



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

---

FACULTAD DE CIENCIAS

Modelo de puntaje estadístico para la evaluación de la  
pobreza en México 2008-2014.

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuaria

PRESENTA:

Jacqueline Ruiz Hernández

TUTOR:

M. en F. Jorge Luis Reyes García





Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**1.Datos del alumno.**

Ruiz

Hernández

Jacqueline

57808100

Universidad Nacional Autónoma de México

Facultad de Ciencias

Actuaría

304195433

**2.Datos del tutor.**

M en F.

Jorge Luis

Reyes

García

**3.Datos del sinodal 1.**

Act.

Ángel Manuel

Godoy

Aguilar

**4.Datos del sinodal 2.**

Act.

María Teresa

Velázquez

Uribe

**5.Datos del sinodal 3.**

Act.

Carlos

Contreras

Cruz

**6.Datos del sinodal 4.**

Act.

Jonathan

González

Salgado

**7.Datos del trabajo escrito.**

Modelo de puntaje estadístico para la evaluación de la pobreza en México 2008-2014.

114 p

2016

# Índice general

<b>1. Marco teórico</b>	<b>1</b>
1.1. ¿Qué es la pobreza?	1
1.1.1. Breve perfil de las políticas de combate de pobreza en México 1988-2012	2
1.2. Definición y medición de la pobreza actualmente en México	7
1.2.1. Enfoques para la medición de la pobreza en México	8
1.2.2. Método para la medición multidimensional de la pobreza con- siderando los enfoques establecidos por el CONEVAL	9
1.3. Criterios para la definición de los indicadores de pobreza	13
1.4. Medición del espacio de carencias sociales	13
1.4.1. Rezago educativo	13
1.4.2. Acceso a los servicios de salud	14
1.4.3. Acceso a la seguridad social	15
1.4.4. Calidad y espacios de la vivienda	17
1.4.5. Acceso a los servicios básicos en la vivienda	19
1.4.6. Acceso a la alimentación	20
1.4.7. Grado de cohesión social	22
1.5. Criterios para la definición de las líneas de bienestar y bienestar mínima	23
1.5.1. Ingreso corriente	24
<b>2. Modelo de regresión logística</b>	<b>25</b>
2.1. Nociones preliminares	25
2.2. Ajuste del modelo	26

2.3.	Modelo logístico Logit . . . . .	27
2.4.	Estimación de los parámetros . . . . .	28
2.5.	ODDS Ratio y Riesgo Relativo . . . . .	30
2.6.	Pruebas estadísticas al modelo logit . . . . .	31
2.6.1.	Deviance . . . . .	31
2.6.2.	Estadístico de Wald . . . . .	33
2.6.3.	Residuos de Pearson . . . . .	33
2.6.4.	Variabes DUMMY en regresión logística . . . . .	34
<b>3.</b>	<b>Análisis y reducción de variables bajo la perspectiva de los modelos de score</b>	<b>36</b>
3.1.	Análisis Bivariado de los datos: Tablas de Contingencia . . . . .	37
3.2.	Prueba Chi-cuadrada . . . . .	39
3.3.	Pesos de Evidencia (WOE) . . . . .	39
3.4.	Valor de la información (IV) . . . . .	40
3.5.	Coefficientes de correlación . . . . .	41
3.6.	Análisis de Cluster . . . . .	43
3.6.1.	Medidas de Similitud . . . . .	43
3.6.2.	Metédos Cluster . . . . .	44
3.7.	Kolmogorov Smirnov . . . . .	49
3.8.	Índice de Gini . . . . .	50
<b>4.</b>	<b>Un modelo de Score para medir Pobreza en México</b>	<b>52</b>
4.1.	Análisis exploratorio de los datos . . . . .	57
4.1.1.	Selección de variables . . . . .	58
4.1.2.	Análisis Bivariado de los datos : Tablas de Contingencia . . . . .	59
4.1.3.	Coefficientes de correlación . . . . .	62
4.1.4.	Análisis de Cluster . . . . .	64
4.2.	Definición de la variable respuesta y variables explicativas . . . . .	70
4.3.	Ajuste del modelo de regresión lógistica . . . . .	72
4.4.	Construcción del Score . . . . .	78
4.5.	Validación del Score . . . . .	79

