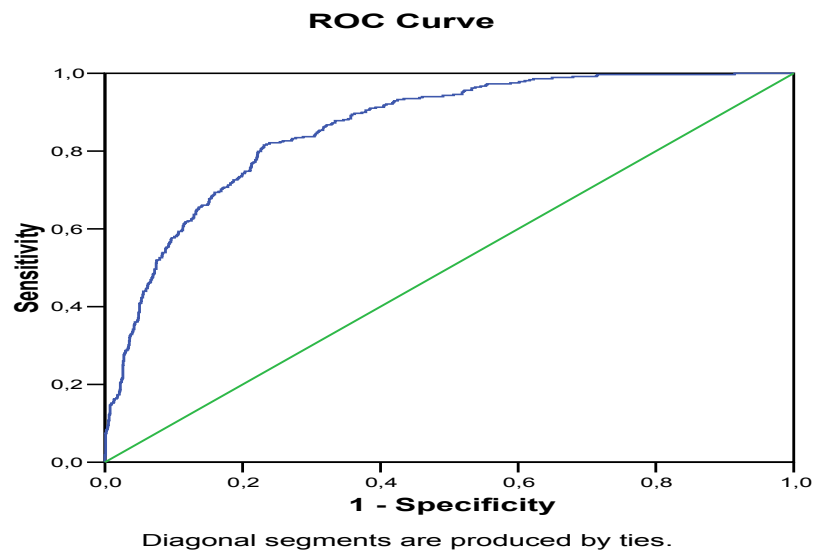
Cuando se utiliza un programa estadístico y se ejecuta una regresión logística, a cada observación se le asigna un valor de índice equivalente previsto de la regresión o también denominado valor predicho. Este valor predicho se utiliza para clasificar las unidades familiares u hogares como pobres o no pobres, los programas computacionales emplean normalmente una mitad como punto de

corte (los hogares que se encuentran por encima del punto de corte se clasifican como pobres). No obstante, este punto de corte puede cambiarse.

Dado a que lo que interesa es clasificar adecuadamente a los pobres , se prefiere tener una menor proporción de coincidencias en la predicción total del modelo, pero una mejor para la categoría “pobre”, por lo cual para determinar el punto de corte se utilizó la curva de ROC basada en los cálculos de la especificidad y sensibilidad del modelo³⁵, Para identificar el mejor punto de clasificaciones utilizó el SPSS graficando la “Curva de ROC” utilizando como variable la probabilidad predicha y la observada.

³⁵ Jovel; Albert J.; “Análisis de Regresión Logística” Centro de Investigaciones Sociológicas; Madrid, 1995 y Cea Dàncona; M. Ángeles; Análisis Multivariable. Teoría y Práctica en la Investigación Social; Editorial Síntesis. Madrid, 2002.

Gráfico nº 5



El eje de las abscisas, que corresponde a la proporción de falsos positivos, aparece etiquetado como 1-Especificidad y el eje de ordenadas, que corresponde a la proporción de aciertos positivos, aparece etiquetado como Sensibilidad. La línea curva escalonada representa los valores de 1-Especificidad y Sensibilidad para cada punto de corte ensayado (es decir, para cada una de las puntuaciones discriminantes incluidas en el análisis). Cada punto de esta curva se corresponde con un valor observado de la función (es decir, con un valor de la variable “pobre” en el archivo de datos).

Cuanto mayor es la curvatura de la línea, mayor es la capacidad de discriminación de la función. La mejor función discriminante posible sería aquella que permitiera obtener un 100% de aciertos positivos (sensibilidad = 1) Con un 0 % de falsos positivos (1-especificidad = 0). Pero en esa situación no existiría incertidumbre

sobre el punto de corte óptimo ya que las distribuciones de frecuencias de cada grupo se encontrarían completamente. Por otra parte, la peor curva posible sería aquella que careciera por completo de curvatura (recorrería la bisectriz del ángulo definido por los dos ejes representados). En una situación como ésta, las distribuciones de frecuencias de los dos grupos se encontrarían completamente solapadas y un aumento de los aciertos positivos llevaría asociado un incremento idéntico de falsos positivos. El gráfico incluye una línea de referencia (la bisectriz) que corresponde a esta situación.

La situación de peor discriminación posible es la representada por la línea de referencia, diagonal, una proporción de área igual a 0,5 (sería una situación en la que se obtendrían tantos aciertos positivos como falsos positivos). Por el contrario, en la situación de mejor discriminación posible, el área existente bajo la curva contendría el 100 % del espacio; es decir, una proporción de área igual a 1. En consecuencia, el área existente bajo la curva ROC no sólo es un indicador de la curvatura de la línea, sino que, además, tiene un mínimo y un máximo conocidos.

La Tabla siguiente entrega una estimación del tamaño del área existente bajo la curva ROC. También ofrece el error típico de esa estimación (error estándar) y el nivel crítico (Sig. asintótica) resultante de contrastar la hipótesis nula de que el área existente bajo la curva vale 0,5. La tabla también incluye los límites del intervalo de confianza (calculados al 95 %) correspondientes a la estimación efectuada. Si el intervalo de confianza incluye el valor 0,5 o, lo que es lo mismo, si

el nivel crítico es mayor que 0,05, no se podrá rechazar la hipótesis nula y, por tanto, no se podrá afirmar que los puntos de corte muestreados generen una curva COR bajo la cual exista un área significativamente mayor que 0,5. Por el contrario, si el intervalo de confianza no incluye el valor 0,5, o lo que es lo mismo, si el nivel crítico es menor que 0,05, se podrá rechazar la hipótesis nula y concluir que el área existente bajo la curva ROC es significativamente mayor que 0,5.

Tabla 14 Área Bajo la Curva

Test Result Variable(s): Probabilidad Predicha

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,860	,010	,000	,841	,879

The test result variable(s): Probabilidad Predicha has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a Under the nonparametric assumption

b Null hypothesis: true area = 0.5

El gráfico de la curva ROC permite apreciar una curvatura muy pronunciada; y el área estimada (0,86) toma un valor próximo a 1. Además, dados los valores del nivel crítico y del intervalo de confianza, se puede concluir que esta área es significativamente mayor que 0,5, por lo cual indica una capacidad o eficacia predictiva del modelo alta.

La Tabla muestra una selección de los puntos representados (máximos y mínimos y las que se encuentran en torno al valor que corresponde al mejor punto de corte) en el gráfico, corresponde a los distintos valores que toma la función ordenados de menor a mayor, y las coordenadas correspondientes a esos valores

en cada uno de los ejes del gráfico. El punto de corte va a estar dado por el cruce en el gráfico entre la especificidad y la sensibilidad.

Tabla 15 Coordenadas de la Curva ROC

Punto	Sensibilidad a	1- Especificidad B	1- Sensibilidad C	Especificidad d	a-d
0,0000	1,0000	1,0000	0,0000	0,0000	1,0000
0,0009	1,0000	0,9994	0,0000	0,0006	0,9994
0,0009	1,0000	0,9988	0,0000	0,0012	0,9988
.....
.....
0,1994	0,7832	0,2210	0,2168	0,7790	0,0042
0,1997	0,7832	0,2204	0,2168	0,7796	0,0036
0,2000	0,7805	0,2204	0,2195	0,7796	0,0009
0,2002	0,7778	0,2204	0,2222	0,7796	-0,0019
0,2004	0,7778	0,2198	0,2222	0,7802	-0,0025
0,2012	0,7778	0,2191	0,2222	0,7809	-0,0031
.....
.....
.....
0,9681	0,0054	0,0000	0,9946	1,0000	-0,9946
0,9734	0,0027	0,0000	0,9973	1,0000	-0,9973
1,0000	0,0000	0,0000	1,0000	1,0000	-1,0000

El mejor punto de corte se encuentra en 0,20 el cual coincide con los niveles de pobreza urbana en Chile es un poco menos de 20%, como lo muestra el gráfico anterior, pero se cortará en 0,25, ya que el objetivo final de este modelo es asignar programas sociales a los hogares pobres y los estudios³⁶ muestran que hay un grupo importante de ellos que se encuentra en el borde de la línea y que transita permanentemente en estas dos situaciones y además que interesa de sobremanera, de que los hogares pobres queden bien clasificados.

