UNIVERSIDAD NACIONAL AUTÓNOMA

DE MÉXICO

FACULTAD DE CIENCIAS

METODOLOGÍA PARA EL ANÁLISIS DE LA PERCEPCIÓN DE BENEFICIARIOS Y NO BENEFICIARIOS SOBRE EL IMPACTO DEL PROGRAMA OPORTUNIDADES

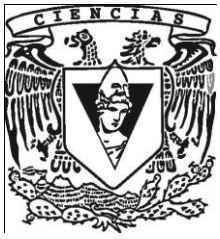
T E S I S

QUE PARA OBTENER EL TÍTULO DE:

A C T U A R I O

P R E S E N T A :

JAIME LÓPEZ MUÑOZ



DIRECTOR DE TESIS:

M. EN D. YVON ANGULO REYES

2011

Hoja de datos del jurado

1. Datos del alumno
   López
   Muñoz
   Jaime
   2381075346
   Universidad Nacional Autónoma de México
   Facultad de Ciencias
   Actuaría
   407058640

2. Datos del tutor
   M. en D.
   Yvon
   Angulo
   Reyes

3. Datos del sinodal 1
   Dra.
   María Edith
   Pacheco
   Gómez Muñoz

4. Datos del sinodal 2
   M. en D.
   Alejandro
   Mina
   Valdés

5. Datos del sinodal 3
   Dr.
   René Alejandro
   Jiménez
   Ornelas

6. Datos del sinodal 4
   M. en D.
   María Teresa
   Velázquez
   Uribe

7. Datos del trabajo escrito
   Metodología para el análisis de la percepción de beneficiarios y no beneficiarios sobre el impacto del Programa Oportunidades.
   153 p.
   2011.

VNIVERSIDAD NACIONAL AVTÓNOMA DE MEXICO

FACULTAD DE CIENCIAS
Secretaría General
División de Estudios Profesionales

Votos Aprobatorios

DR. ISIDRO ÁVILA MARTÍNEZ
Director General
Dirección General de Administración Escolar
P r e s e n t e

Por este medio hacemos de su conocimiento que hemos revisado el trabajo escrito titulado:

Metodología para el análisis de la percepción de beneficiarios y no beneficiarios sobre el impacto del Programa Oportunidades

realizado por **López Muñoz Jaime** con número de cuenta **4-0705864-0** quien ha decidido titularse mediante la opción de **tesis** en la licenciatura en **Actuaría**. Dicho trabajo cuenta con nuestro voto aprobatorio.

Propietario	Dra. María Edith Pacheco Gómez Muñoz	
Propietario	M. en D. Alejandro Mina Valdés	
Propietario Tutora	M. en D. Yvon Angulo Reyes	
Suplente	Dr. René Alejandro Jiménez Ornelas	
Suplente	M. en D. María Teresa Velázquez Uribe	

Atentamente,
"POR MI RAZA HABLARÁ EL ESPÍRITU "
Ciudad Universitaria, D. F., a 15 de junio de 2011
EL JEFE DE LA DIVISIÓN DE ESTUDIOS PROFESIONALES

ACT. MAURICIO AGUILAR GONZÁLEZ

Señor sinodal: antes de firmar este documento, solicite al estudiante que le muestre la versión digital de su trabajo y verifique que la misma incluya todas las observaciones y correcciones que usted hizo sobre el mismo.
MAG/CZS/cigs

A mis padres, Jaime y Cruz Amelia,
porque sólo ellos comprenden el esfuerzo
de este proyecto de vida,
que día con día construyeron con su amor, trabajo y ejemplo.
Esta tesis es suya.

Agradecimientos

Di a aquellos que amas que realmente los amas y en todas las oportunidades
y recuerda siempre que la vida no se mide por la cantidad que respiraste,
sino los momentos que tu corazón palpitó fuerte: de tanto reír, de sorpresa, de éxtasis, de felicidad,
sobre todo de querer sin medida.
Pablo Picasso
(1881-1973)

Agradezco a mis hermanos, Carlos Alberto y Francisco Javier, el apoyo que brindan en todo momento; porque he encontrado en ellos a mis dos mejores amigos.

A mis abuelos Julio López, Olga Sánchez, Juan Muñoz y Emelia Sánchez, son también parte substancial del proyecto.

A mi querida Universidad Nacional Autónoma de México, UNAM, por haberme dado la oportunidad de estudiar y formarme dentro de sus aulas y pasillos en la Facultad de Ciencias, y más tarde en el Instituto de Investigaciones Sociales. Es también ella quien me dio mi primer empleo.

A la Mtra. Yvon Angulo Reyes por ser parte importante en la realización de este trabajo; fueron muchas las horas de dedicación, dirección y consejo. Sin su invaluable ayuda distinta habría sido el fin de ésta tesis.

A los sinodales Dra. María Edith Pacheco, Mtro. Alejandro Mina, Dr. René Jiménez y Mtra. María Teresa Velázquez, por sus atinadas y sabias correcciones durante la revisión de ésta.

A todos los profesores que hicieron de mí un mejor ser humano.

A los amigos que encontré durante mi estancia en la Universidad Nacional: Fali Rivera, Luis López, Carlos Juárez, Ronaldo Pérez, Miriam Vigueras, Thalía Ibarra, Irene Peñuelas, Lilia Rivera, David Ríos, Jonathan Ruiz, Daniel Guillen y Elizabeth Rodríguez.

México, D.F., Septiembre de 2011.

Índice

Índice de Tablas

Índice de Mapas

Índice de Gráficas

Introducción

Generalmente, cuando se hace referencia a la evaluación de programas sociales, se vienen a la mente una serie de indicadores como por ejemplo, cobertura, gasto asignado, etc., aspectos fundamentales para tener idea del desempeño del programa, y relativamente sencillos de medir, pero que únicamente hacen referencia al desempeño en términos de eficiencia (medios) y eficacia (fines). Por otro lado, cuando lo que se quiere analizar es el impacto de algún programa en aspectos sociales, como por ejemplo, cohesión social, capital social o "estructura" comunitaria, la medición y análisis se torna compleja, y por lo tanto, ya no resulta tan fácil ni directa; sin embargo, abordar el impacto que los programas sociales puedan tener en estos aspectos, se vuelve cada vez más importante.

Este trabajo tiene como propósito analizar las herramientas estadísticas y el desarrollo metodológico empleadas para el logro del proyecto general.

El objetivo del proyecto general de investigación del cual se desprende esta tesis, es analizar el impacto que el programa Oportunidades pudiera tener en la creación de un espacio propicio para la generación de capital social, a través de la percepción que las personas tienen del impacto del programa en su comunidad. Con tal fin se propuso la realización de un modelo multivariado en el que se pudiera analizar a beneficiarios y no beneficiarios de acuerdo a factores socioeconómicos, demográficos y contextuales, de manea que se pudiera identificar si la presencia del programa incidía en la percepción que se tiene sobre el impacto del programa en la comunidad. El camino para llegar a un modelo multivariado final en el que se analicen estas relaciones es largo, y requiere de una serie de análisis previos, como son, análisis y evaluación de la información que se va utilizar, construcción de indicadores, análisis univariado y bivariado, etc. Trabajo que en muchas ocasiones se ve opacado por los resultados de las investigaciones, pero que tiene una gran relevancia para el logro de sus objetivos. Para lo cual se realiza una recorrido para hacer visibles estos procedimientos.

La tesis se divide en cuatro apartados. En el primero se describen las principales características del Programa Oportunidades; en el segundo apartado se realiza una caracterización de la población de la *Encuesta lo que dicen los pobres*, que es la fuente de información con la que se propuso el desarrollo del proyecto general; en el tercer apartado se presentan y analizan las ventajas y desventajas de tres técnicas generalmente empleadas para la construcción de índices: Rangos sumados, rangos sumados ponderados, y componentes principales. Finalmente, en el último apartado, se presentan los principales supuestos a cumplir para un modelo de regresión logística multivariado; la secuencia de esta sección es considerar un conjunto de índices e indicadores y modelarlos por medio de una regresión logística univariada, de tal manera que sirva de ejemplo ilustrativo el ajuste de éstas para futuras referencias, luego, con las covariables estadísticamente significativas se ajusta un modelo de regresión logística múltiple, también a manera de ejemplo.

1. El Programa de Desarrollo Humano Oportunidades

1.1 Del origen de Oportunidades

El Programa de Desarrollo Humano Oportunidades, mejor conocido como Oportunidades, tiene su origen en el Programa de Educación, Salud, y Alimentación (PROGRESA) establecido durante el gobierno de Ernesto Zedillo[1], en agosto de 1997.

Durante el Gobierno de Vicente Fox[2] se hace una modificación a la Ley, de tal manera que se crea por decreto del mismo Presidente la Coordinación Nacional del Programa de Desarrollo Humano Oportunidades como un órgano descentralizado de la Secretaría de Desarrollo Social, con autonomía técnica, (DOF, 2002). Tal Coordinación tiene el objeto de formular, coordinar, y evaluar la ejecución del programa especial, anteriormente denominado, PROGRESA. Se buscan fijar prioridades de corto y largo alcance que estén orientadas hacia la cobertura total en educación, erradicar el analfabetismo, garantizar la cobertura universal de los servicios de salud, equilibrar el desarrollo económico y social con respeto y cuidado del medio ambiente, mejorar el nivel de vida y superar la pobreza extrema. Se identifican cinco componentes para las bases de un auténtico desarrollo humano: (a) oportunidad, (b) capacidad, (c) seguridad, (d) patrimonio y (f) equidad. Éstas , combinados con un conjunto de políticas públicas que involucren la participación de los tres órdenes del gobierno, la comunidad, las familias, las organizaciones sociales, el sector privado y la comunidad académica, generarían igualdad de oportunidades para con los grupos más pobres y vulnerables de este país. Es decir, las cinco componentes del desarrollo, y todas estas acciones establecerían un sistema de equidad en la población más pobre. De esta manera, se lograría garantizar el derecho a la educación, mediante la igualdad de oportunidades para el acceso, permanencia y logro de la educación básica, así como ampliar la cobertura de educación media superior y atacar el rezago educativo presente. Esto se vería reflejado en un mejoramiento de la vida personal, familiar y social. También promovería su realización productiva. Por otro lado, se tiene previsto que se mejorarían las condiciones de vida de los mexicanos, se abatirían las desigualdades, y se reducirían los rezagos en materia de salud, (Oportunidades, 2010).

La Coordinación del Programa trabaja mediante:

- Mejoramiento de las condiciones de educación, salud y alimentación.
- La concatenación integral de las acciones de educación, salud y alimentación, con los programas de desarrollo regional y comunitario, fomento económico y empleo temporal en zonas marginadas buscando la generación de oportunidades en éstas zonas y comunidades.
- La participación activa y la corresponsabilidad de los padres y de todos los miembros de la familia y de las comunidades

[1] Ernesto Zedillo Ponce de León, Presidente de México. Del 1 de diciembre de 1994 al 30 de noviembre de 2000.
[2] Vicente Fox Quesada, Presidente de México. Del 1 de diciembre de 2000 al 30 de noviembre de 2006.

- La interrelación con otros programas del sector social y de los gobiernos estatales y municipales.

Cabe mencionar, que la Coordinación del Programa no debe transferir recursos de un programa a otro, sino únicamente fomenta y dirige la vinculación de estrategias y acciones entre ellos.

1.2 Sobre el Programa de Desarrollo Humano Oportunidades

Oportunidades es un programa federal para el desarrollo humano enfocado a la población en pobreza extrema. Para lograrlo brinda apoyos monetarios y en especie en los rubros de educación, salud, nutrición e ingreso. En el programa participan activamente la Secretaría de Desarrollo Social (SEDESOL), La Secretaría de Educación Pública (SEP), la Secretaría de Salud (SSA), el Instituto Mexicano del Seguro Social (IMSS), los gobiernos estatales y municipales. Estos dos últimos en todo el territorio nacional. Su misión es coordinar acciones entre ellas mismas de tal manera que las personas beneficiadas superen su pobreza buscando el desarrollo de sus capacidades básicas, económicas y sociales, (DOF, 2010).

El presupuesto del programa se asigna en tres Secretarías: SEDESOL, SEP y SSA. Su operación está regida por las reglas establecidas por los titulares de estas secretarías, la Secretaría de Hacienda y Crédito Púbico (SHCP) y el IMSS. También se cuenta con un Comité Técnico donde participan Subsecretarías de esas dependencias, así como el Director General del IMSS y un Delegado de la Secretaría de la Función Pública (SFP). El comité supervisa el seguimiento del programa.

Oportunidades es de carácter federal, es decir, tiene presencia en los 31 estados del país y en el Distrito Federal. En cada una de estas entidades existen los Comités Técnicos Estatales donde se involucran los actores federales y estatales relacionados con el programa.

El sistema de selección de los beneficiarios está basado en las características socioeconómicas del hogar, focalizando los recursos en familias que realmente lo necesitan, superando los subsidios y los apoyos discrecionales y definidos con criterios políticos.

Cabe destacar, que el Programa Oportunidades deja a un lado los sistemas paternalistas y el asistencialismo, empleados por el Gobierno Federal desde la segunda mitad del siglo pasado. La ausencia de estos modos de políticas públicas se debe a que en el programa, la corresponsabilidad[3]

[3] La corresponsabilidad es definida, por la Real Academia Española, como la responsabilidad compartida.
Para el Programa Oportunidades, se entiende como, *las responsabilidades que al momento de su incorporación la familia beneficiaria se compromete a cumplir para recibir los apoyos del Programa*: (a) Inscribir a los menores de 18 años que no hayan concluido la educación básica en las escuelas de educación primaria o secundaria; (b) inscribir a los jóvenes de hasta 20 años que hayan concluido la educación básica a la media superior; (c) registrarse en la unidad de salud que les corresponda y cumplir con sus citas periódicas; (d) asistir a las pláticas mensuales de educación para la salud; (e) destinar los apoyos monetarios al mejoramiento del bienestar familiar, en especial a la alimentación de los hijos y para su aprovechamiento escolar.

es un factor de suma importancia. Evidenciando que las familias son parte activa de su propio desarrollo.

1.3 Oportunidades en números de 2008

Los siguientes datos tienen como fecha de corte el 31 de diciembre de 2008, (Oportunidades, 2010).

- El Programa opera a nivel nacional, en más de 92 mil localidades, en los municipios de mayor marginación, en áreas rurales, urbanas y grandes metrópolis.
- Se beneficia a 5 millones de familias, aproximadamente a 25 millones de mexicanos. Lo que equivale a una cuarta parte de la población nacional.
- Contempla ocho modalidades:
  1. Recursos para mujeres, madres de familia, para el ingreso familiar y una mejor alimentación.
  2. Becas para niños y jóvenes, a partir de tercero de primaria y hasta el último grado de educación media superior.
  3. Fondo de ahorro para jóvenes que concluyen su Educación Media Superior.
  4. Apoyo para útiles escolares.
  5. Paquete de servicios médicos y sesiones educativas para la salud.
  6. Complementos alimenticios para niños y niñas entre 6 y 23 meses y con desnutrición entre los 2 y 5 años. También para las mujeres embarazadas o en periodo de lactancia.
  7. Apoyo de $540 bimestrales adicionales por cada adulto mayor, en localidades mayores de 10 mil habitantes.
  8. Apoyo adicional de $100 bimestrales para el consumo energético de cada hogar.
- Durante 2008, el Programa Oportunidades ejerció un presupuesto total de 38, 071 millones de pesos, de acuerdo con el Presupuesto de Egresos de la Federación 2008 y aprobado por el Congreso de la Unión. Dicho presupuesto se gastó de la siguiente manera: SEDESOL 17,431 millones de pesos, SEP 17,350 millones de pesos, SSA 3,289 millones de peso, (Presidencia, 2009).

1.4 Requisitos de participación y permanencia

El Programa Oportunidades aplica un novedoso sistema de identificación de beneficiarios, mediante una encuesta socioeconómica. Las familias que se incorporan al Programa son beneficiarias por sus condiciones de pobreza extrema y su permanencia la determina el cumplimiento de sus corresponsabilidades. Algunas de las causas por las cuales pueden ser suspendidos los apoyos monetarios que otorga el Programa a las familias beneficiarias, son:

1. Cuando en dos ocasiones seguidas, la titular no asiste a recibir los apoyos monetarios.
2. Cuando en cuatro meses seguidos o seis meses no continuos en el curso de los últimos 12, la familia no cumpla con su corresponsabilidad de asistencia a los servicios de salud.
3. Cuando se compruebe que la familia ya no cumple con los criterios de elegibilidad del Programa.
4. Cuando la familia no haya aceptado que se le aplique la encuesta de recertificación o no haya sido recertificada por alguna causa no imputable a ella. En el primer caso, la suspensión es definitiva; en el segundo, puede solicitar al personal de Oportunidades la regularización de su situación.

De acuerdo con las Reglas de Operación vigentes, la concepción, medición e identificación de las familias que viven en condiciones de pobreza extrema, se realiza tomando en cuenta los criterios que el Consejo Nacional de Evaluación de la Política de Desarrollo Social[4] indica por medio de la Ley de Desarrollo Social, (DOF, 2010).

1.5 Del proceso de selección e identificación de beneficiarios

La presente sección es presentada conforme las Reglas de Operación vigentes, publicadas en el Diario Oficial de la Federación el último día de diciembre de 2010; el proceso por el cual se identifica a las familias beneficiarias consta de dos etapas, se describen abajo, (DOF, 2004):

- **Etapa 1. "Selección de Localidades"**

  La selección de localidades nuevas o localidades ya atendidas por el Programa, se realiza con base en el **Índice de Rezago Social**[5] establecido por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), el **Índice de Marginación**[6] establecido por el Consejo Nacional de Población (CONAPO), así como en la información estadística disponible a nivel de localidades, Áreas Geostadísticas Básicas (AGEB), colonias y/o manzanas, generada por el Instituto Nacional de Estadística y Geografía (INEGI), dando prioridad, en la selección y atención, a localidades donde es mayor la concentración de hogares en condiciones de pobreza extrema.

---

[4] Consejo Nacional de Evaluación de la Política de Desarrollo Social, CONEVAL, es un organismo público descentralizado de la Administración Pública Federal, con autonomía y capacidad técnica para generar información objetiva sobre la situación de la política social y la medición de la pobreza en México, que permita mejorar la toma de decisiones en la materia; Tiene como funciones principales:
1. Normar y coordinar la evaluación de la Política Nacional de Desarrollo Social y las políticas, programas y acciones que ejecuten las dependencias públicas;
2. Establecer los lineamientos y criterios para la definición, identificación y medición de la pobreza, garantizando la transparencia, objetividad y rigor técnico en dicha actividad.
Sitio web del CONEVAL. www.coneval.gob.mx, Febrero de 2011.

[5] Índice de Rezago Social, calculado por el CONEVAL. es un estimador de carencias calculado para tres niveles de agregación geográfica: estatal, municipal y local, el cual incorpora indicadores de educación, de acceso a servicios de salud de servicios básicos, de calidad y espacios en la vivienda, y activos en el hogar.

[6] El índice de Marginación, calculado por el CONAPO, es una medida de déficit y de intensidad de las privaciones y carencias de la población en dimensiones relativas a las necesidades básicas establecidas como derechos constitucionales. Más adelante, en el Capítulo IV, se detalla el índice.

Una vez seleccionado el universo de atención, conformado por localidades, AGEBs, colonias y/o manzanas, se procede a validar las condiciones de accesibilidad y capacidad de atención de los servicios de salud y educación para dicho universo, que permitan operar en forma integral los componentes del Programa. La información socioeconómica de los hogares se recolecta mediante la aplicación de **cédulas individuales** para determinar su condición de pobreza extrema.

- **Etapa 2. "Identificación de Familias"**

Para la identificación de las familias susceptibles de ser incorporadas al Programa se utiliza una metodología de puntajes basada en un criterio objetivo, homogéneo, transparente y único a nivel nacional que considera tanto la condición de residencia rural-urbana y regional de las familias, como sus condiciones socioeconómicas y demográficas. Su aplicación evita la discrecionalidad en la identificación de las familias beneficiarias.

La metodología para la identificación de familias beneficiarias observa los lineamientos y criterios emitidos por el CONEVAL para la definición, identificación y medición de la pobreza, de conformidad con lo dispuesto en el artículo 36 de la Ley General de Desarrollo Social y demás disposiciones aplicables, (DOF, 2004).

Por lo tanto, en el proceso de selección de familias beneficiarias del Programa de Desarrollo Humano Oportunidades interactúan dos tipos de filtros. El primero de ellos, el de ***selección de localidades***, permite seleccionar las diferentes localidades con base en sus características tales como pobreza, marginación, rezago social y tamaño de localidad; acto seguido se escoge una localidad, se encuesta a las familias de tal localidad mediante la cédula individual, permitiendo conocer las condiciones en las que viven. Para seleccionar aquellas familias que son candidatas para recibir el beneficio del programa, se construye un índice de selección que permite discriminar técnicamente, mediante un umbral, entre las familias que recibirán Oportunidades y las que no.

Observándolo desde una óptica técnica, el procedimiento de selección de las familias permite elegir a aquellas familias que se encuentran debajo del umbral, asegurando con esto que las familias beneficiarias sean las más pobres.

1.6 Algunos problemas con el procedimiento de selección de familias beneficiarias

Desde el punto de vista social, esta estrategia de selección de las familias beneficiadas presenta algunas dificultades. De todo el espectro de posibilidades que se podría dar entre la situación económica de las familias en México, hay algunas que presentan situaciones particulares,

dado el valor de su indicador de pobreza, el cual se encuentra muy cercano al umbral con el cual se toma la decisión de incluir o no a una familia al programa. Esto es, familias cuya pobreza apenas está por debajo o encima del umbral, (Evelyne Rodríguez, 2005).

A partir de lo anterior, consideremos dos situaciones particulares. Por un lado se tiene a las familias con pobreza apenas por debajo del umbral. Éstas sí reciben Oportunidades de manera periódica, aunque podrían salir por diversas causas como no cumplir con los requisitos de permanencia, corresponsabilidad, educación, salud, edad, etc. Si ellas se mantuvieran durante el todo el tiempo activos en el programa, al terminarse podrían presentar dos situaciones: (a) que la familia haya roto la pobreza que presentaba, y con ello la familia efectivamente pueda valerse por sí misma (lo cual es la misión de Oportunidades); (b) que la familia haya mejorado su estado de pobreza mientras recibía el beneficio, mas al termino de éste, la familia no puede sustentarse por sí misma y recae en su condición de pobreza inicial, más aún la recaída podría ser con mayor intensidad (en este caso, Oportunidades cumple su misión parcialmente).

Para cuando la familia recae en pobreza debajo del indicador de selección, se tiene que, al haber salido del programa, puesto que ya se podía clasificar como una familia no beneficiaria ya no recibe beneficios de Oportunidades, sin embargo, al recaer necesitaría recibir nuevamente beneficios. El mayor problema aparece cuando la familia necesite el beneficio y no pueda inscribirse temporalmente al programa puesto que hay periodos de tiempo bien definidos por la Coordinación Nacional. Por lo tanto, la familia tendrá que supervivir con una situación cada vez más marginada mientras vuelve a recibir apoyo alguno.

Ahora considere el caso contrario, cuando la familia superó el umbral apenas al ras, esto es, tanto la familia del caso anterior como de este son, en términos prácticos, igual de pobres. Luego, por su condición superior al umbral, la familia no recibe beneficio de Oportunidades. Esto quiere decir que ella se tendrá que valerse por sí misma, dejando toda la responsabilidad de superar su pobreza a ella. Desde un punto de vista humanitario, no es justo debido a que ambas familias padecen pobreza con la misma intensidad.

Estos casos ejemplifican dos situaciones extremas en las que se puede incurrir al ser estrictamente técnicos en la selección de familiar beneficiarias.

1.7 Logros de Oportunidades

Los siguientes números del Programa Oportunidades destacan los impactos en las familias beneficiarias, los datos al 31 de diciembre de 2008, (Oportunidades, 2010):

- **Inicio de vida**

i. Durante 2008, más de medio millón de mujeres embarazadas y en lactancia estuvieron en control médico y nutricional.
ii. A mediano plazo, la intervención del Programa reduce entre dos y seis por ciento la probabilidad de muerte infantil[7], y de 11 por ciento en muerte materna.

- **Niñez**

i. Uno de cada cuatro niños mexicanos mejoran su nutrición y salud y crecen su peso y talla.
ii. Niños y niñas mejoran sus habilidades motoras en 15 y 10 por ciento, respectivamente e incrementan de 28 a 44 por ciento la probabilidad de aprobar el primer año de primaria y de avanzar a tiempo en la escuela.

- **Becarios**

i. Cinco millones 200 mil niños, niñas y jóvenes acumulan escolaridad y tienen una expectativa de vida diferente a la de sus padres.
ii. Más de 700 mil becarios son jóvenes que estudian bachillerato.
iii. Prácticamente todos los becarios alcanzan un mayor grado escolar que sus padres.
iv. Los jóvenes están sustituyendo el trabajo por la escuela entre 24 y 48%

- **Mujeres**

i. Cinco millones de mujeres mejoran el consumo, la nutrición y salud de sus familias, tienen mejor atención médica y detectan tempranamente enfermedades.
ii. El 98% de los apoyos de Oportunidades se entregan a mujeres.
iii. Oportunidades invierte mil millones de pesos más en becas de mujeres; hay más mujeres que hombres con beca.
iv. Las niñas y jóvenes de familias en pobreza tienen las mismas oportunidades y más incentivos para estudiar que los varones.

Cabe destacar, que Oportunidades es pionero en el mundo en cuanto al diseño y operación de los programas de transferencias monetarias condicionadas. Actualmente, una treintena de países, en su mayor parte América Latina, cuentan con instrumentos para entregar apoyos a la población en pobreza extrema inspirados en el modelo mexicano, (Beláustegui, 2009).

[7] Según la UNICEF, durante 2007 en México, la tasa de mortalidad infantil fue de 21/1000, por lo que, por cada 1000 niños nacidos únicamente 21 fallecen. La misma tasa para los países más desarrollados del mundo fue de 8/1000 y para los países menos desarrollados es de 59/1000. En la región de África Subsahariana es de 100/1000; (United Nations, 2007).
Por lo tanto, para las familias beneficiarias de Oportunidades, la tasa de mortalidad infantil está en el intervalo [15/1000, 19/1000].

1.8 Otros números de la pobreza

Por otro lado, según el CONEVAL en sus estimaciones de pobreza por ingresos a nivel nacional y en los ámbitos rural y urbano para el año 2006, se tiene, (CONEVAL, 2008):

- En 2008**, 50.6 millones de mexicanos eran pobres de patrimonio**, es decir, no contaban con un ingreso suficiente para satisfacer sus necesidades de salud, de educación, de alimentación, de vivienda, de vestido y de transporte público, aun si dedicaran la totalidad de sus recursos económicos a este propósito.
- En 2008**, 19.5 millones eran pobres alimentarios**, es decir, quienes tienen ingresos insuficientes para adquirir una canasta básica de alimentos, incluso si los destinaran exclusivamente para ese fin.
- Entre 2006 y 2008 **aumentó la cantidad de personas que viven en pobreza de patrimonio**, pasando de 42.6% a 47.4%, esto es un incremento de 4.8% (5.9 millones de personas) en tan sólo dos años.
- Entre 2006 y 2008, cuanto a **la pobreza alimentaria aumentó de 13.8% a 18.2%,** esto es un incremento del 4.4% (5.1 millones de personas) en un bienio
- Entre 2000 y 2008, la incidencia de **la pobreza de patrimonio y de la pobreza alimentaria se redujo 6.2 y 5.9 puntos porcentuales**, respectivamente, lo cual se traduce en una reducción de 2.1 y 4.2 millones de personas pobres de patrimonio y alimentarios, respectivamente.

Es importante recordar, que el Programa Oportunidades tiene sus orígenes en el Programa PROGRESA que data del año 1997; consideremos el último punto de los números anteriores, se observa que entre los años 2000 y 2008 se redujo el número de personas en pobreza de patrimonio y alimentaria, que es precisamente la población a la que está enfocado el Programa Oportunidades.

Bajo las estimaciones de la Coordinación Nacional del Programa Oportunidades y las del CONEVAL, se observa que los logros del programa no son suficientes para erradicar completamente la pobreza en el país. Es cierto que Oportunidades ha hecho mucho al respecto, pero también es bien sabido que hacer falta hacer aún más por los pobres de tal manera que ellos perciban de manera diferente la justicia social de este país. Uno no decide ser pobre, simplemente es una condición en la que uno nace, y, queda en manos propias y de los desarrolladores de políticas públicas el romper la brecha intergeneracional que establece la pobreza.

2. Sobre la encuesta "Lo que dicen los pobres"

Para combatir la pobreza es necesario entender que los pobres
son la parte más importante en la solución de sus problemas.
Son protagonistas, sujetos activos de su desarrollo, no meros receptores pasivos de dádivas.
Los pobres saben mejor que nadie qué necesitan para ser incluidos en el desarrollo.
Josefina Vázquez Mota[8] (1961-)

2.1 Descripción general de sus características y su alcance

En el marco de esta encuesta, el universo de estudio está formado por los hogares del país que se encuentran en condición de pobreza patrimonial[9], tanto en áreas urbanas como en rurales[10].

El objetivo de esta encuesta fue obtener información sobre la población en condición de pobreza que permitiera conocer sus características generales y las percepciones que esta población tenía sobre diversos temas sociales. Como objetivos específicos de esta encuesta se planteó, (Székely, 2005):

I. Conocer las características generales de la población en condición de pobreza.
II. Identificar las opiniones que la población tiene sobre temas como bienestar y justicia social.
III. Conocer la percepción de la población en condición de pobreza sobre vulnerabilidad y discriminación.
IV. Conocer la opinión que tiene la población en condición de pobreza sobre las acciones institucionales y la valoración de los apoyos sociales.

La realización de la encuesta se aplicó en 49 municipios del país en 29 entidades federativas de la República Mexicana. Levantada en el verano de 2003 por el grupo Ipsos-Bimpsa y diseñada por la SEDESOL.

Contó con dos instrumentos de recolección de información: **Cuestionario del Hogar** y **Cuestionario Individual**, diseñados para ser aplicados por encuestadores.

El cuestionario del hogar tenía como función recoger información básica de cada uno de los residentes del hogar y sobre las distintas características de la vivienda, tales como, (Székely, 2005):

[8] Ex-Secretaria de la SEDESOL durante diciembre de 2000 a enero de 2006. En turno mientras la Secretaría organizaba la encuesta *Lo que dicen los pobres*.
[9] El CONEVAL define la pobreza patrimonial, como la insuficiencia de ingreso para invertir en transporte, vivienda, vestido y calzado, aunque sí cuentan con la manera de satisfacer sus necesidades de alimentación, educación y salud. También define otros dos tipos de pobreza: La pobreza de capacidades y la pobreza alimentaria.
[10] La Áreas rurales son definidas, mediante la metodología de medición de la pobreza por ingreso, como localidades de menos de 14,999 habitantes. Por su lado, las Zonas urbanas son localidades de más de 15,000 habitantes.

i. Edad
ii. Sexo
iii. Escolaridad
iv. Lengua indígena
v. Parentesco con el jefe del hogar
vi. Estado civil
vii. Condición de actividad
viii. Acceso a los servicios de salud
ix. Materiales de construcción de la vivienda
x. Servicios de la vivienda
xi. Combustible para cocinar
xii. Existencia de diversos activos del hogar

El cuestionario individual, se aplicó a una persona de 18 o más años en el hogar, seleccionado aleatoriamente; contó con siete secciones que permiten conocer los siguientes tópicos:

i. Características generales del entrevistado.
ii. Trabajo e ingreso.
iii. Bienestar y justicia social.
iv. Pobreza, vulnerabilidad y riesgo.
v. Diferencias y discriminación.
vi. Análisis institucional.
vii. Valoración de apoyos.

La mayoría de las preguntas son cerradas, es decir, se tienen códigos predefinidos en el propio cuestionario y en algunos casos cuentan con la opción de otro en donde se pide especificar la respuesta para su posterior recodificación.

En general sólo se admite una respuesta y las opciones de respuestas no se leen a menos que se dé una instrucción contraria.

Los instrumentos de campo fueron probados mediante un ensayo piloto que se realizó en una muestra de 200 hogares rurales y urbanos.

2.2 Antecedentes de la encuesta

Según (Székely, 2005), el primer estudio científico sobre la pobreza titulado *"Poverty: A Study of Town life"*. Se realizó en Inglaterra a cargo de Seebohm Rowntree. Duró 2 años y concluyó en 1899. Utilizó el método intensivo[11], de tal manera que visitó a más de 11 mil familias en casi 400

[11] El *método intensivo* consiste en estudiar a detalle las características de la población en una comunidad (localidad) para obtener conclusiones que se pueden extrapolar a otras poblaciones.

calles de la Ciudad de York, Inglaterra. La cobertura fue casi total para las familias de dicha ciudad. Es importante mencionar que Rowntree se dedicó a escuchar lo que decía cada una de las familias. Es el primer acercamiento a la pobreza, dejando a un lado la imparcialidad de los estudios externos, que hasta ese momento eran comunes entre la comunidad académica. Las conclusiones eran reveladoras, se estimó que tres de cada diez residentes de la población de York era pobre. Es histórico e innovador este análisis debido a que fue la primera vez que se les preguntaba a los mismos pobres sobre sus condiciones y porqué lo eran.

A raíz de la investigación de Rowntree la comunidad académica, científica y los hacedores de políticas públicas empezaron un debate sobre la pobreza con el fin de diseñar estrategias más eficientes para combatir la pobreza.

Por otro lado, también existen estudios etnográficos de corte antropológico de la pobreza. Estos estudios tienen como premisa la investigación en campo[12], entrevistas y observaciones. También se permite la convivencia por parte del investigador con la población objetivo. En este caso, se tiene el problema de contar con muy pocos casos, los cuales no se pueden generalizar a las demás poblaciones. Un problema recurrente es la subjetividad con la que el investigador ve la realidad.

También existen estudios de medición de la pobreza a través de bases de datos. Con esto se puede realizar cálculos estadísticos descriptivos e inferenciales, permitiendo obtener conclusiones cuantitativas, las cuales sustentan sólidamente los argumentos cualitativos. Es decir, los estudios cuantitativos complementan los estudios cualitativos o viceversa.

Si uno quiere encontrar soluciones a un problema planteado, el mejor aportador de soluciones es el mismo sufriente de los problemas. Él mejor que nadie, sabe lo que necesita. Trasladando esta idea al marco de la pobreza, ésto quiere decir, que las mejores propuestas para combatir la pobreza vienen de los mismos pobres. Por ello, es importante tomar en cuenta sus opiniones, sus percepciones, sus vivencias y sus posibles soluciones.

En la mayoría de países, al hablar de pobreza, siempre se considera la opinión de agentes exógenos a ellos: académicos, investigadores y tomadores de decisiones de políticas públicas, es decir, gente que casi-seguramente no sufre la pobreza.

Hablando de otros países, en 1996, surge el estudio el conocimiento de los pobres a iniciativa del Movimiento del Cuarto Mundo en 1991. Se creó el Grupo Inter-Universitarios de Investigación y Pobreza[13] El objetivo era investigar si los pobres tienen que ser tomados en cuenta al estudiar a la pobreza misma. Concluyen cuatro puntos: (a) Los pobres son parte misma de su estudio. De otra manera, el investigador está sujeto a perder el punto de vista de los pobres; (b) El científico no debe de considerar sus prejuicios de su propia subcultura para captar la subcultura de

[12] El trabajo de campo de corte antropológico es: entrevistas a profundidad, etnografías, etc.
[13] Grupo Grupo Inter-Universitarios de Investigación y Pobreza (GIReP) conformado por investigadores multidisciplinarios provenientes de diversas universidades belgas, francesas y canadienses.

estudio; (c) El punto de vista de los pobres debe ser parte de la investigación, teniendo en mente la significación de sus actores sociales en su comportamiento; (d) La ética comunicacional establece que para garantizar el respeto, la investigación debe de tener el punto de vista del investigado.

Existe otro antecedente, el estudio realizado por Banco Mundial denominado ¿Alguien nos escucha? Voces de 47 países, (Narayan, 1999). Desarrollado en el marco del proyecto Consultas con la pobreza[14] que tuvo el objeto de incluir la voz de los pobres dentro del Reporte Mundial sobre el Desarrollo de la Pobreza 2000-2001. Contó con más de 60 mil voces de hombres y mujeres a lo largo de 47 países. Identifica las líneas de investigación e hipótesis para guiar un trabajo sistemático sobre la pobreza a partir de los mismos sufrientes, (Rahmato, 2000).

Las conclusiones del proyecto del Banco Mundial son:

i. Los pobres describen la pobreza como la carencia de alimentos y activos.
ii. Los pobres no quieren caridad sino oportunidades de empleo.
iii. La estrategia de cambio debe de contar con: partir de la realidad de los pobres, invertir en la organización de ellos, no quieren ser meros receptores sino ser partícipes de las políticas.

Hasta antes de la encuesta "Lo que dicen los pobres", en México, como en la mayoría de los países restantes del mundo, no se les había preguntado a los pobres sobre su condición, de manera sistemática. Rara vez se hace un alto para escuchar sus problemas y las soluciones que ellos proponen.

Con estas bases, el gobierno mexicano hizo un esfuerzo por escuchar lo que dicen los pobres. En julio de 2003, la Secretaria de Desarrollo Social (SEDESOL) realizó una encuesta para documentar lo que dicen los pobres en el México de hoy. Se llama "Lo que dicen los pobres" y tiene representatividad nacional. Incluye las voces de 2939 mexicanos en estado de pobreza patrimonial[15].

En *Lo que dicen los pobres*, existen en la base de datos 2939 casos, es decir, 2939 cuestionarios individuales se levantaron con información de 369 variables, (SEDESOL, 2011). Hay 353 variables provenientes directamente del cuestionario, 12 variables que controlan y dan orden a la base de datos y 1 variable que es el factor de expansión[16].

[14] Rahmato, Dessalegn; Kidanu, Aklilu. *"Consulations with the poor. A study to inform The World Development Report /2000/01 On Poverty and Development"*.

[15] El CONEVAL define la pobreza patrimonial, como la insuficiencia de ingreso para invertir en transporte, vivienda, vestido y calzado, aunque si cuentan con la manera de satisfacer sus necesidades de alimentación, educación y salud. También define otros dos tipos de pobreza: La pobreza de capacidades y la pobreza alimentaria, (CONEVAL, 2008).

[16] En este punto, es importante mencionar que la sección II de *Trabajo e ingreso* está incompleta en la base de datos publicada por la SESEDOL en su sitio web; de manera particular no es pública la información que refiere a los ingresos declarados por el informante, el rango de salario, la frecuencia con la que recibe su salario y dinero extra que recibe por otros medios.
Las preguntas de ingreso que fueron eliminadas son:
2.10 En el trabajo principal de la semana pasada, ¿en qué forma obtiene sus ingresos o le pagan?
2.10 a En el trabajo principal de la semana pasada, ¿cada cuánto obtiene sus ingresos o le pagan?, ¿cuánto ganó o en cuánto calcula sus ingresos?

2.3 Diseño muestral y alcances de la muestra

En este apartado se describe el diseño muestral de la encuesta *Lo que dicen los pobres.* La información presentada está basada en el primer capítulo del libro de *Desmitificación y nuevos mitos sobre la pobreza: "Escuchando lo que dicen los pobres",* (Székely, 2005).

En la encuesta, anteriormente mencionada, el diseño muestral fue un estratificado, por conglomerados y multietápico.

El diseño metodológico de *Lo que dicen los pobres*, se obtuvo de una muestra probabilística de viviendas en localidades urbanas y rurales, que tiene las siguientes características:

- Contiene tres regiones en el país: Norte, Centro y Sur.
- Tiene dos cortes de tamaño por localidad: Rurales de menos de 2500 habitantes y Urbanas 2500 habitantes y más. Por lo que el estudio permite hacer inferencias para las dos cohortes de edad.
- El marco de muestreo está basado en el XII Censo General de Población y vivienda del año 2000[17], (INEGI, 2000).
- Considera una estimación del número de hogares en pobreza patrimonial, para todas las localidades del país, según el Censo 2000.
- El cálculo del tamaño de la muestra se basa en un muestreo aleatorio simple sin remplazo que originalmente establece 384 casos. Debido al diseño de selección en distintas etapas se incrementó el tamaño de la muestra con el fin de compensar el *efecto de conglomeración* que implica este esquema de muestreo, se consideró un Efecto de Diseño (DEFF) de 2.25, con lo que el tamaño de muestra fue de 850 casos. También se consideraron las tres regiones en las que se dividió el país, de tal modo que se obtiene un tamaño de muestra de alrededor de 2,550 entrevistas. Es necesario considerar además que la última unidad de muestreo es la vivienda. Según el marco utilizado se tiene una población en pobreza patrimonial del 80%. Al incrementar nuevamente el tamaño de muestra, esto para compensar la proporción que no es pobre, se tienen 3,060[18] entrevistas. El número utilizado fue de 3,000 casos.

El objetivo de trabajar con un ponderador corresponde a asignar los pesos correctos a los casos de acuerdo al esquema de muestreo empleado.