4.6. Aplicaciones del modelo de puntaje estadístico . . . . .	82
<b>5. Conclusiones</b>	<b>91</b>
<b>Referencias</b>	<b>95</b>
<b>A. Análisis Exploratorio de los datos</b>	<b>98</b>

# Índice de Figuras

1.1. Índice de privación social . . . . .	11
1.2. Población en situación de pobreza multidimensional . . . . .	12
3.1. Gráfico Índice de Gini . . . . .	50
4.1. Población en situación de pobreza multidimensional . . . . .	53
4.2. Dendograma variables empleadas en la construcción por Indicador de Pobreza .	67
4.3. Dendograma variables empleadas en la construcción por Indicador de Pobreza después de selección por IV. . . . .	68
4.4. KS. . . . .	81
4.5. Curva ROC. . . . .	82
4.6. Gasto en funciones de Desarrollo Social en México 1990 - 2012 . . . . .	83
4.7. Porcentaje de personas según tipo de pobreza por ingreso: 1992 - 2014 . . . . .	84
4.8. Distribución de la población acorde al puntaje estadístico . . . . .	86
4.9. Distribución rango 5 por entidad federativa. . . . .	87
4.10. Servicios de Salud. . . . .	88
4.11. Distribución de la población en Puebla rango 5 por grupos de edades 2008- 2014	89
4.12. Distribución de la población en Puebla rango 5 por sexo 2008 - 2014 . . . . .	89



# Índice de Tablas

3.1. Estructura: Tabla de Contingencia . . . . .	38
3.2. Escala de Correlaciones . . . . .	42
4.1. Distribución de la población por ingreso . . . . .	55
4.2. Diagrama distribución base de datos respecto a la variable pobreza. . . . .	56
4.3. Lista de variables empleadas en la construcción por Indicador de Pobreza . . . . .	57
4.4. Tabla de Contingencia: Inasistencia a la escuela, Nivel Educativo . . . . .	60
4.5. Pearson Chi Square . . . . .	60
4.6. Symmetric Measures . . . . .	61
4.7. Coeficientes de Correlación de Pearson. . . . .	63
4.8. Lista de variables empleadas en la construcción por Indicador de Pobreza . . . . .	65
4.9. Tabla de cluster . . . . .	67
4.10. Lista de variables empleadas en la construcción por Indicador de Pobreza . . . . .	70
4.11. Categoría de referencia para cada variable . . . . .	71
4.12. Log likelihood. Primer ajuste al modelo de Regresión . . . . .	72
4.13. Variables en la Ecuación . . . . .	73
4.14. Log likelihood. Ajuste Final al modelo de Regresión . . . . .	73
4.15. Prueba Chi -square sobre los coeficientes del modelo . . . . .	74
4.16. Distribución de individuos Pobres y no Pobres por Rango de Puntaje . . . . .	80
4.17. Diagrama distribución base de datos respecto a la variable pobreza. . . . .	86
A.1. Tabla de Contingencia edad vs pobreza. . . . .	99
A.2. Tabla de Contingencia inasistencia a la escuela vs pobreza. . . . .	99
A.3. Tabla de Contingencia nivel educativo vs pobreza. . . . .	100

A.4. Tabla de Contingencia servicio de salud vs pobreza. . . . .	100
A.5. Tabla de Contingencia seguridad social vs pobreza. . . . .	100
A.6. Tabla de Contingencia acceso directo a la seguridad social a través del jefe del hogar vs pobreza. . . . .	101
A.7. Tabla de Contingencia acceso directo a la seguridad social a través del cónyuge del jefe del hogar vs pobreza. . . . .	101
A.8. Tabla de Contingencia acceso directo a la seguridad social a través de los hijos vs pobreza. . . . .	101
A.9. Tabla de Contingencia programa adulto mayores vs pobreza. . . . .	102
A.10. Tabla de Contingencia servicio médicos por otros núcleos familiares vs pobreza. . . . .	102
A.11. Tabla de Contingencia indicador de carencia del material de piso de la vivienda vs pobreza. . . . .	102
A.12. Tabla de Contingencia indicador de carencia del material del techo de la vivienda vs pobreza. . . . .	102
A.13. Tabla de Contingencia indicador de carencia del material del muros de la vivienda vs pobreza. . . . .	103
A.14. Tabla de Contingencia carencia de acceso al agua vs pobreza. . . . .	103
A.15. Tabla de Contingencia carencia de servicio de drenaje vs pobreza. . . . .	103
A.16. Tabla de Contingencia carencia de servicio de electricidad vs pobreza. . . . .	103
A.17. Tabla de Contingencia grado de inseguridad alimentaria vs pobreza. . . . .	104
A.18. Correlación variable independiente vs la variable Pobreza . . . . .	105