³⁶ Contreras, D.; Cooper, R.; Herman J. Y Neilson C.; Dinámica de la Pobreza y Movilidad Social: Chile 1996-2001; Depto. De Economía; U. De Chile (mimeo) agosto 2004.

A continuación se observa una tabla de 2 por 2; en dicha tabla se muestran los casos observados de la situación de pobreza en los hogares. En esta tabla se enfrentan los valores estimados y observados de la situación de pobreza de los hogares calculando el porcentaje de coincidencias, hay un 80,2%.

Tabla 16 Tabla de Clasificación

Valor Observado		Valor Predicho		Porcentaje coincidencia
		NIVEL DE POBREZA NO	SI	
NIVEL DE POBREZA NO	NO	1334	286	82,3
NIVEL DE POBREZA SI	SI	108	261	70,7
	TOTAL			80,2

Punto de Corte 0.25

5.2.3.5 Matriz de correlación

En la Tabla 17 se observa la matriz de correlación para las variables explicativas (X_1, X_2, \dots, X_{12}) y la variable dependiente Y (pobreza). Esta matriz muestra correlaciones bajas entre las variables.

Tabla 17 Matriz de Correlación

Correlation Matrix

	Constant	NUMPER	ESC	OCUPADO	INACTIVO	EDAD	LAVADORA	ELEFONO	VIDEO	MICROONDAS	CALEFONTO	TVCABLE	ETNIA	
Step 1	Constant	1,000	-.218	-.616	-.544	-.342	-.737	.020	.096	.043	.000	.032	-.009	-.074
	NUMPER	-.218	1,000	.017	-.099	.036	-.092	-.109	-.136	-.057	-.020	.022	-.017	-.025
	ESC	-.616	.017	1,000	.026	.074	.456	-.125	-.075	-.150	-.071	-.154	-.026	.049
	OCUPADO	-.544	-.099	.026	1,000	.753	.135	.002	.032	-.025	.060	.011	.008	-.032
	INACTIVO	-.342	.036	.074	.753	1,000	-.217	.016	.002	-.026	.045	.002	.009	-.025
	EDAD	-.737	-.092	.456	.135	-.217	1,000	-.026	-.127	.021	-.004	-.094	.006	.070
	LAVADORA	.020	-.109	-.125	.002	.016	-.026	1,000	-.085	-.104	-.140	-.189	-.024	-.012
	ELEFONO	.096	-.136	-.075	.032	.002	-.127	-.085	1,000	-.052	-.190	-.047	-.047	.039
	VIDEO	.043	-.057	-.150	-.025	-.026	.021	-.104	-.052	1,000	-.184	-.024	-.085	.034
	MICROONDAS	.000	-.020	-.071	.060	.045	-.004	-.140	-.185	-.184	1,000	-.083	-.060	-.031
	CALEFONTO	.032	.022	-.154	.011	.002	-.094	-.189	-.190	-.024	-.083	1,000	-.058	.020
	TVCABLE	-.009	-.017	-.026	.008	.009	.006	-.024	-.047	-.085	-.060	-.058	1,000	-.054
	ETNIA	-.074	-.025	.049	-.032	-.025	.070	-.012	.039	.034	-.031	.020	-.054	1,000

5.2.3.6 Interacciones

Se analizó si la variable sexo del jefe de hogar tenía un efecto de interacción con el resto de las variables utilizadas, se crearon las variables con el efecto interacción y el modelo fue testado, no presentando en ninguno de los casos interacciones significativas como se observa en la tabla siguiente:

Tabla 18

		B	S.E.	Wald	df	Sig.	Exp(B)
Step a 1	NUMPER	,406	,112	13,236	1	,000	1,501
	ESC	-,125	,046	7,325	1	,007	,883
	OCUPADO	-2,368	,271	76,121	1	,000	,094
	INACTIVO	-1,165	,303	14,772	1	,000	,312
	EDAD	-,039	,010	14,902	1	,000	,962
	LAVADORA	-,365	,159	5,244	1	,022	,695
	TELEFONO	-,785	,179	19,291	1	,000	,456
	VIDEO	-,729	,210	12,087	1	,001	,482
	MICROONDAS	-,462	,216	4,589	1	,032	,630
	CALEFONT	-,707	,159	19,819	1	,000	,493
	TVCABLE	-,719	,259	7,708	1	,005	,487
	ETNIA	1,136	,890	1,629	1	,202	3,114
	snumper	-,023	,077	,091	1	,762	,977
	sedad	-,006	,006	1,008	1	,315	,994
	sesc	,057	,030	3,679	1	,055	1,059
	setnia	-,475	,597	,633	1	,426	,622
	Constant	2,748	,499	30,343	1	,000	15,614

a. Variable(s) entered on step 1: NUMPER, ESC, OCUPADO, INACTIVO, EDAD, LAVADORA, TELEFONO, VIDEO, MICROONDAS, CALEFONT, TVCABLE, ETNIA, snumper, sedad, sesc, setnia.

5.2.3.7 Validación del Modelo Logístico

Para validar el modelo de regresión logística se utilizó la Encuesta CASEN 2003 completa, separada por área geográfica, encontrándose un 82% casos coincidentes, entre la variable original y la estimación del modelo.

Tabla 19 Cruce entre Niveles de Pobreza estimados y originales Zona Urbana

grupo predicho		Medición		Total	%
		POBREZA NO	original SI		
Pobres	NO	2531579	173503	2705082	93,6
	SI	476425	364769	841194	56,6
Total		3008004	538272	3546276	81,7

Fuente: Encuesta CASEN 2003.

5.2.4 Modelo Área Rural

El mismo procedimiento es utilizado en el análisis del área rural, las salidas computacionales se encuentran en el anexo nº4-B. Las siguientes variables son significativas en la determinación de la pobreza en las zonas rurales:

- Número de personas en el hogar
- Escolaridad del Jefe de Hogar
- Jefe de Hogar Ocupado
- Jefe de Hogar Inactivo
- Edad del Jefe de Hogar
- Hogar con refrigerador
- Hogar con Alcantarillado
- Vivienda de Calidad Aceptable

Solamente con ocho variables se puede predecir si un hogar va a ser pobre o no, y se debe destacar que solamente 1 bien (el refrigerador) es determinante y que el acceso al alcantarillado no es significativo en la zona urbana y si en la rural.

Tabla 20 Variables en el modelo rural

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	numper	,547	,065	71,739	1	,000	1,727
	esc	-,097	,036	7,329	1	,007	,907
	ocupado	-2,617	,486	29,036	1	,000	,073
	inactivo	-1,579	,515	9,393	1	,002	,206
	edad	-,060	,009	43,354	1	,000	,942
	Refrigerador	-,808	,213	14,363	1	,000	,446
	alcanta	-,842	,244	11,858	1	,001	,431
	vivienda	-,875	,413	4,484	1	,034	,417
	Constant	3,350	,803	17,412	1	,000	28,509

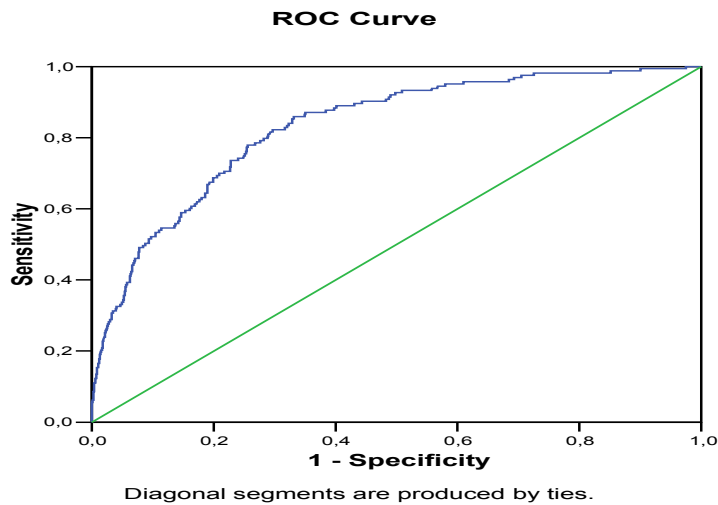
a. Variable(s) entered on step 1: numper, esc, ocupado, inactivo, edad, Refrigerador, alcanta, vivienda.