Con el fin de trabajar con el mismo número de casos que tiene la muestra, pero con la distribución que se le adjudica el ponderador, se construye un ponderador relativizado dado por:

2.10 b Actualmente el salario mínimo mensual es de $1,200 pesos aproximadamente; ¿La cantidad que obtuvo por su trabajo el MES PASADO fue: …

2.10 c ¿Cuántas veces es mayor al salario mínimo?.

[17] El XII Censo de Población y Vivienda fue levantado entre el 7 y 18 de febrero de 2000 en todo el territorio nacional.

[18] El resultado se debe a 2,550(1.20)=3,060.

$$PR = \frac{N_m}{N_p}.FE = \frac{2{,}939}{24{,}451,678}\ .FE$$

Donde

$$FE = FE_R I_{(x \in LR)}\ + FE_U I_{(x \in LU)} = Factor\ de\ expansión$$

LU = Localidad Urbana

LR = Localidad Rural

N_m = Número de informantes encuestados = 2,939

N_p = Número de informantes ponderados = 24,451,678

Con éste último ponderador se trabaja en el resto del proyecto.

Los objetivos de trabajar con el ponderador relativizado son:

- Ajustar la población debido a que proviene de muestras no proporcionales.
- Tener el mismo número de casos que la muestra, es decir, en este caso 2939.
- Mantener la misma distribución que tiene el ponderador (el inverso del factor de expansión multiplicado por el número de residentes en el hogar mayores de edad).

2.4 Descripción y caracterización de la población de estudio de "Lo que dicen los pobres"

La pobreza es uno de los grandes retos a vencer en el mundo.
Muhammad Yunus (1940-)[19]

Para caracterizar a la población objetivo de una encuesta, las características de ésta tienen que estar en función de características e información disponible y válida de otra encuesta que estudie a una población más general, y que evidentemente contenga a la población objetivo que se quiere caracterizar. En este caso, se caracteriza a la encuesta "Lo que dicen los pobres" en función del XII Censo de Población y Vivienda 2000.

El objetivo de caracterizar a la población es conocer y describir a la población estudiada en Lo que dicen los pobres en función de la población nacional, que es la descrita por el Censo 2000. La idea es encontrar una estructura social diferente, de tal modo que los encuestados por Lo que dicen los pobres vivan en condiciones más marginadas que las condiciones en las que se viven en todo el país.

[19] Muhammad Yunus es Premio Nobel de la Paz 2006 por sus notables contribuciones al desarrollo de la economía sustentable "desde abajo". Estima que el 94 por ciento del ingreso económico mundial queda en manos de apenas el 40 por ciento de la población.

Para lograrlo, primero se describen las definiciones de la pobreza por ingresos del Consejo Nacional de Evaluación para la Política de Desarrollo Social (CONEVAL), sus números y su geografía. También se discuten algunos datos de interés arrojados por la encuesta.

La mitad de los habitantes del mundo subsisten con apenas dos dólares al día, en tanto que más de mil millones de personas sobreviven con menos de un dólar diario. Esta situación de desigualdad social es no sólo injusta y terrible sino peligrosa, ya que puede perturbar incluso la paz mundial y traer consigo consecuencias que afecten el crecimiento, el bienestar y la seguridad de los países. En la actualidad, la medición de la pobreza no sólo se hace en términos estrictamente materiales sino también humanos. Comprende bajos índices de ingreso y de consumo, pero también insuficiencias en educación, salud, nutrición, tiempos de vida y otras áreas primordiales que afectan las capacidades humanas básicas, impidiéndole su crecimiento y negándole su derecho a una existencia digna, (Beláustegui, 2009).

La pobreza es un fenómeno complejo y multidimensional, difícil de solucionar pero que es necesario enfrentar con energía, a fin de revertir la marginación y el rezago sociales.

Las definiciones de la pobreza por ingresos establecidas por el Comité Técnico para la Medición de la Pobreza (CTMP)[20]. Según esto, la Ley General de Desarrollo Social (LGDS) indica que para la medición de la pobreza por ingresos una de las variables que se debe utilizar es el ingreso corriente total. De acuerdo con la metodología de cálculo, se definieron tres niveles de pobreza, (CONEVAL, 2008):

- **La pobreza alimentaria:** Incapacidad para obtener una canasta básica alimentaria, aún si se hiciera uso de todo el ingreso disponible en el hogar en comprar sólo los bienes de dicha canasta.
- **La pobreza de capacidades:** Insuficiencia del ingreso disponible para adquirir el valor de la canasta alimentaria y efectuar los gastos necesarios en salud y educación, aún dedicando el ingreso total de los hogares nada más que para estos fines.
- **La pobreza de patrimonio[21]:** insuficiencia del ingreso disponible para adquirir la canasta alimentaria, así como realizar los gastos necesarios en salud, vestido, vivienda, transporte y educación, aunque la totalidad del ingreso del hogar fuera utilizado exclusivamente para la adquisición de estos bienes y servicios.

La categorización de la pobreza por ingresos, en función de tener menos ingresos disponibles para cubrir las necesidades básicas, es: (1) pobreza alimentaria, (2) pobreza de capacidades y (3) pobreza de patrimonio.

Al analizar las líneas de la pobreza[22] establecidas por el CTMP para los años 1992, 2000, 2006 y 2008 son presentas en la siguiente tabla. En el año 2000[23], bastaba tener un ingreso menor

[20] Las definiciones fueron establecidas por el CTMP en el año 2002. Fueron avaladas por el CONEVAL.
[21] Recuérdese que la encuesta "Lo que dicen los pobres" está orientada a este tipo de población en pobreza.
[22] Recuerde que los pobres son aquellos que no tienen suficientes ingresos o consumo, de tal manera que no se les puede clasificar por encima de cierto umbral.

o igual a $1,257.2 y $841.8 pesos mensuales, en localidades urbanas y rurales respectivamente, para catalogar a una persona con pobreza de patrimonio, (tabla 2.1).

Tabla 2.1

Ámbito y tipo de pobreza	Líneas de pobreza[1]			
	1992*	2000	2006	2008
Urbano				
Alimentaria	$167.96	$626.62	$809.87	$949.38
Capacidades	$206.00	$768.55	$993.31	$1,164.41
Patrimonio	$336.99	$1,257.25	$1,624.92	$1,904.84
Rural				
Alimentaria	$124.75	$463.36	$598.70	$706.69
Capacidades	$147.49	$547.83	$707.84	$835.52
Patrimonio	$226.37	$840.81	$1,086.40	$1,282.36

[1] El valor de la línea está en pesos de agosto de cada año. * Para el año de 1992 se realiza el ajuste por el cambio a nuevos pesos
Fuente: estimaciones del CONEVAL con información del Banco de México

El en año 2000, la población nacional era de más de 97 millones, de los cuales el 24.1% correspondía a pobres alimentarios, el 31.8% a pobres de capacidades y el 53.6% a pobres de patrimonio, (tabla 2.2).

Tabla 2.2

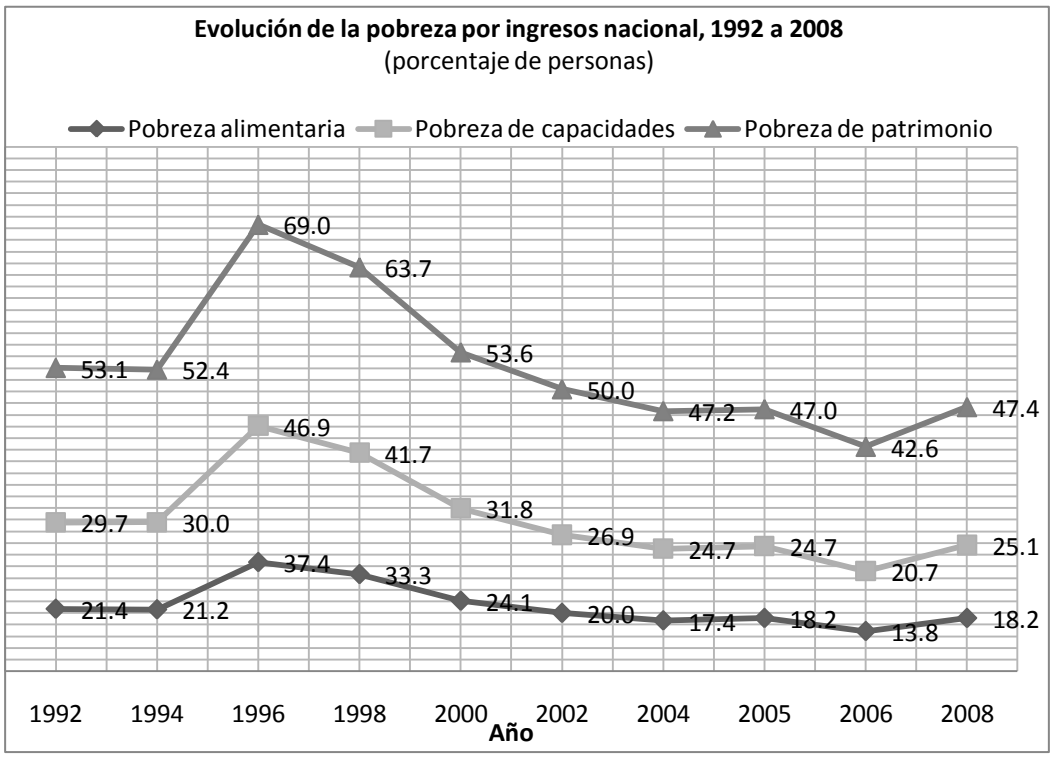
	Población total	Pobreza por ingresos en el 2000		
Entidad	Población 2000	Pobreza alimentaria (%)	Pobreza de capacidades (%)	Pobreza de patrimonio (%)
Nacional	**97,483,412**	**24.1**	**31.8**	**53.6**

Fuente: Mapas de Pobreza por ingresos 2000, CONEVAL.

Al analizar la evolución de la pobreza por ingresos desde el año 1992 hasta el 2008 en porcentaje de personas (gráfica 2.1), se observa que la reducción de la pobreza entre 1996 y 2005 sólo ha permitido disminuir, en pequeña medida, la pobreza prevaleciente en 1994. Otro dato de interés, la pobreza alimentaria se incrementó en casi 15.2 millones de personas entre 1994 y 1996[24], mientras que bajó en 15.3 millones de personas entre 1996 y 2005, (CONEVAL, 2009).

[23] Se considera el año 2000, por ser la estimación anterior al levantamiento de la encuesta *Lo que dicen los pobres*.
[24] El incremento en los diferentes tipos de pobrezas se debió a la crisis que sufrió México en 1994.

Gráfica 2.1



Fuente: estimaciones del CONEVAL con base en las ENIGH de 1992 a 2008

En el año 2002[25], en México, a nivel nacional se tenía que el 20% de la población vivía en pobreza alimentaria, el 26.9% de pobreza de capacidades y el 50% de pobreza patrimonial. Los cuales, al desglosarlos por tipo de localidad, se obtiene para localidades urbanas estimaciones más pequeñas que para las localidades rurales, (tabla 2.3).

Tabla 2.3

Pobreza por ingresos 2002						
	Porcentajes			Número de personas		
Tipo	**Alimentaria**	**Capacidades**	**Patrimonio**	**Alimentaria**	**Capacidades**	**Patrimonio**
Nacional	20.0	26.9	**50.0**	20,139,753	27,085,351	**50,406,024**
Urbano	11.3	17.2	**41.1**	7,062,099	10,696,819	**25,656,394**
Rural	34.0	42.6	**64.3**	13,077,654	16,388,532	**24,749,630**

Fuente: CONEVAL, pobreza por ingresos; Estimaciones en base a la ENIGH 2000.

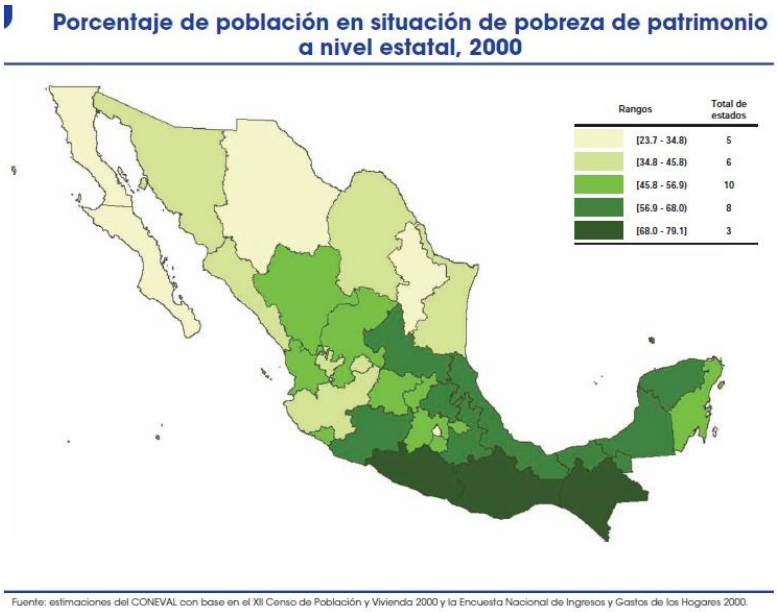
[25] Medición anterior de la evolución de la pobreza por ingresos (2002), anterior al levantamiento de la encuesta "Lo que dicen los pobres" en julio de 2003.

2.5 Pobreza por ingresos

En los siguientes mapas (mapa 1 y mapa2), la información presentada es del 2000 que fue generada a partir del Censo, (INEGI, 2000), y de la ENIGH, (INEGI, 2000 b). La institución encargada del cálculo es el CONEVAL, de acuerdo al marco normativo que establece la Ley General de Desarrollo Social.

Las entidades con mayor porcentaje de pobreza patrimonial (3 estados) entre sus habitantes son: Chiapas (75.7%), Guerrero (70.2%) y Oaxaca (68%). Por el contrario, las entidades con menor presencia de pobreza de patrimonio (5 estados) son: Baja California, Baja California Sur, Sonora, Nuevo León y el Distrito Federal, con un rango de 23.7% a 34.8%, (mapa 1).

Mapa 1

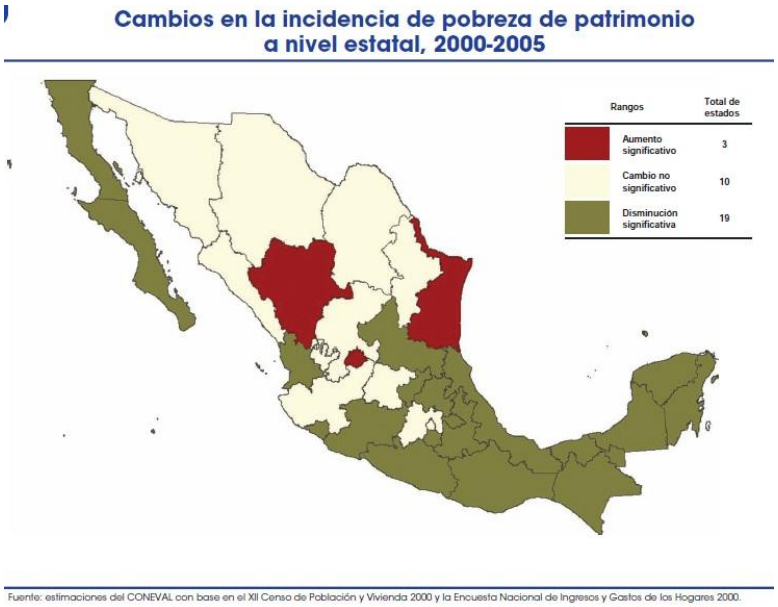


Al analizar la incidencia de pobreza de patrimonio a nivel estatal. Se presentan los cambios presentados en un periodo de cinco años, se observa que en sólo tres entidades federativas el aumento fue significativo: Durango, Tamaulipas y Aguascalientes. Se contabilizaron 10 entidades que mantuvieron su distribución porcentual de pobreza de patrimonio. Es importante mencionar, que durante este periodo, 19 entidades (59.3% del total) tuvieron una disminución significativa. En particular, Chiapas, Oaxaca y Guerrero[26] presentaron una disminución. Por otro lado, Baja California y Baja California Sur también disminuyeron su pobreza de patrimonio.

[26] Entidades con los porcentajes más altos de pobreza de patrimonio en 2000.

Mientras que Sonora, Nuevo León y Distrito Federal[27] mantuvieron los mismos niveles de pobreza patrimonial, (mapa 2).

Mapa 2



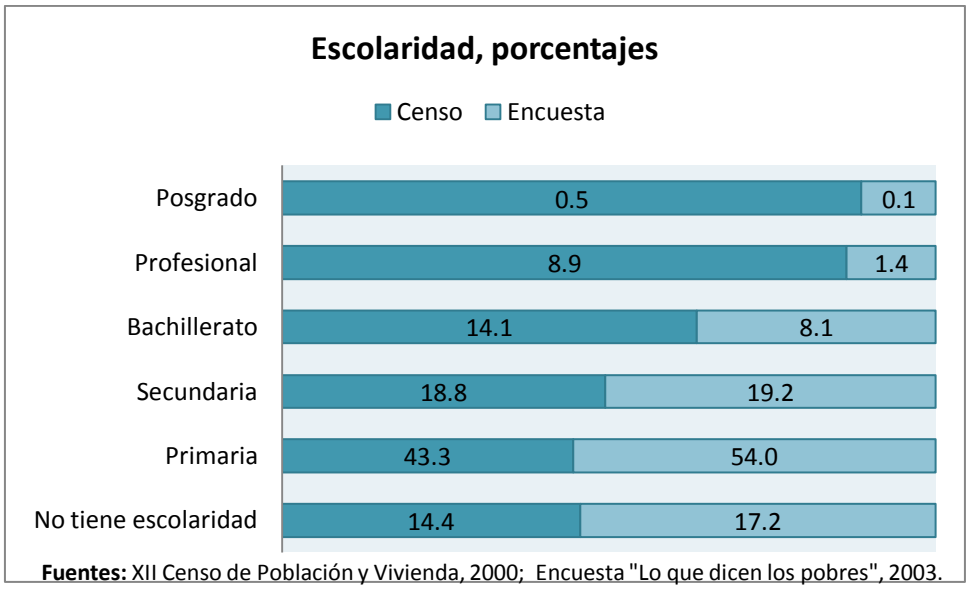
2.6 Algunos hechos de "Lo que dicen los pobres"

Todos los hombres nacen iguales, pero es la última vez que lo son.
Abraham Lincoln
(1808-1865)

Con respecto a la escolaridad, la población de la encuesta muestra menores niveles de escolaridad que los que se tienen a nivel nacional, como muestra se puede observar que el 71.1% tiene primaria o menos comparado con el 57.7% a nivel nacional, (gráfica 2.1).

[27] Entidades con los menores porcentajes de pobreza de patrimonio en 2000.

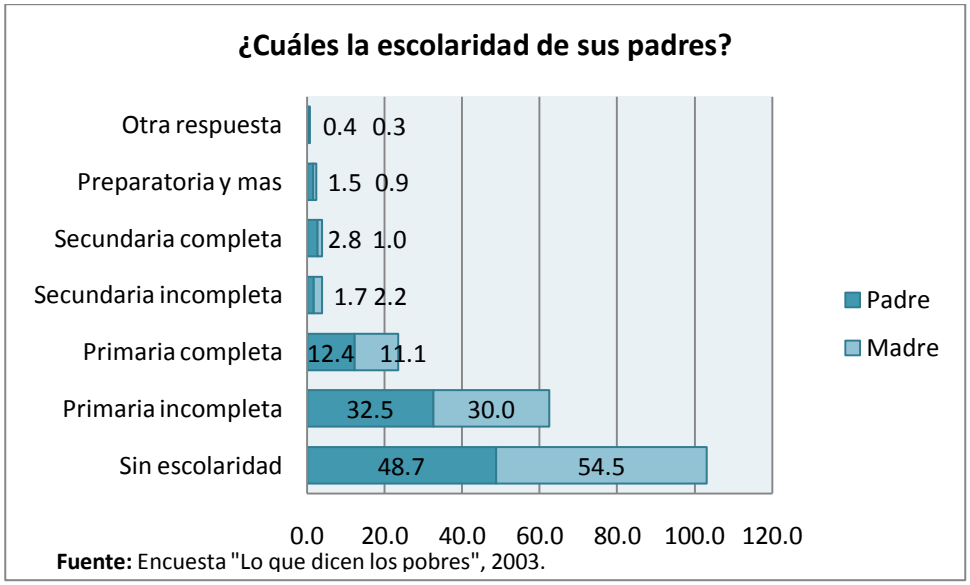
Gráfica 2.1



Fuentes: XII Censo de Población y Vivienda, 2000; Encuesta "Lo que dicen los pobres", 2003.

En cuanto a la escolaridad de los padres del informante (gráfica 2.2), se observa que en el caso del padre, 93.7% tiene primaria o menos, mientras que para el caso de la madre el porcentaje de la madre se incrementa llegando al 95.6%. Ambos porcentajes son están encima del 71.1% mostrado por la encuesta. Nótese que las madres tienen mayor porcentaje de no escolaridad, 54.5%, contra el 48.7% de los hombres.

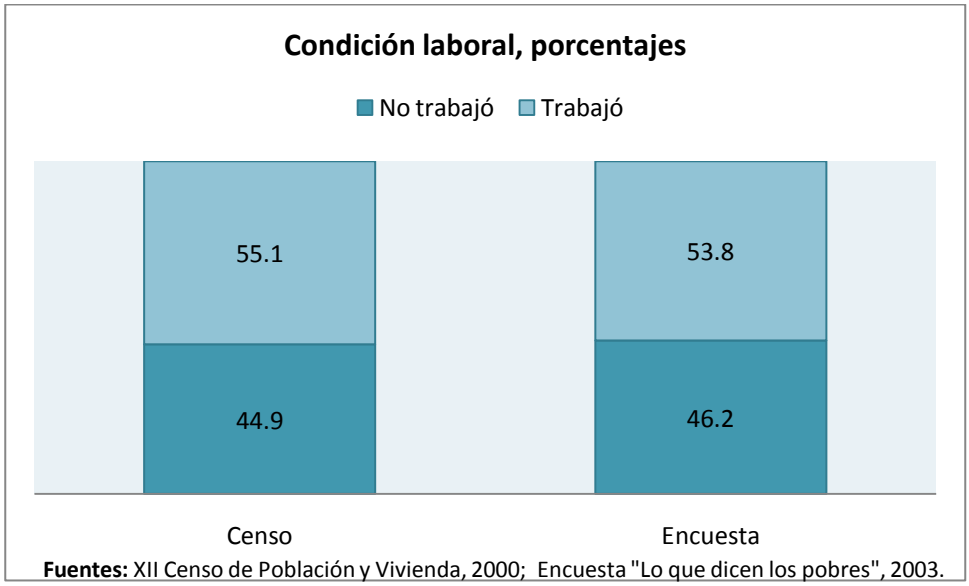
Gráfica 2.2



Fuente: Encuesta "Lo que dicen los pobres", 2003.

Al hablar de la condición laboral entre *Lo que dice los pobres* y *el Censo 2000* (gráfica 2.3) se observa que mantienen porcentajes similares entre las personas que laboraban (46.2% y 44.9%) contra las que no laboraban (53.8% y 55.1%) para pobres de patrimonio y población a nivel nacional, respectivamente. Esto quiere decir, que ambas mantienen la misma distribución en lo que respecta a condición laboral.

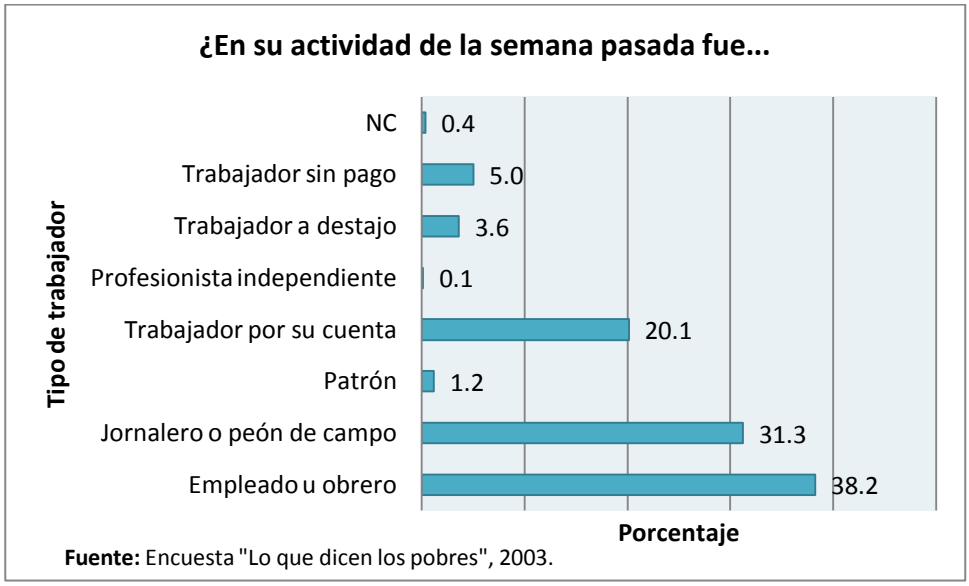
Gráfica 2.3



Fuentes: XII Censo de Población y Vivienda, 2000; Encuesta "Lo que dicen los pobres", 2003.

En cuanto al tipo de trabajo en el que se desempeña la población de la encuesta, (gráfica 2.4), se tiene que poco más de una tercera parte es empleado u obrero, y un porcentaje importante es jornalero o peón de campo, y sólo el 1.2% es patrón. Podría pasar que el pequeño porcentaje de las personas que son patrones no tenga prestaciones de trabajo debido a que ellos mismos no las quieran contratar. También es importante mencionar que el 20% es trabajador por cuenta propia, es decir, es gente que está autoempleada.

Gráfica 2.4



Fuente: Encuesta "Lo que dicen los pobres", 2003.

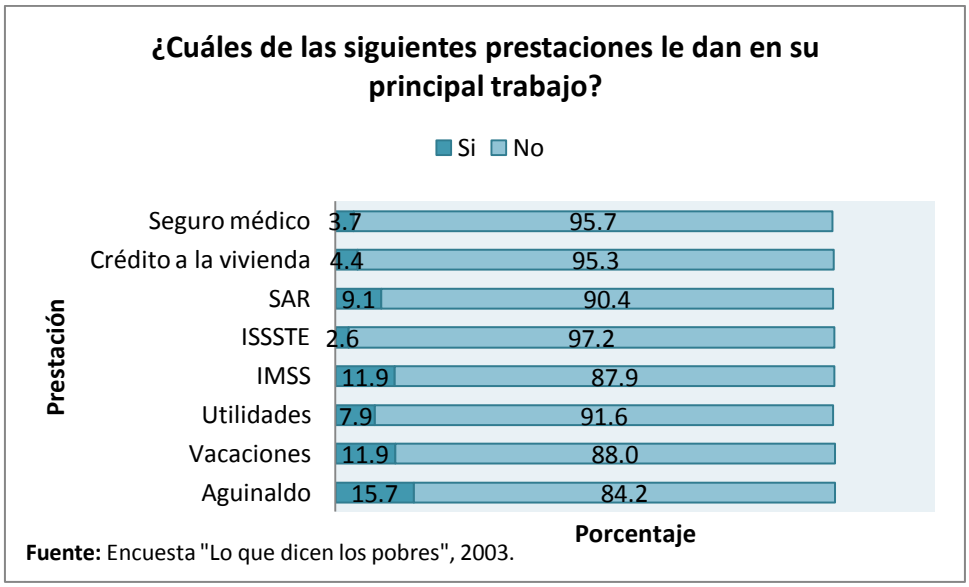
Del total de personas encuestadas, sólo consideremos a las personas asalariadas (gráfica 2.5), es decir, nos quedamos con el 83.7% de las encuestadas: de aquí se concluyen los siguientes porcentajes: el 4% tiene seguro médico, el 4% cuenta con crédito a la vivienda, el 9% tienen Fondo de Ahorro para el Retiro (SAR[28]), el 3% está protegidos por medio de la seguridad social del ISSSTE, el 12% puede tomar vacaciones con goce de sueldo, y 16% recibe aguinaldo.

Los datos anteriores son reveladores para la encuesta debido a que, aunque el 57.4% de ellos declaró tener un trabajo permanente, este hecho no es suficiente para tener prestaciones laborales, incluso cuando algunas de ellas se tienen que recibir con el respaldo de la Ley Federal del Trabajo.

Los trabajadores de *Lo que dicen los pobres* en promedio tienen 6.8 años de antigüedad en el lugar donde trabajan en el momento de la encuesta. Además, el 7.5% de los trabajadores que actualmente están laborando tiene más de 25 años en la misma empresa.

[28] Sistema de Ahorro para el Retiro, SAR.

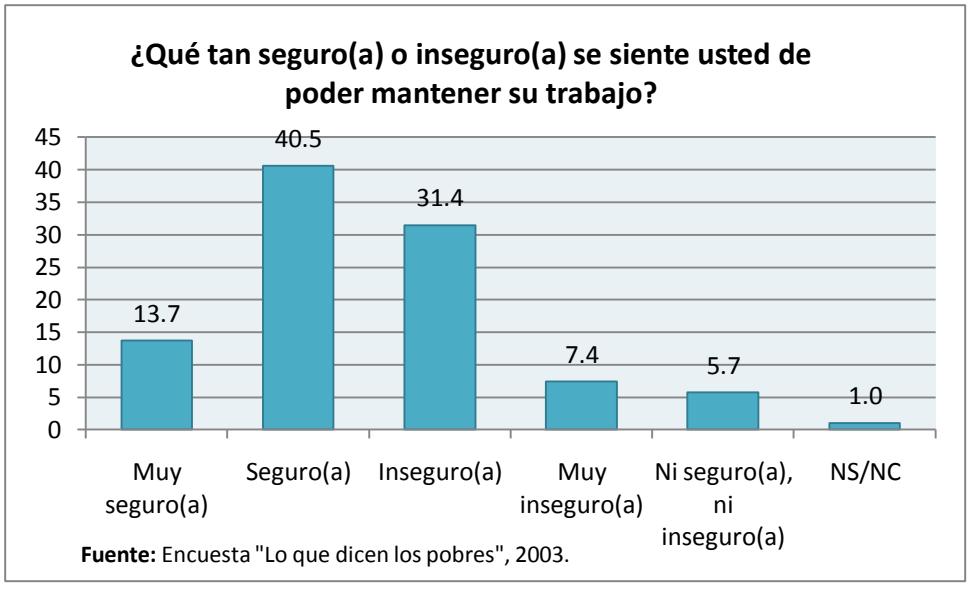
Gráfica 2.5



Fuente: Encuesta "Lo que dicen los pobres", 2003.

Al indagar sobre la seguridad con la que se siente en el trabajo el asalariado, tan sólo uno de cada dos se siente seguro y muy seguro dentro de este, y cuatro de cada diez se sienten inseguros y muy inseguros, (gráfica 2.6).

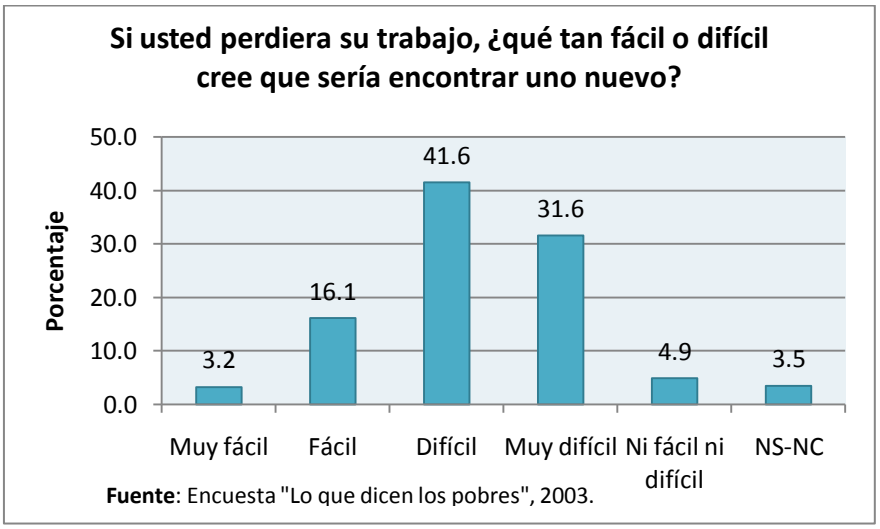
Gráfica 2.6



Fuente: Encuesta "Lo que dicen los pobres", 2003.

Al profundizar en la perspectiva que tienen sobre encontrar un nuevo trabajo en este momento dado que perdieron su trabajo anterior, es revelador que siete de cada diez personas cree que sería difícil o peor, únicamente dos de cada diez cree que sería más que fácil. Esto nos describe la perspectiva que ellos tienen de la situación laboral actual del país. En este sentido, la gente de la encuesta se siente insegura de poder mantener su trabajo, sin embargo, también están conscientes que al perderlo sería difícil y muy difícil encontrar uno nuevo, (gráfica 2.7).

Gráfica 2.7



Fuente: Encuesta "Lo que dicen los pobres", 2003.

Entrando en terreno social, cuando se les pregunta sobre la clase social a la que dirían que pertenecen, nueve de cada diez encuestados mencionaron son de clase social baja, uno de cada diez son de la clase media y menos del uno por ciento dijeron son de la alta, (tabla 2.4).

Tabla 2.4

¿A qué clase social diría que pertenece usted?[29]	
Tipo	**Porcentaje**
Alta	0.2
Media	10.3
Baja	89.5

Fuente: Encuesta "Lo que dicen los pobres", 2003.

[29] La pregunta *¿A qué clase social diría que pertenece usted?* está orientada a medir la percepción de los informantes sobre su situación económica. Aunque para la clasificación de clases sociales una componente importante es el ingreso, al no contar con éste, la pregunta anterior da una orientación sobre su situación. La información de la pregunta podría estar sesgada respecto de la clase social real; esto debido a la manera en que se sitúa el informante.

3. Evaluación del impacto del Programa Oportunidades

A lo largo del capítulo se tratan diferentes temas relacionadas a la evaluación del impacto del programa Oportunidades. Como parte fundamental de la medición de impacto, en la primera parte se presenta la definición de indicadores e índices, enseguida se presentan tres técnicas fundamentales para la construcción de índices, rangos sumados, rangos sumados ponderados y componentes principales. Al final de la sección se construyen algunos índices con el fin de ejemplificar cada una de las metodologías presentadas.

Un aspecto relevante que se aborda en el capítulo, y fundamental en la construcción de índices, es la decisión de dónde realizar los cortes para la categorización de los índices.

3.1 La medición de aspectos sociales

El uso de técnicas cuantitativas para recoger datos económicos y sociales, ha sido una constante que se repite a lo largo de la historia de la humanidad, es más, con frecuencia se asocia el desarrollo de una sociedad por su capacidad en conseguir y generar nuevas formas de información cuantitativa; el uso de indicadores sociales es una aproximación a la medición en las Ciencias Sociales. Los indicadores permiten: (a) describir los fenómenos sociales ; (b) incrementar la comprensión que se tiene en torno a sus interrelaciones; (c) revisar no solo los conceptos utilizados (incorporando variables no consideradas); (d) actualizar las teorías a las que originalmente daban soporte, (Rodríguez Jaume, 2000).

Según Rodríguez Jaume, un indicador es un procedimiento que permite cuantificar alguna dimensión conceptual y que, al aplicarse, produce un número. Son de uso común al medir desempeños, impactos entre períodos y comportamientos sociales.

En torno al concepto de indicador social, se percibe un clima de incertidumbre y ambigüedad, que se extiende al conjunto de aspectos metodológicos y teóricos ligados a la noción de indicador.

El concepto más elemental de indicador es:

"Lo que da señal o cuenta de algo concretándolo", (Ródríguez, Jaume 2000).

En 1977, según Peña Trapero, *"la definición de un indicador social no es única y que depende, en gran medida, de los que se pretende conseguir con su utilización"* (Peña, B., 1977: 256).

El concepto de indicador social, como:

"(…) presentación cuantitativa de información sobre fenómenos sociales, su localización, desarrollo y correlación (…)"[30]*,*(Casas, F. 1989: 25; Olivera, A. 1997: 689).

La definición que presenta Bauer es:

"Los indicadores sociales son los medios por los que una sociedad puede afirmar donde se encuentra en la actualidad o donde estuvo y proporcionan una base de anticipación más que de previsión, en lo que concierne a nuestra evolución en un cierto número de dominios o campos sensibles del bienestar social", (Bauer, 1966).

Carmona establece la siguiente definición en términos del proyecto DORIS[31].

"Un indicador social es la medida estadística de un concepto o de una dimensión de un concepto o de una parte de aquella, basado en un análisis teórico previo e integrado en un sistema coherente de medidas semejantes, que sirvan para describir el estado de la sociedad y la eficacia de las políticas sociales".

También hay definiciones de indicadores sociales que se basan en la recopilación de información sobre aspectos de dinámica social, como el que da la ONU.

"(…) construcciones, basadas en observaciones normalmente cuantitativas, que nos dicen algo acerca de un aspecto de la vida social en el que estamos interesados o acerca de los cambios que están teniendo lugar en él" , (ONU, 1975: 30).

Otro enfoque lo presenta la OCDE[32]:

"Un indicador social es una medida estadística y directa que permite observar el nivel y las variaciones en el tiempo de una preocupación social", (OCDE, 1982: citado por Rodríguez Jaume, 2000).

También hay otras definiciones de indicadores sociales dadas por el INE[33] desde el punto de vista operativo:

"(…) compendios de datos básicos que dan una medida concisa de la situación y cambios relativos a aspectos de las condiciones de vida de la población que son objeto de preocupación social", (INE, 1991: citado por Rodríguez Jaume, 2000).

Mientras que desde una perspectiva metodológica los define como:

[30] Este término de indicador social fue acuñado por la *American Academy of Arts and Science* que en 1962, por encargo de la *NASA* presentó un trabajo cuyo objetivo era medir en la sociedad Americana la repercusión de su programa de exploración espacial.
[31] Gobierno de Quebec: El Proyecto Doris (*Dossiers Regionaux et Indicateurs Sociaux*), citado por Carmona, J.A., 1977, Op. Cit., pp. 30.
[32] La Organización para la Cooperación y el Desarrollo Económicos, OCDE, agrupa a 34 países miembros comprometidos con la democracia y una economía de mercado, cuya finalidad es: (i) Apoyar el desarrollo económico sostenible; (ii) Incrementar el empleo; (iii) Elevar los niveles de vida; (iv) Mantener la estabilidad financiera; (v) Apoyar el desarrollo económico de otros países y (vi) Contribuir al crecimiento del comercio mundial.
[33] Instituto Nacional de Estadística, INE, es la entidad organismo autónomo de carácter administrativo adscrito al Ministerio de Economía y Hacienda, a la **organización estadística en España** y su legislación, al **Sistema Estadístico Europeo** y cómo está organizado y también a la **Oficina del Censo Electoral**, que es el órgano encargado de la formación del censo electoral.

"El indicador aparece como una variable manifiesta, observable o empírica, de la que es posible inferir otra variable, teórica, subyacente o no inmediatamente observable, representada por aquella", (INE, 1991: citado por Rodríguez Jaume, 2000).

La definición de indicador que se tomará como base es la metodológica (definición anterior). Las demás sólo sirven de ayuda para entender un indicador.

Por otro lado, la medición de procesos sociales puede abordarse desde dos ópticas:

- **Teórico-Conceptual:** supone el acercamiento al problema de la medición social ofreciendo un marco conceptual, teórico y metodológico que cubra el vacío que a este respecto se ha manifestado, proporcionando unidad y coherencia.
- **Desarrollo de datos:** busca una aproximación metodológica, y en ella se desarrolla una concepción más empírica en torno al tema de estudio, al centrar el interés en la recogida y análisis de información.

Tradicionalmente el desarrollo empírico de los indicadores sociales se ha venido realizando de dos maneras distintas y complementarias. La primera, presenta exclusivamente una batería de indicadores sociales que evalúan y miden individualmente cada una de las parcelas de las dimensiones de la vida social. En la segunda opción, se puede completar el anterior enfoque presentando índices sintéticos o índices de medición global de la situación o fenómeno sometido a examen.

Según Rodríguez Jaume, "(…) en el momento en el que un hecho social deja de ser una mera observación para pasar a ser un objeto de investigación, necesitamos un concepto que defina esa idea, concepto que a su vez pueda ser medido", (Rodríguez Jaume, 2000).

Una vez seleccionado el concepto, el siguiente paso es exponer una definición teórica del mismo; entendiéndose por definición, explicar en los términos más simples los posibles significados del concepto.

Las dimensiones de un concepto, según González Blasco, son "(…) los distintos aspectos que puede ser considerado un concepto, representando así los <<componentes>> del concepto". Entonces, las dimensiones son conceptualizaciones más precisas del concepto inicial, (Gónzalez, 1994; citado en Rodríguez Jaume, 2000).

Luego de identificar las dimensiones, se seleccionan los indicadores de éstas, con lo que cuantifican. Al elegir el número de indicadores, se pierden algunas características de la población pero se gana la posibilidad de medición. La elección de los indicadores está en función de los propósitos y necesidades de la investigación, manteniendo en mente las posibilidades reales existentes. Los indicadores se obtienen a partir de información procedente de preguntas de algún cuestionario, estadísticas e informes previos.