# INTRODUCCIÓN

Antes del 2002, México no contaba con una metodología oficial de medición de la pobreza. En 2004 con la promulgación de la Ley de Desarrollo Social (LGDS), la medición de la pobreza en México inicio una etapa de institucionalización. En 2009 finalmente se publicó la metodología de medición de pobreza multidimensional, así como las primeras cifras del 2008 por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL).

Esta tesis pretende tomar dicha metodología elaborada por el CONEVAL e investigar si es posible, desarrollar un modelo de regresión que permita reproducir la clasificación de pobreza hecha por el CONEVAL con un menor número de variables de las empleadas actualmente. Con dichos resultados se investigó la viabilidad de asignar un puntaje a cada individuo evaluado por el modelo de regresión, con el objeto de proporcionar una metodología estadísticamente sustentable y operacionalmente viable; es decir construida con los datos actualmente disponibles (inclusive con un número menor) para evaluar los cambios en las condiciones de vida de la población a través del tiempo.

El puntaje estadístico, es una medida consistente y produce resultados comparables a través del tiempo, desarrollar dicho puntaje con un modelo que ocupe un menor número de variables, podría tener un impacto en la evaluación de los monitoreos periódicos de los cambios en los niveles de pobreza, generados por el Instituto Nacional de Estadística y Geografía (INEGI) cada dos años para cada entidad federativa y con información desagregada a nivel municipal cada cinco años. A pesar de que el puntaje estadístico técnicamente implica cierto grado de complejidad, en la práctica su uso puede ser automatizado, de fácil lectura, replicable y además responde al sentido común.

Estadísticamente es posible calibrar el nivel de deterioro sobre la capacidad predictiva del modelo, e inclusive dependiendo de los resultados determinar si se requiere de un nuevo desarrollo o simplemente de un ajuste sobre la estructura de las variables, lo cual podría tener un impacto en términos de tiempo pues un solo modelo puede ser empleado durante varios años.

El documento esta organizado en cinco capítulos. En el primer capítulo titulado “**Marco Teórico**”, se presenta un breve perfil de la situación de la pobreza en México así como la metodología y concepción de la medición multidimensional de la pobreza adoptada en nuestro país por el CONEVAL desde el 2008. En el segundo capítulo “**Teoría Modelo de regresión logística y análisis de reducción de variables**”, se dan a conocer los conceptos generales, objetivos y requerimientos que se deben tener en cuenta al desarrollar un modelo de regresión logística y las técnicas aplicadas bajo la perspectiva del desarrollo de modelos de score, para minimizar el número de variables empleadas actualmente para la clasificación de pobreza. En el tercer capítulo “**Datos y metodología**” se presenta el análisis exploratorio de los datos para la construcción del modelo, la definición de la variable respuesta y variables explicativas; así como la metodología para la Construcción del score y la réplica de dicho modelo con datos 2010, 2012 y 2014. En el cuarto capítulo, que recibe el nombre de “**Resultados**” se dan a conocer los resultados del desempeño del modelo propuesto para reproducir la clasificación de los “hogares pobres” hecha por el CONEVAL con los años 2008, 2010, 2012 y 2014. Finalmente en el capítulo cinco “**Conclusiones**”, se comenta la relevancia de aplicar la metodología del Credit Scoring para asignar un puntaje a cada individuo evaluado por el modelo de regresión con el objetivo de medir los cambios en las condiciones de vida de la población a través del tiempo.

# Capítulo 1

## Marco teórico

### 1.1. ¿Qué es la pobreza?

En un trabajo publicado por el Banco Mundial en el 2004, titulado “*La pobreza en México: una evaluación de las condiciones, tendencias y estrategia de gobierno.*”,<sup>1</sup> se sostiene que a menudo en el análisis de la pobreza en México y otros lugares, ha predominado la concepción de pobreza como el número de personas que vive por debajo de una línea definida en términos de ingresos o gastos; es decir se le ha dado un enfoque unidimensional a la pobreza entendiéndola únicamente como la carencia de bienes materiales o ausencia de elementos básicos que aseguren la subsistencia del ser humano.