Tabla 21 Tabla de Clasificación

Valor	Observado	Valor Predicho		
		NIVEL DE POBREZA		Porcentaje Correcto
		NO	SI	
	NIVEL DE NO POBREZA	677	156	81,3
	SI	58	105	64,4
	Porcentaje			78,5

a El valor de corte es 0,210

El punto de corte seleccionado en la curva ROC es 0,21, lo que permite clasificar correctamente al 78,5% de los casos, a continuación se presenta la curva ROC y el test de significación correspondiente.



El gráfico de la curva ROC permite apreciar una curvatura muy pronunciada; y el área estimada 0,83 toma un valor próximo a 1. Además, dados los valores del nivel crítico y del intervalo de confianza, se puede concluir que esta área es significativamente mayor que 0,5.

Tabla 22 Área Bajo la Curva

Test Resultado: Probabilidad Predicha

Area	Std. Error(a)	Sig.(b)	95% Intervalo de Confianza	
			Minimo	Máximo
,830	,017	,000	,797	,863

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a Under the nonparametric assumption

b Null hypothesis: true area = 0.5

La mejor ecuación de regresión logística para predecir la situación de pobreza de los hogares en la zona rural es:

$$p_i = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k X_k}} = \frac{1}{1 + e^{-z}}$$

Logit P(pobreza) = Z = $\underline{X}' \beta$

con $\underline{X}' =$ (1, NUMPER, ESC, OCUPADO, INACTIVO , EDADJH, REFRIGERADOR, ALCANTARILLADO, VIVIENDA)

y $\beta =$ (3,350; 0,547; -0,097; -2,617; -1,579; -0,060; -0,808; -0,842; -0,875)

El cuál tiene más de un 78% de validación con los datos observados como lo muestra la tabla 21, en el modelo rural también fueron revisadas las interacciones las que no mostraron valores significativos.

Al igual que en el análisis urbano, para validar el modelo de regresión logística se utilizó la Encuesta CASEN 2003 completa para la zona urbana, encontrándose un 81,7% casos coincidentes, entre la variable original y la estimación del modelo.

Tabla 23 Cruce entre Niveles de Pobreza estimados y originales Zona Rural

GRUPO PREDICHO		POBREZA OBSERVADA		Total	Coincidencias %
		NO	SI		
Pobres	NO	351.617	31.320	382.937	91,8
	SI	108.119	59.099	167.218	64,7
Total		459.736	90.419	550.155	81,7

Fuente: Encuesta CASEN 2003

6.- Conclusiones

Esta investigación se ha concentrado en estimar la contribución que distintos factores socioeconómicos de los hogares tienen sobre la pobreza de éstos.

La regresión Logística es una de las herramientas estadísticas con mejor capacidad para el análisis, cuando la variable dependiente es categórica, sirve para determinar los factores de riesgo y factores de prevención frente a la situación de pobreza que enfrentan los hogares. En esta investigación se obtiene un modelo que cumple con los supuestos requeridos y que cumple con todos los tests estadísticos requeridos.

Con el modelo obtenido se alcanza un porcentaje de coincidencia entre la estimación de probabilidad de ser pobre y el valor observado de alrededor de un 80%, tanto en la zona rural como en la zona urbana.

En la zona urbana se obtiene un modelo que considera 12 variables significativas en la determinación de la pobreza. Aumenta la probabilidad de ser pobre: el número de personas por hogar, y la pertenencia o descendencia del jefe de hogar de algún pueblo originario, considerado en la Ley Indígena. Se destaca ésta variable, la que debería ser incluida entre las variables de selección en los programas de pobreza urbana, ya que si se mantienen el resto de las variables constantes y ésta aumenta en 1 unidad, la probabilidad de que

el hogar sea pobre crece en 66%.

Las variables, escolaridad, edad y actividad del Jefe de hogar, tenencia de lavadora, teléfono, video, microondas, calefón y conexión TV cable son factores que disminuyen la probabilidad de que el hogar sea pobre en la zona urbana.

Cabe destacar que cuando se realiza la prueba de medias en el capítulo 5, se demuestra que es significativa la diferencia de edad de los jefes de hogar pobres y no pobres, donde los que están a cargo de hogares pobres son más jóvenes.

En la zona rural, sólo 8 variables de las 23 utilizadas en el modelo inicial son significativas. Al igual que en la zona urbana, aumenta la probabilidad de ser pobre que exista un mayor número de personas en el hogar, y la disminuye el que los jefes de hogar tengan mayor cantidad de años de escolaridad, que este ocupado o sea inactivo.

La diferencia con la zona urbana se observa en el tipo de bienes, ya que en la zona rural hay un solo bien que discrimina entre ser pobre o no, y es la posesión de un refrigerador. En relación a los servicios, corresponde el acceso a alcantarillado y que la calidad de la vivienda sea aceptable.

Esto último demuestra las diferentes características de la pobreza en las distintas zonas geográficas, y la necesidad de tener modelos que predigan o discriminen en forma independiente para cada una de las zonas, esto es relevante, ya que en Chile, la zona rural concentra una mayor proporción de pobres, pero tiene una importancia relativa muy baja, 14% de la población del país vive en la zona rural. Por lo tanto, los modelos generales tienden a mostrar el comportamiento urbano, lo que implica una mala focalización en los recursos y el desarrollo de las políticas sociales en los sectores rurales.

7.- Revisión bibliográfica

Agresti, A.; Categorical data analysis. Wiley Interscience. Canadá 2002.

Altimir, Oscar; La dimensión de la pobreza en América Latina, Naciones Unidas, Cuadernos de la CEPAL, N°27, Santiago, 1979.

Atkinson, A.B.. On the Measurement of Inequality, en Journal of Economic Theory, núm. 2, 1970

BRIONES, G. (1995) Métodos y Técnicas de Investigación para las Ciencias Sociales, Editorial Trillas. México.

Cea Dàncona; M. Ángeles; Análisis Multivariable. Teoría y práctica en la Investigación Social; Editorial Síntesis. Madrid, 2002.

CELADE; Metodología de localización de bolsones de pobreza intracomunales y espacios de inversión. (mimeo, primer informe), Santiago, noviembre 1991.

CEPAL; Determinación de las necesidades de energía y proteínas para la población de nueve países latinoamericanos; (mimeo) Santiago, octubre 1990.

CEPAL; Una estimación de la magnitud de la pobreza en Chile; Santiago, octubre 1990.

CEPAL; Medición y análisis de la pobreza: Notas Técnicas, 26 de abril del 2002 (mimeo).

CEPAL; Uso de Regresiones Categóricas para comprobar el Rendimiento de los Parámetros de Determinación de Beneficiarios.

Chossudovsky; Global Poverty in the Late 20th Century” Journal of International Affairs, Vol 52, n°1, 1998.

Contreras, D.; Cooper, R.; Herman J. Y Neilson C.; Dinámica de la Pobreza y Movilidad Social: Chile 1996-2001; Depto. De Economía; U. De Chile (mimeo) agosto 2004.

Córtés Fernando; Regresión Logística en la investigación social: potencialidades y limitaciones; CES, COLMEX, página web.

Cumsille; Francisco; Análisis de Datos Categóricos; Apuntes de clase, Magíster Bioestadística, Escuela de Salud Pública; Facultad de Medicina; Universidad de Chile.

Feres, J.C. y Mancero X.; El método de las necesidades básicas insatisfechas y sus aplicaciones en América Latina; Estudios Estadísticos y prospectivos, CEPAL, mimeo, Santiago 2001.

Flores Manrique; L.; Análisis Estadístico de los Factores de Riesgo que Influyen en la Enfermedad Angina de Pecho. Tesis Digitales, Universidad Nacional de San Marcos, Lima, Perú, 2000.