En el caso del presente trabajo, la fuente principal es el cuestionario individual (para personas mayores de 18 años) de la encuesta *"Lo que dicen los pobres"* y las fuentes de apoyo son el *XII Censo de Población y Vivienda 2000, el Índice y Grado de Marginación de la CONAPO, índices de pobreza por ingresos del CONEV*AL, etc.

Para construir un indicador social, se necesita tener criterios de selección. El primer paso es tener el objeto el cual sea de interés para medir. Después se desagrega o divide en los distintos aspectos sociales que lo componen. Posteriormente, y en función de las restricciones conceptuales, se le asigna a cada uno de ellos, al menos un indicador que describa su estado. Finalmente, del conjunto de indicadores propuestos, se seleccionan sólo aquellos que cumplan con una serie de criterios lógicos y, en algunas ocasiones, criterios empíricos.

Abajo, se enlistan los criterios de selección de indicadores sociales de la ONU, (ONU, 1975):

- Los indicadores seleccionados deben adecuarse y medir el aspecto de la preocupación social que quiere medir. Para su medición puede utilizarse indistintamente, indicadores directos e indirectos.
- El sistema de indicadores propuestos debe ser mínimo en cuanto a número[34]. Es decir, se busca tener el mínimo número de indicadores que maximicen, tanto como sea posible, la recolección de información.
- El sistema de indicadores debe estar perfectamente coordinado. Describiendo el indicador de la manera más adecuada una mayor cosmovisión del entorno social que pretende describir.
- Los indicadores sociales se obtienen a partir de información estadística confiable, veraz, exacta y con valor comparativo.
- Los indicadores sociales deben de estar disponibles en plazos cortos de tiempo, de manera que cumplan dos metas: reflejar rápidamente el comportamiento de la sociedad y sirvan de base en decisiones políticas.
- Los indicadores sociales deben ser viables, esto es, de inmediata aplicación o a lo más en el futuro más inmediato.

Cambiando ahora al concepto de índice, se define como:

"(…) un número estadístico que intenta resumir la información proporcionada por uno o más indicadores de un concepto", (Díez, 1967; citado en Rodríguez Jaume, 2000).

El objetivo principal de un índice es *"… medir o caracterizar una distribución por una medida única. Sintetiza por ese sistema diferentes variables en una sola resultante"*, (FOESSA., 1970:1958; citado en Rodíguez Jaume, 2000). Luego, un índice cuantifica toda la información útil de un aspecto del fenómeno resumiéndola en un número llamado número índice.

[34] En estadística, a este principio se le conoce como de Parsimonia.

La siguiente **tipología operativa** fue propuesta por (Sevila-Gúzman, 1973; citado en Rodríguez Jaume, 2000). Ellos elaboraron una tipología, nombrados en función del algoritmo utilizado en su construcción. Distinguen dos tipos de índices:

- **Índices que sintetizan información**. Son índices construidos a partir de las dimensiones que lo explican. Ejemplos del estilo son los índices que se construyen en la siguiente sección vía Rangos Sumados, Rangos Sumados Ponderados o Análisis de Componentes Principales. Ejemplos de éstos son: el Índice de Marginación de la CONAPO, el Índice de Rezago Social del CONEVAL, etc.

- **Índices de magnitud temporal**[35]. También son conocidos como números índices y tienen como objetivo convertir la información presentada series cronológicas y/o sincrónicas en magnitudes comparables. Como por ejemplo el Índice Nacional de Precios al Consumidor (INPC), Índice de Precios y Cotizaciones (IPC), etc.

3.2 Técnicas estadísticas para la construcción de índices

En el presente apartado se describe detalladamente las técnicas estadísticas para la construcción de índices, únicamente se consideran tres técnicas, las cuales son de uso más frecuente. Cada descripción está acompañada de la metodología y un ejemplo que ilustra la situación; los ejemplos son obtenidos de preguntas del cuestionario individual de *Lo que dicen los pobres*. Dos de las técnicas utilizan la operación de adición de variables, mientras que la tercera técnica es más abstracta, utiliza la reducción y resumen de datos ofrecido por el Análisis de Componentes Principales (ACP).

Si bien, la metodología de creación de índices es amplia[36]. Para fines específicos del trabajo, sólo se utilizarán tres tipos de técnicas, a saber:

1. Rangos sumados
2. Rangos sumados ponderados
3. Análisis de Componentes Principales

Por fácil que parezcan los métodos, no se puede pasar por alto los detalles y características de cada una de las técnicas anteriores.

[35] La clasificación permite la consideración o no de la temporalidad. De esta manera, los números índices son las series de índices que manifiestan cambios en el tiempo (evolucionan). Del otro lado están aquellos índices que su función es resumir y sintetizar la información en ausencia del tiempo.

[36] Aunque existen distintos métodos de obtener números índice tales como: Suma, Razón (ratio), Crecimiento Relativo, Crecimiento Medio Relativo, Crecimiento Acumulativo, Incremento Absoluto, Incremento Relativo, Incremento Medio Relativo, Prevalencia, Agregación Simple, Agregación Ponderada, Media Aritmética Compuestas sin Ponderar, Agregación Ponderada, Media Aritmética Compuesta Ponderada, Media Aritmética, Correlaciones y Análisis Factorial.

3.2.1 Rangos sumados

Los índices obtenidos por medio de suma de diferentes variables son los más primarios dentro de la investigación social.

La técnica de rangos sumados es bastante sencilla, consiste en sumar dos o más variables categóricas, de tal manera que se obtenga una nueva variable que sintetice la información proporcionada por las variables sumadas. Es importante mencionar que cada una de las variables tiene el mismo peso en la suma, lo que se traduce en que todas las variables que la componen son exactamente iguales, en cuanto a importancia.

Sean X y Y dos variables no métricas, es decir, X y Y pueden ser nominales u ordinales. Supongamos que X tiene n categorías, denotándose cada categoría como x_j, con $j = 1,2, \dots, n$. De la misma manera, supongamos que Y tiene m categorías, denotadas por y_i, i$= 1,2, \dots, m$.

Por lo tanto se tienen las siguientes observaciones. Si X toma el valor x_1, se tiene la categoría 1; si X toma el valor x_2, se tiene la categoría 2; …; si X toma el valor x_n, se tiene como categoría n . En general, cuando X toma el valor x_j , se le asigna la categoría j.

Análogamente para la variable Y se tiene i, si ocurre y_i , con i$= 1,2, \dots, m$.

Consideremos el caso para la suma de dos variables; el paso para k variables es fácil dar por inducción matemática.

Sea $Z = X + Y$ la nueva variable, que también es categórica. Obsérvese que el rango de Z varía entre $x_1 + y_1$ y $x_n + y_m$. Esto debido, a que el valor máximo que pueden tomar las variables X y Y es x_n y y_m, respectivamente. De la misma manera, el mínimo valor que puede tomar x y y es x_1 y y_1, respectivamente.

La suma puede tener nm combinaciones posibles de los sumandos x_j y y_i.

En consecuencia, Z está compuesta por nm categorías, donde cada categoría está dada por $z_j = x_j + y_i$ para j$= 1,2, \dots, n$ e $i = 1,2, \dots, m$.

Este método es conocido como rangos sumados, nombre asignado por el método de construcción de la nueva variable Z.

Para la validación de la confiabilidad que existe entre las variables que se quieren sumar se recomienda utilizar el Alfa de Cronbach, el cual permite cuantificar el nivel de confiabilidad de una escala de medida para la magnitud inobservable construida (índice obtenido por rangos sumados) a partir de las n variables observadas.[37]

[37] Para más detalles véase el anexo correspondiente: Anexo 2, Alfa de Cronbach.

3.2.1.1 Ejemplo:

Para efectos ilustrativos de la metodología de Rangos Sumados, se construye un **Índice de Confianza Generalizada**[38], en donde entenderemos como confianza.

Para la construcción del índice se utilizará la información de las preguntas (a), (b) y (c) de la batería 5.5[39]:

- *5.5 ¿Está usted muy de acuerdo, de acuerdo, en desacuerdo o muy desacuerdo con las siguientes afirmaciones? En el barrio o localidad donde vive …*
  - *(a) la mayoría de la gente es hornada y se puede confiar,*
  - *(b) los líderes de la comunidad nos representan bien ante el gobierno,*
  - *(c) la gente se interesa sólo de su propio bienestar?*[40]

Las respuestas a cada una de las preguntas anteriores: totalmente de acuerdo (1), de acuerdo (2), en desacuerdo (3) y totalmente en desacuerdo (4), otro (5), NS (8) y NC (9). Los valores perdidos son otra, NS y NC.

Observe que las preguntas anteriores miden un aspecto de confianza en diferentes situaciones. El inciso (a) mide la confianza en la mayoría de la gente, es decir, si se considera honrada o no. El inciso (b) mide la confianza en los líderes que se adjudica la comunidad en cuanto a la buena representación ante las autoridades gubernamentales. Y el inciso (c) La indaga sobre la confianza en su propio bienestar y no en los demás.

Para la construcción del índice, considere el inciso (a) como la variable X, el (b) como la variable Y y la pregunta (c) como la variable W. Por lo tanto, el índice creado será el resultado de la suma de las tres variables anteriores. El modelo matemático que describe tal situación es:

$$Z = X + Y + Z$$
$$= Pregunta5.5(a) + Pregunta5.5(b) + Pregunta5.5(c)$$

[38] La **Confianza Generalizada** se refiere básicamente a la creencia por parte del actor o de un conjunto de actores que una persona, colectividad o institución realizarán diversas acciones (proporcionando información, recursos, normas y orientaciones), y por tanto se tendrá un ambiente propicio para generar participación (Angulo, b, p. 9).

[39] Una **batería de preguntas** de un cuestionario hace referencia a un conjunto de preguntas que indagan en aspectos similares, además comparten el mismo fin de medición de eventos. Las preguntas que conforman una batería se preguntan de la misma manera y son presentadas de manera consecutivas. Un ejemplo es la batería 5.5 del cuestionario individual de Lo que dicen los pobres. (véase la ejemplificación de la construcción del índice).

[40] Por ser una batería de preguntas, éstas se leen de la siguiente manera:

(a) ¿Está usted muy de acuerdo, de acuerdo, en desacuerdo o muy desacuerdo con las siguientes afirmaciones? En el barrio o localidad donde vive**… mayoría de la gente es hornada y se puede confiar**.

(b) ¿Está usted muy de acuerdo, de acuerdo, en desacuerdo o muy desacuerdo con las siguientes afirmaciones? En el barrio o localidad donde vive… **los líderes de la comunidad nos representan bien ante el gobierno**.

(c) ¿Está usted muy de acuerdo, de acuerdo, en desacuerdo o muy desacuerdo con las siguientes afirmaciones? En el barrio o localidad donde vive… **la gente se interesa sólo de su propio bienestar**.

La distribución de cada una de las preguntas anteriores se presenta en las tablas 3.1, 3.2 y 3.3.

Tabla 3.1

5.5 ¿ESTÁ USTED ... CON LAS SIGUIENTES AFIRMACIONES " LA MAYORÍA DE LA GENTE ES HONRADA Y SE PUEDE CONFIAR EN ELLA "

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Totalmente de acuerdo**	**715**	24.3	**24.8**	24.8
	De acuerdo	**1205**	41.0	**41.8**	66.5
	En desacuerdo	**759**	25.8	**26.3**	92.8
	Totalmente en desacuerdo	**207**	7.0	**7.2**	100.0
	Total	**2885**	98.2	**100.0**	
Faltantes	Otra	22	.8		
	NS	28	1.0		
	NC	3	.1		
	Total	54	1.8		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Tabla 3.2

5.5 ¿ESTÁ USTED ... CON LAS SIGUIENTES AFIRMACIONES " LOS LÍDERES DE LA COMUNIDAD NOS REPRESENTAN BIEN ANTE EL GOBIERNO "

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Totalmente de acuerdo**	**324**	11.0	**11.7**	11.7
	De acuerdo	**891**	30.3	**32.2**	43.9
	En desacuerdo	**1143**	38.9	**41.3**	85.2
	Totalmente en desacuerdo	**411**	14.0	**14.8**	100.0
	Total	**2768**	94.2	**100.0**	
Faltantes	Otra	38	1.3		
	NS	129	4.4		
	NC	4	.1		
	Total	171	5.8		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Tabla 3.3

5.5 ¿ESTÁ USTED ... CON LAS SIGUIENTES AFIRMACIONES " LA GENTE SE INTERESA SÓLO DE SU PROPIO BIENESTAR "

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Totalmente de acuerdo**	**774**	26.3	**27.2**	27.2
	De acuerdo	**1392**	47.3	**48.9**	76.1
	En desacuerdo	**581**	19.8	**20.4**	96.5
	Totalmente en desacuerdo	**100**	3.4	**3.5**	100.0
	Total	**2847**	96.9	**100.0**	
Faltantes	Otra	25	.8		
	NS	66	2.2		
	NC	2	.1		
	Total	92	3.1		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Entonces las categorías y los pesos que se asignan a cada una de las categorías de las preguntas a considerar en el índice son las siguientes, (tabla 3.4):

Tabla 3.4

Nombre	Peso de la categoría	Categorías para x	Categorías para y	Categorías para w
Totalmente de acuerdo	4	X_4	Y_4	W_4
De acuerdo	3	X_3	Y_3	W_3
En desacuerdo	2	X_2	Y_2	W_2
Totalmente en desacuerdo	1	X_1	Y_1	W_1
Otra, NS y NC	-	-	-	-

Fuente: Encuesta "Lo que dicen los pobres"; 2003. Cálculos propios.

El siguiente paso, es codificar cada una de las preguntas anteriores [a,b y c] de tal manera que se agrupen las categorías[41] que son de interés, (tabla 3.5):

[41] Las categorías disponibles son: Totalmente de acuerdo (4), de acuerdo (3), en desacuerdo (2), totalmente en desacuerdo (1), otra, NS y NC.

Tabla 3.5

Pregunta 5.5 (a)			
a) la mayoría de la gente es honrada y se puede confiar en ella Variable x	Totalmente de acuerdo (4)	Valor: 1	Alta confianza
	De acuerdo (3)		
	En desacuerdo (2)	Valor: 0	Baja confianza
	Totalmente en desacuerdo (1)		
	Otra	Valores perdidos	
	NS		
	NC		
Pregunta 5.5 (b)			
b) los líderes de la comunidad nos representan bien ante el gobierno Variable y	Totalmente de acuerdo (4)	Valor: 1	Alta confianza
	De acuerdo (3)		
	En desacuerdo (2)	Valor: 0	Baja confianza
	Totalmente en desacuerdo (1)		
	Otra	Valores perdidos	
	NS		
	NC		
Pregunta 5.5 (c)			
c) la gente se interesa sólo de su propio bienestar Variable w	Totalmente de acuerdo (4)	Valor: 0	Baja confianza
	De acuerdo (3)		
	En desacuerdo (2)	Valor: 1	Alta confianza
	Totalmente en desacuerdo (1)		
	Otra	Valores perdidos	
	NS		
	NC		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

El cálculo del Alfa de Cronbach vale 0.856, es decir, $\alpha = 0.856$, por lo cual las variables anteriores (opciones (a), (b) y (c)) tiene una buena fiabilidad la escala aditiva[42]. Una vez codificadas de la manera anterior, cada una de las tres variables se suman bajo el modelo $Z = X + Y + W$, dando los siguientes resultados, (tabla 3.6):

Tabla 3.6

Tabla resumen de categorías		
Confianza generalizada		
Preguntas *a*, *b* y *c* *Z=X+Y+W*	Valores asignados suman: 3	Confianza alta
	Valores asignados suman: 1 ó 2	Confianza media
	Valores asignados suman: 0	Confianza baja

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

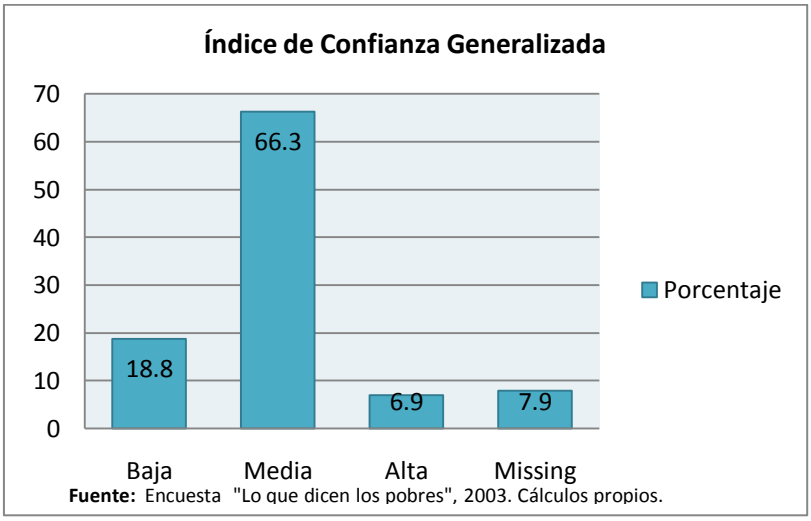
Generalmente en la construcción de índices existen diferentes formas en las que se pierde información. Si bien se tienen ventajas en cuanto a la reducción de información de diversas variables, se tiene el inconveniente de que los casos que no cuentan con información en alguna de las variables utilizadas para la construcción del índice, se queda sin información. Es importante mencionar, que hay 232 valores perdidos que se acumularon entre los valores perdidos de las tres preguntas originales. Esto quiere decir que con la construcción del **Índice de Confianza Generalizada** se pierde información de 232 casos de un total de 2939, (puede entenderse 232 personas encuestadas), que es el *costo de oportunidad* de construir el índice. (Zona Económica, 2011).

Esto refuerza el hecho que al calcular un índice se sintetiza información de diferentes indicadores o variables que lo componen pero a cambio de esto se pierde información que de otra manera no se perdería.

[42] Se considera una buena fiabilidad de una escala cuando el Alfa de Cronbach es al menos de 0.8, esto es, , $\alpha \geq 0.8$.

La distribución del **Índice de Confianza Generalizada es,** (gráfica 3.1):

Gráfica 3.1



Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

En este contexto, se entiende por baja confianza que los encuestados no confían en que la gente es honrada y se puede confiar en ella, no creen que los líderes de la comunidad los representen bien ante el gobierno y creen que la gente se interesa en su propio bienestar. Por otro lado, la alta confianza es entendida como la creencia en que la gente es honrada y se le puede confiar, los líderes de la comunidad son buenos representantes ante las autoridades gubernamentales, y finalmente, que la gente no se preocupa sólo en su propio bienestar. Por lo tanto, según el *Índice de Confianza Generalizada*, el 18.8% de los encuestados tiene confianza baja, el 66.3% tiene confianza media y el 6.9% tiene una confianza alta.

3.2.2 Rangos sumados ponderados

Consideremos las mismas variables no métricas del método anterior, es decir, sean X y Y estas variables. x_j, con $j = 1,2,\dots,n$ y y_i, i$= 1,2,\dots,m$ sus categorías, respectivamente.

Sean w y p, números distintos del cero. Es decir, $a \neq 0$ y $b \neq 0$. A estos números se les llaman los ponderadores de las variables X y Y, respectivamente. La manera de calcular estos ponderadores puede ser de diversas maneras, ya sea por puntuaciones factoriales[43], proporciones estimadas, medias, experiencia del investigador, etc. Es recomendable que los coeficientes de ponderación representar lo más fidedignamente posible a la variable que pondera.

[43] Las puntuaciones factoriales son valores arrojados por el Análisis Factorial y denotan el peso que cada caso aporta al ACP. En la siguiente sección se detallan.

Según Rodríguez Jaume (2000), los coeficientes de ponderación deben cumplir: (1) El coeficiente ha de ser sencillo, por lo que es recomendable que sea un número entero pequeño. (2) El coeficiente puede ser positivo y/o negativo en aquellos casos en los que las categorías de las respuestas así lo precisen.

El método consiste en multiplicar cada ponderador con su respectiva variable y el resultarlo sumarlo, esto es:

$$Z = aX + bY$$

La idea de ponderar las variables, supone que cada una de ellas tiene diferentes pesos en la variable Z. Esto es, que cada variable aporta al índice diferentes porcentajes, en función de la importancia de ésta. Mientras que en el caso anterior (rangos sumados), se asume que todas las variables aportan lo mismo en la variable sumada Z. A diferencia del primer método, en este se puede controlar la participación de cada una de las variables dentro de Z, lográndolo por medio de los ponderadores.

Ahora, supóngase que se tienen $X_1, X_2, \dots, X_r$ variables no métricas. Sean, $p_1, p_2, \dots, p_r$, números distintos del cero, sus respectivos ponderadores. La variable generada por la suma de rangos ponderados es:

$$Z = p_1X_1 + p_2X_2 + \cdots + p_rX_r = \sum_{k=1}^{r} p_kX_k$$

Obsérvese, que al suponer que todos los ponderadores $p_k$valen lo mismo, entonces la suma de rangos ponderados es equivalente a la suma de rangos.

Nótese que Z es una **combinación lineal** de las variables $X_1, X_2, \dots, X_r$, donde los números escalares son $p_1, p_2, \dots, p_r$.

Es recomendable utilizar el Alfa de Cronbach para medir la confiabilidad la nueva escala, en este caso, una escala aditiva ponderada.

3.2.2.1 Ejemplo:

Se ejemplifica la sección por medio de la construcción de un índice de confianza generalizado obtenido por este método. Si bien, se obtuvo el *Índice de Confianza Generalizada* por el **método de rangos sumados**, ahora se hace por medio de **rangos sumados ponderados**. La idea es contrastar ambos índices.

Según la tabla 3.7, la pregunta 5.5(a) L*a mayoría de la gente es honrada y se puede confiar en ella* tiene una ponderación de 3, con lo que las categorías *Totalmente de acuerdo y De acuerdo* valen 3, y en otro caso vale 0. Por otro lado, la pregunta 5.5(b) *Los líderes de la comunidad nos*

representan bien ante las autoridades de gobierno tiene un coeficiente de ponderación de 2, con lo que las categorías *Totalmente de acuerdo y De acuerdo* valen 2 y los demás casos valen cero. Finalmente, la pregunta 5.5(c) La gente se interesa sólo de su propio bienestar es ponderado por la unidad, 1, por lo que *Totalmente en desacuerdo y En Desacuerdo* valen 1, y en las demás categorías vale 0.

Se observa que bajo tal ponderación en la construcción del índice se da mayor peso al componente de confianza en las demás personas en general (*pregunta 5.5a. La mayoría de la gente es honrada y se puede confiar en ella),* al cual se le asigna un peso de 3; seguida del componente de confianza en los líderes y representantes (*pregunta 5.5b. los líderes de la comunidad nos representan bien ante el gobierno)* con peso de 2, y finalmente también se considera un componente de confianza basado en la solidaridad percibida ante las acciones de las demás personas (*pregunta 5.5c. la gente se interesa de su propio bienestar*), con un peso de 1.

Nuevamente las preguntas (a), (b) y (c) son denotadas por las variables $X, Y\ y\ Z$ respectivamente. Los coeficientes de ponderación son: $a = 3,\ b = 2\ y\ c = 1$. Por lo tanto, el modelo matemático para el Índice de confianza Generalizada Ponderada es:

$$Z = aX + bY + cX = 3X + 2Y + 1W$$

$$= 3 * pregunta5.5(a) + 2 * pregunta5.5(b) + pregunta5.5(c)$$

El siguiente paso es sumar las variables ponderadas. Obsérvese que el mayor valor que podría obtener es 6, y el de menor valor es 0. La imagen del índice es el conjunto {0,1,2,3,4,5,6}[44]. Cuando $Z = 6$, entonces la persona confía en las tres situaciones anteriores (a, b y c), mientras que al tomar el valor mínimo, el $Z = 0$, significa que la persona no tiene confianza en ninguna de las situaciones a, b y c.

[44] Cuando se construía el Índice de Confianza Generalizada por medio del método de rangos sumados a imagen era {0,1,2,3} y el valor mayor 3.

Tabla 3.7

Tabla resumen de categorías (ponderadas)			
Pregunta 5.5 (a)			
a) la mayoría de la gente es honrada y se puede confiar en ella Variable x	Totalmente de acuerdo	Valor: 3	Alta confianza
	De acuerdo		
	En desacuerdo	Valor: 0	Baja confianza
	Totalmente en desacuerdo		
	Otra	Valores perdidos	
	NS		
	NC		
Pregunta 5.5 (b)			
b) los líderes de la comunidad nos representan bien ante el gobierno Variable y	Totalmente de acuerdo	Valor: 2	Alta confianza
	De acuerdo		
	En desacuerdo	Valor: 0	Baja confianza
	Totalmente en desacuerdo		
	Otra	Valores perdidos	
	NS		
	NC		
Pregunta 5.5 (c)			
c) la gente se interesa sólo de su propio bienestar Variable w	Totalmente de acuerdo	Valor: 0	Baja confianza
	De acuerdo		
	En desacuerdo	Valor: 1	Alta confianza
	Totalmente en desacuerdo		
	Otra	Valores perdidos	
	NS		
	NC		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

La distribución del **Índice de Confianza Generalizada Ponderada,** (tabla 3.8), es:

Tabla 3.8
Índice de Confianza Generalizada (ponderado)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje Acumulado
Validos	**0**	**554**	18.8	**20.5**	20.5
	1	**180**	6.1	**6.7**	27.1
	2	**131**	4.4	**4.8**	31.9
	3	**671**	22.8	**24.8**	56.7
	4	**189**	6.4	**7.0**	63.7
	5	**779**	26.5	**28.8**	92.5
	6	**203**	6.9	**7.5**	100.0
	Total	**2707**	92.1	**100.0**	
Faltantes	**Sistema**	**232**	7.9		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Nuevamente hay 232 casos faltantes, dejando 2707 casos válidos (en total hay 2939). Por lo tanto, se pierde información de 232 personas encuestadas que de otra manera no se hubieran perdido, sin embargo, recuerde que el índice sintetizó información de las tres preguntas a, b y c. Por lo tanto, el costo de oportunidad de construir el índice es perder información de 232 casos.

Una manera transformar un índice numérico, es decir, un índice que se representa con una variable métrica (de intervalo o razón) es por medio de la ***técnica de Estratificación Óptima de Dalenius-Hodges,*** [(Dalenius, 1959); véase anexo 3]. Ésta se basa en analizar la distribución del índice de tal manera que se encuentren estratos (categorías) con una varianza lo más homogénea posible. Esto es, que los elementos pertenecientes a un mismo estrato tenga varianzas muy parecidas mientras que para elementos pertenecientes a diferentes estratos la varianza sea considerablemente diferente. La idea básica es agrupar elementos parecidos en el mismo grupo y elementos diferentes en grupos diferentes.

Luego, para obtener un índice de confianza generalizada con tres categorías se utiliza la técnica de Dalenius-Hodges. El procedimiento anterior nos conduce al **Índice de Confianza Generalizada Ponderado (Dalenius-Hodges)**, [tabla 3.9], cuya distribución es:

Tabla 3.9
Índice de Confianza Ponderado (Dalenius--Hodges)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Baja**	**734**	25.0	**27.1**	27.1
	Media	**802**	27.3	**29.6**	56.7
	Alta	**1171**	39.8	**43.3**	100.0
	Total	2707	92.1	100.0	
Faltantes	Sistema	232	7.9		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

En este contexto, se entiende por baja confianza que los encuestados no confían en que la gente es honrada y se puede confiar en ella, no creen que los líderes de la comunidad los representen bien ante el gobierno y creen que la gente se interesa en su propio bienestar. Por otro lado, la alta confianza es entendida como la creencia en que la gente es honrada y se le puede confiar, los líderes de la comunidad son buenos representantes ante las autoridades gubernamentales, y finalmente, que la gente no se preocupa sólo en su propio bienestar. Por lo tanto, según el *Índice de Confianza Ponderado*, el 27.1% de los encuestados tiene confianza baja, el 29.6% tiene confianza media y el 43.3% tiene una confianza alta.

3.2.3 Análisis de Componentes Principales[45] (ACP)

El ACP fue introducido por Hotelling en 1932. Se define como una técnica estadística que permite transformar un conjunto de variables, intercorrelacionadas, en otros conjuntos de variables no correlacionadas denominados factores. El objetivo de la técnica es explicar la mayor cantidad de varianza de las variables originales a través del menor número de factores.

Este tipo de análisis es una técnica e interdependencia en la que se consideran todas las variables simultáneamente, cada una relacionada con todas las demás y utilizando aún el concepto de valor teórico, el compuesto lineal de las variables. El valor teórico es también conocido como factor.

Entonces se puede usar el ACP para dos fines: como análisis exploratorio de las variables y como manera de construir un índice. A lo largo del trabajo, se utiliza únicamente para la construcción de índices.

[45] El éxito de esta técnica viene dado por cumplir dos requisitos básicos: (a) El principio de parsimonia, que establece que "todo modelo debe ser más simple que los datos en los que se basan". (b) el número de factores elegidos deben ser interpretables.

3.2.3.1 Metodología del Análisis por Componentes Principales

La metodología expuesta corresponde a un resumen del Capítulo III de (Hair, 1999).

Se divide en seis pasos, con el fin de trabajar detalladamente en cada uno de ellos sin perder detalle alguno.

1. **Objetivos del Análisis de Componentes Principales**: Se tienen dos objetivos específicos, las cuales son *el resumen de datos y la reducción de datos*[46]. El propósito general es condensar o resumir un conjunto de datos en una serie de factores buscando tener la mínima pérdida de información.

2. **Diseño.** El diseño implica tres decisiones básicas, a saber, (a) el cálculo de la matriz de correlación[47] de las variables de interés, (b) el diseño del estudio en términos de número de variables, las propiedades de las variables y el tipo de variables permisible, (c) el tamaño de muestra necesario.

El análisis factorial supone variables del tipo métrico (de intervalo o de razón). En caso de tener variables que sean no métricas, es decir, categóricas, se tienen que convertir en variables ficticias o de diseño[48].

3. **Supuestos:** existen dos tipos de supuestos, los estadísticos y los del tipo conceptual. Los primeros están formados por **normalidad, homocedasticidad y linealialidad**[49]. Cuando de violan estos supuestos se provoca una disminución en las correlaciones observadas.

[46] El **resumen de datos** busca una estructura de las relaciones entre las variables o los encuestados mediante la investigación de las correlaciones entre las variables o bien entre los encuestados.
La **reducción de datos** sirve para: (a) identificar un conjunto de variables suplentes de un conjunto de variables más grande para su utilización en análisis posteriores; (b) crea un conjunto de variables completamente nueva, más pequeña y fácil de manejar, para reemplazar el conjunto original.
En el caso (b) es particularmente donde se crea un índice que está en función de las variables originales. Las estimaciones de los factores y las contribuciones de cada variable a los factores, llamados carga de los factores, son todos los componentes necesarios para el análisis.

[47] El cálculo de la matriz de correlaciones se hará sobre la matriz de datos de entrada.

[48] Véase el anexo de variables de diseño.

[49] La **linealidad** es una propiedad importante de los métodos utilizados para efectuar mediciones en un intervalo de concentraciones. La linealidad de la respuesta a patrones puros y a muestras realistas puede determinarse. Generalmente no es cuantificada pero es comprobada mediante inspección o utilizando pruebas de significancia de la no-linealidad. La **no-linealidad** significativa es usualmente corregida mediante el uso de funciones de calibración no-lineal o eliminada seleccionando un intervalo de operación más restringido. (Metas y Metrólogos Asociados, 2008).
La **Normalidad** significa que las variables tienen una distribución Normal con Media μ y Varianza σ^2, es decir, si X es la variable, entonces $X \sim N(\mu, \sigma^2)$, (Kreyszin, 1973).
La **Homocedasticidad** es el nombre que se le adjudica a una Varianza Constante, (Kreyszin, 1973).

3.2.3.2 Pruebas para validar supuestos de ACP

Por el lado de los supuestos de tipo conceptual, se debe asegurar que **la matriz de correlaciones tiene suficientes correlaciones estadísticamente significativas**, con el fin de justificar el ACP. Esto se logra con una prueba de correlación para cada par de variables incluidas en el modelo. Si A y B dos variables métricas, las hipótesis son:

Ho: A y B están no correlacionados vs **Ha:** A y B están correlacionados

Teniendo en cuenta, que si el valor-p del estadístico de prueba es menor que el valor α (error asociado a la prueba), se rechaza la hipótesis nula.

También se utiliza **la matriz de correlaciones anti-imagen**[50]. Para esto, si las correlaciones anti-imagen son grandes, es indicativo de que esas variables no son ideales para el análisis factorial.

El **Contraste de Esfericidad de Bartlett**[51] es una prueba estadística para la presencia de correlaciones entre variables. Proporciona la probabilidad de que la matriz de correlación de las variables sea una matriz identidad[52].

La **Medida de Suficiencia de Muestreo**[53] (MSA), es un índice con rango entre 0 y 1, tomando 1 cuando la variable es perfectamente predicha y sin error por las otras variables. El punto de corte para el MSA es ½, buscando siempre valores superiores.

Una vez, cumplidos los supuestos anteriormente mencionados, se puede iniciar con el análisis factorial.

4. **La estimación de factores y la valoración del ajuste en general**: en este paso se tienen que tomar decisiones sobre el método de extracción de los factores, el número de factores seleccionados para representar al conjunto de variables iniciales. Recuerde que, para la construcción de los índices de este trabajo, se obtienen mediante el Análisis de Componentes Principales (ACP). Esto porque siguiendo el interés específico de la sección, buscamos el mínimo número de factores necesarios para explicar la proporción máxima de varianza representada en el conjunto de variables inicial; es decir, la crear un índice mediante ACP se busca sintetizar información de las variables (componentes) por medio de un número mínimo de factores y que éstos expliquen la mayor variabilidad posible.

---

[50] La Matriz de Correlaciones Anti-imagen es el valor negativo de la correlación parcial.

[51] El contraste de esfericidad de Bartlett es una prueba sensible al tamaño de muestra, arrojando resultados imprecisos cuando el número de casos crece.

[52] La matriz identidad es cuadrada con los elementos de la diagonal con 1 y ceros en las demás entradas, esto es, $I = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$. (Brickell, 1974).

[53] El MSA está en función del tamaño muestral, aumentan las correlaciones medias, aumenta el número de variables (o desciende el número de factores). También se puede hacer un análisis de MSA para las variables individuales incluidas en el modelo.

El ACP considera la **varianza total**[54] y estima los factores que contienen proporciones bajas de varianza única, y en algunos casos, los de **varianza de error**. Por lo tanto, los primeros factores no contienen la suficiente varianza única o de error como para distorsionar la estructura de factores en su conjunto. Por medio del ACP, se insertan las unidades en la diagonal de la matriz de correlación, para que se extraiga la varianza completa en la matriz de los factores.

Las comunalidades son estimaciones de la varianza compartida o común entre las variables.

El método de extracción utilizado en este trabajo, será el de **la raíz latente**[55]. Solamente se consideran Eigen-valores estrictamente mayores a 1[56]. La traducción de esto, significa que cada factor extraído por este medio explica al menos una variable.

5. **Interpretación de los factores**.

Se deben seguir tres pasos: (a) calcular la matriz inicial de factores no rotados con rotación de factores si es permisible, (b) criterios para la significación de las cargas factoriales y (c) interpretación de la matriz de factores.

(a) La matriz de factores no rotados[57] contiene las cargas factoriales[58] para cada variable sobre cada factor.

Para que los factores sean ortogonales, es necesario que el segundo factor se derive de la varianza restante después de extraerse el primer factor. Los factores siguientes se definen de forma análoga hasta haber agotado la varianza de los datos.

La **carga factorial** es el medio para explicar la aportación de cada variable dentro del factor, es decir, son las correlaciones entre cada variable y el factor.

Se pueden rotar los ejes de manera ortogonal, cuando los ejes se rotan de tal manera que se mantiene un ángulo de $\pi/2$ entre ellos. En el presente trabajo, se limita a rotar ortogonalmente, por medio de la técnica **VARIMAX**[59].

[54] En el análisis factorial existen tres tipos de varianza: (a) **Varianza común** es aquella varianza en una variable que se comparte con el resto de las variables en el análisis; (b) **Varianza específica** es aquella varianza asociada a una variable específica, es decir, es la varianza que únicamente pertenece a una variable individual; (c) **Varianza de error** es aquella varianza que se debe a la poca fiabilidad en el proceso de recolección de datos, al error de medición o a un componente aleatorio en el fenómeno estudiado.

[55] El **Método de la Raíz Latente**, también conocida como Regla de Kaiser, establece la condición de extracción: cualquier factor debería individual debería justificar la varianza de al menos una variable. Cada variable aporta el valor de 1 para el autovalor o Eigen-valor total.

[56] Los **Egienvalores** pueden ser interpretados como la variabilidad total explicada por el factor; (Hosmer, 1989).

[57] Al calcularse esta matriz, únicamente se está interesado en buscar la mejor combinación lineal de variables, esto es, en encontrar aquella combinación particular de las variables originales que cuenta con el mayor porcentaje de varianza de los datos. Por lo tanto, el orden de aparición de los factores es equivalente a la mejor combinación lineal de variables. El primer factor es mejor combinación lineal que el segundo factor, debido a que explica una mayor proporción de la varianza. El segundo a su vez, es mejor combinación que el tercer factor, y así sucesivamente.

[58] Las **soluciones factoriales** no rotadas por sí mismas alcanzan el objetivo de reducción de datos, pero se debe profundizar sobre que la solución ofrecida por el análisis de factorial, si es la manera más fácil de ofrecer una interpretación a los factores.

[59] El criterio de rotación VARIMAX se centra en simplificar las columnas de la matriz de los factores, alcanzando la máxima simplificación posible cuando hay ceros y unos en una columna. Eso significa que, maximiza la suma de las varianzas de las cargas requeridas de la matriz de los factores. Tiende a haber altas cargas de factores (cercanos a 1 o -1) y algunas cargas próximas a cero en cada columna de la matriz. Una carga alta indica asociación entre la variable y el factor, y una carga cercana a cero indica falta de asociación entre la variable y el factor. Este método es una mejor aproximación para lograr una rotación ortogonal de factores, (Hosmer, 1989).

(b) Los **criterios para la significación de cargas factoriales** son: (I) asegurar la significación práctica, (II) valoración de la significación estadística y (III) ajustes basados en el número de variables.

(c**)** La **Interpretación de la matriz de factores**: conlleva las siguientes partes, (i) examen de la matriz de cargas factoriales, (ii) identificación de la mayor carga para cada variable, (iii) valoración de la comunalidad y (iv) etiquetación de los factores.

Para realizar (i), en la matriz de cargas de factores se identifica que cada columna de números en la matriz de factores representa un factor aislado. Las entradas de cada vector columna son las cargas factoriales de cada variable sobre cada factor. En (ii) debe de identificarse la mayor carga para cada variable sobre cada factor. Una vez identificado debe de probarse si es significativa (mayor a ½), y luego analizar la segunda variable comprobando la carga para cada variable sobre cada factor y probar su significancia. Continuar este procedimiento hasta terminar con todas las variables analizadas. En (iii), valoración de la comunalidad. Validación del análisis factorial[60]. Este paso es para validar el análisis realizado en los pasos anteriores.

6. **Usos adicionales de los resultados del ACP**[61]. En el último paso, las alternativas son: (I) examinar la matriz de factores y seleccionar la variable con mayor carga factorial como un representante de una dimensión factorial particular, o (II) reemplazar el conjunto original de variables por otro conjunto de variables totalmente nuevo, evidentemente, con menos variables, creado a partir de escalas aditivas o de la puntación de factores.

La **puntuación de factores** es la medida compuesta creada para cada observación sobre cada factor extraído en el ACP. Se usan las ponderaciones factoriales en relación con los valores de la variable original para calcular la puntuación de cada observación. Así la puntuación de los factores puede ser utilizada para representar el(los) factor(es) en los análisis posteriores. La puntuación de factores está estandarizada para tener una media de cero y una desviación estándar de uno.

Después del marco teórico anterior, en el sexto paso, por fin se obtienen un índice o índices de un conjunto de variables correlacionadas. El inciso (II) del sexto paso será de gran ayuda, a la hora de crear índices de las peguntas contenidas en la encuesta *Lo que dicen los pobres.*

[60] Lo más común es comprobarlo por medio de otra muestra, o al inicio dividir la muestra en dos, una para el análisis y otra para la validación. Lo importante del punto es observar la generalización de las variables por medio de los factores extraídos en los 5 pasos anteriores.

[61] (I) Selección de variables suplentes para representar al conjunto inicial de variables: para lograr esto, en cada factor, se busca la variable que tenga la mayor carga factorial y se elige como representante del factor. En el caso que existan dos factores muy cercanos entre sí (numéricamente), se procede a elegir aquel variable que afecte más al factor según la experiencia del investigador.

(II) Creación de una escala aditiva: para realizarla, se combinan todas las variables que cargan alto sobre un factor y se promedian. El promedio será la nueva variable de sustitución.

3.2.3.3 Ejemplo de la creación de un índice mediante el ACP

En las siguientes páginas, se ejemplifica un caso, el ***Índice de Percepción de Impacto comunitario de los programas de combate a la pobreza,*** por medio de la técnica anteriormente detallada. Se hace uso del Análisis de Componentes Principales, para la batería de preguntas 7.12 casos (b), (c) y (f)[62].