French-Davis, Ricardo, Distribución y Pobreza en Chile; ICHEH, 1996 (mimeo).

Guerrero, Miguel, Metodología para la caracterización socioeconómica, págs 11-28 de "Como ha cambiado la vida de los chilenos"; INE, Chile 2004.

Gujarati, Damodar; "Econometría"; INE, Chile 2004. Ed. Mc Graw Hill, 3ª edición, Colombia, 1997.

Hair J.F; Anderson R.E.; Tatham R.L. y Black W.C. Análisis Multivariante; Ed. Prentice Hall; 1999.

Hosmer D. Y Lemeshow S.; Applied Logistic Regression; Wiley Series in Probability and Statistics; 2ª edición, Canadá, 2000.

Hosmer D. Y Lemeshow S.; Applied Logistic Regression; Wiley Interscience Publication; Solutions Manual to Accompany, Canadá, 2001.

Jovel; Albert J.; "Análisis de Regresión Logística" Centro de Investigaciones Sociológicas; Madrid, 1995.

INE; Censo de Población y Vivienda. Chile 2002; Santiago; 2003.

INE; Metodología de Clasificación Socioeconómica de los Hogares Chilenos; Santiago; agosto 2003.

Kaztman, Rubén; La heterogeneidad de la pobreza. El caso de Montevideo. Revista de la CEPAL N°37. Santiago, 1990.

Larrañaga, Osvaldo y equipo; Mejoramiento del Factor Discriminatorio de la Ficha CAS; Informe Final; Santiago, diciembre 2002.

Levy J.P. y Varela Jesús; Análisis Multivariante para Ciencias Sociales; Ed. Prentice Hall; 2003.

López O.; P. Evaluación de la Precisión del Diagnóstico de Edad Gestacional Fetal basado en los Datos de la Mujer; Tesis de Grado Magíster en Bioestadística; Escuela de salud Pública, Facultad de Medicina; U. De Chile, Santiago, 2002.

LOHR S. L. (2000): Muestreo: diseño y análisis. Thomson 1ª. Edición.

MIDEPLAN; Volumen 1: "Pobreza, Distribución del Ingreso e Impacto Distributivo del Gasto Social"; Santiago; agosto 2004.

MIDEPLAN; Distribución e Impacto Distributivo del Gasto Social en los Hogares, 1996. Santiago, mayo 1998. Mimeo.

MIDEPLAN - PNUD, Desarrollo Humano en las Comunas de Chile; Temas de Desarrollo Humano Sustentable, N°3, Santiago, 2000.

Miller S.M.; Rein M., Roby M. y Cross B., Poverty, Inequality and Conflict en Annals of the American Academy of Political Science, 1967.

Moon, Bruce, The Political Economy of Basic Human Needs; Cornell University Press, 1991.

Ortúzar, Juan de Dios; Modelos Econonétricos de Elección Discreta. Ediciones Universidad Católica de Chile; 2000.

Pagano, M. Y Gauvreau, K. Principles of Biostatistics; Duxbury, Thomson Learning; USA, 2000.

Pérez, César; Análisis Estadístico Multivariado con SPSS; Ed. Prentice Hall, Madrid, España, 2004.

Pérez, César; Técnicas Estadísticas con SPSS; Ed. Prentice Hall, Madrid, España, 2003.

PNUD, Desarrollo sin Pobreza. II Conferencia Regional sobre la Pobreza en América Latina y el Caribe, Quito, Ecuador, 1990.

Raczynski, Dagmar, Políticas Sociales y Programas de Combate a la Pobreza en Chile: balance y desafíos, Colección Estudios CIEPLAN, 39, Santiago, Chile, 1994.

Rajaraman, Indira; Constructing the Poverty Line: Rural Punjab, 1960-1961, Discussion Paper, núm. 43, Programa de Investigación en Desarrollo Económico, Universidad de Princeton.

Rein, M. ; Problems in the Definition and Measurement of Poverty, en Peter Townsend, The Concept of Poverty, Heineman, Londres, 1971, p. 46.

Sen, Amartya y Nussbaum M.; La calidad de vida. ; Fondo de Cultura Económica; México, 1993.

Sen, Amartya, Sobre conceptos y medidas de Pobreza. Revista de Comercio Exterior, vl 42, número 4; México, abril 1992.

Sen, Amartya, Sobre conceptos y medidas de Pobreza. Revista de Comercio Exterior, vl 42, número 4; México, abril 1992.

Silva Aycaguer; L.C., , Excursión a la Regresión Logística en Ciencias de la Salud..Ediciones Diaz de Santos; Madrid 1995.

Skoufias, Emmanuel; Evaluación de la selección de hogares beneficiarios en PROGRESA. Programa de Educación, Salud y Alimentación. PROGRESA; México 2004; mimeo.

Stigler, G. J. ; The Cost of Subsistence, en Journal of Farm Economics, núm. 27,1945;

Townsend, Peter; Poverty as Relative Deprivation: Resources and Styles of Living, 3 (ed.), en Dorothy Wedderburn; Poverty, Inequality and Class Structure, Cambridge University Press, Cambridge, 1974

Sen, Amartya, Sobre conceptos y medidas de Pobreza. Revista de Comercio Exterior, vl 42, número 4 México, abril 1992.

Vivanco, M., Análisis Estadístico Multivariado, Teoría y práctica. Ed. Universitaria, Santiago, Chile, 1999.

8.- ANEXOS

AnexoN° 1 Metodología de medición de la pobreza: El enfoque del ingreso o Línea de Pobreza (LP)

Para calcular el método de la Línea de Pobreza (LP) se procede de la siguiente manera:

- a) Definición de las necesidades básicas y de sus componentes.
- b) Definición de una canasta de satisfactores esenciales para cada hogar.
- c) Cálculo del costo de la Canasta que se constituye en la línea de Pobreza.
- d) Comparación de la Línea de Pobreza con el Ingreso del Hogar.
- e) Clasificación de todos los hogares cuyo ingreso es menor que la línea de pobreza, como pobres. Todas las personas que pertenecen a un hogar pobre se consideran pobres.

Para determinar la composición de la Canasta Básica de Alimentos se consideraron los siguientes aspectos:

- Los requerimientos nutricionales de la población
- La estructura y el costo de la canasta básica de alimentos
- La cuantía de los recursos necesarios para satisfacer las necesidades alimentarias.

Para calcular los requerimientos nutricionales de la población se consideran³⁷:

- La estructura sociodemográfica según área urbana y rural
- Categorías de actividad física de las personas

En el siguiente cuadro se muestran los requerimientos medios estimados tanto en materia de energía como de proteínas necesarias en Chile:

³⁷ CEPAL: Determinación de los requerimientos nutricionales en Chile; mimeo; Santiago, 1990.

Cuadro N°1

	Calorías	Proteínas
Requerimiento Promedio Nacional	2187	37
Requerimiento Población Urbana	2176	37
Requerimiento Población Rural	2236	38

En síntesis:

La forma habitual de medir pobreza utilizada en América Latina consiste en:

1. Definir una Canasta Alimentaria que satisfaga los requerimientos mínimos (calorías, proteínas, etc.) en base al consumo real de la población de cada país.
2. Se calcula el costo de esta Canasta de Alimentos y éste es considerado como la "Línea de Indigencia" que corresponde al costo mensual per cápita de una canasta básica alimentaria, cuya composición cubre las necesidades nutricionales promedio de la población y considera sus hábitos de consumo predominantes.

Se estiman dos líneas de indigencia, una correspondiente a las zonas urbanas y la segunda a las zonas rurales. Las diferencias entre ambas se derivan del menor costo de los alimentos y del mayor requerimiento promedio de energía que, según estudios previos, caracteriza a las zonas rurales. En definitiva, la línea de indigencia rural representa el 77% del valor de la urbana.

3. Para definir la Línea de Pobreza se multiplica el costo de la Canasta Alimentaria por un factor, con el fin de obtener el costo de la Canasta necesaria para satisfacer las necesidades básicas (incluyendo las no alimentarias).
4. Este factor se elige con base en el cociente observado entre el gasto total de consumo del hogar y el gasto en alimentos, del primer estrato de hogares que satisfaga sus requerimientos nutricionales, este factor se obtiene de las Encuestas de Presupuestos Familiares (habitualmente este factor es 2).

5. Luego se calcula el porcentaje de hogares en situación de indigencia que corresponde a aquellos cuyo ingreso mensual per cápita es inferior a la Línea de Indigencia en relación al total de hogares. Una vez identificados estos hogares, es posible obtener el porcentaje de población indigente, que corresponde a las personas que son miembros de dichos hogares con relación al total de población.
6. La Línea de Pobreza en las zonas urbanas se calcula como el doble del valor de la línea de indigencia urbana, en tanto que el de las zonas rurales aquélla asciende a 1.75 vez el valor de la línea de indigencia rural.
Estas relaciones entre gastos básicos alimentarios y no alimentarios se derivan también de estudios previos realizados por CEPAL.

Los valores de la línea de pobreza son los siguientes:

Cuadro : Línea de Indigencia y Pobreza (\$nominal y real)

Zona	Tipo de Canasta	2000		2003
		\$ nominal	\$ real (Nov 2003)	\$ real (Nov 2003)
Urbana	Indigente	20.281	21.826	21.856
	Pobre no Indigente	40.562	43.653	43.712
Rural	Indigente	15.616	16.806	16.842
	Pobre no Indigente	27.328	29.410	29.473

Fuente: División Social MIDEPLAN en base a Casen 2000, 2003

Anexo N°2 – Ficha técnica de la Encuesta CASEN 2003³⁸

Universo

La encuesta es representativa de la población que habita en hogares particulares del país. Es posible realizar estimaciones a nivel nacional, por zona, regional y comunal. El año 2003, es posible realizar estimaciones en 302 de un total de 341 comunas.

Diseño Muestral

El tipo de muestreo utilizado es:

- **Estratificado:** La estratificación utilizada es de tipo geográfico. El país se dividió en estratos, entendiéndose como tal a la conjunción de división político administrativa (comuna o agrupación de comunas) y área geográfica (urbana o rural).
- **Por conglomerados:** En el área urbana están constituidos por los sectores de empadronamiento censal y en la rural como un conjunto de viviendas próximas.
- **Probabilístico:** En cada conglomerado se seleccionan sectores censales con probabilidad proporcional al tamaño de éste, medido por el número de viviendas. Las viviendas a encuestar se eligen dentro de cada sector seleccionado.

Tamaño de la muestra

El año 2003, la muestra consideró 68.400 viviendas, se encuestaron 68.150 hogares, esto es, aproximadamente, 272.000 personas.

³⁸ MIDEPLAN; Volumen 1: "Pobreza, Distribución del Ingreso e Impacto Distributivo del Gasto Social"; Santiago; agosto 2004.

El error muestral a nivel nacional (hogares) es de 0,4%, a nivel urbano, 0,5% y a nivel rural, 0, 8%. El máximo error a nivel de región y zona es de 5%.

Hogares muestrales

Región	Zona			Zona		
	Urbana	%	Muestra	Rural	%	Muestra
I	95.822	2,7	54	5.156	0,9	9
II	110.688	3,1	62	1.678	0,3	3
III	66.428	1,9	37	5.775	1,1	10
IV	117.172	3,3	66	39.984	7,3	72
V	400.522	11,3	225	31.600	5,7	57
VI	155.168	4,4	87	66.823	12,1	121
VII	158.785	4,5	89	83.688	15,2	152
VIII	426.433	12,0	239	96.674	17,5	175
IX	159.795	4,5	90	71.399	12,9	129
X	187.781	5,3	105	95.442	17,3	173
XI	19.810	0,6	11	6.464	1,2	12
XII	43.107	1,2	24	3.578	0,7	6
R.M.	1.619.601	45,5	910	43.465	7,9	79
Total	3.561.112	100,0	2.000	551.726	100,0	1.000

Fecha del levantamiento de la encuesta

Entre el 8 de noviembre y el 20 de diciembre del 2003.

Definiciones

“Línea de pobreza”: ingreso mínimo por persona para cubrir el costo de una canasta mínima individual para la satisfacción de las necesidades alimentarias y no alimentarias. Los hogares pobres son aquellos cuyos ingresos no alcanzan para satisfacer las necesidades básicas de sus miembros (\$43.712 en zona urbana, y \$29.473 en zona rural). La diferencia de líneas de corte urbano y rural se deben a la diferencia de disponibilidad de ingresos monetarios en cada una.

“Línea de indigencia”: ingreso mínimo por persona para cubrir el costo de una canasta alimentaria. Son indigentes los hogares que, aun cuando destinan todos sus ingresos a satisfacer las necesidades alimentarias de sus miembros, no logran cubrirlas adecuadamente (\$21.856 urbano y \$16.842 rural).

Hogar: aquel constituido por una persona o un grupo de personas, con o sin vínculos de parentesco, que comparten vivienda y tienen presupuesto de alimentación común.

Anexo 3: Frecuencias y Gráficos de las variables

a) Zona Urbana

Región en la que vive

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid I	54	2,7	2,7	2,7
II	62	3,1	3,1	5,8
III	37	1,9	1,9	7,7
IV	66	3,3	3,3	11,0
V	225	11,3	11,3	22,2
VI	87	4,4	4,4	26,6
VII	89	4,5	4,5	31,0
VIII	239	12,0	12,0	43,0
IX	90	4,5	4,5	47,5
X	105	5,3	5,3	52,7
XI	11	,6	,6	53,3
XII	24	1,2	1,2	54,5
R.M.	910	45,5	45,5	100,0
Total	1999	100,0	100,0	

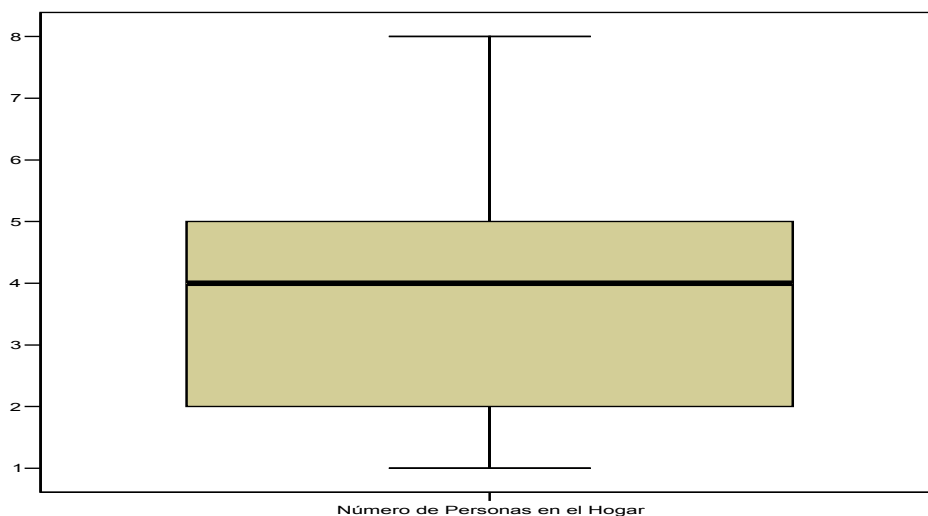
¿Hogar es Pobre:?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid NO	1629	81,5	81,5	81,5
SI	370	18,5	18,5	100,0
Total	1999	100,0	100,0	

Número de personas en el hogar

Descriptives

		Statistic	Std. Error
Número de Personas en el Hogar	Mean	3,67	,038
	95% Confidence Interval for Mean	3,59	
	Lower Bound	3,74	
	Upper Bound		
	5% Trimmed Mean	3,60	
	Median	4,00	
	Variance	2,888	
	Std. Deviation	1,699	
	Minimum	1	
	Maximum	8	
	Range	7	
	Interquartile Range	3	
	Skewness	,372	,055
	Kurtosis	-,271	,109