6. 12 ¿Qué tan de acuerdo (totalmente de acuerdo, de acuerdo, en desacuerdo, totalmente en desacuerdo) está usted con las siguientes frases: Los programas de combate a la pobreza (leer cada frase):

(b) hacen a la gente dependiente del gobierno?
(c) se usan para fines electorales?
(f) crean conflictos entre las comunidades?

Aunque la batería de preguntas 7.12 esté compuesta de 6 incisos, se consideran únicamente tres de ellos (b, c y f). Al estudiar cada una de las preguntas por medio de la misma técnica de ACP, se observa que en conjunto (los seis) sólo se extrae un factor que explica el 40.65% de la varianza, mientras que al considerar únicamente tres incisos, se extrae un factor pero que explica el 56.65% de la varianza, es decir, un 15% más que cuando se consideraban todas las variables de la pregunta. Al analizar los seis incisos de 7.12, se puede observar que hay dos diferentes tipos de incisos, por un lado los que analizan los conflictos de la comunidad (incisos b, c y f), los que analizan las consecuencias del programa ante el gobierno y ante la solución del problema de pobreza (incisos a, d y e).

Obsérvese que las preguntas anteriores indagan en la opinión que tiene la gente de los problemas que podrían generarse desde que se entregan programas de combate a la pobreza[63]. En la primera pregunta, *"Los programas de combate a la pobreza hacen a la gente dependiente del gobierno*", se esperaría que las personas que son beneficiarias de algún programa social estén en contra de dicha frase, mientras las que no reciben beneficios si lo estén. En *"Los programas de combate a la pobreza se usan para fines electorales"* se esperaría nuevamente que los beneficiarios de éstos estén en contra de la frase, mientras que aquellos que no son beneficiarios estén a favor. Para la última pregunta, *"Los programas de combate a la pobreza crean conflictos en las comunidades"* se esperaría que los beneficiarios de éstos consideren que no generan problema alguno, mientras que los no beneficiarios consideran exactamente lo contrario, esto es, que se inclinen en favorecer la frase anterior. Por lo tanto, para las tres preguntas anteriores, los

[62] La batería completa de preguntas (7.12) es:
¿Qué tan de acuerdo (totalmente de acuerdo, de acuerdo, en desacuerdo o totalmente en desacuerdo) está usted con las siguientes frases: Los programas de combate a la pobreza…?: (a) Hacen a la gente dependiente del gobierno?; (b) ***Crean desigualdades entre la gente de la comunidad?; (c) Se usan para fines electorales?; (d)*** *Sólo disminuyen pero no solucionan el problema?; (e) Acostumbran a la gente a no trabajar lo suficiente?; (f)* ***Crean conflictos entre las comunidades?.***
Cada una tiene como posibles respuestas: Totalmente de acuerdo (1); De acuerdo (2); En desacuerdo (3); Totalmente en desacuerdo (4).
[63] Recuerde que los programas de combate a la pobreza mencionados en el cuestionario individual de *Lo que dicen los pobres* son seis: (a) DICONSA, (b) Empleo Temporal, (c) LICONSA, (d) Microrregiones, (e) Oportunidades Productivas y (f) Oportunidades-PROGRESA.

beneficiarios de alguno de los programas sociales deberían de estar en contra de cada una de las frases, mientras que las personas que no son beneficiarios deberían estar a favor.

Las distribuciones de las tres preguntas anteriores son presentadas diferentes tablas a continuación, (tablas 3.10, 3.11, 3.12).

En el caso (b) *"Los programas de pobreza de combate a la pobreza crean desigualdades entre la gente de la comunidad"*, (tabla 3.10), el 10.6% está totalmente de acuerdo, el 41% está de acuerdo, el 39.4% está en desacuerdo y el 9% está totalmente en descuerdo. Hay 154 casos faltantes (5.2%) de 2939 casos, éstos están conformados por las respuestas NS y NC.

Tabla 3.10

7.12 ¿ QUÉ TAN DE ACUERDO ESTÁ USTED CON LAS SIGUIENTES FRASES LOS PROGRAMAS DE COMBATE A LA POBREZA " CREAN DESIGUALDADES ENTRE LA GENTE DE LA COMUNIDAD " ?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Validos	**Totalmente de acuerdo**	**294**	10.0	**10.6**	10.6
	De acuerdo	**1142**	38.8	**41.0**	51.5
	En desacuerdo	**1099**	37.4	**39.4**	91.0
	Totalmente en desacuerdo	**251**	8.5	**9.0**	100.0
	Total	**2785**	94.8	**100.0**	
Faltantes	NS	147	5.0		
	NC	6	.2		
	Total	154	5.2		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

En el caso (c) "*Los programas de pobreza de combate a la pobreza se usan para fines electorales"*, (tabla 3.11), se tiene que el 12.7% está totalmente de acuerdo, el 37.1% está de acuerdo, el 38.1% está en desacuerdo y el 12% está totalmente en desacuerdo. Hay 235 casos (8%) faltantes de 2939 casos.

Tabla 3.11

7.12 ¿QUÉ TAN DE ACUERDO ESTÁ USTED CON LAS SIGUIENTES FRASES LOS PROGRAMAS DE COMBATE A LA POBREZA " SE USAN PARA FINES ELECTORALES " ?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Validos	**Totalmente de acuerdo**	**344**	11.7	**12.7**	12.7
	De acuerdo	**1004**	34.2	**37.1**	49.9
	En desacuerdo	**1031**	35.1	**38.1**	88.0
	Totalmente en desacuerdo	**324**	11.0	**12.0**	100.0
	Total	**2704**	92.0	**100.0**	
Faltantes	NS	229	7.8		
	NC	6	.2		
	Total	235	8.0		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

En el caso (f) "*Los programas de pobreza de combate a la pobreza crean conflictos entre las comunidades*", (tabla 3.12), se tiene que el 11.9% está totalmente de acuerdo, el 38.1% está de acuerdo, el 38.3% está en desacuerdo y 11.7% está totalmente en desacuerdo. Hay 206 (7%) valores perdidos de 2939 casos.

Tabla 3.12

7.12 ¿QUÉ TAN DE ACUERDO ESTÁ USTED CON LAS SIGUIENTES FRASES LOS PROGRAMAS DE COMBATE A LA POBREZA " CREAN CONFLICTOS EN LAS COMUNIDADES"?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Validos	**Totalmente de acuerdo**	**326**	11.1	**11.9**	11.9
	De acuerdo	**1040**	35.4	**38.1**	50.0
	En desacuerdo	**1048**	35.7	**38.3**	88.3
	Totalmente en desacuerdo	**319**	10.9	**11.7**	100.0
	Total	**2733**	93.0	**100.0**	
Faltantes	NS	197	6.7		
	NC	9	.3		
	Total	206	7.0		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

3.2.3.4 Índice de Percepción de Impacto por medio de puntaciones factoriales

Consideremos las tres preguntas anteriores en la creación de un índice mediante la técnica estadística de Análisis de Componentes Principales. Observe que cada una de las variables es no métrica, lo cual resulta frustrante en este punto puesto que el ACP es utilizado sobre variables de tipo métrico. Sin embargo, al estudiar detalladamente cada una de las variables se concluye lo siguiente:

- Las respuestas válidas son: Totalmente de acuerdo (1), de acuerdo (2), en desacuerdo (3) y totalmente en desacuerdo (4). Cada respuesta tiene asignado un valor entre la unidad y el cuatro, de manera tal que están espaciadas equitativamente, manteniendo como distancia una unidad.
- Por la manera en la que están diseñadas las preguntas (¿qué tan de acuerdo está usted con las siguientes frases: los programas de combate a la pobreza…) y también debido al tipo de respuesta, se establece un orden de respuestas, teniendo como valor mínimo el *Totalmente de acuerdo* (1) y como valor máximo *Totalmente en desacuerdo* (4).

En consecuencia, podemos asumir el supuesto de que tales preguntas que corresponden a variables no métricas se podrían considerar como variables métricas. Ahora podríamos comenzar a estimar el Análisis de Componentes Principales.

El primer paso es la validación de los supuestos, para esto se tiene que calcular la matriz de correlaciones, la matriz anti-imagen de correlaciones y la condición de esfericidad de Bartlett, las cuales, evidentemente, son para variables métricas.

La matriz de correlaciones es de 3x3 y tiene todas las entradas positivas, variando las correlaciones bivariadas entre 0.295 y 1. Los p-valores de las pruebas asociadas de correlación valen 0.000, por lo cual se rechaza la hipótesis nula de correlación igual a cero, $\rho(X, Y) = 0$. Y por lo tanto, las correlaciones bivariadas entre las tres variables anteriores son estadísticamente significativas.

La prueba de esfericidad de Bartlett vale 1160.53 para una χ_3^2, con tres grados de libertad. Esto genera un p-valor de 0.000, por lo cual se rechaza la hipótesis nula de igualdad de matrices entre la de matriz de correlaciones y la identidad. En consecuencia, la matriz de correlaciones es estadísticamente diferente a una matriz identidad de 3x3.

La Prueba de Adecuación Muestral de Kaiser-Meyer-Olkin vale 0.612, el cual es mayor a la cota inferior pedida de ½. Al revisar la matriz anti-imagen de correlaciones, en las entradas de la diagonal se tienen los valores 0.584, 0.744 y 0.588, los cuales, nuevamente son mayores a ½.

Debido a los tres argumentos anteriores, es decir, a que se tienen variables estadísticamente correlacionadas, la matriz de correlaciones es estadísticamente diferente a una matriz identidad y a que la prueba de Asociación de K-M-O es mayor a ½, entonces se concluye que

se cumplen los supuestos en los que descansa la teoría del ACP[64], y en consecuencia, se puede continuar con el desarrollo de la técnica.

Al analizar la tabla 3.13, se observa que sólo se extrae una componente, a saber, el primero de ellos, esto ocasionado al Eigenvalor que tienen asociado de 1.766 que evidentemente es estrictamente mayor que 1. Con la extracción de la primer componente, se explica el 58.85% de la variabilidad de las tres variables iniciales. Suponga hipotéticamente que se extraen no una componente, sino las dos primeras, con ello se explicarían el 84.21% de la variabilidad; finalmente, si se extrajeran hipotéticamente las tres componentes se explicaría el 100% de la varianza, siendo este modelo el ideal; Pero esto no sucede así en este modelo, y nos quedamos únicamente con la primer componente, sí, aquella que explica el 58.85% de la varianza.

Tabla 3.13
Total de Varianza Explicada por el Modelo

	Eigenvalores iniciales			Extracción de la suma de las cargas al cuadrado.		
Componente	Total	% de Varianza	% Acumulado	Total	% de Varianza	% acumulado
1	**1.766**	**58.858**	**58.858**	**1.766**	**58.858**	**58.858**
2	.761	25.360	84.219			
3	.473	15.781	100.000			

Método de extracción: Análisis de Componentes Principales.
Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Por otro lado, al analizar las comunalidades en la tabla 3.14 se expresan la proporción con la que cada una de las variables apoya en la explicación de la varianza del modelo. Cuando se tiene la pregunta *"crean desigualdad entre la gente de la comunidad*" la extracción es **0.679,** cuando es el la pregunta es *"se usan para fines electorales"* la extracción es la menor de todas **(0.421)** y cuando es *"crean conflictos entre las comunidades*" la extracción es del **0.666**.

Tabla 3.14
Communalidades[65]

Los programas de combate a la pobreza …	Inicial	Extracción
… crean desigualdad entre la gente de la comunidad.	1.000	**.679**
… se usan para fines electorales.	1.000	**.421**
… crean conflictos entre las comunidades.	1.000	**.666**

Método de extracción: Análisis de Componentes Principales.
Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

[64] Por tratarse de variables no métricas, que bajo las observaciones hechas al inicio de la sección se consideran como variables métricas, no se verifica la normalidad de las variables. Esto únicamente altera la explicación de la variabilidad del modelo. Aún así, se está interesado en aplicar la técnica con el único fin de obtener un índice.
[65] El vector de extracción de las comunalidades sirve como vector de coeficientes de ponderación para las tres variables en juego. Es una manera más de obtener coeficientes de ponderación para la creación de un índice por medio del método de Rangos Sumados Ponderados.

Debido a que se extrae una sola componente, la técnica del ACP no puede rotar el sistema[66].

Finalmente el índice se crea a partir de las **puntuaciones factoriales (Scores)**. Debido a la extracción de una componente, el índice será exactamente igual a la puntuación factorial de esa componente. Si se hubieran extraído más componentes, se tendrían dos puntuaciones factoriales (una por cada factor), y en ese caso el índice sería igual a la suma de las puntuaciones.

Se observa que hay 372 casos faltantes de 2939 casos totales, por lo cual se pierde información de 372 casos que no cuentan con información para las preguntas consideradas. Todo el procedimiento del ACP condujo al índice[67] anterior, denominado **Índice de Percepción de Impacto (puntuaciones factoriales)**. Las puntuaciones del índice están correlacionadas y tienen una media igual a cero y una varianza igual a la unidad.

3.2.3.5 Índice de Percepción de Impacto (Dalenius-Hodges)[68]

También se puede transformar el índice anterior por medio de la *Técnica de Estratificación Óptima de Dalenius y Hodges,* (véase anexo 3)*,* de manera tal que se categorice según puntos de corte estratégicos. El índice obtenido es una variable nominal de dos categorías y es llamado **Índice de Percepción de Impacto (Dalenius-Hodges).**

Considere cuatro categorías. Los puntos de corte son -1.004957, -0.040696 y 0.927919. Las frecuencias del **Índice de Percepción de Impacto (Dalenius-Hodges), tabla 3.15,** son:

Tabla 3.15
Índice de Percepción de Impacto (Dalenius-Hodges)

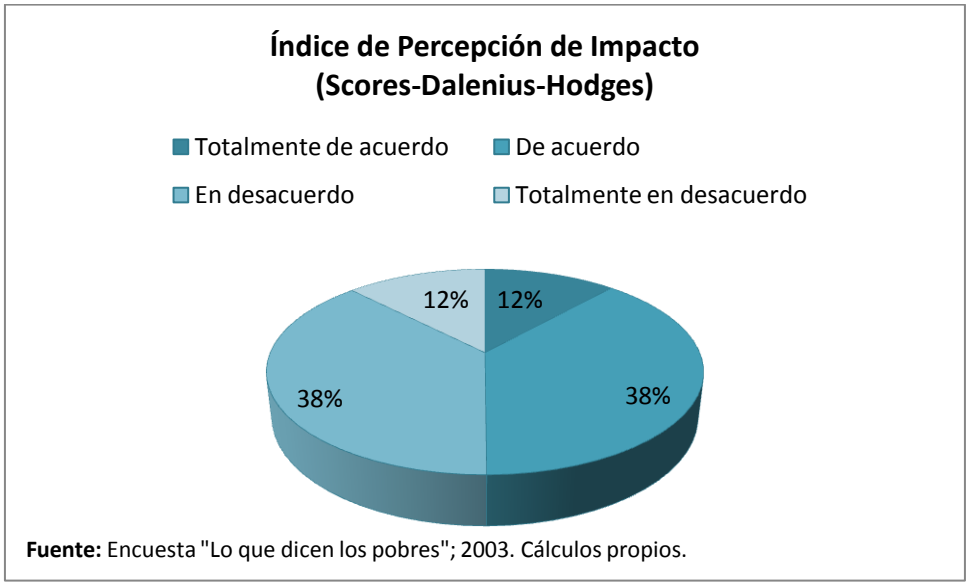
		Frecuencia	Porcentaje	**Porcentaje válido**	Porcentaje acumulado
Válidos	**Totalmente de acuerdo**	**301**	10.3	**11.7**	11.7
	De acuerdo	**980**	33.3	**38.2**	49.9
	En desacuerdo	**968**	32.9	**37.7**	87.6
	Totalmente en desacuerdo	**318**	10.8	**12.4**	100.0
	Total	2567	87.4	100.0	
Faltantes	Sistema	372	12.6		
Total		**2939**	**100.0**		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

[66] La rotación supone la existencia de al menos dos componentes. Recuerde que la rotación sirve para homogeneizar la participación de cada variable en las componentes extraídas.
[67] Este índice es obtenido por las puntaciones factoriales de la técnica de Análisis de Componentes Principales (ACP). Este índice es una variable métrica de razón.
[68] El respectivo índice se puede considerar como la siguiente pregunta: ¿***Qué hacer cuando se tiene un índice con una medida continua?, ¿Se puede construir algún índice con medida discreta a partir del índice de medida continua?***

Observe que en la gráfica 3.3, una de cada dos personas esta al menos de acuerdo con el **Índice de Percepción de Impacto**, mientras que el resto, uno de cada dos está desacuerdo o totalmente desacuerdo con el índice. En el contexto de éste, la mitad de las personas encuestadas se inclinan a favor de que los programas de combate a la pobreza hacen a la gente dependiente del gobierno, se usan para fines electorales y crean conflictos en las comunidades en los que se entregan; la mitad complementaria de estas personas mantiene una posición en contra a tales situaciones.

Gráfica 3.2



Fuente: Encuesta "Lo que dicen los pobres"; 2003. Cálculos propios.

A diferencia del primer índice [Percepción de Impacto (puntuaciones factoriales)] que tenía de una media continua, pues puede tomar todos los valores positivos. Éste [Índice de Percepción de Impacto (Dalenius-Hodges)] tiene una medida discreta, que tiene cuatro categorías: Totalmente de acuerdo, de acuerdo, en desacuerdo y totalmente desacuerdo.

3.2.3.6 Una aplicación del ACP en la construcción de índices en México

El Grado de Marginación (GM) es calculado por el Consejo Nacional de Población (CONAPO) cada 5 años; esta vez por medio de un conjunto de variables obtenidas a través del *II Conteo de Población y* Vivienda del INEGI, (INEGI, 2000). Los cálculos son hechos para diferentes

estratos, la sección se centra en los datos a nivel localidad[69]. Las primeras estimaciones de marginación datan de 1990.

Hablando de Metodología seguida por el CONAPO, (CONAPO C, 2005), las variables que forman parte del Grado de Marginación son: (i) Indicador de condición de alfabetismo, es decir, porcentaje de la población de 15 años o más analfabeta; (ii) Indicador de nivel educativo, refiriéndose al porcentaje de población de 15 años o más sin primaria completa; (iii) Indicador de viviendas particulares sin agua entubada; (iv) Indicador de viviendas particulares sin energía eléctrica (porcentaje); (v) Indicador de viviendas con algún nivel de hacinamiento; (vi) Indicador de viviendas particulares habitadas con piso de tierra; (vii) Indicador de viviendas particulares sin refrigerador; (viii) Indicador de viviendas particulares sin drenaje ni excusado. Los cuales se construyeron mediante el siguiente cociente:

$$I_{ij} = \frac{Número\ de\ casos\ favorables}{Número\ total\ de\ casos} x\ 100$$

Donde i es el número de indicador, i=1,2,…,8[70]

Donde j es el número de localidad, j=1,2,…,104 359.

También existe un indicador del tamaño de la población dado por una codificación en rangos la población que reside en las diversas localidades a lo largo y ancho del país.

Una vez que el CONAPO obtuvo los 9 indicadores anteriores, construyen un indicador con los objetivos de: *reducir la dimensionalidad original y al mismo tiempo retener y reflejar al máximo la información referida a la dispersión de los datos en cada uno de los nueve indicadores, así como las relaciones entre ellos*, y establecer una ordenación entre las unidades de observación (localidades). Por lo tanto, el CONAPO recurrió a la técnica de *Análisis de Componentes Principales*. Los cálculos que obtuvo el son los siguientes:

Las correlaciones[71] fueron estadísticamente significativas. Por otro lado, la medida de adecuación Keiser-Meyer-Olkin (KMO) vale 0.837. Lo que indica que sí se puede utilizar el ACP como técnica de creación de un índice. En la tabla 3.16, se presentan los resultados de la aplicación del ACP a los indicadores anteriores.

El CONAPO seleccionó únicamente la primera componente, debido a que el Eigen-valor asociado vale 3.89109 y con una explicación de la varianza de 48.63860%. Las outuaciones factoriales dan origen al **Índice de Marginación**.

[69] El cálculo de los indicadores de las variables anteriores se hace para 104,359 localidades del país en las que tiene sentido medir los indicadores básicos.

[70] Recuérdese que se tienen únicamente 8 indicadores socioeconómicos.

[71] La correlación lineal entre las variables tiene rangos diversos: el más grande es de 0.72 entre población analfabeta y sin primaria completa; el más chico es de 0.19 entre viviendas sin agua entubada en el ámbito de la vivienda y viviendas con algún nivel de hacinamiento.

Para obtener **el Grado de Marginación** de las localidades[72], el procedimiento lo dividieron en dos etapas, con 20 intervalos cada una, esto debido a que buscaron una estratificación óptima, en cinco grupos, que fuera consistente con los valores municipales y estatales, ordenando las 104 359 localidades en forma ascendente de acuerdo con el valor del *índice de marginación*. El procedimiento que utilizó para nivel municipio y entidad fue *la técnica de estratificación de óptimo de Dalenius-Hodges.*Por lo tanto, el Grado de Marginación por localidad está segmentado en cinco categorías, las cuales son: *Muy bajo, Bajo, Medio, Alto y Muy alto.*

Tabla 3.16

Valores propios de la matriz de correlaciones y porcentaje de varianza explicada a nivel localidad, 2005			
Componentes principales	**Valores propios (Eigen-valores)**		
	Total	**% de varianza**	**% de var acumulada**
1	3.89109	48.63868	48.63868
2	0.9485	11.85621	60.49489
3	0.85309	10.66357	71.15846
4	0.70642	8.83031	79.98877
5	0.64667	8.08336	88.07213
6	0.40384	5.04806	93.12019
7	0.28171	3.52137	96.64155
8	0.26868	3.35845	100

Fuente: Estimaciones del CONAPO con base en el II Conteo de Población y Vivienda 2005

Las distribuciones del **Grado de Marginación** para los años **2000 y 2005** se presentan en las tablas 3.17, 3.16 y 3.19.

Tabla 3.17
Grado de Marginación 2000

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Alto**	**49200**	45.9	**45.9**	45.9
	Bajo	**7030**	6.6	**6.6**	52.4
	Medio	**14825**	13.8	**13.8**	66.3
	Muy alto	**33896**	31.6	**31.6**	97.9
	Muy bajo	**2267**	2.1	**2.1**	100.0
	Total	**107218**	100.0	**100.0**	

Fuente: Grado de Marginación 2000, CONAPO.

[72] El CONAPO calcula el Grado de Marginación en tres niveles: Estatal, Municipal y por Localidad. Primeramente se realizaba a nivel estatal (32 entidades), más tarde fue a nivel Municipal y a últimas fechas, desde 1990, por localidad. La ventaja de contar con estimaciones en diferentes niveles es tener estimaciones más heterogéneas y reales en unidades cada vez más pequeñas. De esta manera, la varianza se estratifica y se distribuye mejor entre los niveles.

Tabla 3.18
Grado de Marginación 2005

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Alto**	**47239**	45.3	**45.3**	45.3
	Bajo	**10730**	10.3	**10.3**	55.5
	Medio	**13616**	13.0	**13.0**	68.6
	Muy alto	**27365**	26.2	**26.2**	94.8
	Muy bajo	**5409**	5.2	**5.2**	100.0
	Total	104359	100.0	100.0	

Fuente: Grado de Marginación 2005, CONAPO.

Hasta aquí, los cálculos y estimaciones fueron hechos por el CONAPO. Aprovechando de la existencia del *índice de marginación* y *del grado de marginación* para todas las localidades del país, procedo a enlazar los índices con las localidades en las que se levantó la encuesta *Lo que dicen los pobres*. Anexo la distribución del grado de marginación para la encuesta.

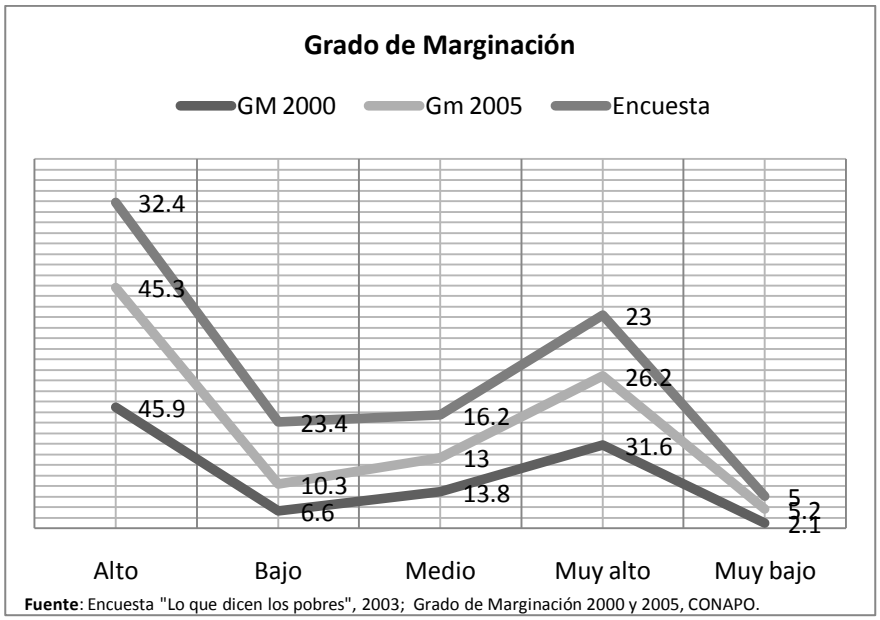
Tabla 3.19
Grado de Marginación 2005
Encuesta "Lo que dicen los pobres"

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Muy bajo**	**952**	32.4	**32.4**	32.4
	Bajo	**687**	23.4	**23.4**	55.8
	Medio	**477**	16.2	**16.2**	72.0
	Alto	**677**	23.0	**23.0**	95.0
	Muy alto	**147**	5.0	**5.0**	100.0
	Total	**2939**	100.0	**100.0**	

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos con base en los datos publicados por el CONAPO de Grado de Marginación 2005.

Se presenta la gráfica 3.3 con los porcentajes del Grado de Marginación para todas las localidades (GM 2000 y GM 2005) y para las localidades de *Lo que dicen los pobres.*

Gráfica 3.3



Fuente: Encuesta "Lo que dicen los pobres", 2003; Grado de Marginación 2000 y 2005, CONAPO.

Sin embargo, para los fines que persigue el presente trabajo, codifiqué dicha variable de la manera siguiente, (tabla 3.20):

Tabla 3.20

Grado de Marginación		
Variable anterior	Codificación	Nueva variable
Muy bajo	Valor: 1	Bajo
Bajo		
Medio	Valor: 2	Medio
Alto	Valor: 3	Alto
Muy alto		

Fuente: Grado de Marginación 2005, CONAPO.

Con lo cual, la distribución del Grado de Marginación para la encuesta queda de la siguiente manera, (tabla 3.21):

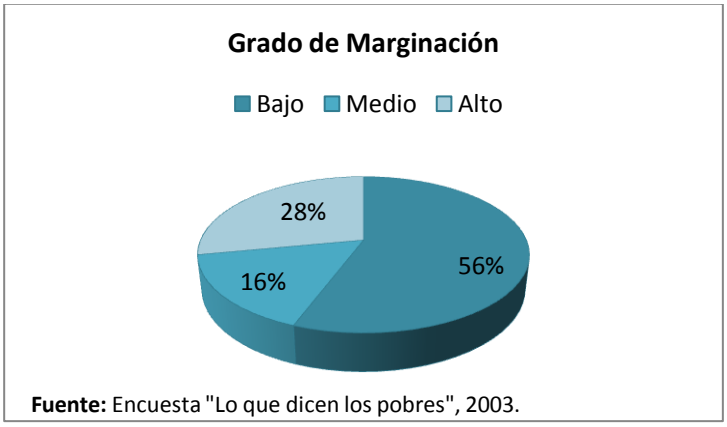
Tabla 3.21
Grado de Marginación

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Bajo**	**1639**	55.8	**55.8**	55.8
	Medio	**477**	16.2	**16.2**	72.0
	Alto	**823**	28.0	**28.0**	100.0
	Total	2939	100.0	100.0	

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios con base en el G.M. del CONAPO.

En la encuesta *Lo que dicen los pobres,* (gráfica 3.4), el 56% de las localidades tiene un Grado de Marginación Bajo, el 16% es de marginación media y el 28% de marginación alta. En este sentido, una localidad con un Grado de Marginación Alto significa ser una localidad pobre, mientras que un grado bajo hace referencia a una localidad no pobre.

Gráfica 3.4



Fuente: Encuesta "Lo que dicen los pobres", 2003.

3.3 Construcción de indicadores e índices de acuerdo al modelo para la evaluación del impacto

Recuérdese que el objetivo que persigue la presente tesis es describir las herramientas estadísticas y el desarrollo metodológico empleadas para medir el logro del proyecto general[73], en el cual fue inscrita ésta.

[73] El propósito del Proyecto General es analizar si los programas sociales inciden en la percepción que los beneficiarios y no beneficiarios tienen en las comunidades donde son implementados.

3.3.1 Otros índices basados en las técnicas estadísticas anteriores

La definición de **Percepción de Impacto** resulta necesaria para aterrizar las ideas del índice construido en la sección anterior *(índice de percepción de impacto por puntuaciones factoriales).*

La **Percepción de Impacto** es definida como: *la opinión de los ciudadanos sobre la utilidad e influencia en la realidad de los programas sociales dirigidos a ellos o implementados en las comunidades en donde viven, así como la política social en general.*

Considere nuevamente los *Indicadores de Percepción de Impacto*, los cuales son: (1) *"crean desigualdades entre la gente de la comunidad",* (2) *"se usan para fines electorales" y* (3) *"crean conflictos entre las comunidades"*).

Otro *Índice de Percepción del Impacto* se obtiene mediante el método de suma de rangos ponderados combinado con la estratificación de Dalenius-Hodges. Los coeficientes de ponderación están dados por los valores de las comunalidades obtenidos en el ejercicio respectivo de Análisis de Componentes Principales[74]. Por lo tanto, los coeficientes de ponderación son, (tabla 3.22):

Tabla 3.22
Ponderadores[75]

Pregunta	Variable	Coeficiente de ponderación
Indicador de Percepción de Impacto (crean desigualdad entre la gente de la comunidad)	X_b	**.679**
Indicador de Percepción de Impacto (se usan para fines electorales)	X_c	**.421**
Indicador de Percepción de Impacto (crean conflictos entre las comunidades)	X_f	**.666**

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

El Índice de Percepción de Impacto Ponderado (y) está dado por la siguiente expresión matemática:

$$y = 0.679x_b + 0.421x_c + 0.666x_f$$

[74] Véase la sección anterior, particularmente la construcción de los tres Indicadores de Percepción de Impacto y su Análisis de Componentes Principales asociado.
[75] La presente tabla de Ponderadores está en función de la tabla de comunalidades.

Una vez obtenido el índice anterior, se procede a hacer puntos de corte de tal manera que se tengan tres categorías, la técnica utilizada es la de Estratificación Óptima de Dalenius-Hodges, (tabla 3.23).

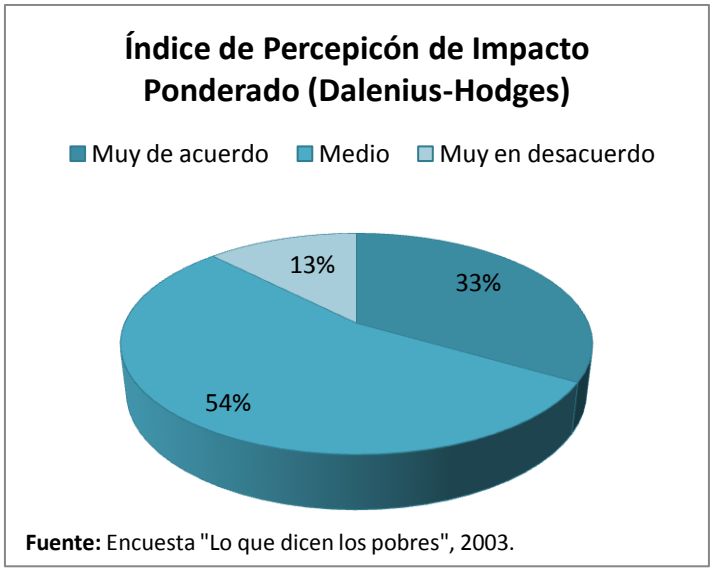
Tabla 3.23
Índice de Percepción de Impacto Ponderado (Dalenius-Hodges)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Muy de acuerdo**	**859**	29.2	**33.4**	33.4
	Medio	**1391**	47.3	**54.2**	87.6
	Muy en desacuerdo	**318**	10.8	**12.4**	100.0
	Total	**2567**	87.4	**100.0**	
Faltantes	**Sistema**	**372**	12.6		
	Total	2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Según el *Índice de Percepción de Impacto Ponderado de Dalenius-Hodges*, (gráfica 3.5), más de la mitad de personas encuestadas (54%) tienen una opinión en el índice de Muy de acuerdo, un tercio (33%) tienen una postura en desacuerdo y las demás personas (13%) mantienen una posición neutra dentro del índice (medio).

Gráfica 3.5



Fuente: Encuesta "Lo que dicen los pobres", 2003.

3.3.2 Índice de Confianza Institucional

A continuación se construye un índice de confianza utilizando el método de rangos sumados. Como **Confianza Institucional**[76] entenderemos una manera de confianza que no depende de la familiaridad interpersonal o un pasado común, sino que se basa en estructuras formales, producidas y legitimadas socialmente; es el tipo de confianza que generan las instituciones reguladoras, encargadas de sancionar a las personas u organizaciones que no cooperen o actúen irresponsablemente, de acuerdo a la definición de (Luna y Velasco, 2005, p. 131).

De esta manera la ***confianza institucional*** es entendida como una confianza basada en estructuras formales, siendo algunos ejemplos de ella: la ley como marco regulador y normativo de la sociedad, instituciones de gobierno en sus diferentes formas, Organizaciones no gubernamentales, etc. Los actores institucionales pueden ser el Presidente de la República, diputados y Cenadores, Autoridades municipales, estatales y federales.

Una vez, entendidas las definiciones anteriores de confianza, se procede a buscar preguntas en el cuestionario de *Lo que dicen los pobres* que al menos den una aproximación de la definición con la realidad.

Las preguntas 3.24[77] y 6.12[78] del cuestionario individual manejan en el fondo del la confianza institucional. A *priori* las preguntas anteriores dan origen a indicadores de confianza institucional, por lo que hay tres indicadores. Éstos se unen para formar el *Índice de Confianza Institucional.* Se presentan las distribuciones de las preguntas 3.24 y 6.12, (tablas 3.23, 3.24 y 3.25).

En la tabla 3.23, (pregunta 3.24), se observa que hay 75 valores perdidos, teniendo 2864 casos válidos.

[76] La confianza es entendida básicamente a la creencia por parte del actor o de un conjunto de actores que una persona, colectividad o institución realizarán diversas acciones (proporcionando información, recursos, normas y orientaciones), y por tanto se tendrá un ambiente propicio para generar participación, de acuerdo a la definición de (Angulo, b, p. 9).
En su versión más elemental, la **confianza** puede ser definida como: un conjunto de expectativas positivas sobre los demás o, más específicamente, sobre las acciones de los demás (Luna y Velasco, 2005, p. 129)

[77] *3.24 ¿Qué tan de acuerdo o en desacuerdo está usted con la siguiente afirmación, "Para que en este país haya justicia es necesario que la gente tome la ley en sus propias manos"?.* Con las respuestas: (a) muy de acuerdo, (b) de acuerdo, (c) en desacuerdo, (d) muy en desacuerdo, ni de acuerdo ni en desacuerdo, NS y NC.

[78] *6.12 ¿Cuál es la forma más efectiva para influir en lo que hace el gobierno de México?.* Con respuestas: (01) votar en elecciones, (02) escribir cartas a las autoridades, (03) hablar con los diputados, (04) hablar con el Presidente de la República, (05) tener amigos entre los funcionarios, (06) acudir a alguna organización, (07) presentarse en algún medio de comunicación, (08) dar dinero a los políticos, (09) protestar en manifestaciones públicas, (97) otra, NS (98) y NC (99).

Tabla 3.23
3.24 ¿QUÉ TAN DE ACUERDO O EN DESACUERDO ESTÁ CON LA SIGUIENTE AFIRMACIÓN " PARA QUE EN ESTE PAÍS HAYA JUSTICIA ES NECESARIO TOME LA LEY EN SUS PROPIAS MANOS?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Muy de acuerdo**	**163**	5.5	**5.7**	5.7
	De acuerdo	**617**	21.0	**21.6**	27.2
	En desacuerdo	**1167**	39.7	**40.7**	68.0
	Muy en desacuerdo	**755**	25.7	**26.4**	94.3
	Ni de acuerdo, ni en desacuerdo	**162**	5.5	**5.7**	100.0
	Total	**2864**	97.5	**100.0**	
Faltantes	NS	66	2.3		
	NC	8	.3		
	Total	75	2.5		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

En la ***primera mención de*** la pregunta 6.12 hay 244. (tabla 3.24), (8.3%) casos faltantes, por lo cual hay 2965 casos válidos. De forma tal, que no se tiene información de 244 respuestas a esta pregunta.

Tabla 3.24
6.12 EN SU OPINIÓN, ¿CUÁL ES LA FORMA MÁS EFECTIVA PARA INFLUIR EN LO QUE HACE EL GOBIERNO DE MÉXICO?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Votar en las elecciones**	**1259**	42.8	**46.7**	46.7
	Escribir cartas a autoridades	**293**	10.0	**10.9**	57.6
	Hablar con los diputados	**216**	7.4	**8.0**	65.6
	Hablar con el presidente de la republica	**334**	11.3	**12.4**	78.0
	Tener amigos entre los funcionarios	**105**	3.6	**3.9**	81.9
	Acudir a alguna organización	**131**	4.4	**4.8**	86.7
	Presentarse en algún medio de comunicación	**205**	7.0	**7.6**	94.3
	Dar dinero a los políticos	**29**	1.0	**1.1**	95.4
	Protestar en las manifestaciones publicas	**124**	4.2	**4.6**	100.0
	Total	**2695**	91.7	**100.0**	
Faltantes	Otra	14	.5		
	NS	211	7.2		
	NC	20	.7		
	Total	244	8.3		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

En la ***segunda mención*** de la pregunta 6.12, (tabla 3.25), hay 601 (20.4%) casos faltantes los cuales dejan 2338 casos válidos. Esto representa una de cada cinco respuestas como faltantes, mientras que 4 de cada cinco fueron contestadas. Por tal motivo, no se utiliza la segunda mención de la pregunta 6.12 en la construcción del indicador respectivo y en consecuencia no se utiliza en la construcción del índice.

Tabla 3.25
6.12 EN SU OPINIÓN, ¿CUÁL ES LA FORMA MÁS EFECTIVA PARA INFLUIR EN LO QUE HACE EL GOBIERNO DE MÉXICO?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Votar en las elecciones**	**329**	11.2	**14.1**	14.1
	Escribir cartas a autoridades	**330**	11.2	**14.1**	28.2
	Hablar con los diputados	**205**	7.0	**8.8**	37.0
	Hablar con el presidente de la republica	**385**	13.1	**16.5**	53.4
	Tener amigos entre los funcionarios	**180**	6.1	**7.7**	61.1
	Acudir a alguna organización	**335**	11.4	**14.3**	75.5
	Presentarse en algún medio de comunicación	**288**	9.8	**12.3**	87.8
	Dar dinero a los políticos	**70**	2.4	**3.0**	90.8
	Protestar en las manifestaciones publicas	**215**	7.3	**9.2**	100.0
	Total	**2338**	79.6	**100.0**	
Faltantes	0	595	20.3		
	Otra	1	.0		
	NS	4	.1		
	NC	0	.0		
	Total	601	20.4		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

El siguiente paso es construir el **Indicador de Confianza Institucional** y el **Indicador de Confianza Institucional (mecanismos de participación en el gobierno)**. Para esto se codifican las variables en función del evento que se desea medir, (tablas 2.36 y 3.27).

- **Indicador Confianza Institucional (Justicia)**

Tabla 3.26

3.24 ¿Qué tan de acuerdo o en desacuerdo está usted con la siguiente afirmación: 'Para que en este país haya justicia es necesario que la gente tome la ley en sus propias manos'?	
a) Muy de acuerdo	Valor: 0
b) De acuerdo	
c) En desacuerdo	Valor: 1
d) Muy en desacuerdo	

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

- **Indicador de Confianza Institucional (Mecanismos de participación en el gobierno)**

Tabla 3.27

6.12 ¿Cuál es la forma más efectiva para influir en lo que hace el gobierno de México?		
Votar en las elecciones	Valor: 1	Alta
Escribir cartas a autoridades		
Hablar con los diputados		
Hablar con el presidente de la República		
Tener amigos entre los funcionarios	Valor: 0	Baja
Presentarse en algún medio de comunicación		
Dar dinero a los políticos		
Protestar en las manifestaciones públicas		
Otra (esp.)	Valores perdidos	
NS		
NC		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

En cuanto al *Indicador de Confianza Institucional de Justicia de personas,* (tabla 3.28), mantiene una baja confianza, dejando dos de cada tres personas con una confianza alta. En este contexto, la baja confianza hace referencia a que la gente está de acuerdo o muy de acuerdo en hacer justicia por medio de tomar la ley en sus propias manos, mientras que tener una alta confianza significa exactamente lo contrario, es decir, creer en que la existencia de la justicia no necesariamente necesita de tomar la ley en sus manos.