Promedio de Edad de Escolaridad de los Jefes de Hogar

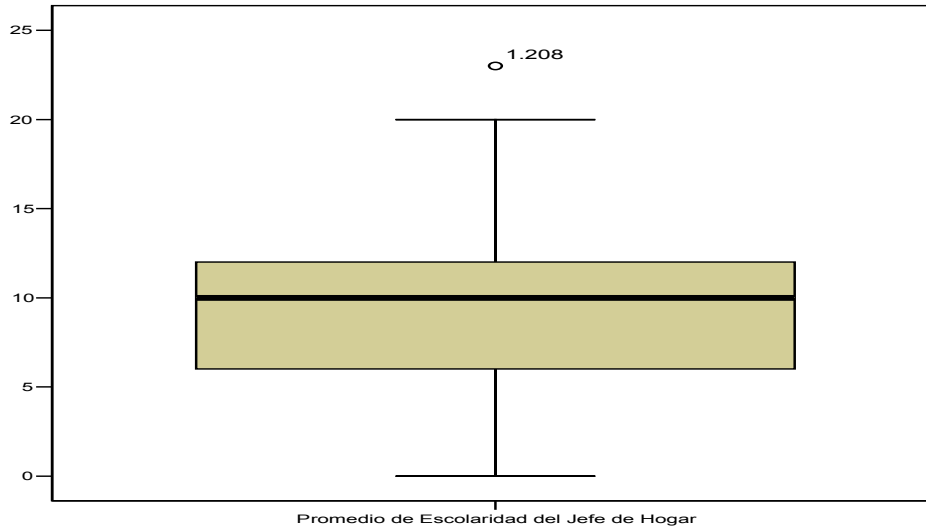
Descriptives

			Statistic	Std. Error
Promedio de Escolaridad del Jefe de Hogar	Mean		9,30	,102
	95% Confidence Interval for Mean	Lower Bound	9,11	
		Upper Bound	9,50	
	5% Trimmed Mean		9,33	
	Median		10,00	
	Variance		20,496	
	Std. Deviation		4,527	
	Minimum		0	
	Maximum		23	
	Range		23	
	Interquartile Range		6	
	Skewness		-,084	,055
	Kurtosis		-,480	,110

Percentiles

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average	Promedio de Escolaridad del Jefe de Hogar	1,50	3,00	6,00	10,00	12,00	15,00	17,00
Tukey's Hinges	Promedio de Escolaridad del Jefe de Hogar			6,00	10,00	12,00		



¿Jefe de Hogar está Ocupado?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	619	31,0	31,0	31,0
	SI	1380	69,0	69,0	100,0
	Total	1999	100,0	100,0	

Jefe de Hogar está Inactivo?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	1481	74,1	74,1	74,1
	SI	518	25,9	25,9	100,0
	Total	1999	100,0	100,0	

Sexo del Jefe de Hogar

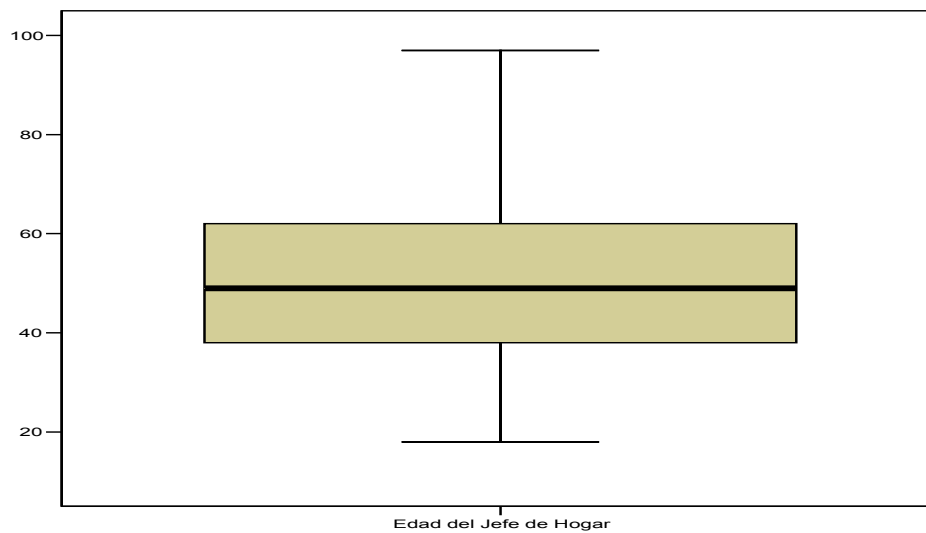
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MUJER	542	27,1	27,1	27,1
	HOMBRE	1457	72,9	72,9	100,0
	Total	1999	100,0	100,0	

Descriptives

			Statistic	Std. Error
Edad del Jefe de Hogar	Mean		50,37	,353
	95% Confidence Interval for Mean	Lower Bound	49,68	
		Upper Bound	51,07	
	5% Trimmed Mean		50,02	
	Median		49,00	
	Variance		249,576	
	Std. Deviation		15,798	
	Minimum		18	
	Maximum		97	
	Range		79	
	Interquartile Range		24	
	Skewness		,340	,055
	Kurtosis		-,617	,109

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average(Definition)	Edad del Jefe de Ho	27,00	31,00	38,00	49,00	62,00	73,00	78,00
Tukey's Hinges	Edad del Jefe de Ho			38,00	49,00	62,00		



Lavadora Automática

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	876	43,8	43,8	43,8
	SI	1123	56,2	56,2	100,0
	Total	1999	100,0	100,0	

Refrigerador

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	322	16,1	16,1	16,1
	SI	1677	83,9	83,9	100,0
	Total	1999	100,0	100,0	

Teléfono fijo

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	1015	50,8	50,8	50,8
	SI	984	49,2	49,2	100,0
	Total	1999	100,0	100,0	

Videograbador

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	1333	66,7	66,7	66,7
	SI	666	33,3	33,3	100,0
	Total	1999	100,0	100,0	

Microondas

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	1327	66,4	66,4	66,4
	SI	672	33,6	33,6	100,0
	Total	1999	100,0	100,0	

Computador

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	1552	77,6	77,6	77,6
	SI	447	22,4	22,4	100,0
	Total	1999	100,0	100,0	

Cálefont

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	783	39,2	39,2	39,2
	SI	1216	60,8	60,8	100,0
	Total	1999	100,0	100,0	

Teléfono Móvil

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	1145	57,3	57,3	57,3
	SI	854	42,7	42,7	100,0
	Total	1999	100,0	100,0	

INTERNET

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	1815	90,8	90,8	90,8
	SI	184	9,2	9,2	100,0
	Total	1999	100,0	100,0	

TVCABLE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	1575	78,8	78,8	78,8
	SI	424	21,2	21,2	100,0
	Total	1999	100,0	100,0	

ETNIA

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	1922	96,1	96,1	96,1
	SI	77	3,9	3,9	100,0
	Total	1999	100,0	100,0	

AGUA

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	103	5,2	5,2	5,2
	SI	1896	94,8	94,8	100,0
	Total	1999	100,0	100,0	

ALCANTA

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	154	7,7	7,7	7,7
	SI	1845	92,3	92,3	100,0
	Total	1999	100,0	100,0	

VIVIENDA

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	54	2,7	2,7	2,7
	SI	1945	97,3	97,3	100,0
	Total	1999	100,0	100,0	

NHOGAR

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	1920	96,0	96,0	96,0
	SI	79	4,0	4,0	100,0
	Total	1999	100,0	100,0	

Hacinamiento

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	1892	94,6	96,3	96,3
	SI	72	3,6	3,7	100,0
	Total	1964	98,2	100,0	
Missing	System	35	1,8		
Total		1999	100,0		

b) Frecuencias de las variables Zona Rural

Región

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid I	5	1,0	1,0	1,0
II	2	,4	,4	1,4
III	5	1,0	1,0	2,4
IV	36	7,2	7,2	9,6
V	29	5,8	5,8	15,4
VI	61	12,2	12,2	27,5
VII	76	15,2	15,2	42,7
VIII	88	17,6	17,6	60,3
IX	65	13,0	13,0	73,3
X	86	17,2	17,2	90,4
XI	6	1,2	1,2	91,6
XII	3	,6	,6	92,2
R.M.	39	7,8	7,8	100,0
Total	501	100,0	100,0	