Tabla 3.28
Indicador de Confianza Institucional (justicia)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Baja**	**942**	32.1	**32.9**	32.9
	Alta	**1922**	65.4	**67.1**	100.0
	Total	**2864**	97.5	**100.0**	
Faltantes	Sistema	75	2.5		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

En cuanto al *Indicador de Confianza Institucional de Mecanismos de Participación en el Gobierno*, (tabla 3.29), el 22% de la gente se mantiene con una baja confianza y el 78% con una alta confianza. Entendiéndose por baja confianza el no confiar en el gobierno, esto debido a que se cree que la mejor manera de influir en las acciones del gobierno por medio de tener amigos entre los funcionarios, presentarse en algún medio de comunicación, dar dinero a los políticos o protestar en manifestaciones públicas. En contraparte a la baja confianza, se encuentra la alta confianza entendida como confiar en el gobierno, mediante el voto en las elecciones, escribir cartas a las autoridades, hablar con diputados y hablar con el Presidente de la República.

Que en otras palabras significa que la gente que tiene alta confianza en el gobierno cree en la influencia para con el gobierno mediante mecanismos institucionales ya establecidos, a decir verdad, ellos creen que el camino de la pluralidad, legalidad y transparencia. La baja confianza se refiere a la antítesis del argumento anterior.

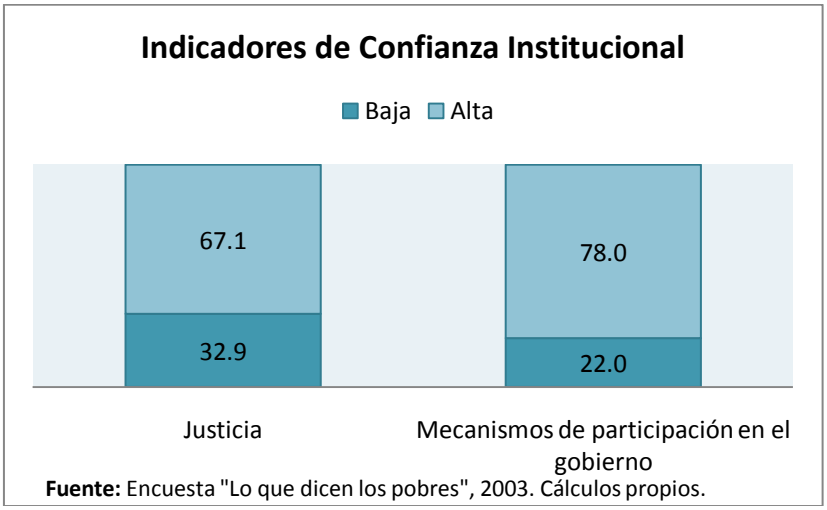
Tabla 3.29
Indicador de Confianza Institucional
(Mecanismos de participación en el gobierno)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Baja**	**593**	20.2	**22.0**	22.0
	Alta	**2101**	71.5	**78.0**	100.0
	Total	**2695**	91.7	**100.0**	
Faltantes	Sistema	244	8.3		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

En el *Indicador de Confianza Institucional (justicia)* hay 75 valores perdidos, mientras que en el *Indicador de Confianza Institucional (mecanismos de participación en el gobierno)* hay 244 valores perdidos, [gráfica 3.6]. Recuerde que estos valores perdidos son por falta de respuestas y por lo tanto, no se pueden adjudicar a la construcción de los indicadores.

Gráfica 3.6



Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Ya obtenidos los indicadores de confianza, el siguiente paso es combinarlos de tal manera que formen un índice. La manera en la que se agrupan es por el método de *rangos sumados*. Entonces el índice estará dado por la expresión:

$$Y = X_J + X_{MPG}$$

Donde

Y = Índice de Confianza Institucional

X_J = Indicador de Confianza Institucional (justicia)

X_{MPG} = Indicador de Confianza Institucional (mecanismos de participación en el gobierno).

Debido a que los indicadores anteriores tienen una imagen en el conjunto {0,1}, el índice tiene una imagen en el conjunto {0,1,2}, tomando como valor mínimo el 0 (cuando ambos tienen confianza baja) y como máximo el valor 2 (ambos tienen confianza alta).En el caso que tome el valor 1 significa que sólo uno de los dos indicadores tiene confianza alta y el otro indicador confianza baja.

Por lo tanto, el ***Índice de Confianza Institucional*** [79] , (tabla 3.30), tiene los siguientes porcentajes: baja confianza 7.8%, media confianza 38.8% y alta confianza 53.4%. En este sentido, la baja confianza está ligada a tomar la ley por sus propias manos e influir en las acciones del gobierno mediante mecanismos de participación no institucionales; la confianza media se refiere a tomar la ley en sus propias manos o elegir un mecanismos de participación no institucional; finalmente, la

[79] El Alfa de Cronbach de los indicadores de *Confianza Institucional para Justicia y Mecanismos de Participación* es de 0.92.

alta confianza considera no propio el tomar la ley en sus manos y creer que la mejor manera de influir en las acciones del gobierno es mediante la vía institucional del participación.

Tabla 3.30
Índice de Confianza Institucional

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Baja**	**208**	7.1	**7.8**	7.8
	Media	**1027**	34.9	**38.8**	46.6
	Alta	**1415**	48.1	**53.4**	100.0
	Total	**2649**	90.1	**100.0**	
Faltantes	Sistema	290	9.9		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Sin embargo, para trabajar con un índice con dos categorías se utiliza la técnica de estratificación óptima de Dalenius-Hodges, (véase anexo 3). El índice derivado de la técnica anterior tiene la siguiente distribución, (tabla 3.31):

Tabla 3.31
Índice de Confianza Institucional
(Dalenius-Hodges)

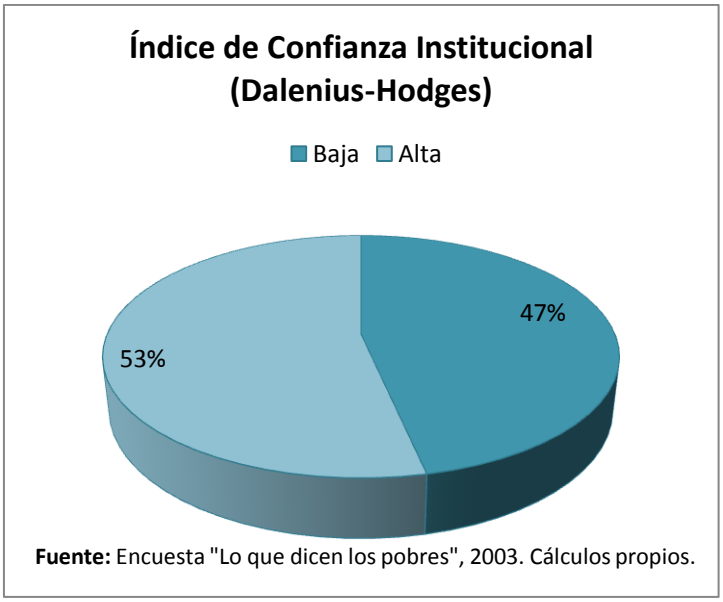
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Baja**	**1235**	42.0	**46.6**	46.6
	Alta	**1415**	48.1	**53.4**	100.0
	Total	**2649**	90.1	**100.0**	
Faltantes	Sistema	290	9.9		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

El cual mantiene los 290 casos no válidos, es decir el 9.9% se pierde mediante la construcción del índice.

El **Índice de Confianza Institucional de Dalenius-Hodges,** (gráfica 3.7), tiene la distribución de: baja confianza 46.6% y alta confianza de 53.4%. La baja confianza hace referencia a tomar la ley en sus propias manos y elegir mecanismos de participación no institucionales, mientras que la alta confianza considera no hacer justicia por su propia mano y elegir la participación por mecanismos institucionales.

Gráfica 3.7



Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

3.4 Algunos indicadores útiles

3.4.1 Indicador del Tipo de Localidad

En el Programa de Desarrollo Humano Oportunidades, la metodología de selección de familias beneficiarias tiene dos pasos: la selección de localidades y la selección de familias beneficiarias. Para la selección de localidades, tienen mayor probabilidad de selección las localidades que se encuentran en pobreza extrema. Sin embargo, generalmente las localidades que lo están, son aquellas que tienen una población pequeña. Por lo tanto, se incluye la variable *tipo de localidad,* la cual nos permite saber la condición de la localidad en a que se encuestó a la persona. El tamaño de localidad es obtenido por medio de la información que publicó el CONAPO en el año 2005, (CONAPO, 2005).

Cabe mencionar que de las localidades en donde se levantó la encuesta la de menor población es de 3 habitantes y la de mayor población es de1,820,888 habitantes.

Para el presente trabajo se consideran dos puntos de corte: 2,500 y 15,000 habitantes. Por lo tanto, se tienen tres intervalos de número de habitantes:

$$[0,2500)\ Localidades\ rurales$$

$$[2500,15000)\ localidades\ semirurales$$

$$[15000,\infty)\ localidades\ urbanas$$

La distribución de las localidades en las que se levantó la encuesta *Lo que dicen los pobres* es, (tabla 3.32 y gráfica 3.10):

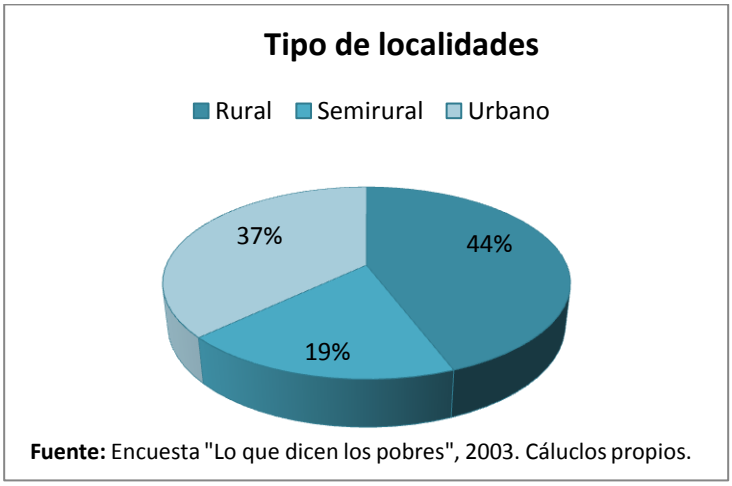
Tabla 3.32
Tipo de localidad

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**Rural**	**1297**	44.1	**44.1**	44.1
	Semi-rural	**555**	18.9	**18.9**	63.0
	Urbano	**1086**	37.0	**37.0**	100.0
	Total	2939	100.0	100.0	

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Por lo tanto, las localidades de *Lo que dicen los pobres,* (gráfica 3.8), tienen la siguiente distribución en cuanto al Grado de tipo de localidad (tamaño de localidad): localidades urbanas 37%, decir, con más de 15 mil habitantes; localidades semi-rural 19% con al menos 2500 habitantes pero menos de 15 mil habitantes; y localidades rurales 44% con menos de 2500 habitantes.

Gráfica 3.8



Fuente: Encuesta "Lo que dicen los pobres", 2003. Cáluclos propios.

3.4.2 Indicador de Beneficiarios de Oportunidades

En el cuestionario individual de *Lo que dicen los pobres* se pregunta (7.8[80]) por la condición de beneficencia de los distintos programas sociales que se manejan a lo largo de la encuesta.

Recuerde que objetivos del proyecto general en el que se inscribe el presente trabajo es evaluar si los programas sociales inciden en la percepción que los beneficiarios y no beneficiarios tienen en las comunidades donde son implementados, estos beneficiarios son el programa Oportunidades.

Por lo tanto, se construye un indicador que discrimine a las familias beneficiarias del Programa de Desarrollo Humano Oportunidades (PROGRESA) de las que no lo son. Es importante mencionar, que la pregunta que da origen al indicador está orienta a indagar en si al menos un miembro del hogar en el que vive el encuestado es beneficiario del Programa Oportunidades, no importando si el beneficiario es él u otra persona con la que comparta el hogar.

En la tabla 3.33 se resume la codificación del indicador de beneficiario de Oportunidades-PROGRESA.

Tabla 3.33

7.8 ¿Usted o algunos de los que viven en su hogar son beneficiarios de los programas sociales?						
Programa	Si		No	NS	NC	**Viejos valores**
	Espontánea	con ayuda				
f) Oportunidades Progresa	1	2	3	8	9	
Indicador de Beneficiario de Oportunidades-Progresa	1	1	0	Valores perdidos		**Nuevos valores**
	Si		No			

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

[80] La batería de preguntas es:
7.8 ¿Usted o algunos de los que viven en este hogar son beneficiarios de programas sociales? (Encuestador: registrar primero los programas que el entrevistado señale espontáneamente. Para los no mencionados lea los programas y si la respuesta es afirmativa registrar en "con ayuda"). Los programas son: (a) DICONSA, (b) Empleo temporal, (c) LICONSA, (d) Microrregiones, (e) Oportunidades Productivas, ***(f) Oportunidades-PROGRESA***.
Las respuestas son: Sí [espontánea] (1); Sí [con ayuda] (2); No (3); NS (8) y NC (9).

Cuya distribución se muestra en la tabla 3.34.

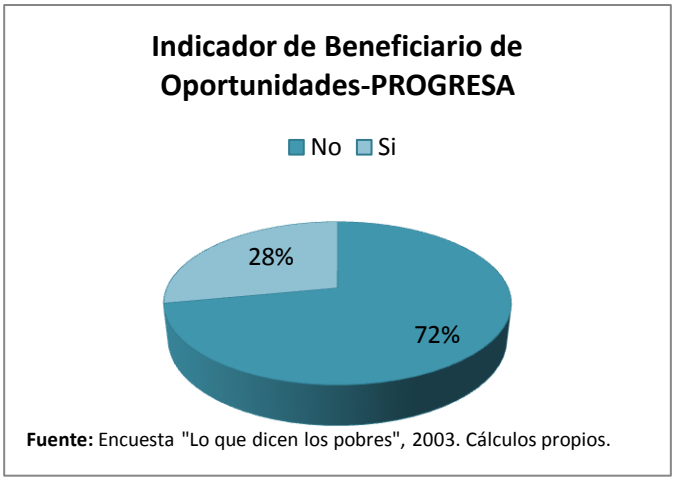
Tabla 3.34
Indicador de Beneficiario de OPORTUNIDADES/PROGRESA[81]

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	**No**	**2108**	71.7	**72.0**	72.0
	Si	**820**	27.9	**28.0**	100.0
	Total	**2927**	99.6	**100.0**	
Faltantes	Sistema	12	.4		
Total		2939	100.0		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

En *Lo que dicen los pobres*, (gráfica 3.9), el 72% de los encuestados pertenece a un hogar no beneficiario del Programa Oportunidades, mientras que el 28% sí recibe en su hogar el programa.

Gráfica 3.9



Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

[81] En la encuesta se pregunta por Oportunidades-PROGRESA debido a que en julio de 2003, fecha en la que se levantó *Lo que dicen los pobres*, el Programa de Desarrollo Humano Oportunidades era de reciente creación y el programa social que lo sustentaba era el Programa de Educación, Salud y Alimentación (PROGRESA).

4. Análisis Explicativo[82]

Essentially all models are wrong, but some are useful.
(Esencialmente todos los modelos son erróneos, pero algunos son útiles)
George E. P. Box[83] (1919-)

Finalmente se llega a la construcción de un modelo de Regresión Logística Multivariado basado en algunos de los índices e indicadores construidos en el capítulo anterior, los cuales son: *Índice de Percepción de Impacto, Índice de Confianza Institucional, Indicador de Grado de Marginación e Indicador de Tipo de Localidad.* La variable dependiente será el Indicador de Beneficiario de Oportunidades. Tal regresión logística multivariada está sustentada en modelos de regresión logística univariadas con la misma variable dependiente.
El modelo de regresión logística se utilizará en el capítulo para predecir las probabilidades de pertenencia a un grupo (ser beneficiario o no de Oportunidades) sujeto a un conjunto de variables independientes (índices en indicadores).

En este capítulo se introducen los modelos explicativos, con particular énfasis en la Regresión Logística Univariada y Múltiple, pues son éstas las que se ajustaran de acuerdo a los índices e indicadores construidos en el capítulo anterior. La primera sección es la introducción a los modelos explicativos más comunes; la segunda expone la teoría de la regresión logística de una variable; en la tercera ejemplifica la aplicación de la regresión a partir del ajuste de modelos univariados; la cuarta sección está dedicada al desarrollo de la teoría de la regresión logística múltiple; y para finalizar, en la última sección, se ajusta un modelo de regresión logística múltiple.

4.1 Modelos explicativos

Son llamados modelos explicativos a aquellas expresiones matemáticas que sirven para modelar el efecto que tiene un conjunto de variables sobre otra variable. Las variables que causan al modelo son las variables independientes, también son llamadas variables predictoras, puesto que su función dentro de los modelos es redecir otra variable. Mientras que la variable sobre la que causan efectos todas éstas es llamada variable dependiente, evidentemente, la variable dependiente mantiene una relación de dependencia con las variables independientes.

Ejemplos de modelos explicativos son la Regresión Lineal Simple, la Regresión Lineal Múltiple, el Análisis Discriminante, los Modelos de Probabilidad Lineal, y por supuesto, la Regresión

[82] Los modelos matemáticos explicativos más frecuentes son: Modelo de Regresión, Modelo Econométrico, Método de Encuestas de Intenciones de Compras y Modelo de Insumo-Producto.
[83] George E. P. Box es un estadístico, quien ha hecho importantes contribuciones en las áreas de Control de Calidad, Análisis de Series de Tiempo, Diseño de Experimentos e Inferencia Bayesiana.

Logística Univariada y Multivariada. Cada uno de los modelos anteriores tiene supuestos y usos diferentes. Su finalidad está en función del tipo de variables independientes que participan y la variable dependiente. Así, por ejemplo, la regresión lineal simple y múltiple son útiles cuando se tiene una variable dependiente de tipo métrico y una variable o un conjunto de variables independientes de cualquier tipo, ya sean categóricas o métricas. Mientras que los demás modelos necesitan una variable dependiente no métrica, con variables independientes de cualquier tipo.

En las Ciencias Sociales es común modelar las situaciones propias de tal ciencia mediante el Análisis de Discriminante (AD) y la Regresión Logística (RL), siendo de mayor uso ésta última. Este uso común está influenciado por la necesidad que tienen los investigadores sociales de discriminar entre diferentes grupos o eventos sujetos a algunas variables y con ello se estiman las probabilidades de ocurrencia de eventos. Aunque ambos modelos estadísticos necesitan una variable dependiente categórica (no métrica), existen diferencias que las hacen tener propósitos y aplicaciones diferentes.

El Análisis Discriminante supone un modelo lineal que tiene variable dependiente de tipo no métrico con al menos dos categorías que son conocidas y variables independientes métricas. Los supuestos asociados a la técnica son los mismos que tiene la regresión lineal, es decir, las variables individuales deben de cumplir normalidad, linealidad, homocedasticidad e independencia de los términos de error e igualdad de matrices de varianzas y covarianzas entre los grupos. Los objetivos primarios son entender las diferencias de los grupos y predecir la verosimilitud de que una entidad (persona u objeto) pertenezca a una clase o grupo particular basándose en varias variables métricas independientes. Es apropiada para contrastar la hipótesis de que las medias de los grupos de un conjunto de variables independientes para dos o más grupos son iguales, (Hair, 1999).

Por otro lado, la Regresión Logística (RL) tiene una variable dependiente dicotómica y las variables independientes son cualitativas o cuantitativas; cuando la variable es cualitativa, se construyen variables de diseño (dummy-variables) con el fin de incluirlas en el modelo. También es conocida como Análisis Logit. Es de uso muy común, incluso más que el AD debido a que la interpretación de los coeficientes es equivalente a la regresión lineal y tiene supuestos más fáciles de cumplir que el AD. Las variables independientes también son conocidas como covariables.

La RL predice directamente la probabilidad de ocurrencia de un suceso. El término de error tiene una distribución binomial, su varianza no es constante. Bajo algunas modificaciones, la RL puede tener una variable dependiente con más de dos categorías.

El Análisis Discriminante y la Regresión Logística cuentan con la capacidad para incorporar efectos no lineales y una amplia variedad de diagnósticos. Cuando el AD presenta una variable dicotómica como variable dependiente, la mayoría de las veces es utilizado en su lugar el modelo de RL, (Hair, 1999).

4.2 Modelo de Regresión Logística

En las siguientes secciones se describe la teoría del modelo de Regresión Logística Univariada y Multivariado. También se ajustan modelos de los dos tipos con base en los índices e indicadores del capítulo anterior.

Cabe destacar que la Regresión Logística tiene dos objetivos específicos, siendo estos:

- Conocer los pesos de ciertos factores o covariables.
- Estimar la probabilidad de pertenencia a un grupo.

Ambos objetivos, son caracterizados mediante la variable dependiente dicotómica, puesto que puede entenderse como la ocurrencia o no de un evento, o de manera análoga la pertenencia a uno de los dos grupos que la conforman.

Considérense las variables e índices del capítulo anterior a modo de realizar ejercicios de Regresión Logística Univariada (RLU). Al final del capítulo se presentará un modelo de Regresión Logística Múltiple (RLM) con las variables e índices que sean estadísticamente significativos.

Los modelos propuestos en este capítulo son un ejercicio práctico con el único fin de mostrar la aplicación de estos modelos.

Suponga que se lo que se desea es identificar los principales factores que determinan el ser o no ser beneficiario del Programa Oportunidades. Para ello se utilizan los siguientes indicadores e índices, (tabla 4.1):

Tabla 4.1

Tabla Resumen de los índices y variables para modelo de RLS y RLM			
Índice o Variable	**Tipo de variable**	**No. De categorías**	**Función de la variable en la RLU**
Índice de Percepción de Impacto (Dalenius-Hodges)	Nominal	2	**Covariable**
Índice de Confianza Institucional (Dalenius-Hodges)	Nominal	2	**Covariable**
Tipo de Localidad	Nominal	3	**Covariable**
Grado de Marginación	Nominal	3	**Covariable**
Condición de Beneficiario de Oportunidades	Nominal	2	**Variable dependiente**

Fuente: Encuesta "Lo que dicen los pobres", 2003.

A lo largo del capítulo se hace uso **del *Índice de Percepción de Impacto (Dalenius-Hodges)*** como covariable. De la misma manera es usado **el Índice *de Confianza Institucional (Dalenius-Hodges).***

El indicador de ***tipo de localidad*** está divido en tres categorías: rural, semi-rural y urbano. Se utiliza como covariable en los modelos propuestos. Regresando al capítulo anterior, (capítulo 3), las localidades rurales tienen de 1 a 2499 habitantes; las localidades semi-rurales están conformadas por 2500 a 14999 habitantes; finalmente, las localidades urbanas tienen 15 mil y más habitantes.

El **Grado de Marginación** también está dividido en tres categorías: alto, medio y bajo; es usado como posible variable independiente. La idea intuitiva del porque sí usar el *Grado de Marginación* está basada en el hecho que el Programa de Desarrollo Humano Oportunidades otorga beneficios a las personas que padecen pobreza de algún tipo, y generalmente, éstas personas se encuentran en localidades de alta marginación.

El indicador de la ***condición de beneficiario de Oportunidades*** es una variable nominal dicotómica, en la que se discrimina[84] si alguna de las personas que viven en el mismo hogar que la persona encuestada es beneficiario del Programa de Desarrollo Humano Oportunidades. Siguiendo el propósito[85] del Proyecto General de Investigación, se decide usar dicha variable como dependiente en los ejercicios de regresión logística.

4.3 Análisis bivariado

Una primera aproximación acerca del comportamiento de las variables independientes respecto de la variable dependiente (Condición de beneficiario de Oportunidades) es cruzarlas con la variable dependiente, permitiendo calcular el estadístico de la χ^2, el cual sirve para medir el grado de asociación entre dos variables no métricas[86]. Las hipótesis de la Prueba de Asociación son:

Ho: X no está asociada con Y vs **Ha:** X está asociada con Y

Se rechaza la hipótesis nula cuando se cumple la desigualdad $\alpha > p - valor$, para algún nivel α establecido. Para más información, véase el anexo técnico de la prueba χ^2.

[84] En este contexto, el discriminar se entiende como seleccionar excluyendo. Se desecha la otra definición: dar trato de inferioridad a una persona o colectividad por motivos raciales, religiosos, políticos, etc.
(RAE, 2011).

[85] El propósito de dicho Proyecto de Investigación es analizar si los programas sociales inciden en la percepción que los beneficiarios y no beneficiarios del Programa Oportunidades tienen en las comunidades donde son implementados. Véase la Introducción del presente trabajo.

[86] Véase anexo 4 de la Prueba de Asociación de la χ^2.

El nivel α [87]establecido para la prueba, es $\alpha = \frac{1}{4}$. Considerar un nivel α tan flexible permite no ser tan exigente en las pruebas, incluso logrando detectar asociaciones débiles entre las covariables y la variable dependiente. Recuerde que el análisis Bivariado de asociación únicamente tiene la finalidad de ser un análisis exploratorio de los datos.

Este método se utiliza como un análisis exploratorio de los datos y variables anterior a la estimación y ajuste del modelo de regresión logística. Rechazar la hipótesis nula significa que las variables están asociadas y con ello, al menos conceptualmente, deberían servir como variables independientes en el modelo causal.

La tabla 4.2, resume los resultados de las pruebas de asociación. Obsérvese con atención que en tres de los cuatro índices o variables el estadístico de prueba χ^2 es suficientemente grande como para rechazar la hipótesis nula. Además los respectivos p-valores son 0.000. Por lo tanto, para éstos se puede concluir:

- El ***Índice de Percepción de Impacto*** está asociado con la ***Condición de beneficiarios de Oportunidades*.**
- De la misma manera, el ***Grado de Marginación*** está asociado con la ***Condición de beneficiarios de Oportunidades.***
- El ***tipo de localidad*** está asociado con la ***Condición de beneficiarios de Oportunidades.***

Por otro lado, el *Índice de Confianza Institucional* tiene un estadístico de prueba χ^2 de 1.24, el cual es muy pequeño, generando un p-valor de 0.264. Por lo tanto, ni si quiera con el nivel $\alpha = \frac{1}{4}$ propuesto se logra hacer la variable estadísticamente significativa. En conclusión, la hipótesis nula de no asociación se rechaza con un 25% de significancia. Esto quiere decir, que el Índice de Confianza Institucional y la Condición de beneficiario de Oportunidades no están asociadas[88].

Finalmente, sabemos *a priori* que el *Índice de Percepción de Impacto,* el *Grado de Marginación y* el *tipo de localidad* sirven conceptualmente de variables independientes en los ejercicios que se propondrán más adelante. En el caso del *Índice de Confianza Institucional* se estudiará con especial cuidado debido a la independencia que tiene con la variable dependiente (Condición de beneficiario de Oportunidades).

[87] La estimación del estadístico χ^2 es un método exploratorio de variables que es altamente recomendado realizar previo al ajuste de modelos de Regresión Logística. Se aconseja ser flexible (no tan exigente) en el nivel α establecido para a prueba. Según algunos autores, $\alpha = \frac{1}{4}$. Véase (Hair, 2009).

[88] La misma conclusión se puede entender: dichas variables son independientes.

Tabla 4.2

Pruebas de asociación Variable Dependiente vs variables independientes			
Variable	Chi-cuadrada	Grados de Libertad	P-valor
Índice de Percepción de Impacto	28.61	1	0.000
Índice de Confianza Institucional	1.24	1	0.264
Grado de Marginación	262.67	2	0.000
Tipo de localidad	137.68	2	0.000

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

4.4 Modelo de Regresión Logística Univariado (RLU)

La presente sección es un resumen del libro de Hosmer, (Hosmer, 1989).

En cualquier método de modelos causales se pretende encontrar el "mejor ajuste", es decir, un modelo parsimonioso y razonable en términos económicos, sociales o biológicos.

La principal diferencia entre un modelo de regresión lineal y uno de regresión logística es el tipo de variable dependiente, en el primero tiene que ser una variable de razón mientras que en el segundo, tiene que ser una variable dicotómica.

En cualquier regresión la clave numérica está en el valor promedio de la variable dependiente. Esta cantidad es llamada la esperanza condicional y es expresada como:

$$E[Y|x]^{89}$$

Donde Y denota la variable resultado **(variable dependiente)** y x denota el valor de la **variable independiente**, también llamado **covariado**.

Con una variable dependiente dicotómica, se debe de cumplir las siguientes desigualdades:

$$0 \leq E[Y|x] \leq 1$$

Denótese la esperanza condicional por $\pi(x)$, es decir, $\pi(x) = E[Y|x]$. En el caso de una RLU, se tiene

[89] En la **regresión lineal simple** se tiene:

$$\boldsymbol{E[Y|x] = B_0 + B_1 x}$$

La cual es posible para el conjunto de los Números Reales es decir, $x \in \mathbb{R}$; (Kreyszing, 1973).

$$\pi(x) = \frac{e^{B_0 + B_1 x}}{1 + e^{B_0 + B_1 x}}$$

Que al obtener su función inversa, se convierte en la función $g(x)$, llamada Función Logit[90]:

$$g(x) = Ln\left[\frac{\pi(x)}{1 - \pi(x)}\right] = \beta_0 + \beta_1 x$$

En la regresión lineal se supone que una observación de la variable dependiente se puede expresar como $y = E[Y|x] + \varepsilon$. La cantidad ε[91], es llamada el error y se expresa como una desviación de la observación de la esperanza condicional. Pero en una regresión logística, se expresa el valor de la variable dependiente dado x, $y|x$ como $y = \pi(x) + \varepsilon$; donde ε tiene dos posibles valores:

- Si $y = 1$ entonces $\varepsilon = 1 - \pi(x)$ con probabilidad $\pi(x)$.
- Si $y = 0$ entonces $\varepsilon = -\pi(x)$ con probabilidad $1 - \pi(x)$.

Por lo tanto, ε tiene una distribución con Media cero, $E[x] = 0$, y varianza igual a $Var(x) = \pi(x)[1 - \pi(x)]$. Luego la distribución condicional de la variable dependiente tiene una distribución Binomial con probabilidad de éxito dada por la media condicional, $\pi(x)$. Esto es:

$$Y|x \sim Binomial[\pi(x)]$$

Ahora suponga que se tiene una muestra de n observaciones de parejas (x_i, y_i), para $i = 1,2, \dots, n$. Donde y_i denota el valor de la variable dicotómica para el i-ésimo sujeto. Además suponga que la variable dependiente y_i ha sido codificada en 0 y 1, representando la ausencia y la presencia de una característica, respectivamente.

El método de estimación para los parámetros, $\beta = (\beta_0, \beta_1)$ de la Regresión Logística, es el de **Máxima-verosimilitud**[92]. Parte de la **Función de Verosimilitud,** la cual expresa la probabilidad de que un conjunto de datos observados como una función de parámetros desconocidos. Los estimadores de Máxima-Verosimilitud de estos parámetros son escogidos de tal manera que maximizan esta función.

Si Y está codificada como cero o uno entonces,

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

[90] La función Logit, $g(x)$, es lineal en sus parámetros, puede ser continua o no, en un rango de $(-\infty, \infty)$.
[91] Generalmente, ε tiene una distribución Normal con Media cero y Varianza constante entre los niveles de la variable dependiente. Por lo tanto, la distribución condicional de la variable dependiente $y|x$ tiene una distribución que es Normal con Media igual a $E[Y|x]$ y Varianza constante.
[92] El método de Máxima-verosimilitud radica en establecer valores para los parámetros desconocidos los cuales maximizan la probabilidad de obtener el conjunto observado de datos.

La cual provee la probabilidad condicional de $Y = 1|x$, esto es $P[Y = 1|x]$, entonces es $P[Y = 0|x] = 1 - \pi(x)$. Por lo tanto, la contribución en la verosimilitud de aquellas parejas (x_i, y_i) donde $y_i = 1$ es $\pi(x_i)$, y de aquellas parejas en las que $y_i = 0$ es $1 - \pi(x_i)$, donde $\pi(x_i) = \pi(x)|_{x=x_i}$.

Debido a que $Y_i|x_i \sim Binomial[\pi(x_i)]$

La función de verosimilitud está dada por:

$$\varsigma(x_i) = \pi(x_i)^{Y_i}(1 - \pi(x_i))^{1-Y_i}$$

Debido al supuesto de independencia entre las observaciones, la función de verosimilitud es el producto de las funciones $\varsigma(x_i)$ para $i = 1,2, \dots, n$.

$$l(\beta) = \prod_{i=1}^{n} \varsigma(x_i)$$

Si en vez de trabajar con la función de verosimilitud $l(\beta)$, se trabaja con el logaritmo de ésta, se obtiene la función de Log-Verosimilitud:

$$L(\beta) = Ln(l(\beta)) = Ln[\prod_{i=1}^{n} \varsigma(x_i)] = \sum_{i=1}^{n} \{y_i Ln[\pi(x_i)] + (1 - y_i) Ln[1 - \pi(x_i)]\}$$

Para encontrar el valor de β que maximiza la función $L(\beta)$ con respecto a β_0 y β_1 y se iguala el resultado a cero. Las ecuaciones, que son llamadas **Ecuaciones de Verosimilitud**, son:

$$\sum_{i=1}^{n} [y_i - \pi(x_i)] = 0$$

y

$$\sum_{i=1}^{n} x_i [y_i - \pi(x_i)] = 0$$

Para una regresión logística las ecuaciones anteriores son no lineales en β_0 y β_1, luego requieren de un método especial para su estimación, este método es iterativo[93]. El vector de soluciones de las ecuaciones anteriores, es llamado el **vector de parámetros estimados**, y es denotado por $B = (B_0, B_1)$.

Note que en general, para una función $f(x_1, \dots, x_n)$, con $f: \mathbb{R}^n \to \mathbb{R}$, y f al menos de clase C^2, es decir, con derivadas parciales continuas. La **optimización** (maximizar, minimizar, etc.) se hace mediante:

[93]Este método ha sido programado en la mayoría de los paquetes de estadística. No es la excepción para SPSS.

(1) Calcular el Gradiente de f , $\boldsymbol{\nabla f(x_1,\dots,x_n)} = \left(\frac{\partial f}{\partial x_1},\dots,\frac{\partial f}{\partial x_n}\right).(\boldsymbol{x_1},\dots,\boldsymbol{x_n})$. Esto se logra al diferenciar f respecto de sus n derivadas.

(2) Igualar el Gradiente de f con el vector cero, esto es, $\nabla f(x_1,\dots,x_n) = 0$.

(3) Resolver el sistema de ecuaciones para cada una de las funciones diferenciadas e igualadas a cero, $f(x_i) = 0$, para $i = 1,\dots,n$. Las soluciones de las ecuaciones, $\boldsymbol{p} = (\boldsymbol{p_1},\dots,\boldsymbol{p_n})$ corresponden a los puntos críticos para la función f.

(4) Para determinar si un punto crítico, p, es un máximo o mínimo local se calcula el determinante de la Matriz Hessiana, esto es, $|\boldsymbol{H(x_1,\dots,x_n)}| = Det\left(\begin{bmatrix} \frac{\partial^2 f}{\partial {x_1}^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial {x_n}^2} \end{bmatrix}\right)|_{(x_1,\dots,x_n)=(p_1,\dots,p_n)}$. Si los Eigen-valores de $H(x_1,\dots,x_n)$ son estrictamente positivos, entonces $\boldsymbol{p} = (p_1,\dots,p_n)$ es un **Mínimo Local** de $f(x_1,\dots,x_n)$; Si los Eigen-valores de $H(x_1,\dots,x_n)$ son estrictamente negativos, entonces $\boldsymbol{p}$ es un **Máximo Local** de $f(x_1,\dots,x_n)$; Si los Eigen-valores de $H(x_1,\dots,x_n)$ son estrictamente tanto negativos como positivos, entonces $\boldsymbol{p}$ es un **Punto Silla** de $f(x_1,\dots,x_n)$; [Marsden, Tromba; 2004].

En el caso particular de la estimación de los coeficientes de una regresión logística univariada se tiene un coeficiente y una constante a estimar, por lo cual la matriz Hessiana $\boldsymbol{H(x_1,\dots,x_n)}$ es de dos por dos.

Por otro lado, un **modelo saturado** es aquel que contiene tantos parámetros como datos existentes. La comparación de los valores estimados usando la función de Verosimilitud está basada en la siguiente expresión:

$$D = -2Ln\left[\frac{(Verosimilitud\ del\ modelo\ nulo)}{(Verosimilitud\ del\ modelo\ saturado)}\right]$$

El cociente encerrado entre los corchetes de la expresión anterior[94] es llamado el **Cociente de Verosimilitudes** y sirve para probar el test con el mismo nombre. También se puede reescribir de la siguiente manera:

$$D = -2\sum_{i=1}^{n} y_i Ln\left(\frac{\hat{\pi}(x_i)}{y_i}\right) + (1-y_i)Ln\left(\frac{1-\hat{\pi}(x_i)}{1-y_i}\right)$$

El estadístico anterior, D, es llamado la ***Deviance***[95]. Para propósitos de asegurar la significancia de una variable independiente, se compara el estadístico D con y sin la variable independiente en la ecuación, mediante el estadístico G[96]:

[94] La razón de usar menos dos veces el Logaritmo Natural del Cociente de Verosimilitudes es matemática y es necesaria para obtener una cantidad cuya distribución es conocida y por lo tanto sirve para probar hipótesis.

[95] El estadístico Deviance (D) juega un papel similar a la suma de cuadrados en la regresión lineal.

[96] El estadístico G juega el mismo papel que el test del Numerador del Parcial de F en la regresión lineal.

$$G = D(Para\ el\ modelo\ nulo) - D(Para\ el\ modelo\ con\ la\ variable)$$

Debido a que la Verosimilitud del modelo saturado es común a ambos en los valores de D, se puede expresar como:

$$G = -2Ln\left[\frac{(Verosimilitud\ del\ modelo\ nulo)}{(Verosimilitud\ del\ modelo\ con\ la\ variable)}\right]$$

Para el caso de una sola variable independiente, la Máxima-Verosimilitud estimada para β_0 es $B_0 = Ln\left(\frac{n_1}{n_0}\right) = Ln\left[\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} 1-y_i}\right]$ y el valor ajustado es constante, $\frac{n_1}{n}$. El estadístico G es:

$$G = 2\left\{\sum_{i=1}^{n}[y_i Ln(\hat{\pi}(x_i) + \ (1-y_i)Ln(1-\hat{\pi}(x_i)] - [n_1 Ln(n_1) + \ n_0 Ln(n_0) - nLn(n)]\right\}$$

Bajo la hipótesis nula de que $B_1 = 0$, el estadístico G sigue una distribución Chi-cuadrada con un grado de libertad, $G \sim \chi_1^2$. Las hipótesis asociadas al estadístico anterior son:

Ho: $B_i = 0$ vs **Ha**: $B_i \neq 0$

El P-valor asociado a la prueba del Cociente de Verosimilitudes es $P[\chi_1^2 > G]$. Para i=0,1.

La **prueba de Wald**[97] es obtenida al comparar la Máxima Verosimilitud estimada de el parámetro de la pendiente, B_1, con el estimado del Error Estándar. El cociente resultante, bajo la hipótesis nula que $B_1 = 0$, sigue una distribución Normal Estándar.

$$W = \frac{B_1}{\widehat{SE(B_1)}}$$

El P-valor, de dos colas, asociado a la prueba de Wald es $P[|Z| > \frac{B_1}{\widehat{SE(B_1)}}]$.

El estadístico de prueba para el **Test del Score** es:

$$ST = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sqrt{\bar{y}(1-\bar{y})\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

El P-valor, de dos colas, asociado al Test del Score es $P[|Z| > ST]$ y tiene las mismas hipótesis de la Prueba de Wald y del Cociente de Verosimilitudes.

El estadístico D sirve para comparar entre dos modelos de RL, a saber el modelo nulo (aquel que no tiene covariables) y el modelo saturado (el que tiene todas las covariables), permitiendo saber si tiene sentido o no el ajuste del modelo de RL; el estadístico G, permite

[97] La prueba de Wald para el coeficiente B_1 es de gran importancia debido a que le da o no sentido a la Regresión Logística (univariada o multivariada). En caso de que la hipótesis nula no se rechace, se concluye que el coeficiente es igual a cero y por lo tanto, a lo más se podría establecer el modelo nulo. En caso contrario, cuando se rechaza la hipótesis nula, entonces el coeficiente es estadísticamente significativo y luego se puede seguir con los demás cálculos del modelo.

comparar entre dos modelos, los cuales son el modelo nulo y el modelo con una variable, probando así la permanencia de la variable; por último, la prueba de Wald sirve para probar la permanencia de una variable dentro del modelo, en función de su error estándar.

En un modelo de RL Univariado, si las pruebas indican que no es pertinente la inclusión de una variable, entonces la Regresión Logística no tendría sentido, y por lo tanto no se estima. En este caso, sería más recomendable lanzar una moneda al aire y escoger aleatoriamente una cara de ella, puesto que se aseguraría al menos una probabilidad de éxito de ½.

4.4.1 Estimación de los modelos univariados

Una vez expuesta la teoría de la Regresión Logística Univariada, procedemos a la estimación de los modelos con las variables propuestas al inicio del este capítulo. Es decir, lo que se desea es identificar los factores que influyen en la ocurrencia del evento de ser beneficiario del Programa Oportunidades (o no serlo).

Como se recordará, el Índice de Percepción de Impacto, el Tipo de localidad y el Grado de Marginación fueron estadísticamente significativos al cruzarlos con la variable dependiente, la condición de beneficiarios de Oportunidades. Por otro lado, el Índice de Confianza Institucional no lo fue, se tendrá especial cuidado en la estimación del modelo de regresión logística para éste.

4.4.1.1 Modelo para Índice de Percepción de Impacto

El modelo teórico que se busca es:

$$\pi(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}$$

Donde X es el Índice de Percepción de Impacto, β_1[98] es el coeficiente teórico (no estimado) del índice de percepción, y β_0 es la constante.

En la tabla 4.3, se observa que el análisis considera 2545 casos (86.6%) dejando 394 fuera (13.6%). La tabla de clasificación del modelo sin ajustar es:

[98] El modelo anterior es teórico, si fuera un modelo estimado, los coeficientes β_0 y β_1 serían estimados y se cambiarían por B_0 y B_1. La variable X se cambiaría por la observación x.

Tabla 4.3

Tabla de clasificación para: Índice de Percepción de Impacto (observaciones)				
	Condición de beneficiario Oportunidades			
Condición de beneficiario Oportunidades		No	Si	Porcentaje correcto
	No	1820	0	100
	Si	741	0	0
	Porcentaje sobre todos			71.1

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

El modelo nulo tiene la expresión:

$$g(x) = -0.899$$

La única variable que no está en la ecuación es el *Índice de Percepción de Impacto*. Recuerde que dicho índice tiene dos categorías: ***De acuerdo (1)*** y ***En desacuerdo (0)***, por lo que el índice se tiene que convertir en una variable de diseño[99] con la categoría *En desacuerdo* de referencia.

Al estimar el modelo con el único covariado[100] posible, el cálculo se detiene en la cuarta iteración debido a que el estimado del parámetro ha cambiado menos de 0.001. La tabla 4.4 indica que:

- $-2Ln(Verosimilitud) = 3111.21$
- La R^2 de Nagelkerke vale 0.016

El valor $D = -2Ln(Verosimilitud) = 3111.2$ es la Deviance y es establecido de esa manera para tener un número comparable con una distribución estadística conocida, en este caso, la Chi-cuadrada con un grado de libertad, χ_1^2. Para considerar únicamente la verosimilitud, se aplica su función inversa, es decir, $Verosimilitud = e^{-x/2}$.

Tabla 4.4
Resumen del modelo

Paso	-2 Log Verosimilitud	Nagelkerke R^2
1	3111.212(a)	.016

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

[99] Véase el anexo correspondiente de Categorías de Diseño.
[100] Las variables independientes también son conocidas como covariados.

Las variables en la ecuación se resumen en la tabla 4.5. Se observa que los parámetros estimados valen -0.670 y -0.468, para β_0 y β_1, respectivamente. La prueba de Wald arroja que lambas son estadísticamente significativas pues tiene un p-valor de 0.000. Lo cual hace que se rechace la hipótesis nula que establece $\beta_0 = 0\ y\ \beta_1 = 0$ y por lo tanto, existe evidencia para suponer que simultáneamente se cumple $\beta_0 \neq 0\ y\ \beta_1 \neq 0$, es decir, que los coeficientes estimados β_0 y β_1 no se anulan y con ello la Regresión Logística tiene sentido, por lo tanto, el modelo estimado es:

$$g(x) = B_0 + B_1 x = -0.670 - 0.468x$$

Tabla 4.5
Variables en la ecuación

		B	S.E.	Wald	GL	Sig.	Exp(B)	95.0% I.C. para EXP(B)	
								L.I.	L.S.
Paso 1(a)	**Índice de Percepción de Impacto**	**-.468**	.088	28.3	1	.000	.626	.527	.744
	Constante	**-.670**	.060	124.9	1	.000	.512		

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Por lo tanto, el modelo estimado de la regresión logística univariada para el Índice de Percepción de Impacto es:

$$\pi(x) = \frac{e^{B_0 + B_1 x}}{1 + e^{B_0 + B_1 x}} = \frac{e^{-0.670 - 0.468x}}{1 + e^{-0.670 - 0.468x}}$$

Consideremos el valor $x = 1$, entonces $\pi(1) = \frac{e^{-0.670 - 0.468 * 1}}{1 + e^{-0.670 - 0.468 * 1}} = 0.5503$, o lo que es equivalente, 0.2426 es la probabilidad condicional de que una persona **sea beneficiaria** del Programa Oportunidades y que tiene un **Índice de Percepción de Impacto catalogado como de acuerdo**, $(x = 1)$, es 0.5503.

Por otro lado, la probabilidad condicional de que la persona **sea beneficiaria** de Oportunidades dado que su **Índice de Percepción de Impacto es en desacuerdo** es exactamente el complemento de ésta, es decir, $1 - \pi(1) = 0.4496$.

Además, la probabilidad condicional de que la persona **no sea beneficiaria** de Oportunidades, dado que tiene **de acuerdo en su índice de percepción** es, $\pi(0) = \frac{e^{-0.670}}{1 + e^{-0.670}} = 0.6615$.

La contraparte de ésta probabilidad, esto es, $0.3384 = 1 - \pi(0)$ hace referencia a la probabilidad condicional de que la persona **no sea beneficiaria** dado que está **en desacuerdo en el índice de percepción.**

Las cuatro probabilidades anteriores, son los cuatro posibles casos que presenta un modelo de regresión logística univariado. La interpretación de tales probabilidades condicionales fue descrito anteriormente. La tabla 4.6 resume las probabilidades.

Tabla 4.6

(x,y)	π(x)
(0,0)	0.338
(0,1)	0.662
(1,0)	0.450
(1,1)	0.550

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Ahora al exponenciar el coeficiente $B_1 = -0.468$ se tiene, $e^{B_1} = 0.626$. El cual tiene un **Intervalo de Confianza** (IC) al 95% para e^{B_1} es [0.527,0.744], que por cierto, no incluye a la unidad.

Si el intervalo incluyera a la unidad, significaría que el IC para B_1 podría tomar el valor de cero, es decir, se anularía, y en consecuencia el modelo no tendría sentido.

El coeficiente exponenciado $e^{B_1} = 0.626$ es también conocido como función de ***Radio de Momios,*** la cual es una aproximación a que tan probable es la ocurrencia de un resultad cuando $\mathrm{x} = 1$ respecto de que no ocurra $\mathrm{x} = 0$. Cuando $e^{B_1} > 1$, entonces la probabilidad de ocurrencia de evento respecto la no ocurrencia es mayor; cuando $e^{B_1} = 1$ tienen la misma probabilidad de ocurrencia; cuando $e^{B_1} < 0$, entonces la probabilidad de ocurrencia del evento es menor que la probabilidad de no ocurrencia.

En este caso, e^{B_1} es menor que la unidad, por lo cual es menor la probabilidad de ocurrir (*De acuerdo*) respecto de no ocurrir (*en desacuerdo*) y el cociente es justamente 0.626.

Para los demás modelos univariados, se presenta la tabla 4.7 se resume las características de los demás modelos univariados[101].

[101] Los demás modelos univariados son regresiones logísticas para: el *Índice de Confianza Institucional, el Tipo de localidad y el Grado de Marginación.*

Tabla 4.7

Resumen de los modelos de Regresión Logística Univariados									
Modelo	Covariable	R^2 Nagerlkerke	B_o	B_1	$SE(B_1)$	Wald	% Correctamente clasificados		
							No Beneficiarios	Beneficiario	Total
1	**Índice de Percepción de Impacto**	0.016	-0.67	-0.47	0.88	28.34	100	0	71.1
2	**Índice de Confianza Institucional**	0	-1.01	0.10	0.06	1.23	100	0	72.4
3	**Tipo de localidad**	0.07	-1.80	-	-	130.69	100	0	72
	Tipo de localidad (01)	-	-	1.15	0.10	125.88	-	-	-
	Tipo de localidad (02)	-	-	1.01	0.12	66.09	-	-	-
4	**Grado de Marginación**	0.12	-0.05	-	-	245.77	100	0	72
	Grado de Marginación (01)	-	-	-1.49	0.10	245.09	-	-	-
	Grado de Marginación (02)	-	-	-0.89	0.12	51.85	-	-	-

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

La explicación de cada una de las regresiones logísticas de la tabla 4.7 se detalla en ñas siguientes páginas.

4.4.1.2 Regresión Logística Univariada para el Índice de Confianza Institucional

El modelo teórico que se busca es:

$$\pi(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$$

Donde X es el Índice de Confianza Generalizada, β_1 es el coeficiente teórico (no estimado) del índice de confianza, y β_0 es la constante.

La variable *Índice de Confianza Institucional* tiene dos categorías: *Baja (1) y Alta (0).* Debido a esta complicación se construye una variable de diseño que tiene como categoría de referencia el segundo estrato[102].

El modelo para el **Índice de Confianza Institucional** no es estadísticamente significativo, esto debido a que al calcular el estadístico de Wald tiene un valor demasiado pequeño, W= 1.23 con un p-valor de 0.264, lo cual hace que no se rechace la hipótesis que establece que la pendiente es igual a cero. Por lo tanto, el coeficiente estimado B_1 no es diferente de cero y por lo cual no tiene sentido que se establezca este modelo.

Desde que el *Índice de Confianza Institucional* se cruzó con la variable dependiente (*Condición de beneficiarios de Oportunidades*), se notó que no existía asociación entre las variables[103].

Observe que la R^2 de Nagelkerke es 0, esto es, la proporción de la variabilidad explicada bajo este modelo es 0. Esta conclusión no contradice a la no significancia de la pendiente del modelo.

Por otro lado, al calcular e^{B_1} se tiene 1.102, el cual es muy cercano a $1 = e^0$. El Intervalo de confianza para e^{B_1} al 95% es [0.929,1.307] que contiene al 1, esto es, que el coeficiente B_1 podría ser cero[104]. Nuevamente se llega a la misma conclusión que con la prueba de Wald.

4.4.1.3 Regresión Logística Univariada para el Tipo de localidad

La variable ***Tipo de localidad*** tiene tres categorías: *Rural, Semirural, y urbano*; por lo cual se crean dos variables de llamadas *Tipo de localidad (01) y (02)*[105], siendo la categoría de referencia el tipo urbano. Por lo tanto se tienen estimaciones para cada una de las variables de diseño.

El modelo teórico que se busca es:

$$\pi(X) = \frac{e^{\beta_0+\beta_{01}X_{01}+\beta_{02}X_{02}}}{1+e^{\beta_0+\beta_{01}X_{01}+\beta_{02}X_{02}}}$$

Donde X_{01} es el *Indicador de localidad cuando es rural* con un coeficiente β_{01} ; X_{02} es el *Indicador de localidad cuando es semi-rural* y tiene un coeficiente de β_{02}. β_0 es la constante.

La prueba de Wald simultánea para ambos coeficientes tiene un valor de 130.68 con un p-valor de 0.000, por lo tanto se rechaza la hipótesis nula que establece que ambos coeficientes son

[102] La variable de diseño para el Índice de Confianza Institucional es: (1) primer estrato y (0) segundo estrato.
[103] Véase el inicio del Capítulo V para más detalles.
[104] Que el Intervalo de Confianza para e^{B_i} contenga a la unidad, significa que podría pasar $e^{B_i} = 1$ y por lo tanto $B_i = 0$.
[105] Las variables de diseño para el tipo de localidad son:
Tipo de localidad (01) que tiene codificada: (1) rural, (0) semirural y urbano.
Tipo de localidad (02) que tiene codificada: (1) semirural, (0) rural y urbano.

iguales a cero, concluyendo que al menos uno de los coeficientes es estadísticamente diferente de cero. Se revisan cada una de las pendientes o coeficientes para saber cual es estadísticamente diferente de cero. Los coeficientes de las variables de diseño del índice son 1.15 y 1.01 para la primera y segunda variable de diseño, respectivamente; numéricamente son diferentes de cero y estadísticamente también lo son puesto que los estadísticos de Wald para cada una de ellas son 125.88 y 66.09 (con los p-valor de 0.000 para ambas), por lo que se continua rechazando la hipótesis nula. Luego no existe evidencia para suponer que los coeficientes de las variables de diseño son estadísticamente iguales a cero, permitiendo suponer ambas pendientes son distintas de cero (estadísticamente significativas).

Al exponenciar cada uno de los coeficientes se obtienen $e^{B_{10}}$ y $e^{B_{11}}$[106] valiendo 3.150 y 2.733, con intervalos de confianza al 95% son [2.578,3.849] y [2.144,3.482] para las variables de diseño 1 y 2 de índice de confianza, respectivamente. Ambos intervalos no contienen al 1; de tal manera que se corrobora una vez más, de manera indirecta, el resultado obtenido por la prueba de Wald.

El coeficiente del interceptor (B_0, también conocido como la ordenada al origen) vale -1.707 y también es estadísticamente significativa[107].

La R^2 de Nagelkerke vale 0.07, por lo que con el modelo propuesto se explica la variabilidad al 7%.

La expresión matemática del modelo es:

$$g(x) = \beta_0 + \beta_{10}x + \beta_{11}x = -1.707 + 1.005x_{10} + 1.147x_{11}$$

Por lo tanto, el modelo ajustado para el tipo de localidad es:

$$\pi(x) = \frac{e^{-1.707+1.005x_{10}+1.147x_{11}}}{1 + e^{-1.707+1.005x_{10}+1.147x_{11}}}$$

4.4.1.4 Regresión Logística Univariada para el Grado de Marginación

El *Grado de Marginación* calculado por el CONAPO tiene cinco categorías, sin embargo, debido a los fines que persigue este trabajo dentro del Proyecto General de investigación en el que está inscrito, se decidió trabajar con un Grado de marginación con tres categorías dadas por *Alto, Bajo y Medio*[108]. Se construyen dos variables de diseño respecto de la categoría de referencia Alto. Las nuevas variables son denominadas *Grado de Marginación (01) y (02)*[109].

[106] Los coeficientes para la primera y segunda variables de diseño son B_{10} y B_{11}, respectivamente.

[107] Hasta este momento se ha utilizado demasiadas veces el concepto de "***estadísticamente significativo***". Si el lector no lo recuerda o no lo ha descubierto, significa que mediante una prueba de hipótesis se establece que la hipótesis nula H_0 se rechaza, de tal manera que no existe evidencia para suponer ésta y entonces se supone válida H_a.

[108] Para más detalles véase la construcción de la variable Grado de Marginación en la última sección del capítulo anterior (Capítulo IV).

[109] Las variables de diseño para el Grado de Marginación son:
Grado de Marginación (01) que tiene codificada: (1) bajo, (0) medio y alto.

El modelo teórico que se busca es:

$$\pi(X) = \frac{e^{\beta_0+\beta_{01}X_{01}+\beta_{02}X_{02}}}{1+e^{\beta_0+\beta_{01}X_{01}+\beta_{02}X_{02}}}$$

Donde X_{01} es el *grado de marginación bajo*, con un coeficiente β_{01} ; X_{02} es el *grado de marginación medio* y tiene un coeficiente de β_{02}. β_0 es la constante.

La prueba de Wald simultánea para ambos coeficientes tiene un valor de 245.767 con un p-valor de 0.000, por lo tanto se rechaza la hipótesis nula (H_0) que establece que ambos coeficientes son iguales a cero, entonces al menos uno de los coeficientes es estadísticamente significativo. Para saber cuál de los dos coeficientes es diferente de cero, se analizan individualmente; los valores que tienen B_{10} y B_{11} son -1.493 y -0.891, con estadísticos de Wald de 245.09 y 51.849 (p-valor de 0.000 para ambos) para la primera y segunda variable de diseño, respectivamente. Por lo tanto, se rechazan las dos hipótesis nulas (H_0) y se puede suponer que los coeficientes B_{10} y B_{11} son estadísticamente diferentes de cero.

Al exponenciar estos coeficientes se obtienen los valores 0.225 y 0.410, con Intervalos de Confianza al 95% de [0.186,0.271] y [0.322,0.523]. Ninguno de los IC[110] contienen a la unidad, por lo que los coeficientes no pueden valer cero; esto indica que no se contradice el resultado de las pruebas de Wald.

La constante para el modelo vale -0.052 con un estadístico de Wald de 0.560 (p-valor de 0.454), indicando que la hipótesis nula no se rechaza; finalmente se concluye que la constante no es estadísticamente significativa.

La R^2 de Nagelkerke es 0.12; explicando con el modelo de el Grado de Marginación un 12% de la variabilidad.

La expresión matemática del modelo es:

$$g(x) = B_{10}x + B_{11}x = -1.493x_{10} - 0.891x_{11}$$

Escribiéndolo de manera diferentes, se convierte en

$$\pi(x) = \frac{e^{B_1x}}{1+e^{B_1x}} = \frac{e^{-1.493x_{10}-0.891x_{11}}}{1+e^{-1.493x_{10}-0.891x_{11}}}$$

Grado de Marginación (02) que tiene codificada: (1) medio, (0) bajo y alto.

[110] Intervalo de confianza, IC.

4.4.2 Conclusiones de los modelos

De los cuatro modelos univariados de Regresión Logística presentados anteriormente, únicamente tres fueron estadísticamente significativos, estos son: ***Modelo de Índice de Percepción de Impacto, Modelo del Tipo de localidad y Modelo del Grado de Marginación.*** El modelo que no fue significativo es el del ***Índice de Confianza Institucional*.**

La conclusión anterior se veía venir desde que se cruzaron dichas covariables con la condición de beneficiario de Oportunidades (variable dependiente de todos los modelos de RLU). En las pruebas de asociación de la Chi-cuadrada se establecían los mismos resultados. Por lo tanto, los métodos de análisis exploratorio de las variables (en este caso las tablas cruzadas) es de gran ayuda para evaluar las variables a considerar en el modelo antes de ajustar algún modelo de RLU.

En la última sección del presente capítulo se tratará de ajustar una Regresión Logística Multivariada con las covariables significativas de la sección presente. Con esto, se estará proponiendo el modelo del que se habló al inicio del trabajo.

4.5 Modelo de Regresión Logística Multivariado (RLM)

Hosmer y Lemeshow exponen los modelos de regresión logística múltiples de la siguiente manera, (Hosmer, 1989: II, III, IV, V y VI):

Sean $(x_1, x_2, \dots, x_p)$ una muestra de variables independientes las cuales forman el vector $x = (x_1, x_2, \dots, x_p) \epsilon \mathbb{R}^p$. También suponga que estas variables son métricas, es decir, variables de escala de intervalo o de razón.

En la sección anterior se definió $P(Y = 1|x) = \pi(x)$, la *probabilidad condicional del resultado obtenido.*

La función logit del modelo de regresión logística múltiple o multivariado está dada por $g(x) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$, del cual se obtiene:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

Ahora suponga que la j-ésima variable es no métrica[111], x_j, con k categorías. Entonces se crean k_j-1 variables de diseño denotadas como D_{ju} y los coeficientes para estas variables de diseño son denotadas por β_{ju}, u = 1, 2, …, k_j-1. Por lo tanto, el modelo logit con p variables y la j-ésima variable no métrica está dado por:

$$g(x) = \beta_0 + \beta_1 X_1 + \cdots + \beta_{j-1} X_{j-1} + \sum_{u=1}^{k_j-1} \beta_{ju}\, x_{ju} + \beta_{j+1} X_{j+1} + \cdots + \beta_p x_p$$

Suponga que se tiene una muestra de n observaciones independientes de la pareja (x_i, y_i), i = 1,…,n. Se necesita obtener p+1 coeficientes estimados a partir de estas observaciones, $\beta' = (\beta_0, \beta_1, \ldots, \beta_p)$[112].

El método de estimación es el de **Máxima verosimilitud**. Habrá p+1 **ecuaciones de verosimilitud** las cuales son obtenidas al diferenciar cada una de las **funciones de Log Verosimilitud** con respecto a los p+1 coeficientes.

$$\sum_{i=1}^{n} \left(y_i - \pi(x_i)\right) = 0$$

y

$$\sum_{i=1}^{n} [x_{ij}\, (y_i - \pi(x_i)] = 0$$

Para j = 1, …, p.

Denote $\beta' = (\beta_0, \beta_1, \ldots, \beta_p)$ el vector solución de estas ecuaciones.

La estimación de las varianzas y covarianzas de los coeficientes estimados β' siguen el mismo método de máxima verosimilitud. Los estimadores son obtenidos de la matriz de segundas derivadas parciales de las funciones de log verosimilitud; son de la siguiente forma:

$$\frac{\partial^2 L(\beta)}{\partial \beta j^2} = -\sum_{i=1}^{n} x_{ij}^2\, \pi_i(1-\pi_i)$$

Y

$$\frac{\partial^2 L(\beta)}{\partial \beta j \partial \beta u} = -\sum_{i=1}^{n} x_{ij}\; x_{iu}\, \pi_i(1-\pi_i)$$

[111] Si algunas de las variables independientes son no métricas, es decir, las variables son nominales u ordinales, y son necesarias ser incluidas en el modelo, es necesario transformarlas en un conjunto de variables de diseño (dummy variables). Véase el anexo correspondiente.

[112] Los cuales corresponden a la ordenada al origen del hiperplano estimado y las pendientes de cada una de las p variables $x_1, \ldots, x_p$.

Para j, u= 0, 1, …, p. Donde π_i denota $\pi(x_i)$. La matriz de p+1 por p+1 que contiene los términos negativos de las ecuaciones anteriores es denotada como $I(\beta)$[113]. Las varianzas y covarianzas son estimadas al invertir esta matriz, la cual se denota $\Sigma(\beta) = I^{-1}(\beta)$ [114]. Los estimados de las varianzas y covarianzas son obtenidos al evaluar B_j (estimada) en $\sigma^2(\beta_j)$ y B_j , B_u (estimadas) en $\sigma^2(\beta_j, \beta_u)$ en las entradas de la matriz $\Sigma(\beta)$[115]. Esto para j,u = 1, …, p[116].

En el caso particular de la estimación de los coeficientes de una regresión logística univariada se tiene un coeficiente y una constante a estimar, por lo cual la matriz Hessiana $\boldsymbol{H}(\boldsymbol{x_1}, \dots, \boldsymbol{x_n})$ es de p+1 por p+1.

Para probar la significancia del modelo, la prueba del Cociente de **Máxima Verosimilitud** (*Log-likelihood Ratio Test*) es para probar la significancia de los p coeficientes del modelo (sin la ordenada al origen). Las hipótesis[117] son:

Ho: Todos los coeficientes (pendientes) son iguales a cero

Vs

Ha: Alguno de los coeficientes es diferente de cero

La hipótesis nula, en pocas palabras establece que el modelo de RLM no tiene sentido, puesto que todos los coeficientes son cero. En otras palabras, sería mejor dejarlo a la suerte, echando una moneda al aire y escogiendo aleatoriamente la cara de ésta. Por otro lado, la hipótesis nula establece la negación de Ho, es decir que ajustar un modelo de RLM sí tiene sentido, debido a que al menos uno de los coeficientes es diferente de cero.

El estadístico de prueba es:

$$G = -2\,\mathrm{Ln}\left(\frac{\text{Verosimilitud sin las variable s}}{\text{Verosimilitud con las variables}}\right)$$

$$= -2\,\mathrm{Ln}\left(\frac{\text{Verosimilitud del modelo nulo}}{\text{Verosimilitud del modelo bajo Ha}}\right)$$

[113] Se le conoce como la matriz de información.

[114] La varianza se denota como $\sigma^2(\beta_j)$ y corresponde al j-ésimo elemento de la diagonal de esta matriz, la cual es la varianza de B_j y $\sigma^2(\beta_j, \beta_u)$ es cualquier elemento que no está en la diagonal de la matriz y denota la covarianza entre los elementos B_j y B_u. $\sigma^2(\beta_j)$ y $\sigma^2(\beta_j, \beta_u)$ son las varianza y covarianza estimadas de los coeficientes.

[115] Es decir, se tiene $\sigma^2(B_j)$ y $\sigma^2(B_j, B_u)$ en el arreglo cuadrangular $\Sigma(B)$.

[116] En consecuencia, la desviación estándar de los coeficientes estimados está dada por:

$$SE(B_j) = \sqrt{\sigma^2(B_j)}$$

Para j=0,1, … ,p.

[117] Obsérvese, que en la hipótesis alternativa, el que alguno sea diferente de cero, puede incluir que todos sean diferentes de cero. Se utiliza ese enunciado por ser la negación de la hipótesis nula. También se puede escribir las hipótesis de la manera siguiente: **Ho:** $\beta_j = 0$ para todo j = 1,…,p vs **Ha:** $\beta_j \neq 0$ para algún j = 1,…,p

El cual tiene una distribución Chi-cuadrada con p grados de libertad y una significación α, es decir, $G \sim \chi^2_p$. Si $G > \chi^2_{(\alpha)p}$, se rechaza la hipótesis nula H_o.

Otra prueba es mediante el **Estadístico de Wald (univariado)[118].**Considerando los estimados para el coeficiente j y para su desviación estándar. El estadístico tiene una distribución Normal Estándar. Se rechaza Ho si $W_j > Z_\alpha$.

Una prueba más es el **Test de Comparación de la Verosimilitud entre dos modelos:** Se utiliza para comparar el ajuste de dos modelos. Las hipótesis están dadas por:

Ho: Los coeficientes de las variables excluidas son iguales a cero

Vs

Ha: Al menos uno de los coeficientes de las variables excluidas es diferente de cero

Tiene el estadístico de prueba:

$$G = -2\text{Ln}[(\text{Verosimilitud del modelo con variables exlcuidas})] + 2\text{Ln}\,(\text{verosimiltud del modelo con todas las variables})]$$

Éste tiene una distribución Chi-cuadrada con Grados de Libertad igual al número de variables excluidas. Es decir, $G \sim \chi^2_{(\#\ de\ variables\ \ excluidas\ )}$[119].

La Prueba Multivariada de Wald[120] es obtenida a partir el siguiente producto matricial:

$$W = B'[\Sigma(B)]^{-1}B = B'(X'VX)B$$

El cual tiene una distribución Chi-cuadrada con p+1 Grados de Libertad, $G \sim \chi^2_{(p+1)}$; Tiene las hipótesis:

Ho: $\beta_j = 0$ para todo j = 0,1,…,p vs **Ha:** $\beta_j \neq 0$ para algún j = 0,1,…,p

Una vez ajustado el modelo de Regresión Logística Múltiple, y que los coeficientes de las variables contenidas en el modelo son estadísticamente significativas (o empíricamente tienen sentido), se procede a la interpretación de los coeficientes[121].

---

[118] La prueba de Wald se mencionó en la sección anterior (RLU). En este caso, después de rechazar la hipótesis nula con la prueba de Razón de Verosimilitudes, se procede a calcular la significación de cada uno de los coeficientes, que tiene las hipótesis:
Ho: $\beta_j = 0$ vs **Ha:** $\beta_j \neq 0$
Esto para toda j = 0,1, …, p. El estadístico de prueba es: $W_j = \frac{B_j}{SE(Bj)}$

[119] Cuando una covariable no métrica sea incluida (o excluida) del modelo, todas las variables de diseño deben ser incluidas (o excluidas). De lo contrario se estaría cometiendo un grave error. Si k es el número de categorías de la variable no métrica, entonces su contribución en los grados de libertad de la prueba del cociente de verosimilitudes para la exclusión de la variable es k-1. Se debe ser cuidadoso en la prueba de Wald debido a los grados de libertad múltiples, para asegurar la significancia del coeficiente.

[120] Las pruebas son únicamente para los coeficientes que son interpretados como pendientes, consisten en eliminar B_0 de el vector B, así como el primer renglón y la primera columna de X'VX.

Recuerde que los coeficientes incluidos en el modelo representan la pendiente o la tasa de cambio de en función de la variable dependiente por unidad de cambio en la variable independiente. Una manera de estudiar los coeficientes del modelo es considerar el modelo como una ecuación que genera un hiperplano de p dimensiones en un espacio de p+1 dimensiones. La interpretación de los coeficientes conlleva dos problemas:

- Determinar la relación funcional entre la variable dependiente y la variable independiente.
- Definir apropiadamente la unidad de cambio en la variable independiente.

El primer problema se soluciona, al recordar que la función establecida entre la variable dependiente y las variables independientes es *lineal,* mediante una transformación. Esto es llamado en los modelos lineales generalizados como función de enlace (*link function*). Por ejemplo, en la RLU, la función enlace está dada por la transformación logit[122]:

$$g(x) = \ln\left\{\frac{\pi(x)}{1-\pi(x)}\right\} = \beta_0 + \beta_1 x$$

4.5.1 Interpretación de los coeficientes

A continuación se realiza la interpretación de los coeficientes de la regresión logística en diferentes situaciones. Primeramente se trabaja con la RLU en los casos en la que la variable independiente es dicotómica y luego politómica. Finalmente se trabaja con la RLM.

Cuando la **covariable es dicotómica**[123], supóngase que X es la variable independiente o covariable de una RLU, entonces X es no métrica, así que se codifica como una variable de diseño (1 y 0). Bajo este modelo hay dos valores de $\pi(X)$ y su equivalente $1-\pi(X)$, explicadas con mayor detalle en la sección dedicada al marco teórico de la RLU. Además, sea Y la variable dependiente, que toma sólo dos resultados (1 y 0), luego, las combinaciones de los resultados posibles de la regresión son 4[124]. En consecuencia, se tienen los siguientes casos[125]:

- Si (0,0), entonces $1-\pi(0) = \frac{1}{1+e^{\beta_0}}$
- Si (0,1), entonces $\pi(0) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$

---

[121] Según los autores Hosmer y Lemeshow, una pregunta que es útil realizar cuando se busca una interpretación de los coeficientes del modelo de RLM, es ¿*qué nos dicen los coeficientes del modelo acerca de las preguntas de investigación que motivaron el estudio?*; (Hosmer, 1989).

[122] Véase la sección teórica de la Regresión Logística Univariada (RLU), 4.4.

[123] Recuerde que **los *Índices de Confianza Institucional y de Percepción de Impacto*** son variables dicotómicas.

[124] Este resultado se debe a que la cardinalidad del producto cartesiano de dos elementos es 4. A saber, son posibles las siguientes parejas (0,0), (0,1), (1,0) y (1,1), donde (x,y).

[125] Cuando el parámetro es β_1 o β_0 se habla de una interpretación teórica; sin embargo, cuando los parámetros han sido estimados, B_1 o B_0 se habla de una interpretación sobre las estimaciones.

- Si (1,0), entonces $1-\pi(1)=\frac{1}{1+e^{\beta_0+\beta_1}}$
- Si (1,1), entonces $\pi(1)=\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$

Sobre las bases anteriores, únicamente hay dos posibilidades de las **Funciones Odds**[126], cuando toma el valor de 0 y 1, es decir:

$$\frac{\pi(1)}{1-\pi(1)} \qquad \text{y} \qquad \frac{\pi(0)}{1-\pi(0)}$$

La **Función Log-Odds** está definida por el Logaritmo Natural de la función logit:

$$g(0)=\ln\left\{\frac{\pi(0)}{1-\pi(0)}\right\} \quad \text{y} \quad g(1)=\ln\left\{\frac{\pi(1)}{1-\pi(1)}\right\}$$

La **Función de Radio de los Odds** (*Odds Ratio Function*), denotado por ψ, está definido como el cociente de las funciones de odds, cuando el numerador es x=1 y el denominador x=0. Por lo tanto, se tiene:

$$\psi=\frac{\pi(1)/(1-\pi(1))}{\pi(0)/(1-\pi(0))}$$

Que al tomar el Logaritmo Natural se convierte en la función **Log Odds Ratio**, que es la diferencia de las funciones logit evaluadas en 1 y 0.

$$\mathrm{Ln}(\psi)=\mathrm{Ln}\left\{\frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}}\right\}=\mathrm{g}(1)-\mathrm{g}(0)$$

En consecuencia, para una RLU con una variable independiente dicotómica, el ***Odds Ratio***, es: $\psi=e^{B_1}$.

Y la función ***Log Odds Ratio*** es: $\mathrm{Ln}(\psi)=B_1$

El *Odds Ratio* es una medida de asociación que tiene una gran variedad de aplicaciones, y que aproxima a que tan probable es la ocurrencia de un resultado (cuando x=1) respecto de que no ocurra (x=0).

El ***Odds Ratio, ψ ,*** es un parámetro de interés fácil de interpretar. También se le puede asociar un intervalo de confianza, cuando ψ es estimado[127]. El intervalo de confianza al (1-α)100% para β_1, está dado por: $e^{B_1 \pm z_{1-\frac{\alpha}{2}} SE(B_1)}$[128].

[126] La **Función de Odds** está definida como la probabilidad de un evento entre la probabilidad de su complemento.
Un ejemplo está dado por una variable dicotómica, es decir con Distribución Bernoulli (p). Suponga que $x \sim Bernoulli(p)$, entonces x=1 con probabilidad p y x=0 con probabilidad 1-p. La función de Odds está dada por:

$$\frac{p}{1-p}$$

Es decir, la probabilidad del evento favorable (éxito) entre la probabilidad del evento desfavorable (fracaso), [Ross, 2000].

Al tomar la función logaritmo natural de ambos lados, se obtiene el intervalo de confianza para el **Log-Odds Ratio:** $B_1 \pm z_{1-\frac{\alpha}{2}} SE(B_1)$.

En el caso que **la covariable sea politómica**[129], supóngase que x es una variable no métrica que tiene k>2 categorías. Entonces se construye un conjunto de k-1 variables de diseño. Consideremos el caso cuando k=4.

Dado que k=4, entonces se tienen 3 variables de diseño que son generadas por una categoría de referencia. Sean las categorías C_1, C_2, C_3 y C_4, siendo ésta última la categoría de referencia. Las variables de diseño son d_1,d_2 y d_3[130].

Una vez ajustado el modelo de RLU, se tienen estimaciones[131] para cada una de las variables de diseño, sean éstas β_{11}, β_{12} y β_{13}. Por lo tanto,

- $Ln(\psi(c_1, c_4)) = B_{11}$
- $Ln(\psi(c_2, c_4)) = B_{12}$
- $Ln(\psi(c_3, c_4)) = B_{13}$

Las anteriores B_{1j} son estimadas, para j=1,2,3.

Lo Intervalos de Confianza son obtenidos usando la misma aproximación que en el caso de una variable independiente dicotómica. Primero se calcula el IC al (1-α)100% para $\beta_{ij,}$ de tal manera que $B_{ij} \pm z_{1-\frac{\alpha}{2}} SE(B_{ij})$, esto para i=1 y j=1,2,3. Paso seguido de exponenciar de ambos lados el IC, por lo tanto: $e^{B_{ij} \pm z_{1-\frac{\alpha}{2}} SE(B_{ij})}$, que corresponde al Intervalo de Confianza para el Odds Ratio de B_{ij}.

Es importante recordar que rara vez, un modelo univariado es suficiente para explicar un problema, es por esto, que la mejor alternativa recae en los modelos multivariados.

En una **Regresión Logística Múltiple** se asume que cada coeficiente estimado conlleva a diferentes estimados para la función Log Odds ajustado para todas las demás variables que incluye

---

[127] En este punto se presenta un problema, en IC estará sesgado a la derecha debido a que la distribución de ψ está acotada en cero (sólo valores no negativos). Para grandes muestras, ψ (estimado) debería de comportarse como una distribución Normal. Por los motivos anteriores, las inferencias de ψ (estimado) se hacen en base a $\mathrm{Ln}(\psi) = \beta_1$ (estimados), el cual tiende a una distribución Normal para tamaños de muestra más pequeños.

[128] Otro punto importante, es que si $\beta_1 > 1$, entonces el IC está sesgado a la derecha. Esto siempre es válido cuando x está codificada como una variable de diseño. Cuando no pasa lo anterior, existen otros métodos para tratar el IC, los cuales no se detallan en este trabajo.

[129] Cuando la covariables es politómica es de especial interés para el trabajo, puesto que las variables Tipo de localidad y el Grado de Marginación son politómicas con tres categorías.

[130] Véase el anexo correspondiente a las variables de diseño.

[131] La razón de las tres ecuaciones siguientes debe a (lo mismo aplica para las otras dos ecuaciones.):

$$Ln\big(\psi(c_1, c_4)\big) = g(c_1) - g(c_4) = [\beta_0 + \beta_{11}(d_1 = 1) + \beta_{12}(d_2 = 0) + \beta_{13}(d_3 = 0)] +$$

$$- [\beta_0 + \beta_{11}(d_1 = 0) + \beta_{12}(d_2 = 0) + \beta_{13}(d_3 = 0)] = \beta_{11}$$

el modelo. Se analiza la RLM desarrollando un modelo bivariado que contiene una variable continua y una variable dicotómica.

Suponga que existen dos grupos, el 1 y el 2. El modelo establecido por las ideas anteriores es:

$$g(x, a) = \beta_0 + \beta_1 x + \beta_2 a$$

Donde

- g(x,a) es la función logit del modelo.
- X, toma los valores x=1 si está en el grupo 2 y x=0 si está en el grupo 1.
- a es una variable continua. Dado que hay dos grupos, a tiene diferentes distribuciones para cada grupo.
- β_0 es el interceptor.
- β_1 representa la diferencia en las unidades de a entre los grupos 1 y 2.
- β_2 es la tasa de cambio de la variable a.

Si consideramos una tabla de 2x2 de clasificaciones (una tabla cruzada), se obtendría una función de *Log Odds Ratio* aproximadamente igual a $\beta_1 + \beta_2(A_2 - A_1)$, el cual es una estimación incorrecta del efecto del grupo en la diferencia de la distribución de la variable a. Para corregir esta diferencia, incluimos la variable a en el modelo y se calcula la diferencia de la función logit en un valor común de la variable a. Un ejemplo de esta idea es la media combinada de la variable a, denotada por A Por lo tanto, la diferencia de la función logit evaluada en A, x=1 y x=0, es:

$$g(x = 1, A) - g(x = 0, A) = \beta_1$$

Luego, el coeficiente $\beta_1$es la función *log odds ratio* que se esperaría obtener de una comparación univariada, esto si los dos grupos tuvieran la misma distribución de la variable a[132].

El procedimiento sería equivalente para cualquier número y mezcla de variables. Las funciones *Odds Ratio* ajustadas son obtenidas al comparar individuos quienes difieren únicamente en la característica de interés y tienen los valores constantes para todas las demás variables.

Con estas consideraciones técnicas se debe ajustar un Modelo de Regresión Logística Múltiple con los índices y variables mencionados anteriormente. Recuérdese que las variables han sido elegidas por medio de preguntas de la encuesta "*Lo que dicen los pobres*" que aproximan a definiciones teóricas de la Sociología. La encuesta no fue hecha con la intención de medir percepción de impacto, confianza, etc. Sin embargo, se insiste que el comportamiento de las

[132] El método de ajuste cuando las variables son todas dicotómicas, politómicas, continuas o una mezcla de ellas son idénticos al explicado en este paso.
Por ejemplo, supongamos que a es una variable continua y que tiene un punto de corte en z, es decir, el soporte de la variable a está divido en dos exactamente en el punto z. Para obtener el efecto de la variable a ajustada, se obtiene un modelo bivariado que contiene dos variables dicotómicas y se calcula la diferencia de la función logit en los dos niveles del grupo y en un valor común de la variable dicotómica a (recuerde que a fue dicotomizada).

variables propuestas y la falta de variables que miden el ingreso es determinante en la aplicación de los modelos expuestos en esta sección.

4.5.2 Estimación del modelo Multivariado

Considere las siguientes covariables: *(1) Índice de Percepción de Impacto, (2) Índice de Confianza Institucional, (3) Tipo de localidad y (4) Grado de Marginación*. Entonces **lo que se quiere hacer es identificar si existen diferencias entre beneficiarios y no beneficiarios entre las covariables anteriores**. Cabe mencionar que, el ajuste del modelo no es determinar las variables que intervienen en la ocurrencia de un evento, sino en identificar la diferencias entre dos grupos.

Por lo tanto, a priori el modelo de RLM debería tener cuatro variables independientes o covariables; sin embargo, el número de variables que contenga el modelo estará en función de los coeficientes de las variables que sean estadísticamente significativas[133].

Como al inicio del capítulo 4, se tiene:

- El **Índice de Percepción de Impacto** tiene dos categorías (0 y1), por lo que ya no se calculó una variable de diseño.
- De la misma manera, se trató *el* **Índice de Confianza Institucional** pues tienen tiene dos categorías (0 y 1).
- El **Tipo de localidad** tiene tres categorías, por lo que se calcularon dos variables de diseño con categoría de referencia "Tipo urbano". Las categorías tienen el nombre de *Tipo de localidad (01) y (02).*
- Para el **Grado de Marginación** también se calcularon dos variables de diseño con categoría de referencia el "Alto". Las nuevas variables se denominaron *Grado de Marginación (01) y (02).*
- La variable dependiente sigue siendo la **Condición de beneficiario de Oportunidades**, que ya está codificada en 0 y 1.