NIVEL DE POBREZA

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid NO	419	83,6	83,6	83,6
SI	82	16,4	16,4	100,0
Total	501	100,0	100,0	

NUMPER

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	43	8,6	8,6	8,6
2	63	12,6	12,6	21,2
3	112	22,4	22,4	43,5
4	130	25,9	25,9	69,5
5	82	16,4	16,4	85,8
6	34	6,8	6,8	92,6
7	17	3,4	3,4	96,0
8	9	1,8	1,8	97,8
9	6	1,2	1,2	99,0
10	1	,2	,2	99,2
11	3	,6	,6	99,8
13	1	,2	,2	100,0
Total	501	100,0	100,0	

Escolaridad

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	56	11,2	11,2	11,2
	1	17	3,4	3,4	14,6
	2	29	5,8	5,8	20,4
	3	51	10,2	10,2	30,5
	4	51	10,2	10,2	40,7
	5	29	5,8	5,8	46,5
	6	70	14,0	14,0	60,5
	7	23	4,6	4,6	65,1
	8	80	16,0	16,0	81,0
	9	19	3,8	3,8	84,8
	10	16	3,2	3,2	88,0
	11	9	1,8	1,8	89,8
	12	32	6,4	6,4	96,2
	13	3	,6	,6	96,8
	14	1	,2	,2	97,0
	15	3	,6	,6	97,6
	16	4	,8	,8	98,4
	17	4	,8	,8	99,2
	18	1	,2	,2	99,4
	19	3	,6	,6	100,0
	Total	501	100,0	100,0	

OCUPADO

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	146	29,1	29,1	29,1
	SI	355	70,9	70,9	100,0
	Total	501	100,0	100,0	

INACTIVO

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	367	73,3	73,3	73,3
	SI	134	26,7	26,7	100,0
	Total	501	100,0	100,0	

Sexo

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MUJER	87	17,4	17,4	17,4
	HOMBRE	414	82,6	82,6	100,0
	Total	501	100,0	100,0	

Estadísticas Descriptivas

		Statistic	Std. Error
Numero de Personas por Hogar	Mean	3,79	,054
	95% Confidence Interval for Mean	3,69	
	Lower Bound		
	Upper Bound	3,90	
	5% Trimmed Mean	3,73	
	Median	4,00	
	Variance	2,860	
	Std. Deviation	1,691	
	Minimum	1	
	Maximum	14	
	Range	13	
	Interquartile Range	2	
	Skewness	,492	,077
	Kurtosis	,857	,155
Escolaridad del Jefe de Hogar	Mean	5,79	,123
	95% Confidence Interval for Mean	5,55	
	Lower Bound		
	Upper Bound	6,03	
	5% Trimmed Mean	5,62	
	Median	6,00	
	Variance	15,094	
	Std. Deviation	3,885	
	Minimum	0	
	Maximum	20	
	Range	20	
	Interquartile Range	5	
	Skewness	,523	,077
	Kurtosis	,111	,155
Edad del Jefe de Hogar	Mean	52,67	,511
	95% Confidence Interval for Mean	51,67	
	Lower Bound		
	Upper Bound	53,67	
	5% Trimmed Mean	52,40	
	Median	51,00	
	Variance	260,430	
	Std. Deviation	16,138	
	Minimum	18	
	Maximum	103	
	Range	85	
	Interquartile Range	25	
	Skewness	,250	,077
	Kurtosis	-,745	,155

Lavadora Automática

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	371	74,1	74,1	74,1
	SI	130	25,9	25,9	100,0
	Total	501	100,0	100,0	

Refrigerador

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	181	36,1	36,1	36,1
	SI	320	63,9	63,9	100,0
	Total	501	100,0	100,0	

Teléfono fijo

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	466	93,0	93,0	93,0
	SI	35	7,0	7,0	100,0
	Total	501	100,0	100,0	

Videgrabador

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	454	90,6	90,6	90,6
	SI	47	9,4	9,4	100,0
	Total	501	100,0	100,0	

Microondas

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	463	92,4	92,4	92,4
	SI	38	7,6	7,6	100,0
	Total	501	100,0	100,0	

Computador

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	481	96,0	96,0	96,0
	SI	20	4,0	4,0	100,0
	Total	501	100,0	100,0	

Cálefont

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	409	81,6	81,6	81,6
	SI	92	18,4	18,4	100,0
	Total	501	100,0	100,0	

Teléfono Móvil

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	347	69,3	69,3	69,3
	SI	154	30,7	30,7	100,0
	Total	501	100,0	100,0	

INTERNET

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	500	99,8	99,8	99,8
	SI	1	,2	,2	100,0
	Total	501	100,0	100,0	

TVCABLE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	479	95,6	95,6	95,6
	SI	22	4,4	4,4	100,0
	Total	501	100,0	100,0	

ETNIA

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	424	84,6	84,6	84,6
	SI	77	15,4	15,4	100,0
	Total	501	100,0	100,0	

AGUA

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	318	63,5	63,5	63,5
	SI	183	36,5	36,5	100,0
	Total	501	100,0	100,0	

ALCANTA

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	296	59,1	59,1	59,1
	SI	205	40,9	40,9	100,0
	Total	501	100,0	100,0	

VIVIENDA

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	22	4,4	4,4	4,4
	SI	479	95,6	95,6	100,0
	Total	501	100,0	100,0	

NHOGAR

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	493	98,4	98,4	98,4
	SI	8	1,6	1,6	100,0
	Total	501	100,0	100,0	

Hacinamiento

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NO	473	94,4	95,4	95,4
	SI	23	4,6	4,6	100,0
	Total	496	99,0	100,0	
Missing	System	5	1,0		
Total		501	100,0		

Anexo 4a Modelo de Regresión Logística Zona Urbana

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	1989	99,5
	Missing Cases	10	,5
	Total	1999	100,0
Unselected Cases		0	,0
Total		1999	100,0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
NO	0
SI	1

Block 0: Beginning Block

Iteration History^{a,b,c}

Iteration		-2 Log likelihood	Coefficients
			Constant
Step 1		1923,524	-1,258
0	2	1908,151	-1,466
	3	1908,095	-1,479
	4	1908,095	-1,479

- a. Constant is included in the model.
 b. Initial -2 Log Likelihood: 1908,095
 c. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Classification Table^{a,b}

Observed			Predicted		
			NIVEL DE POBREZA DEL HOGAR		Percentage Correct
			NO	SI	
Step 0	NIVEL DE POBREZA NO		1620	0	100,0
	DEL HOGAR SI		369	0	,0
Overall Percentage					81,4

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

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-1,479	,058	657,762	1	,000	,228

Variables not in the Equation

Step	Variables	Score	df	Sig.
0	NUMPER	82,090	1	,000
	ESC	44,500	1	,000
	OCUPADO	37,257	1	,000
	INACTIVO	,481	1	,488
	EDAD	40,825	1	,000
	LAVADORA	110,980	1	,000
	TELEFONO	169,411	1	,000
	VIDEO	96,199	1	,000
	MICROONDAS	98,826	1	,000
	CALEFONT	197,939	1	,000
	TVCABLE	63,379	1	,000
ETNIA	19,360	1	,000	
Overall Statistics		513,859	12	,000

Block 1: Method = Enter

Iteration History^{a,c,d}

Iteration	-2 Log Likelihood	Coefficients												
		Constant	NUMPER	ESC	OCUPADO	INACTIVO	EDAD	LAVADORA	TELEFONO	VIDEO	MICROONDAS	CALEFONT	TVCABLE	ETNIA
Step 1	527,802	1,173	,182	-,010	-1,611	-1,056	-,019	-,210	-,305	-,257	-,123	-,434	-,137	,370
1	378,072	2,008	,302	-,028	-2,081	-1,133	-,035	-,323	-,563	-,498	-,275	-,626	-,362	,462
3	352,480	2,548	,363	-,043	-2,324	-1,180	-,044	-,366	-,731	-,667	-,414	-,690	-,614	,496
4	350,958	2,695	,378	-,048	-2,389	-1,199	-,046	-,375	-,781	-,722	-,471	-,703	-,730	,508
5	350,949	2,705	,379	-,048	-2,394	-1,201	-,046	-,376	-,785	-,727	-,476	-,703	-,742	,509
6	350,949	2,705	,379	-,048	-2,394	-1,201	-,046	-,376	-,785	-,727	-,476	-,703	-,742	,509

a.Method: Enter

b.Constant is included in the model.