El **modelo teórico de Regresión Logística Múltiple** que se quiere probar, es:

$$\pi(x) = \frac{e^{\beta_0+\beta_{GM(01)}X_{GM(01)}+\beta_{GM(02)}X_{GM(02)}+\beta_{TL(01)}X_{TL(01)}+\beta_{TL(02)}X_{TL(02)}+\beta_{IPI}X_{IPI}+\beta_{ICI}X_{ICI}}}{1+e^{\beta_0+\beta_{GM(01)}X_{GM(01)}+\beta_{GM(02)}X_{GM(02)}+\beta_{TL(01)}X_{TL(01)}+\beta_{TL(02)}X_{TL(02)}+\beta_{IPI}X_{IPI}+\beta_{ICI}X_{ICI}}}$$

[133] Según la sección anterior, 4.5: el Modelo de Regresión Logístcia Multivariado, se pueden usar **las pruebas de Razón de Verosimilitudes, de Cambio en la Verosimilitud, de Wald Univariada y de Wald Multivariada para probar la significancia del modelo de RLM** y la de cada uno de los coeficientes por separado.

En donde:

β_0 es la constante del modelo.

$\beta_{GM(01)}$ es el coeficiente de la variable de diseño 01 (GM bajo) del Grado de Marginación.

$X_{GM(01)}$ es el la variable de diseño 01 (GM Bajo) del Grado de Marginación

$\beta_{GM(02)}$ es el coeficiente de la variable de diseño 02 (GM medio) del Grado de Marginación.

$X_{GM(02)}$ es el la variable de diseño 02 (GM medio) del Grado de Marginación.

$\beta_{TL(01)}$ es el coeficiente de la variable de diseño 01 (tipo rural) del Tipo de localidad.

$X_{TL(01)}$ es el la variable de diseño 01 (tipo rural) del Tipo de localidad.

$\beta_{TL(02)}$ es el coeficiente de la variable de diseño 02 (tipo semi-rural) del Tipo de localidad.

$X_{TL(02)}$ es el la variable de diseño 02 (tipo semi-rural) del Tipo de localidad.

β_{IPI} es el coeficiente del Índice de Percepción de Impacto.

X_{IPI} es la variable del Índice de Percepción de Impacto.

β_{ICI} es el coeficiente del Índice de Confianza Institucional.

X_{ICI} es la variable de Índice de Confianza Institucional.

Aunque el modelo teórico anterior pareciera tener muchas variables, **la realidad es que sólo tres índices intervienen en él, así como un indicador**. Note, que hay cuatro variables de diseño, dos para el *Grado de Marginación* y dos para el *Tipo de localidad*.

Hechas las aclaraciones anteriores, se comienza la construcción del modelo multivariado propuesto. La estimación fue realizada por **Máxima Verosimilitud** y con el **Método de paso por paso hacia adelante** (***Stepwise Forward Method***), [Hosmer, 1989]. Este método de estimación de un modelo logístico multivariado es de gran ayuda cuando no se tiene información anterior de modelos con variables similares. También ayuda a entender las variables que están asociadas con la variable dependiente (y que no son fáciles de explicar). Este procedimiento equivale a un algoritmo en el cual se va construyendo el modelo. En cada paso se analiza la importancia de incluir (o excluir) variables que son estadísticamente significativas (o no los son). En la regresión logística se asume que los errores tienen una distribución binomial y su significancia es probada vía el cociente de probabilidades de la χ^2 . Cabe mencionar que existen paquetes estadísticos que incluyen opciones y programas para ajustar un modelo multivariado por pasos (hacia adelante o hacia atrás).

- **Paso 0**

El algoritmo de ajuste comienza con la estimación del modelo nulo, el cual tiene un coeficiente $B_0 = -0.919$ y su estadístico de $-2Ln(Verosimilitud) = 2818.83$. Su estadístico de Wald vale $W = 405.92$ para una $\chi^2_{(1)}$ (p-valor de 0.000), lo cual conduce a que se rechace la hipótesis nula. Por lo tanto, se puede suponer que B_0 es estadísticamente significativo; la tabla de clasificación con la constate incluida en el modelo es, (tabla 4.8):

Tabla 4.8
Tabla de clasificación [a,b]

			Observados		
			Beneficiario OPORTUNIDADES/PROGRESA		Porcentaje correcto
			No	Si	
Paso 0	Beneficiario OPORTUNIDADES/PROGRESA	No	1685	0	100.0
		Si	672	0	.0
	Porcentaje total				71.5

a La constante es incluida en el modelo.
b El valor de corte es ½.
Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Que se interpreta de la siguiente manera: bajo el modelo nulo[134] se clasifican correctamente 1685 casos de los no beneficiarios (100%) y 0 casos de los beneficiarios (0%), dando un porcentaje total de 71.5% de casos bien clasificados.

Las variables que no están en la ecuación están en la tabla 4.9. Obsérvese que aparece cada una de las covariables de manera conjunta con las variables de diseño que generan éstas. Según la prueba del Score, la variable candidata a entrar en el siguiente paso (paso 1) en la estimación del modelo es aquella que tenga el mayor estadístico del Socre (o que tenga el menor p-valor), condición que cumplen el *Grado de Marginación y el Tipo de localidad*; se considera el primero de ellos para entrar en el modelo.

[134] El modelo nulo sólo tiene la constante estimada.

Tabla 4.9
Variables no presentes en la ecuación

			Score	GL	Sig.
Paso 0	Variables	**Grado de Marginación**	**233.938**	**2**	**.000**
		marg(1)	185.993	1	.000
		marg(2)	.919	1	.338
		Tipo de localidad	**135.662**	**2**	**.000**
		tipour(1)	97.185	1	.000
		tipour(2)	2.371	1	.124
		Índice de Confianza Institucional	**2.147**	**1**	**.143**
		Índice de Percepción de Impacto	**22.315**	**1**	**.000**
	Estadísticas totales		264.775	6	.000

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

- **Paso 1**

El modelo a probar es:

$$\pi(x) = \frac{e^{\beta_0 + \beta_{GM(01)} X_{GM(01)}}}{1 + e^{\beta_0 + \beta_{GM(01)} X_{GM(01)}}}$$

Por lo tanto, las variables que se desean probar su significancia son el Grado de Marginación (01 y 02).

En el primer paso, $B_0 = 0.021$, $B_{GM(01)} = -1.33$, $B_{GM(02)} = -0.808$, con un estadístico de $-2Ln(Verosimilitud) = 2592.07$.

Al probar la hipótesis del cambio entre verosimilitudes se tiene:

Las hipótesis para tal prueba establecen:

Ho: las covariables fuera del modelo son iguales a cero

Vs

Ha: Al menos una de las covariables que está fuera del modelo es diferente de cero

Y el cálculo del estadístico es:

$$G = -2\text{Ln}[(\text{Verosimilitud del modelo con variables exlcuidas})] + 2\text{Ln}\,(\text{verosimiltud del modelo con todas las variables})] = 2818.833 - 2592.074 = 226.759$$

Que tiene una distribución $\chi^2_{(2)}$, genera un p-valor de 0.000, por lo que se rechaza la hipótesis nula, *Ho, que establece que las variables (con dos grados de libertad porque hay dos variables de diseño) que no están en el modelo son iguales a cero*. Por lo tanto, al menos una de las variables es estadísticamente significativa[135].

Por otro lado, La **prueba de Wald** tiene las siguientes hipótesis:

Ho: $B_i = 0$ vs **Ha**: $B_i \neq 0$ para i=0,1,2…

El estadístico asociado es $W = 217.49$ para una $\chi^2_{(2)}$ y con un p-valor=0.000 (se rechaza la hipótesis nula de coeficiente estimado igual a cero). Por lo tanto, por medio de dos maneras diferentes llegamos al mismo resultado, y se puede afirmar que el coeficiente del Grado de Marginación es estadísticamente significativo.

La prueba de Wald para la constante concluye que no es estadísticamente significativa, p-valor= 0.714.

La expresión matemática para este modelo es:

$$g(x) = B_{GM(01)}x_{GM(01)} + B_{GM(02)}x_{GM(02)} = -1.582x_{GM(01)} - 0.851x_{GM(02)}$$

La **R^2 de Nagelrkerke** es 0.131, con esto, el modelo explica el 13.1% de la variabilidad.

Por otro lado, las hipótesis de la prueba de cambio en $-2Ln(Verosimilitud)$ son:

Ho: Los coeficientes estimados son cero
vs
Ha: Al menos uno de los coeficientes estimados es diferente de cero.

En este paso, supóngase que se elimina del modelo la variable Grado de Marginación[136] , de manera tal que se provoca un cambio de 226.759 en el valor de $-2Ln(Verosimilitud)$, generando un p-valor de 0.000 para una distribución $\chi^2_{(2)}$. Por lo tanto, se rechaza la hipótesis nula Ho, que establece que los coeficientes estimados son cero. Y por lo tanto, nuevamente las variables de diseño del GM son estadísticamente significativas.

[135] Recuerde que el Grado de Marginación generó dos variables de diseño, por lo que el estadístico G se tiene que comparar con una distribución $\chi^2_{(2)}$, esto con el fin de hacer una comparación exacta.
[136] Eliminar la variable Grado de Marginación significa que se eliminan las variables de diseño asociadas a tal variable. En este caso, GM (01) y GM (02).

- **Paso 2**

Aquí, entran las variables de diseño del **Tipo de localidad** y el modelo a probar es el siguiente:

$$\pi(x) = \frac{e^{\beta_0 + \beta_{GM(01)} X_{GM(01)} + \beta_{GM(02)} X_{GM(02)}}}{1 + e^{\beta_0 + \beta_{GM(01)} X_{GM(01)} + \beta_{GM(02)} X_{GM(02)}}}$$

A partir de lo anterior, la constante y los coeficientes estimados son: $B_0 = 0.494$, $B_{GM(01)} = -1.312$, $B_{GM(02)} = -0.8$, $B_{TL(01)} = 0.521$, $B_{TL(02)} = 0.639$, además $-2Ln(Verosimilitud) = 2571.176$. Al calcular el estadístico de **prueba de la Deviance** se tiene:

$$G = 2592.074 - 2571.176 = 247.65$$

Que tiene una distribución $\chi^2_{(4)}$, y genera un p-valor de 0.000, por lo tanto se rechaza la hipótesis nula, y se puede suponer que al menos uno de todos los coeficientes estimados es diferente de cero. Por otro lado, la prueba de Wald establece $W_o = 10.251$, $W_{GM} = 99.46$ y $W_{TL} = 20.725$, con p-valores de 0.001, 0.000 y 0.000 para la constante, el *Grado de Marginación y el Tipo de localidad*, respectivamente. Por dos vías diferentes, se probó las significancia estadística de los coeficientes y la constante.

La **prueba de Wald** para cada uno de los coeficientes de las variables de diseño son, (tabla 4.10):

Tabla 4.10

Resumen de la prueba de Wald				
Variable	**Variables de diseño**	**Wald**	**Sig.**	**Resultado**
Constante	-	10.251	0.001	**Significativa**
Grado de Marginación	1	98.507	0	**Significativa**
	2	31.425	0	**Significativa**
Tipo de localidad	1	13.789	0	**Significativa**
	2	18.314	0	**Significativa**

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Por lo tanto, todos los coeficientes de las variables de diseño ara el Grado de Marginación y el Tipo de localidad son estadísticamente significativos, y su expresión matemática, del modelo Loggit, es:

$$g(x) = B_0 + B_{GM(01)} x_{GM(01)} + B_{GM(02)} x_{GM(02)} + B_{TL(01)} x_{TL(01)} + B_{TL(02)} x_{TL(02)}$$

$$= -0.49 - 1.31 x_{GM(01)} - 0.8 x_{GM(02)} + 0.52 x_{TL(01)} + 0.64 x_{TL(02)}$$

La **R^2 de Nagelkerke** para el modelo anterior es 0.143, con ello se explica el 14.3% de la variabilidad.

Supóngase que del modelo ajustado, se elimina la variable *Tipo de localidad*[137], y con ello sólo nos quedamos con la variable *Grado de Marginación*. Debido a la corrección anterior, el cambio en el estadístico $-2Ln(Verosimilitud)$ se ve alterado en 20.897 de una distribución $\chi^2_{(2)}$, con un p-valor de 0.000; esto conduce a la conclusión que el *Tipo de localidad* es estadísticamente significativo.

Ahora supóngase que la variable eliminada del modelo de RLM es el *Grado de Marginación,* y se mantiene el *Tipo de localidad*. El cambio en $-2Ln(Verosimilitud)$ es 104.596 para una distribución $\chi^2_{(2)}$ (con un p-valor de 0.000), por lo tanto, los coeficientes de las variables de diseño (y con ello el de la variable que los origina, GM) es estadísticamente significativo, ie, los coeficientes estimados no pueden ser cero.

- **Paso 3**

En el **tercer paso**, entra como covariable el *Índice de Percepción de Impacto*. Bajo esto, el modelo que se desea probar es:

$$\pi(x) = \frac{e^{\beta_0+\beta_{GM(01)}X_{GM(01)}+\beta_{GM(02)}X_{GM(02)}+\beta_{TL(01)}X_{TL(01)}+\beta_{TL(02)}X_{TL(02)}+\beta_{IPI}X_{IPI}}}{1+e^{\beta_0+\beta_{GM(01)}X_{GM(01)}+\beta_{GM(02)}X_{GM(02)}+\beta_{TL(01)}X_{TL(01)}+\beta_{TL(02)}X_{TL(02)}+\beta_{IPI}X_{IPI}}}$$

Bajo el modelo anterior, la constante y los coeficientes estimados son $B_0 = 0.325$, $B_{GM(01)} = -1.289$, $B_{GM(02)} = -0.781$, $B_{TL(01)} = 0.53$, $B_{TL(02)} = 0.64$, 0.781 y $B_{IPI} = -0.380$, además $-2Ln(Verosimilitud) = 2555.769$. Al calcular el **estadístico de *Deviance*** para la significancia del modelo, se obtiene:

$$G = 2592.074 - 2555.769 = 263.064$$

Que se compara con una distribución $\chi^2_{(5)}$, y su p-valor asociado es 0.000, por lo que se puede suponer que al menos uno de los coeficientes estimados es diferente de cero, y con ello la Regresión Logística Múltiple es significativa, es decir, tiene sentido ajustarse. Por otro lado, las pruebas de Wald establecen $W_o = 4.081$, $W_{GM} = 94.746$ y $W_{TL} = 20.932$ y $W_{IPI} = 15.352$, con p-valores 0.043, 0.000, 0.000 y 0.000 para la constante, el Grado de Marginación, el Tipo de Localidad y el Índice de Percepción de Impacto, respectivamente.

[137] Al eliminar el Tipo de localidad se eliminan automáticamente las variables de diseño asociados a ésta. Al mantener el Grado de Marginación, se mantienen las dos variables de diseño que originaron dicha variable.

La **prueba de Wald** genera los siguientes valores, (tabla 4.1). Al observarla, se deduce que según la prueba de Wald univariada todos los coeficientes de las variables de diseño del *Grado de Marginación, el Tipo de localidad y el Índice de Percepción de Impacto* son estadísticamente significativos, esto es, que todos los coeficientes son estadísticamente diferentes de cero.

Tabla 4.11

Resumen de la prueba de Wald				
Variable	Var. de diseño	Wald	Sig.	Resultado
Constante	-	4.081	0.043	**Significativa**
Grado de Marginación	1	93.923	0	**Significativa**
	2	29.682	0	**Significativa**
Tipo de localidad	1	14.21	0	**Significativa**
	2	18.356	0	**Significativa**
Índice de Percepción de Impacto	-	15.352	0	**Significativa**

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

La expresión matemática para de la función loggit del modelo de RLM es:

$$g(x) = B_0 + B_{GM(01)}x_{GM(01)} + B_{GM(02)}x_{GM(02)} + B_{TL(01)}x_{TL(01)} + B_{TL(02)}x_{TL(02)} + B_{IPI}x_{IPI}$$

$$= -0.32 - 1.29x_{GM(01)} - 0.78x_{GM(02)} + 0.53x_{TL(01)} + 0.64x_{TL(02)} - 0.38x_{IPI}$$

La **R^2 de Nagelkerke** para el modelo anterior es 0.151, con ello se explica el 15.1% de la variabilidad total.

El método de Máxima Verosimilitud sólo calculó cuatro pasos, debido a que en cada paso, la restricción es que el estadístico $-2Ln(Verosimilitud)$ cambiara más de 1/1000[138];

La tabla de clasificación del modelo de RLM en el último paso de estimación se presenta en la tabla 4.12. En el modelo final de la RLM se clasifica correctamente 1534 casos de no beneficiarios (91%) y 180 casos de beneficiarios (26.8%), obteniendo el 72.7% de los casos bien clasificados.

[138] Lo cual significa que el estadístico debería cumplir: $-2Ln(Verosimilitud) > 0.001$. La cota de 1/1000 fue la establecida por default en el programa de cálculo, SPSS.

Tabla 4.12
Tabla de clasificación del modelo RLM [(a)139]

			Observados		
			Beneficiario OPORTUNIDADES/PROGRESA		Porcentaje correcto
			No	Si	
Paso 0	**Beneficiario OPORTUNIDADES/PROGRESA**	No	**1534**	**151**	91.0
		Si	**492**	**180**	26.8
	Porcentaje total				72.7

a El valor de corte es ½.

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

Observe que el Índice de Confianza Institucional (ICI) no fue incluido en el modelo de RLM final**,** esto se debe a que la variable no es estadísticamente significativa, esto es, que los coeficientes de las variables de diseño que genera dicho índice valen siempre cero. Esto es reforzado por dos hechos: el primero de ellos es que el ICI no estaba asociado con la variable dependiente (indicador de beneficiario del Programa Oportunidades); también en el modelo de RLU no fueron estadísticamente significativas sus variables de diseño.

Ahora, supóngase que del modelo ajustado, se elimina la variable *Tipo de localidad*[140], y con ello sólo nos quedamos con la variable *Grado de Marginación y el Índice de Percepción de Impacto*. Debido a la corrección anterior, el cambio en el estadístico $-2Ln(Verosimilitud)$ se altera en 21.120, de una distribución $\chi^2_{(2)}$, con un p-valor de 0.000; esto conduce directamente a la conclusión que el *Tipo de localidad* es estadísticamente significativo.

Ahora supóngase que la variable eliminada del modelo de RLM es el *Grado de Marginación*, y se mantiene el *Tipo de localidad y el Índice de Percepción de Impacto*. El cambio en $-2Ln(Verosimilitud)$ es 99.305 para una distribución $\chi^2_{(2)}$ (con un p-valor de 0.000), por lo tanto, el coeficiente del *GM* es estadísticamente significativo, ie, los coeficientes estimados de las variables de diseño no pueden ser cero.

Finalmente, supóngase que se elimina el **Índice de Percepción de Impacto** y se mantienen el *Grado de Marginación y el Tipo de localidad*. El cambio en el estadístico $2Ln(Verosimilitud)$ es

[139] Recuerde que para el modelo nulo, la tabla de clasificiación discriminaba correctamente en un 71.9%, luego, para tres variables después, el modelo de RLM clasificaba bien el 72.7%. Esto implica que hubo un aumento del 0.08%.
Empero, el modelo de RLM distibuye de mejor manera la clasificación; en éste, ya se clasifican correctamente uno de cada cuatro beneficiarios del Programa Oportunidades.

[140] Al eliminar el Tipo de localidad se eliminan automáticamente las variables de diseño asociados a ésta. Al mantener el Grado de Marginación, se mantienen las dos variables de diseño que originaron dicha variable. También se debe mantener la variable que da origen al Índice de Percepción de Impacto.

15.408, para una distribución $\chi^2_{(1)}$, con un p-valor de 0.000. Por lo tanto, el eliminar este índice altera significativamente el resultado. Se concluye que es mejor, mantener tal índice.

Finalmente, el modelo de Regresión Logística Múltiple, conduce a la **probabilidad de que el resultado sea favorable dados los valores presentados en $\mathbf{x}$**, matemáticamente esto es:

$$\pi(\mathrm{x}) = \frac{\mathrm{e}^{\mathrm{g}(\mathrm{x})}}{1 + \mathrm{e}^{\mathrm{g}(\mathrm{x})}}$$

$$= \frac{\mathrm{e}^{-0.32-1.29x_{GM(01)}-0.78x_{GM(02)}+0.53x_{TL(01)}+0.64x_{TL(02)}-0.38x_{IPI}}}{1 + \mathrm{e}^{-0.32-1.29x_{GM(01)}-0.78x_{GM(02)}+0.53x_{TL(01)}+0.64x_{TL(02)}-0.38x_{IPI}}}$$

$$= \mathrm{P}[\mathrm{Y} = 1|\mathrm{x}]$$

Recuerde que el modelo tiene tres covariables: el *Grado de Marginación, el Tipo de Localidad y el Índice de Percepción de Impacto*. Las primeras dos generan, cada una, dos variables de diseño. Por lo tanto, el modelo anterior tiene **seis parámetros estimados**, *cinco de ellos coeficientes de las variables y una constante*. Establecidas las ideas anteriores, sea

$$x = (x_{GM(01)}, x_{GM(02)}, x_{TL(01)}, x_{TL(02)}, x_{IPI})$$

El **vector de datos de entrada del modelo de RLM**. Note que $x \in \mathbb{R}^5$. Al inicio del capítulo se analizan las variables de diseño, ahora el análisis se torna para cada una de las covariables[141] contenidas en el modelo.

- $x_{GM(01)}$ es **la primera variable de diseño del Grado de Marginación** (GM). El evento de interés es "GM-Bajo"; la categoría de referencia es "GM-Alto".
- $x_{GM(02)}$ es **la segunda variable de diseño del Grado de Marginación**. El evento de interés es un "GM-Medio" ; la categoría de referencia es " GM-Alto".
- $x_{TL(01)}$ es **la primera variable de diseño del Tipo de localidad**. El evento de interés es " tipo-Rural"; la categoría de referencia es "tipo-Urbano".
- $x_{TL(02)}$ es **la segunda variable de diseño del Tipo de localidad**. El evento de interés es "tipo-Semirural"; la categoría de referencia es "tipo-Urbano".
- x_{IPI} es **el Índice de Confianza Institucional,** tiene como categoría de interés "Baja confianza" y como categoría de referencia el "Alta confianza".

Como las variables anteriores son todas de diseño, luego están codificadas de manera tal que el evento de interés es 1 y el evento que no es de interés con 0. Por lo tanto, la imagen de cada una de ellas es el conjunto {0,1}.

[141] En este punto, las covariables presentadas son estadísticamente significativas.

Es turno de hablar de los coeficientes del modelo de RLM ajustado, para ello se anexa la tabla 4.13. Los Intervalos de Confianza para los Coeficientes ajustados exponenciados, e^{B_i} con $i\epsilon\{GM(01), GM(02), TL(01), TL(02), IPI\}$, están al 95%. Ninguno de los cinco contiene a la unidad, lo cual no contradice la significancia de los coeficientes; si tuviera a la unidad (1) dentro de algún intervalo, significaría que e^{B_i} podría tomar el valor 1, con lo cual al calcularle el Logaritmo Natural se obtendría el 0, implicando la no significancia estadística del coeficiente[142].

Tabla 4.13
Tabla de Intervalos de Confianza para los Coeficientes Estimados

	B	Exp(B)	95.0% I.C. para EXP(B)	
			L.I	L.S..
Grado de Marginación				
Grado de Marginación (1)	-1.289	.276	.212	.358
Grado de Marginación (2)	-.781	.458	.346	.607
Tipo de localidad				
Tipo de localidad 1)	.530	1.699	1.289	2.240
Tipo de localidad (2)	.640	1.897	1.415	2.542
Índice de Percepción de Impacto	**-.380**	.684	.566	.827
Constante	**-.325**	.723		

Fuente: Encuesta "Lo que dicen los pobres", 2003.
Cálculos propios.

Las combinaciones[143] que pueden ocurrir en el vector $x = (x_{GM(01)}, x_{GM(02)}, x_{TL(01)}, x_{TL(02)}, x_{IPI})$ son 32. Esto es, entre todas las posibles combinaciones del vector x, únicamente hay 32 maneras posibles de producir resultados en la variable independiente. Es importante mencionar, que el número de casos posibles 32 es bastante reducido en comparación con todos los que se podrían formar con las 2939 personas encuestadas por Lo que dicen los pobres y con un modelo de RLM que considere covariables continuas.

Por lo tanto, la probabilidad condicional del resultado favorable dados los valores de la muestra:

$$\pi(x_{GM(01)}, x_{GM(02)}, x_{TL(01)}, x_{TL(02)}, x_{IPI}) = \frac{e^{g((x_{GM(01)}, x_{GM(02)}, x_{TL(01)}, x_{TL(02)}, x_{IPI}))}}{1 + e^{g((x_{GM(01)}, x_{GM(02)}, x_{TL(01)}, x_{TL(02)}, x_{IPI}))}}$$

Si bien, en el marco teórico de la RLM se da una ligera explicación de cómo tratar y explicar los coeficientes de un modelo multivariado, la explicación no es suficiente, pues sólo se describen los casos cuando hay dos variables, una de ellas dicotómica y la otra continua. En el

[142] Este punto se había detallado en la estimación de los Modelos Univariados, mas se vuelve a mencionar con el fin de refrescar la mente del lector

[143] El 32 es el resultado de contar las **Ordenaciones con Repetición de 2 en 5**, o dicho de otra manera, **32= 2^5.**
Algunas de estas posibles combinaciones son las quíntuplas:
(0,0,0,0,0), (0,0,0,0,1), (0,0,0,1,0), (0,0,1,0,0), (0,1,0,0,0), (1,0,0,0,0), (1,0,0,0,1), (1,0,0,1,0), (1,0,1,0,0), (1,1,0,0,0), (0,1,1,0,0)…

modelo presentado anteriormente, se tienen 5 variables dicotómicas, por lo que se dificulta considerablemente la comprensión del modelo.

4.5.2.1 Ejemplo:

Asuma que el vector $x = (1,0,1,0,1)$. Entonces:

$$\pi(1,0,1,0,1) = \frac{e^{g(1,0,1,0,1)}}{1 + e^{g(1,0,1,0,1)}} = \frac{e^{-0.32-1.29x_{GM(01)}+0.53x_{TL(01)}-0.38x_{IPI}}}{1 + e^{-0.32-1.29x_{GM(01)}+0.53x_{TL(01)}-0.38x_{IPI}}} = 0.188$$

En consecuencia, $P[Y = 1|(1,0,1,0,1)] = 0.188$. Este vector describe a una persona que vive en una localidad rural, con un Grado de Marginación bajo y que tiene una opinión que está dentro del primer estrato del Índice de Percepción de Impacto[144].

4.5.3 Probabilidades asociadas al modelo de estimado

De manera general, el resumen de la información del modelo de RLM es a través de la siguiente tabla que calcula todos los casos (32 posibles) del modelo de Regresión Logística Multivariado propuesto en la presente sección. La primera columna de la tabla expresa el identificador del número de caso; de la segunda a la sexta columna, las cuales tienen escritas un número romano, sitúan los datos de entrada al modelo, es decir, cumplen la función de $x = (x_{GM(01)}, x_{GM(02)}, x_{TL(01)}, x_{TL(02)}, x_{IPI})$; la columna de la función Loggit está calculada para cada caso; las últimas dos columnas Pi(x) y 1-Pi(x) son las probabilidades condicionales de obtener un evento favorable[145] (o desfavorable) condicionado al vector de datos, x.

La función loggit es el Logaritmo Natural del cociente de probabilidad de la ocurrencia de un evento entre la probabilidad de no ocurrencia. Es decir:

$$Loggit(x) = Ln\left(\frac{p}{1-p}\right)$$

En el caso de nuestro modelo, la función loggit equivale al Logaritmo Natural del cociente entre la probabilidad de ser beneficiario del Programa Oportunidades y la probabilidad de no serlo, (tabla 4.14).

[144] Nuevamente es fácil deducir las características de la persona en cuestión si se considera la codificación de las variables.
[145] Un resultado favorable es entendido, en este contexto, como ser beneficiario del Programa Oportunidades. No serlo, representa la negación del hecho.

Tabla 4.14

Tabla resumen del modelo de Regresión Logística Multivariado (todos los casos)								
Combinaciones del vector x						Función Loggit g(x)[146]	Probabilidades	
#	I	II	III	IV	V		Pi(x)	1-Pi(x)
1	**0**	**0**	**0**	**0**	**0**	**-0.320**	**0.421**	**0.579**
2	0	0	0	0	1	-0.700	0.332	0.668
3	0	0	0	1	0	0.320	0.579	0.421
4	0	0	1	0	0	0.210	0.552	0.448
5	0	1	0	0	0	-1.100	0.250	0.750
6	1	0	0	0	0	-1.610	0.167	0.833
7	1	1	0	0	0	-2.390	0.084	0.916
8	1	0	1	0	0	-1.080	0.254	0.746
9	1	0	0	0	1	-1.990	0.120	0.880
10	0	1	1	0	0	-0.570	0.361	0.639
11	0	0	1	1	0	0.850	0.701	0.299
12	0	0	0	1	1	-0.060	0.485	0.515
13	0	1	0	1	0	-0.460	0.387	0.613
14	0	0	1	0	1	-0.170	0.458	0.542
15	0	1	0	0	1	-1.480	0.185	0.815
16	1	0	0	1	0	-0.970	0.275	0.725
17	0	0	1	1	1	0.470	0.615	0.385
18	0	1	0	1	1	-0.840	0.302	0.698
19	0	1	1	1	0	0.070	0.517	0.483
20	1	0	0	1	1	-1.350	0.206	0.794
21	1	1	0	0	1	-2.770	0.059	0.941
22	1	1	1	0	0	-1.860	0.135	0.865
23	**1**	**0**	**1**	**0**	**1**	**-1.460**	**0.188**	**0.812**
24	1	1	0	1	0	-1.750	0.148	0.852
25	1	0	1	1	0	-0.440	0.392	0.608
26	0	1	1	0	1	-0.950	0.279	0.721
27	1	1	1	1	0	-1.220	0.228	0.772
28	1	1	1	0	0	-1.860	0.135	0.865
29	1	1	0	0	0	-2.390	0.084	0.916
30	1	0	0	0	0	-1.610	0.167	0.833
31	1	0	0	0	0	-1.610	0.167	0.833
32	**1**	**1**	**1**	**1**	**1**	**-1.600**	**0.168**	**0.832**

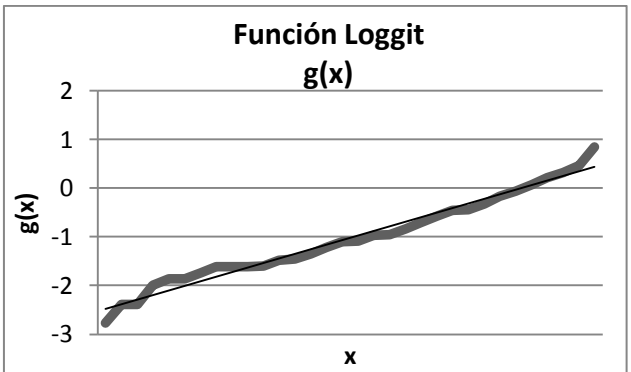
Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

[146] La función Loggit está dada por:

$$g(x) = B_0 + B_{GM(01)}x_{GM(01)} + B_{GM(02)}x_{GM(02)} + B_{TL(01)}x_{TL(01)} + B_{TL(02)}x_{TL(02)} + B_{IPI}x_{IPI}$$

Se muestra la gráfica 4.1 de la función Loggit, para construirla se consideran los 32 casos posibles del modelo de RLM. Estos se ordenan de manera ascendente y se grafican. Observe el comportamiento lineal que sigue la función, lo cual es un supuesto básico en los modelos de RL[147].

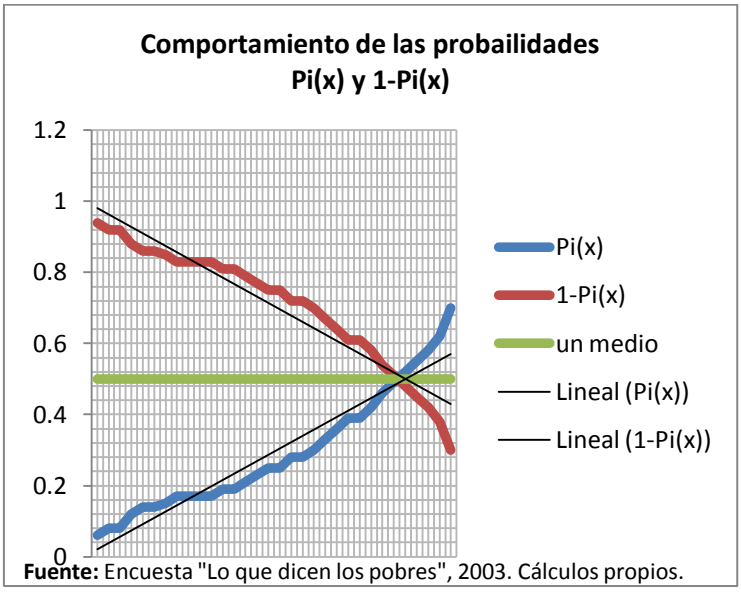
Gráfica 4.1



Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos

También se presenta la gráfica 4.2 del comportamiento de las probabilidades $\pi(x)$ y $1-\pi(x)$[148] con base en todos los casos posibles para el modelo de RLM. Debido a que las probabilidades anteriores son complementos una de otra, entonces la gráfica tiene una simetría agradable a la vista. También se anexaron las rectas que más se ajustan por medio del método de mínimos cuadrados. La recta horizontal verde está situada exactamente en ½ y coincide exactamente con el eje de simetría de las curvas.

Gráfica 4.2



Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

[147] Véase el marco teórico de la RLU y RLM, secciones 4.4 y 4.5.

[148] Recuerde que $\pi(x)$ es la probabilidad condicional del resultado favorable dados los valores de la muestra, x. Análogamente, $1-\pi(x)$ es la probabilidad condicional de que el resultado no sea favorable dado los resultados de la muestra, x.

Se analizan las **funciones de Momios (Odds Function)** y de **Cociente de Momios (Odds-Ratio Function)** para cada uno de las covariables significativas del modelo de RLM. Luego, hay cinco casos de análisis, de tal manera que sólo varía la variable que se quiere medir y las demás permanecen constantes. Esto se logra cuando la variable de interés toma el valor de 1 y las demás permanecen constantes. Los casos presentados son:

(1,0,0,0,0), (0,1,0,0,0), (0,0,1,0,0), (0,0,0,1,0), (0,0,0,0,1)

Los cuales se comparan con el **vector de referencia (0,0,0,0,0**), es decir, cuando no se presenta ninguna de los eventos de interés; puesto en términos de os resultados del modelo, se hace referencia al caso **cuando el Grado de Marginación es Alto[149], el Tipo de localidad es Urbano**[150] y el **Índice de Percepción de Impacto** está en la *desacuerdo*.

- **Caso (1,0,0,0,0):** se presenta cuando *el **Grado de Marginación es Bajo*** manteniendo constantes las demás variables. Entonces la **función de Momios**[151] es:

$$\frac{\pi(1,0,0,0,0)}{1-\pi(1,0,0,0,0)} = \frac{0.167}{0.833} = 0.199$$

La probabilidad condicional de **ser beneficiario del Programa Oportunidades** dado que se pertenece a una localidad con **grado de marginación bajo, teniendo las demás variables constantes** es: $\pi(1,0,0,0,0) = 0.167$. La probabilidad de **no ser beneficiario de Oportunidades** dado que se pertenece a una localidad con las condiciones anteriores es $1-\pi(1,0,0,0,0) = 0.833$. Luego, $\frac{\pi(1,0,0,0,0)}{1-\pi(1,0,0,0,0)}$ representa el cociente de ser beneficiario de Oportunidades respecto de no serlo dado que se pertenece a una localidad con GM- bajo.

Y para el vector (0,0,0,0,0) es

$$\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)} = \frac{0.421}{0.579} = 0.7261$$

La probabilidad condicional de ser beneficiario de Oportunidades dado que se vive en una localidad urbana con grado de marginación Alto y que tiene un índice de percepción de impacto en desacuerdo es $\pi(0,0,0,0,0) = 0.421$. Entonces, el complemento de tal probabilidad es el no ser beneficiario de oportunidades dadas las condiciones anteriores es $1-\pi(0,0,0,0,0) = 0.579$. Análogamente, $\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)} = 0.7261$ que en el cociente anterior, representa el cociente de ser beneficiario de Oportunidades respecto de no serlo sujeto a vivir en una localidad urbana con GM-alto y en desacuerdo con el IPI.

[149] El Grado de Marginación resultó ser Alto porque para la primera variable de Diseño se tiene (0) queriendo indicar que su GM no es bajo. En la segunda variable de diseño también se tiene (0), mencionando que su GM no es Medio. Por el *principio matemático del tercero excluido*, tiene que ser un GM Alto.

[150] Bajo el mismo razonamiento que el GM, el Tipo de localidad tiene que ser Urbano.

[151] La función de Momios también es conocida como la función de Odds (Odss function).

Ahora, considere el Logaritmo natural de las funciones anteriores, es decir, las funciones Log-Odds:

$$\mathrm{Ln}\left\{\frac{\pi(1,0,0,0,0)}{1-\pi(1,0,0,0,0)}\right\} = \mathrm{Ln}\{0.1.999\} = -1.61 = g(1,0,0,0,0)$$

Y

$$\mathrm{Ln}\left\{\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)}\right\} = \mathrm{Ln}\{0.7261\} = -0.32 = g(0,0,0,0,0)$$

Obsérvese que las funciones de Log-Odds anteriores corresponden a las funciones loggit (logaritmo natural de las probabilidades de ser beneficiario del Programa Oportunidades o no serlo bajo las situación de los vectores $x = (1,0,0,0,0)$ y $x = (0,0,0,0,0)$), anteriormente calculadas[152].

Ahora considere la función del **Cociente de Momios** (*Odds-Ratio Function*) es:

$$\psi = \frac{\left\{\frac{\pi(1,0,0,0,0)}{1-\pi(1,0,0,0,0)}\right\}}{\left\{\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)}\right\}} = \frac{0.199}{0.7261} = 0.275$$

El cociente anterior, no es más que una forma de expresar la proporción de veces que un suceso ocurra frente a que no ocurra; en este caso es la proporción de veces que hay entre ser beneficiario de oportunidades y no serlo, el valor que toma es $\psi = 0.275$, y se interpreta como 0.275:1, es decir, dado un efecto aparece ante la presencia de otra variable es 275/1000 menos que si esta variable no está presente.

Y la respectiva función Log-Cociente de Verosimilitudes (*Log-Odds-Ratio Function*)*:*

$$\mathrm{Ln}(\psi) = \mathrm{Ln}\left[\frac{\left\{\frac{\pi(1,0,0,0,0)}{1-\pi(1,0,0,0,0)}\right\}}{\left\{\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)}\right\}}\right] = \mathrm{Ln}(0.275) = -1.29$$

$$= -1.61 - (-0.32) = g(1,0,0,0,0) - g(0,0,0,0,0)$$

Es necesario remarcar que se cumple la igualdad, que se habló en el marco teórico de la RLM, $\mathrm{Ln}(\psi) = g(1,1,1,1,1) - g(0,0,0,0,0) = -1.29$[153].

La misma interpretación sigue los siguientes casos.

[152] Véase de la tabla resumen del modelo de RLM ajustado.
[153] Dicha cantidad es idéntica a la obtenida en la tabla de resumen del modelo RLM ajustado.

- **Caso (0,1,0,0,0):** se presenta cuando **el *Grado de Marginación es Medio*** manteniendo constantes las demás variables. Entonces la función de Momios es:

$$\frac{\pi(0,1,0,0,0)}{1-\pi(0,1,0,0,0)} = \frac{0.25}{0.75} = 0.332$$

La probabilidad condicional de ser beneficiario de Oportunidades dado que se vive en con grado de marginación Medio, manteniendo las demás variables constantes es $\pi(0,1,0,0,0) = 0.25$. Entonces, el complemento de tal probabilidad es el no ser beneficiario de oportunidades dadas las condiciones anteriores es $1-\pi(0,1,0,0,0) = 0.75$. Análogamente, $\frac{\pi(0,1,0,0,0)}{1-\pi(0,1,0,0,0)} = 0.332$ que en el cociente anterior, representa el cociente de ser beneficiario de Oportunidades respecto de no serlo sujeto a vivir en una localidad con GM-Medio, con las demás variables constates.

Ahora, considere el Logaritmo natural de las funciones anteriores, es decir, las funciones Log-Odds:

$$\mathrm{Ln}\left\{\frac{\pi(0,1,0,0,0)}{1-\pi(0,1,0,0,0)}\right\} = \mathrm{Ln}\{0.332\} = -1.1 = g(0,1,0,0,0)$$

Ahora considere la función del Cociente de Odds es:

$$\psi = \frac{\left\{\frac{\pi(0,1,0,0,0)}{1-\pi(0,1,0,0,0)}\right\}}{\left\{\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)}\right\}} = \frac{0.332}{0.7261} = 0.4584$$

Como $\psi = 0.4584$, y se interpreta como 0.4584:1, es decir, dado un efecto aparece ante la presencia de otra variable es 458/1000 menos que si esta variable no está presente.

Y la respectiva función Log-Cociente de Verosimilitudes*:*

$$\mathrm{Ln}(\psi) = \mathrm{Ln}\left[\frac{\left\{\frac{\pi(0,1,0,0,0)}{1-\pi(0,1,0,0,0)}\right\}}{\left\{\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)}\right\}}\right] = \mathrm{Ln}(0.4584) = -0.78$$

$$= -0.78 - (-0.32) = g(0,1,0,0,0) - g(0,0,0,0,0)$$

- **Caso (0,0,1,0,0):** se presenta cuando *el* ***Tipo de localidad es Rural*** manteniendo constantes las demás variables. Entonces la función de Momios es:

$$\frac{\pi(0,0,1,0,0)}{1-\pi(0,0,1,0,0)} = \frac{0.552}{0.448} = 1.233$$

La probabilidad condicional de ser beneficiario de Oportunidades dado que se vive en una localidad rural, manteniendo lo demás constante es $\pi(0,0,1,0,0) = 0.552$. Entonces, el complemento de tal probabilidad es el no ser beneficiario de oportunidades dadas las condiciones anteriores es $1 - \pi(0,0,1,0,0) = 0.448$. Análogamente, $\frac{\pi(0,0,1,0,0)}{1-\pi(0,0,1,0,0)} = 1.233$ que en el cociente anterior, representa el cociente de ser beneficiario de Oportunidades respecto de no serlo sujeto a vivir en una localidad rural, con las demás variables constantes.

El Logaritmo natural de las funciones anteriores, es decir, las funciones Log-Odds:

$$\mathrm{Ln}\left\{\frac{\pi(0,0,1,0,0)}{1-\pi(0,0,1,0,0)}\right\} = \mathrm{Ln}\{1.233\} = 0.21 = g(0,0,1,0,0)$$

Y considere la función del Cociente de Odds es:

$$\psi = \frac{\left\{\frac{\pi(0,0,1,0,0)}{1-\pi(0,0,1,0,0)}\right\}}{\left\{\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)}\right\}} = \frac{1.233}{0.7261} = 1.6989$$

Como $\psi = 1.6989$, y se interpreta como1.6989:1, es decir, dado un efecto aparece ante la presencia de otra variable es 1.6989 más que si esta variable no está presente.

La respectiva función Log-Cociente de Verosimilitudes*:*

$$\mathrm{Ln}(\psi) = \mathrm{Ln}\left[\frac{\left\{\frac{\pi(0,0,1,0,0)}{1-\pi(0,0,1,0,0)}\right\}}{\left\{\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)}\right\}}\right] = \mathrm{Ln}(1.6989) = -0.53$$

$$= 0.21 - (-0.32) = g(0,0,1,0,0) - g(0,0,0,0,0)$$

- **Caso (0,0,0,1,0)**: se presenta cuando **el *Tipo de localidad es Semirural*** manteniendo constantes las demás variables. Entonces la función de Momios es:

$$\frac{\pi(0,0,0,1,0)}{1-\pi(0,0,0,1,0)} = \frac{0.579}{0.421} = 1.377$$

La probabilidad condicional de ser beneficiario de Oportunidades dado que se vive en una localidad semi-rural manteniendo lo demás constante es $\pi(0,0,0,1,0) = 0.579$. Entonces, el complemento de tal probabilidad es el no ser beneficiario de oportunidades dadas las condiciones anteriores es $1 - \pi(0,0,0,1,0) = 0.421$. Análogamente, $\frac{\pi(0,0,0,1,0)}{1-\pi(0,0,0,1,0)} = 1.377$ que en el cociente anterior, representa el cociente de ser beneficiario de Oportunidades respecto de no serlo sujeto a vivir en una localidad semi-rural manteniendo todo constante.

El Logaritmo natural de las funciones anteriores, es:

$$\text{Ln}\left\{\frac{\pi(0,0,0,1,0)}{1-\pi(0,0,0,1,0)}\right\} = \text{Ln}\{1.377\} = 0.32 = g(0,0,0,1,0)$$

Y la función del Cociente de Odds es:

$$\psi = \frac{\left\{\frac{\pi(0,0,0,1,0)}{1-\pi(0,0,0,1,0)}\right\}}{\left\{\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)}\right\}} = \frac{1.377}{0.7261} = 1.8964$$

Ya que $\psi = 1.8964$, y se interpreta como 1.8954:1, es decir, dado un efecto aparece ante la presencia de otra variable es casi dos veces más que si esta variable no está presente.

La respectiva función Log-Cociente de Verosimilitudes*:*

$$\text{Ln}(\psi) = \text{Ln}\left[\frac{\left\{\frac{\pi(0,0,0,1,0)}{1-\pi(0,0,0,1,0)}\right\}}{\left\{\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)}\right\}}\right] = \text{Ln}(1.8964) = 0.64$$

$$= 0.32 - (-0.32) = g(0,0,0,1,0) - g(0,0,0,0,0)$$

- **Caso (0,0,0,0,1):** se presenta cuando el ***índice de Percepción de Impacto está en la De acuerdo*** manteniendo constantes las demás variables. Por lo tanto la función de Momios es:

$$\frac{\pi(0,0,0,0,1)}{1-\pi(0,0,0,0,1)} = \frac{0.332}{0.668} = 0.4966$$

La probabilidad condicional de ser beneficiario de Oportunidades dado que está de acuerdo con el IPI, manteniendo lo demás constante es $\pi(0,0,0,0,1) = 0.332$. Entonces, el complemento de tal probabilidad es el no ser beneficiario de oportunidades dadas las condiciones anteriores es $1-\pi(0,0,0,0,1) = 0.668$. Análogamente, $\frac{\pi(0,0,0,0,1)}{1-\pi(0,0,0,0,1)} = 0.4966$ que en el cociente anterior, representa el cociente de ser beneficiario de Oportunidades respecto de no serlo sujeto a estar de acuerdo en el IPI, manteniendo todo constante.

Luego, el Logaritmo natural de las funciones anteriores, es:

$$\text{Ln}\left\{\frac{\pi(0,0,0,0,1)}{1-\pi(0,0,0,0,1)}\right\} = \text{Ln}\{0.4966\} = -0.7 = g(0,0,0,0,1)$$

Y la función del Cociente de Odds es:

$$\psi = \frac{\left\{\frac{\pi(0,0,0,0,1)}{1-\pi(0,0,0,0,1)}\right\}}{\left\{\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)}\right\}} = \frac{0.4966}{0.7261} = 0.6838$$

Debido a que $\psi = 0.6838$, y se interpreta como 0.6838:1, es decir, dado un efecto aparece ante la presencia de otra variable es 683/1000 menor de veces si esta variable no está presente.

La respectiva función Log-Cociente de Verosimilitudes:

$$\text{Ln}(\psi) = \text{Ln}\left[\frac{\left\{\frac{\pi(0,0,0,0,1)}{1-\pi(0,0,0,0,1)}\right\}}{\left\{\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)}\right\}}\right] = \text{Ln}(0.6838) = -0.38$$

$$= -0.7 - (-0.32) = g(0,0,0,1,0) - g(0,0,0,0,0)$$

La siguiente tabla resume las funciones anteriores para todos los casos.

Tabla 4.15

x	Coeficiente	Momio	Ln(Momio)	Cociente de Momios	Ln(Cociente)	Diferencia De Loggit
(0,0,0,0,0)	-	0.726	-0.32	-	-	-
(1,0,0,0,0)	$B_{(GM(01))}$ = -0.38	0.497	-0.7	0.684	-0.38	-0.38
(0,1,0,0,0)	$B_{(GM(02))}$ = 0.64	1.377	0.32	1.896	0.64	0.64
(0,0,1,0,0)	$B_{(TL(01))}$ = 0.53	1.234	0.21	1.699	0.53	0.53
(0,0,0,1,0)	$B_{(TL(02))}$ = -0.78	0.333	-1.1	0.458	-0.78	-0.78
(0,0,0,0,1)	$B_{(IPI)}$ = -1.29	0.200	-1.61	0.275	-1.29	-1.29

Fuente: Encuesta "Lo que dicen los pobres", 2003. Cálculos propios.

La primera columna establece los valores del vector de entradas $x = \left(x_{GM(01)}, x_{GM(02)}, x_{TL(01)}, x_{TL(02)}, x_{IPI}\right)$; la segunda establece el coeficiente de la variable estimado vía Máxima Verosimilitud en la RLM; la tercera columna tiene calculado el momio (Odds function), $\frac{\pi(x)}{1-\pi(x)}$; la cuarta columna computa el logaritmo natural del Momio; el Cociente de momios es $\psi = \frac{\left\{\frac{\pi(x)}{1-\pi(x)}\right\}}{\left\{\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)}\right\}}$; luego se localiza el logaritmo natural del cociente de momios; finalmente aparece la diferencia de las funciones loggit, $g(x) - g(0,0,0,0,0)$.

Obsérvese que las columnas del *Coeficiente*, el *Logaritmo del Cociente* y la *Diferencia Loggit* son idénticas, esto se debe a la igualdad $B_i = g(x) - g(0,0,0,0,0) = \mathrm{Ln}\left[\frac{\left\{\frac{\pi(x)}{1-\pi(x)}\right\}}{\left\{\frac{\pi(0,0,0,0,0)}{1-\pi(0,0,0,0,0)}\right\}}\right]$, donde $i = GM(01), Gm(02), TL(01), TL(02), IPI$.

4.5.4 Conclusiones del modelo final

El modelo final de Regresión Logística Múltiple es:

$$\pi(\mathrm{x}) = \frac{\mathrm{e}^{\mathrm{g(x)}}}{1+\mathrm{e}^{\mathrm{g(x)}}} = \frac{\mathrm{e}^{-0.32-1.29x_{GM\,(01)}-0.78x_{GM\,(02)}+0.53x_{TL\,(01)}+0.64x_{TL\,(02)}-0.38x_{IPI}}}{1+\mathrm{e}^{-0.32-1.29x_{GM\,(01)}-0.78x_{GM\,(02)}+0.53x_{TL\,(01)}+0.64x_{TL\,(02)}-0.38x_{IPI}}}$$

Los coeficientes del modelo son:

$B_0 = -0.32$, es la constante del modelo. La cual únicamente nos proporciona el término constante de la función Loggit.

$B_{GM(01)} = -1.29$, es el coeficiente de la variable de diseño 01 (GM bajo) del **Grado de Marginación.**

$B_{GM(02)} = -0.78$, es el coeficiente de la variable de diseño 02 (GM medio) del **Grado de Marginación.**

$B_{TL(01)} = 0.53$, es el coeficiente de la variable de diseño 01 (tipo rural) del **Tipo de localidad.**

$B_{TL(02)} = 0.64$, es el coeficiente de la variable de diseño 02 (tipo semi-rural) del **Tipo de localidad.**

$B_{IPI} = -0.38$, es el coeficiente del **Índice de Percepción de Impacto**.

Con estos resultados se termina el capítulo de Modelos Causales. El autor espera que haya sido de ayuda y sirva de referente en la estimación de Regresiones Logísticas Univariadas y Multivariadas. Hago énfasis en que estas últimas aumentan la eficiencia de los datos, llegando a mejores estimaciones que con los modelos que sólo consideran una variable.

Por otro lado, $\pi(\mathrm{x}) = \mathrm{P}[\mathrm{Y} = 1|\mathrm{x}]$ es la probabilidad condicional de ser beneficiario del Programa de Desarrollo Humano Oportunidades dado el vector $x = \left(x_{GM\,(01)}, x_{GM\,(02)}, x_{TL\,(01)}, x_{TL\,(02)}, x_{IPI}\right)$.

Conclusiones Generales

Describir la metodología de la evaluación del impacto de beneficiarios y no beneficiarios de Oportunidades es una tarea innovadora y por lo tanto delicada, pero a la vez reconfortante. Presento en las siguientes líneas las conclusiones generales de trabajo.

El Programa de Desarrollo Humano Oportunidades tiene bien definidas sus características y funciones desde su origen, ocurrido en el año 2002. Tiene unos objetivos claros: la cobertura total en educación, erradicar el analfabetismo, garantizar la cobertura universal de los servicios de salud, equilibrar el desarrollo económico y social con respeto y cuidado del medio ambiente, mejorar el nivel de vida y superar la pobreza extrema. Aún más, Oportunidades es pionero en el mundo en cuanto al diseño y operación de los programas de transferencias monetarias condicionadas bajo la corresponsabilidad del beneficiario.

Sin embargo, como se mencionó en el primer capítulo del presente trabajo, los programas sociales basados únicamente en la selección técnica de beneficiarios pueden acarrear conflictos en las comunidades, al grado que pudieran modificar la estructura social de las comunidades en las que se entregan dichos beneficios.

En particular, la metodología del programa Oportunidades puede generar problemas en aquellos casos en los que se encuentren al umbral de selección[154] establecido por la Coordinación Nacional del programa, ya que familias con pequeñas diferencias en el ingreso pueden tener clasificación diferente, es decir, cercano al umbral por abajo, considerado como potencial beneficiario (pobre) y cercano al umbral por arriba, como no beneficiario (no pobre). Evidentemente, este esquema de selección no afecta a las familias que tienen un ingreso muy superior al umbral, es decir, aquellas familias que ciertamente no necesitan los recursos otorgados por el programa[155].

La encuesta *Lo que dicen los pobres* fue diseñada para conocer las características generales los pobres de patrimonio en México, sus inquietudes, sus opiniones acerca de temas como: bienestar, justicia social, vulnerabilidad, discriminación, sobre las acciones institucionales y la valoración de los apoyos sociales. Los cual resulta excelente, puesto que en este país no se había hecho un esfuerzo de tal magnitud para caracterizar a la población de *Lo que dicen los pobres*.

Si bien, *Lo que dicen los pobres* representa un hito en la investigación de la pobreza en México, las condiciones en las que se vive la pobreza de patrimonio no cambiaran hasta que se comprendan mejor las necesidades que yacen en ellos, de forma tal que el Gobierno mejore el diseño de los programas sociales. Entre las mejoras, es necesario considerar realmente las

[154] No se conoce el umbral de selección de beneficiarios del Programa de Desarrollo Humano Oportunidades.
[155] Las situaciones expuestas anteriormente se encuentran detalladas en "*I. El Programa de Desarrollo Humano Oportunidades*", del presente trabajo.

características, situación, opinión y propuestas de la población a la que están dirigidos los beneficios.

Si bien, la encuesta tiene objetivos muy específicos que coadyuvan al desarrollo del país, no fue diseñada con el objeto de medir temas tales como capital social, percepción de impacto, confianza, etc. Por lo tanto, resulta difícil evaluar el impacto que representa la existencia de un programa de beneficios, como lo es Oportunidades, en la estructura social de las comunidades en las que se ofrecen éstos. En consecuencia, la dificultad adjudicada al trabajo es debida a la carencia de fuentes de información y de referencia[156], restringiendo el alcance de los objetivos del proyecto en el que se inscribe esta tesis. No obstante, mi trabajo está limitado únicamente a la metodología cuantitativa que puede ser utilizada para la evaluación de los impactos.

Aquí es importante mencionar que la metodología es establecida sobre el Cuestionario Individual de *Lo que dicen los pobres* y no sobre el Cuestionario del Hogar. Aunque, incluir variables relacionadas al Hogar en los modelos de Regresión Logística (univariados y multivariados) no sería difícil, debido a que la metodología de la evaluación ya está descrita a lo largo de la presente tesis.

Al caracterizar a la población de *Lo que dicen los pobres* respecto de la población nacional en el año 2000, (INEGI, 2000), se encuentran diferentes hechos:

Los años estudiados por la población estudiada en la encuesta, en promedio, es menor que los años estudiados de la población nacional, nueve de cada diez encuestados no tiene estudios o cursó hasta la secundaria, mientras que a nivel nacional se mantiene en tres de cada cuatro personas; definitivamente la escolaridad alcanzada por una persona está relacionada con el grado de estudios del padre y de la madre. Para los pobres de patrimonio la madre tiene menor escolaridad que el padre.

La condición laboral entre ambas poblaciones es similar, sin embargo, se establece en contextos diferentes, teniendo lo pobres desventajas más acentuadas, por ejemplo: en un plazo de diez años, el principal temor de los pobres es el desempleo, incluso encima de la misma muerte; siete de cada diez pobres son empleados, obreros, jornaleros o peones. En general, el 90% de los trabajadores no cuentan con algún tipo de prestación de su empleo. Están conscientes que al perder su trabajo sería al menos difícil (73%) encontrar uno nuevo aún cuando el 50% e siente seguro en él.

En cuanto a los indicadores e índices fue necesario establecer definiciones y diferencias entre ambos. Así un indicador es una variable observable, que depende de un conjunto de otras variables, y que tiene como finalidad hacer observables las características de éstas. Muchas de las veces tales características no son evidentes, mientras que se entiende por índice, un número que resume la información proporcionada por un conjunto de indicadores.

[156] Ejemplo de fuentes de información no disponibles de manera pública son: la parte de ingresos del Cuestionario Individual y la Base de Datos den Cuestionario del Hogar de *Lo que dicen los pobres.*
Por el lado de las fuentes de referencia, no se conoce de manera pública el umbral de ingreso que considera la Coordinación Nacional para clasificar a las familias beneficiarias y no beneficiarias de Oportunidades.

Se expusieron tres técnicas estadísticas para la construcción de índices: **Rangos Sumados, Rangos Sumados Ponderados y Análisis de Componentes Principales**. El uso de cada una está en función del tipo de variables que se tenga, la información disponible y los objetivos que persiga el índice. Al utilizar los métodos de suma (los primeros dos) se puede dar seguimiento a las transformaciones que sufren los indicadores al sumarlos, logrando saber fácilmente los cambios ocurridos en éstas. Por el contrario, al trabajar con un índice generado a través del ACP se complica la identificación de cada uno de los indicadores (componentes) que interactúan en el proceso; de manera general, bajo esta técnica se obtienen índices en escala continua, los cuales socialmente conllevan una interpretación difícil. Para solucionar el problema anterior, se decidió utilizar **la técnica de Estratificación Óptima de Dalenius y Hodges**, que no es más que construir un índice categórico donde cada categoría tiene homogeneidad de la variabilidad entre sus elementos, y mantenga una heterogeneidad de la variabilidad entre las demás categorías.

Los índices e indicadores construidos en el capítulo 3 son: **Índice de Confianza Generalizada, Índice de Confianza Generalizada Ponderado, Índice de Percepción de Impacto, Indicador de Tipo de localidad, Índice de Grado de Marginación e Indicador de Beneficiario de Oportunidades**. El último de la lista sirve como variable dependiente en los modelos causales expuestos en el capítulo V del presente trabajo.

Se decide utilizar el ser o no beneficiario de Oportunidades como variable dependiente de los modelos de los modelos causales ajustados en el capítulo anterior debido a los objetivos perseguidos por el Proyecto de Investigación del cual se describe su metodología.

Al estudiar los modelos causales disponibles para la evaluación de la percepción de los beneficiarios y no beneficiarios, y bajo la variable dependiente ya expuesta (Indicador de Beneficiario de Oportunidades), se encuentran dos posibles soluciones estadísticas al problema: el Análisis Discriminante y el Análisis de Regresión Logística. La primera opción para ser tentadora, pero resulta de difícil aplicación en este trabajo debido al no cumplimiento de los supuestos que tal técnica necesita (principalmente la Normalidad de las variables).

Finalmente la decisión radica en el Análisis de Regresión Logística Univariado y Multivariado por la facilidad en el cumplimiento de sus supuestos. Por lo tanto, con las variables e índices generados en el Capítulo 3, se ajustan modelos univariados (uno para cada índice o indicador) y un modelo multivariado al final, bajo la precaución de los resultados producidos por el análisis univariado.

La interpretación del modelo está basada en estimar la probabilidad de pertenencia a un grupo, en este caso ser beneficiario de Oportunidades o no serlo, restringido a algunas variables independientes. También existe otra interpretación del modelo por medio de encontrar los pesos adecuados para los factores que determinan la ocurrencia de un evento (en este caso ser beneficiario de Oportunidades), sin embargo no se utiliza ya que no se está determinando las condiciones que debe satisfacer una familia para ser beneficiaria del programa (es responsabilidad de la coordinación Nacional del Programa hacerlo). Luego, la interpretación del modelo es la

probabilidad de ser beneficiario o no beneficiario bajo algunas circunstancias como tipo de localidad, grado de marginación y Percepción de Impacto.

Anexos técnicos

1 Variables ficticias o de diseño

Sea X una variable no métrica, nominal u ordinal, con 2 categorías. Denótese la j-ésima categoría por x_j con j=1,2.

También suponga que si ocurre $x_1$no puede ocurrir x_2 y viceversa, es decir, son complementos una de otra, pero además forman una partición de la variable X.

Supóngase que se escoge una categoría k de referencia, para alguna k=1,2, es decir, k = 1 ó k=2. Para construir una variable de diseño, se considera la categoría de referencia x_k, se le asigna el valor de cero y a la categoría que no es de referencia (la que no es x_k, en este caso) se le asigna el valor de 1.

Supongamos sin pérdida de generalidad, que k=2. Por lo tanto, la categoría que es de referencia es X_2. En consecuencia, se codifica la variable X en otra variable Y tal que: $x_1 = 1$ y $x_2 = 0$. En lenguaje matemático:

$$Y = x_1 I_A + x_2 I_B$$

Donde I_A es a función indicadora sobre el evento A.

A= es el evento de que ocurra 1

B = es el evento de que ocurra 2

Obsérvese, que no puede ocurrir simultáneamente A y B, porque 1 y 2 por hipótesis generan categorías que son una partición de la variable X.

En consecuencia, $0 \leq Y = x_1 I_A + x_2 I_B \leq 1$, tomando en la imagen sólo los valores 1 y 0, es decir, se tiene una variable Y que es dicotómica.

La probabilidad de que ocurra x_1 es p y la de x_2 es 1-p (por ser complementos). Por lo tanto, Y tiene una distribución de probabilidad Bernoulli[157] con parámetro p.

Obsérvese que X, variable dicotómica, genera una variable Y dicotómica.

Al generalizar estás ideas, se tiene que, si X es una variable no métrica, con n categorías denotadas por $x_1, x_2, \ldots, x_n$.

[157] La distribución Bernoulli es denotada por Ber(p), donde p es el único parámetro y está dado por la probabilidad de éxito. Su media esperanza es E[X]=p y la varianza Var(X)=p(1-p), esto siempre que x~ Ber(p). Es utilizada para modelar variables no métricas de tipo dicotómico.

Considérese $k_1, k_2, \dots, k_{n-1}$ índices de las categorías de$x_1, x_2, \dots, x_{n-1}$, en donde k_j cumple las condiciones de abajo. Evidentemente $k < n$.

Suponga sin pérdida de generalidad, que la categoría de referencia es n.

Para cada k_i, se construye la variable ficticia Y_{k_i}, de la misma forma que en el caso anterior. Entonces, $x_{k_i} = 1$, para alguna i= 1,…,n-1, y $x_{k_j} = 0$ para i≠j .

Nótese lo siguiente, como $x_1, x_2, \dots, x_n$ son una partición de X, entonces x_{k_1}, $x_{k_2}, \dots$, x_{k_n}, también es una partición de Y. Por lo tanto, Y_{k_i} está dada por

$$Y_{k_i} = x_{k_1} + x_{k_2} + \dots, + x_{k_n}$$

Y de aquí se concluye que, sólo puede ocurrir $x_{k_i} = 1$ ó $x_{k_j} = 0$ para toda $i \neq j$, entonces al sumar, cada una de lasx_{k_i} y la x_{k_j}, se tiene $0 \leq Y_{k_j} \leq 1$ tomando los extremos en su imagen. Es decir, $x_{k_j} = 0$ ó $x_{k_j} = 1$.

Esto se hace para cada i=1,…, n-1. Se obtienen Y_{k_1}, $Y_{k_2}, \dots$, $Y_{k_{n-1}}$ variables ficticias.

Obsérvese que si X tiene n categorías, se obtienen n-1 variables ficticias respecto una variable de referencia, en este caso, supuesta sin pérdida de generalidad, n.

2 Alfa de Cronbach[158]

Permite cuantificar el nivel de fiabilidad de una escala de medida para la magnitud inobservable construida a partir de las k variables observadas; no deja de ser una media ponderada de las correlaciones entre las variables que forman parte de la escala.

Se calcula de dos formas: a partir de las varianzas, Alfa de Cronbach, o de las correlaciones de los ítems, Alfa de Cronbach Estandarizado. Ambas fórmulas son equivalentes.

La fórmula del Alfa, en función de las varianzas, es:

$$\alpha = \left(\frac{k}{k-1}\right)\left[1 - \frac{\sum_{i=1}^{n} S_i^2}{S_t^2}\right]$$

Donde

S_i^2 es la varianza de la variable i.

S_t^2 es la varianza de la suma de todas las variables.

[158] Bogliaccini, Juan; Cardoso, Manuel; Rodríguez, Federico; "Construcción de índices de Confiabilidad: Alfa de Cronbach"; Facultad de Ciencias Humanas, Universidad Católica de Uruguay; Autotexto No. 5, Serie Estadística, 13 páginas, 2005.

k es el número de variables.

Por otro lado, la fórmula del Alfa, en función de las correlaciones, es:

$$\alpha_{Est.} = \frac{kp}{1 + p(k-1)}$$

Donde

k es el número de variables.

p es el promedio de las correlaciones lineales entre cada una de las variables.

Es necesario, para crear una escala confiable, que las variables estén altamente correlacionadas entre sí. Observe que el nivel máximo de correlación se alcanza cuando las variables $X_1, X_2, \dots, X_k$ son iguales, es decir, $X_1 = X_2 = \dots = X_k$. Luego se tiene:

$$S_t^2 = k^2 S_1^2$$

En consecuencia, el valor del Alfa de Cronbach es igual a 1, $\alpha = 1$.

Por otro lado, si las variables fuesen independientes entre sí (por lo que no podrían constituir conjuntamente una escala fiable), entonces se tendría que:

$$S_t^2 = \sum_{i=1}^{k} S_1^2$$

Logrando un valor de Alfa de 0, $\alpha = 0$.

El Alfa de Cronbach puede tener valores negativo, esto en caso de existir parejas de variables negativamente correlacionadas. También, recuerde que esta Alfa no es un estadístico asociado a una prueba de hipótesis, por lo tanto no tiene asociado algún p-valor que permita rechazar la hipótesis de fiabilidad en la escala. No obstante, cuanto más se aproxime a su valor máximo, 1, mayor es la fiabilidad de la escala. Según la experiencia, se recomiendan valores del alfa superiores a 0.7 o 0.8, con lo cual se garantiza la fiabilidad de la escala..

3 Método de Estratificación Óptima de Dalenius y Hodges[159]

El objetivo de la técnica es establecer rangos a partir de un conjunto de datos estandarizados por medio de varias operaciones: ordenar, clasificar y agrupar un conjunto de datos dispersos. También sirve para obtener desviaciones y factores de proyección, y así determinar tamaños de muestra, proyecciones estadísticas, etc.

Por medio de la técnica se garantiza una buena agrupación, minimizando con esto la varianza. Sea x un conjunto de observaciones con $x = \{x_1, x_2, \dots, x_n\}$. Los pasos a seguir son los siguientes:

1. Ordenar ascendentemente las observaciones x_i.
2. Transformar los datos ordenados a una escala de 0-100, obteniéndose los valores de la variable Y mediante la transformación:

$$y_i = \frac{100}{Máx(x) - Mín(x)}[x_i - Mín(x)]$$

Donde

$Máx(x)$ = valor máximo de las observaciones $x_i, i = 1,2, \dots, n.$

$Mín(x)$ = valor mínimo de las observaciones $x_i, i = 1,2, \dots, n.$

x_i = observación i-ésima con $i = 1,2, \dots, n.$

Evidentemente se pide que $Máx(x) - Mín(x) \neq 0.$

3. Dividir el rango de o a 100 en 10 intervalos de igual longitud, esto es,

$$[0,10], [10,20], \dots, [80,90], [90,100]$$

4. Obtener las frecuencias de observaciones de cada uno de los intervalos

$$F_1, F_2, \dots, F_{9,} F_{10}$$

5. Calcular la raíz cúbica de la frecuencia para cada intervalo.

$$\sqrt[3]{F_i} \quad para\ i = 1,2, \dots ,10$$

6. Acumular los valores obtenidos en el punto anterior, sea T su acumulación:

$$T = \sum_{i=1}^{10} \sqrt[3]{F_i}$$

7. Dividir el valor acumulado, T, del paso anterior entre el número de estratos N que se desean formar para obtener la longitud del límite óptimo para los estratos

$$A = \frac{T}{N}$$

Donde

[159] Instituto Nacional de Estadística y Geografía; Información Referenciada Geoespacialmente Integrada en un Sistema, IRIS 3.0. Anexo de Estratificación Óptima de Dalenius-Hodges.

A = Longitud del límite óptimo de los estratos

N = Número de estratos que se desean formar (para N >2).

8. Los límites óptimos de los estratos son:

$$L_1 = A, L_2 = 2A, \dots, L_N = T$$

9. Para marcar los puntos de corte para los estratos, se comparan los valores de los límites óptimos con los de la acumulada T; se inicia marcando el valor de T, más cercano al primer límite óptimo, luego el más cercano al segundo límite omitiendo los marcados anteriormente, y así sucesivamente.
10. Una vez extraídos los intervalos, los cuales son cerrados, se puede regresar a los valores originales y así obtener los límites superior e inferior de cada estrato, además de las frecuencias de cada una de los estratos.

4 Prueba de la χ^2 para la asociación

Considérese una tabla de contingencia (también llamada tabla cruzada) de rxc originada por las variables no métricas x y y.

La variable x tiene las categorías $x_1, x_2, \dots, x_r$. En total hay r categorías. La variable y tiene las categorías $y_1, y_2, \dots, y_c$. En total hay c categorías. Entonces, la tabla cruzada tiene rc celdas.

Las hipótesis asociadas a la prueba de asociación son:

Ho: x no está asociada con y vs **Ha:** x está asociada con y

Las cuales son equivalentes a:

Ho: x es independiente de y vs **Ha:** x no es independiente de y

En general, si θ_{ij} es la probabilidad de que un elemento quede en la celda perteneciente al i-ésismo renglón y a la j-ésima columna. θ_i. es la probabilidad de que un elemento quede en el i-ésimo renglón. $\theta_{.j}$ es la probabilidad de que un elemento quede en la j-ésima columna.

La hipótesis nula que se desea probar es $\theta_{ij} = \theta_{i.}\theta_{.j}$. Con lo cual, las hipótesis de la prueba son:

Ho: $\theta_{ij} = \theta_{i.}\theta_{.j}\ \forall i, j$ vs **Ha:** $\theta_{ij} \neq \theta_{i.}\theta_{.j}\ para\ algunas\ i, j$

Donde $i = 1,2, \dots, r$ y $j = 1,2, \dots, c$.

La frecuencia observada de la celda del i-ésimo renglón y la j-ésima columna se denota por f_{ij}. Los totales por renglones son $f_{i.}$. Los totales por columnas con $f_{.j}$. El gran total es denotado por f.

$$f = \sum_{i=1}^{r} f_{i.} = \sum_{j=1}^{c} f_{j.}$$

Con esta notación, las probabilidades estimadas de $\theta_{i.}$ y $\theta_{.j}$ son:

$$\dot{\theta}_{.j} = \frac{f_{.j}}{f} \qquad \text{y} \qquad \dot{\theta}_{i.} = \frac{f_{i.}}{f}$$

Las probabilidades anteriores también son conocidas como las marginales para las columnas y renglones, respectivamente.

Y con la hipótesis nula de independencia se obtiene,

$$e_{ij} = \dot{\theta}_{.j} \, . \, \dot{\theta}_{i.} \, . f = \frac{f_{.j}}{f} \, . \frac{f_{i.}}{f} \, . f = \frac{f_{i.} f_{.j}}{f}$$

Para la frecuencia esperada de la celda esperada en el i-ésimo renglón y j-ésima columna. Obsérvese que e_{ij} se obtiene multiplicando el total de la columna por el total del renglón y dividiéndolo entre el gran total.

El estadístico χ^2 se define como:

$$\chi^2 = \sum_{\mathrm{i}=1}^{\mathrm{r}} \sum_{\mathrm{j}=1}^{\mathrm{c}} \frac{\left(f_{ij} - e_{ij}\right)^2}{e_{ij}}$$

$$= \sum_{\mathrm{i}=1}^{\mathrm{r}} \sum_{\mathrm{j}=1}^{\mathrm{c}} \frac{\left(f_{ij} - \frac{f_{i.} f_{.j}}{f}\right)^2}{\frac{f_{i.} f_{.j}}{f}}$$

Resulta que $\chi^2 \sim \chi^2_{\alpha,(r-1)(c-1)}$. Donde α es el error tipo I, y $(r-1)(c-1)$ son los grados de libertad de la distribución de probabilidad de la χ^2.

Se rechaza la hipótesis nula si $\chi^2 > \chi^2_{\alpha,(r-1)(c-1)}$.

Como la estadística de prueba descrita sólo tiene, aproximadamente, una distribución $\chi^2_{(r-1)(c-1)}$ con $(r-1)(c-1)$ grados de libertad, se acostumbra a utilizar la prueba sólo cuando ninguna de las e_{ij} sea menor que 5, es decir, si $e_{ij} \geq 5$ para todo $i = 1,2, \ldots, r$ y $j = 1,2, \ldots, c$. En el caso que no se cumpla la condición anterior, se sugiere combinar algunas de las celdas con una pérdida correspondiente al número de grados de libertad.

Bibliografía

ANGULO REYES, YVON. “Evaluación del impacto del programa Oportunidades en la confianza y generación de redes sociales (proyecto de investigación)”, mimeo, s/f. (Angulo, a)

ANGULO REYES, YVON. “Marginalidad, redes sociales y políticas públicas, sustentado en la elaboración de modelos estadísticos jerárquicos (macro-micro) de explicación multivariable”, mimeo, s/f. (Angulo, b)

BELÁUSTEGUI, Carrera, Mario. “*Oportunidades: para vivir mejor*”, Julio 2009. Coordinación Nacional del Programa de Desarrollo Humano Oportunidades. (Beláustegui, 2009)

BAUER, RAYMOND “1966: Social Indicators”, Cambridge, Mass: the M.I.T. Press. (Bauer, 1966)

BOGLIACCINI, Juan; Cardoso, Manuel; Rodríguez, Federico; “Construcción de índices de Confiabilidad: Alfa de Cronbach”; Facultad de Ciencias Humanas, Universidad Católica de Uruguay; Autotexto No. 5, Serie Estadística, 13 páginas, 2005. (Bogliaccini, 2005)

BRICKELL, F.; 1974. “Matrices y Espacios Vectoriales”. Editorial LIMUSA. (Brickell, 1974).

CAMACHO ROSALES, JUAN (2009). “Estadística con SPSS para Windows, versión 12”, Alfaomega Grupo Editor. México, D.F. (Camacho, 2009)

CONSEJO DE EVALUACIÓN PARA LA POLÍTICA DE DESARROLLO SOCIAL; “Evolución de la pobreza en México 2009”, CONEVAL, 24 páginas. México, D.F. (CONEVAL, 2009)

CONSEJO NACIONAL DE POBLACIÓN (2005). “Anexo C de la Metodología de Estimación del Índice de Marginación a nivel localidad”, México, D.F. (CONAPO, 2005)

DALENIUS, TORE; HODGES, JOSEPH (1959). “Minimun Variance Stratification”, 5ª Journal of the American Statistical Association 88. March 1959. (Dalenius, 1959)

DIARIO OFICIAL DE LA FEDERACIÓN, “Decreto por el que se reforma el diverso por el que se crea la Coordinación Nacional del Programa de Educación, Salud, y Alimentación como órgano desconcentrado de la Secretaría de Desarrollo Social”, 6 de marzo de 2002, México, D.F. (DOF, 2002)

DIARIO OFICIAL DE LA FEDERACIÓN, “Ley General de Desarrollo Social”, 20 de enero de 2004, México, D.F. (DOF, 2004)

DIARIO OFICIAL DE LA FEDERACIÓN, “Acuerdo por el que se emiten las reglas de operación del Programa de Desarrollo Humano oportunidades”, 31 de diciembre de 2010, México, D.F. (DOF, 2010)

DURSTON, JOHN (2000). “¿Qué es el capital social comunitario?”, Serie Políticas sociales, no. 38, CEPAL, Chile, julio. (Durston, 2000)

GROSSMAN, STANLEY (1994). “Elementary Linear Algebra with Applications”. Saunders College Publishing. (Grossman, 1994)

HOSMER, DAVID; LEMESHOW, STANLEY (1989). “Applied Logistic Regression”, John Wiley & Sons Inc. United States of America. (Hosmer, 1989)

J.F. HAIR, JR. R.E, ANDERSON, R.L. TATHAM, W.C. BLACK (1999). “Análisis Multivariante”, 5a ed., Prentice Hall Iberia, Madrid. (Hair, 1999)

KHANDER, SHAHIDUR; GAYATRI, KOOLWAL; HUSSAIN, SAMAD (2010). “Handbook on impact evaluation: quantitative methods and practices”. The World Bank. Washington, D.C. (Khander, 2010)

KREYSZING, ERWIN (1973). “Introducción a la Estadística Matemática, Principios y Métodos”, Editorial LIMUSA, Novena reimpresión 1987. (Kreyszing, 1973)

LANE, CHRISTINE (Reinhard Bachman, comps.) (1998). “Trust within and between Organizations: Conceptual Issues and Empirical Applications”, Oxford University Press, Inglaterra. Citado en LUNA y VELASCO, 2005. (Lane, 2005)

LAZARFELD, P.F., (1958). “Evidence and Inference in “Social Research” Daedalus”, vol. 87, no. 4, pp. 99-130. (Lazarfield, 1958)

LUJÁN PONCE, NOEMÍ (1999). La construcción de confianza política, IFE: colección “Temas de la Democracia”, serie “Ensayos”, no.6, México. (Luján, 1999)

LUNA, MATILDE (José Luis Velasco) (2005). “Confianza y desempeño en las redes sociales”, Revista Mexicana de Sociología, vol. 67, no. 1, enero-marzo. (Luna y Velasco, 2005)

MARSDEN, JEROLD; TROMBA, ANTHONY (2004). “Cálculo Vectorial”. Quinta edición, Pearson educación, S.A., Madrid. (Marsden, 2004)

MÉNDEZ, IGNACIO; ESLAVA, GUILLERMINA; ROMERO, PATRICIA (2004). “Conceptos básicos de muestreo”. Monografías, Volumen 12, No. 27. (Méndez, 2004)

NARAYAN, 1999; Deepa; Patel, Raj; Schafft, Anne; Koch-Schulte, Sarah*. “Can anyone hear us? Voices from 47 countries”.* Poverty Group, PREM. The World Bank. (Narayan, 1999)

ORGANIZACIÓN DE LAS NACIONES UNIDAS (1975). “1975: Sistema de Estadísticas Sociales y Demográficas (SESD). Proyecto de normas sobre los indicadores sociales”. Versión en castellano del documento de Naciones Unidas, Towards a System of Social and Demographic Statistic. (ONU, 1975)

(Oportunidades, 2011)

SITIO ELECTRÓNICO DE LA REAL ACADEMIA DE LA LENGUA ESPAÑOLA; www.rae.es
(RAE, 2011)

SITIO ELECTRÓNICO DE ZONA ECONÓMICA; http://www.zonaeconomica.com/costo-de-oportunidad
(Zona económica, 2011)

Bases de datos

XII CENSO DE POBLACIÓN Y VIVIENDA 2000. Instituto Nacional de Información, Estadística y Geografía. Levantado entre el 7 y 18 de febrero de 2000. (INEGI, 2000)

ENCUESTA NACIONAL DE INGRESO GASTO DEL HOGAR 2000 (ENIGH 2000). Levantada por INEGI del 10 de agosto al 17 de noviembre de 2000. (INEGI b, 2000)

II CONTEO DE POBLACIÓN Y VIVIENDA 2005. Instituto Nacional de Información, Estadística y Geografía, 2005. (INEGI, 2005)

ENCUESTA "LO QUE DICEN LOS POBRES, <<CUESTIONARIO INDIVIDUAL>>" (2003), Secretaría de Desarrollo Social. Diseñada por la SEDESOL, levantada por Ipsos-Bimpsa, Julio de 2003, México, D.F. (SEDESOL, 2003)

GRADO DE MARGINACIÓN POR LOCALIDAD 2005. Consejo Nacional de Población. Estimaciones hechas por la CONAPO con información del II Conteo de Población y Vivienda 2005. (CONAPO, 2005)

Software utilizado

STASTISTICAL PACKAGES FOR THE SOCIAL SCIENCES para Windows, Versión 13. SPSS Inc., 1989-2004. 1 de septiembre de 2004.

MICROSOFT OFFICE, EXCEL 2007.

MICROSOFT OFFICE, WORD 2007.

ADOBE READER 9.