c.Initial -2 Log Likelihood: 1908,095

d.Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	557,145	12	,000
Block	557,145	12	,000
Model	557,145	12	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1350,949 ^a	,244	,396

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8,844	8	,356

Contingency Table for Hosmer and Lemeshow Test

		NIVEL DE POBREZA DEL HOGAR = NO		NIVEL DE POBREZA DEL HOGAR = SI		Total
		Observed	Expected	Observed	Expected	
Step 1	1	198	197,670	1	1,330	199
	2	199	195,654	0	3,346	199
	3	195	193,180	4	5,820	199
	4	186	189,242	13	9,758	199
	5	186	183,681	13	15,319	199
	6	170	174,317	29	24,683	199
	7	163	162,542	36	36,458	199
	8	145	144,404	54	54,596	199
	9	110	117,986	89	81,014	199
	10	68	61,324	130	136,676	198

Classification Table^a

Observed		Predicted		
		NIVEL DE POBREZA DEL HOGAR		Percentage Correct
		NO	SI	
Step 1	NIVEL DE POBREZA DEL HOGAR NO	1303	317	80,4
	SI	98	271	73,4
	Overall Percentage			79,1

a. The cut value is ,230

ANEXO 4B Modelo Regresión Logística – Zona Rural

Block 1: Method = Enter

Iteration History^{a,b,c,d}

Iteration	-2 Log likelihood	Coefficients								Dependent Variable: ^1
		Constant	numper	esc	ocupado	inactivo	edad	Refrigerador	alcanta	
Step 1	754,225	1,558	,240	-,034	-1,735	-1,282	-,023	-,424	-,275	-,515
1 2	684,429	2,577	,425	-,068	-2,265	-1,452	-,044	-,686	-,578	-,762
3	674,444	3,194	,524	-,091	-2,545	-1,549	-,057	-,789	-,787	-,858
4	674,087	3,343	,546	-,097	-2,613	-1,577	-,060	-,807	-,839	-,875
5	674,086	3,350	,547	-,097	-2,617	-1,579	-,060	-,808	-,842	-,875
6	674,086	3,350	,547	-,097	-2,617	-1,579	-,060	-,808	-,842	-,875

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 887,796

d. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	213,710	8	,000
Block	213,710	8	,000
Model	213,710	8	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	674,086 ^a	,193	,327

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	4,401	8	,819

Contingency Table for Hosmer and Lemeshow Test

	NIVEL DE POBREZA = NO		NIVEL DE POBREZA = SI		Total
	Observed	Expected	Observed	Expected	
Step 1	98	99,064	2	,936	100
2	99	97,663	1	2,337	100
3	96	96,221	4	3,779	100
4	96	94,548	4	5,452	100
5	93	92,249	7	7,751	100
6	89	88,264	11	11,736	100
7	81	83,844	19	16,156	100
8	74	77,759	26	22,241	100
9	66	65,487	34	34,513	100
10	41	37,902	55	58,098	96

Classification Table^a

Observed	Predicted		
	NIVEL DE POBREZA		Percentage Correct
	NO	SI	
Step 1 NIVEL DE POBREZA NO	677	156	81,3
SI	58	105	64,4
Overall Percentage			78,5

a. The cut value is ,210

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 number	,547	,065	71,739	1	,000	1,727
esc	-,097	,036	7,329	1	,007	,907
ocupado	-2,617	,486	29,036	1	,000	,073
inactivo	-1,579	,515	9,393	1	,002	,206
edad	-,060	,009	43,354	1	,000	,942
Refrigerador	-,808	,213	14,363	1	,000	,446
alcanta	-,842	,244	11,858	1	,001	,431
vivienda	-,875	,413	4,484	1	,034	,417
Constant	3,350	,803	17,412	1	,000	28,509

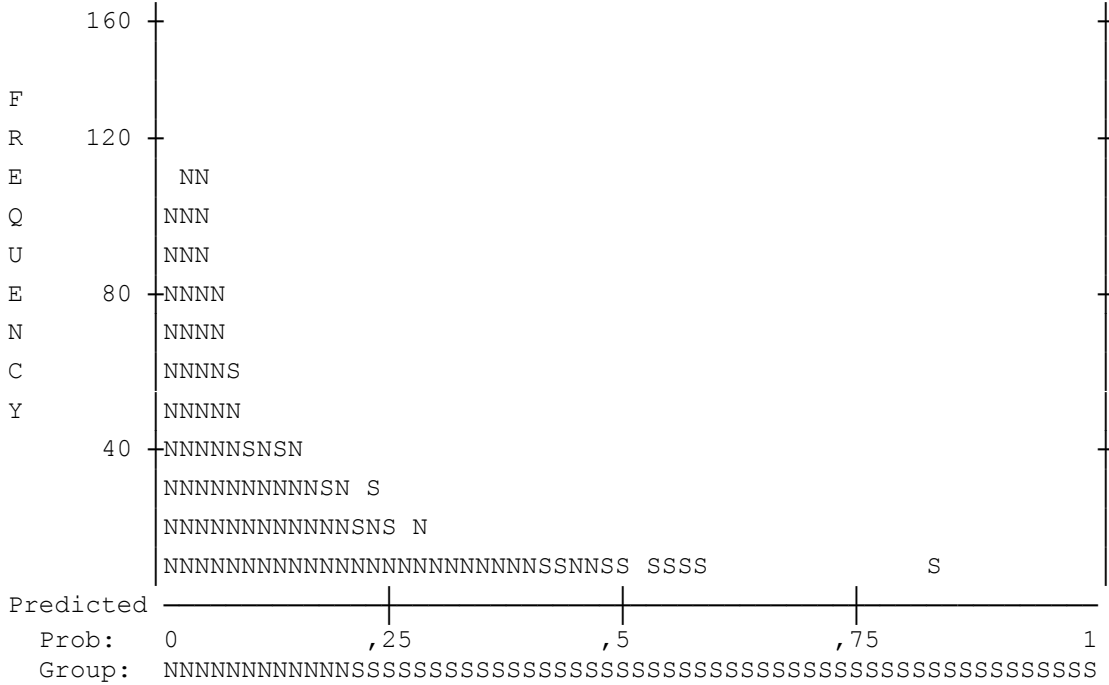
a. Variable(s) entered on step 1: number, esc, ocupado, inactivo, edad, Refrigerador, alcanta, vivienda.

Correlation Matrix

		Constant	numper	esc	ocupado	inactivo	edad	Refrigerador	alcanta	vivienda
Step 1	Constant	1,000	-,121	-,504	-,608	-,516	-,576	,007	-,003	-,344
	numper	-,121	1,000	,024	-,079	,027	-,180	-,186	-,066	-,187
	esc	-,504	,024	1,000	,074	,083	,509	-,151	-,127	-,088
	ocupado	-,608	-,079	,074	1,000	,868	,091	,070	,091	-,023
	inactivo	-,516	,027	,083	,868	1,000	-,130	,075	,072	,011
	edad	-,576	-,180	,509	,091	-,130	1,000	-,092	-,002	-,058
	Refrigerador	,007	-,186	-,151	,070	,075	-,092	1,000	-,224	-,030
	alcanta	-,003	-,066	-,127	,091	,072	-,002	-,224	1,000	-,072
	vivienda	-,344	-,187	-,088	-,023	,011	-,058	-,030	-,072	1,000

Step number: 1

Observed Groups and Predicted Probabilities



Predicted Probability is of Membership for SI
 The Cut Value is ,21
 Symbols: N - NO
 S - SI
 Each Symbol Represents 10 Cases.