Sin embargo, de acuerdo al Banco Mundial (2004) esto es sólo una parte de lo que la pobreza significa en una sociedad. Estudios empíricos sobre lo que la gente piensa y consideraciones conceptuales apoyan la noción de que la pobreza abarca múltiples dimensiones, como son las capacidades humanas, en especial la situación educativa y de salud, el acceso a la infraestructura, los ingresos y la inclusión social.

Adoptar una perspectiva más amplia del bienestar y la pobreza tiene fuertes raíces conceptuales. Una influencia muy importante en el pensamiento sobre la pobreza ha sido el trabajo

---

<sup>1</sup>Banco Mundial de México, *La pobreza en México: una evaluación de las condiciones, tendencias y estrategia de gobierno*, México: Banco Mundial, 2004, Traducción Marcela Pimentel Lusarreta, página 2-4.

de Amartya Sen (2000) <sup>2</sup> donde se sostiene que los logros de la gente en términos de hacer o ser, o sus “*funcionamientos*”, es lo que debería evaluarse como éxitos en bienestar. Tales logros pueden ir desde funcionamientos demasiado básicos, entre ellos estar bien alimentado, prevenir enfermedades que se puedan evitar, hasta los más sofisticados como tener respeto de sí mismo, ser capaz de participar en la vida de la comunidad, por mencionar algunos. Sen asegura que las condiciones materiales, entre ellas los ingresos, pueden contribuir a la realización de dichos logros.

Con base en lo expresado anteriormente, se aprecia una de las principales limitaciones del concepto de pobreza en términos unidimensionales, dado que las medidas basadas exclusivamente en el ingreso o el consumo pierden muchos casos de pobreza evaluada desde otros indicadores.

### **1.1.1. Breve perfil de las políticas de combate de pobreza en México 1988-2012**

De acuerdo a Julio Boltvinik (2001), todavía en la segunda mitad de los años ochenta la pobreza no era un tema del que se ocupara la prensa y otros medios de difusión masiva. Hoy en cambio el panorama es diferente a escala mundial, por ejemplo la Organización de las Naciones Unidas (ONU) llevó a cabo la Cumbre Mundial sobre la pobreza y Desarrollo Social (Copenhague, 1995), que pretendió poner el tema de la pobreza en el centro del interés internacional como lo está ya, el del medio ambiente. Por su parte el Banco Mundial ha renovado esfuerzos de décadas anteriores y viene presionando a gobiernos de Tercer Mundo para que lleven a cabo programas de reducción de pobreza.

En América Latina la inmensa mayoría de los gobiernos, después de años de abandono relativo de lo social, tienen actualmente programas de lucha contra la pobreza. México es un buen ejemplo de este ciclo.

---

<sup>2</sup>Amartya Sen, *Desarrollo y Libertad*, Buenos Aires, Editorial Planeta S.A, 2000, Recuperado de: <http://www.cdh.uchile.cl/media/publicaciones/pdf/5/260.pdf>

En efecto, la pobreza se ha transformado en el México de hoy en una vivencia cotidiana para más de cincuenta millones de mexicanos lo que representa el 46,2% de la población según el último reporte presentado por el CONEVAL, razón por la cual en los últimos años se ha hecho más frecuente el desarrollo y la implementación de políticas y programas sociales que ayuden a enfrentar y superar los desafíos inherentes a la pobreza. A continuación se presentan los programas de combate a la pobreza, propuestos e implementados en México desde 1988 hasta 2012.

### **Programas sociales de combate a la pobreza en el sexenio de Carlos Salinas de Gortari (1988-1994)**

A partir de diciembre de 1988, la política gubernamental mexicana de combate a la pobreza registró un importante giro, pues como parte de las primeras medidas de gobierno tomadas por Carlos Salinas de Gortari, fue creado el Programa Nacional de Solidaridad (PRONASOL), un instrumento de política social que estaba encaminado a la superación de los rezagos sociales más lacerantes de los grupos más vulnerables de la sociedad mexicana. Pronasol fue instrumentado como respuesta al entorno social adverso que dio lugar la crisis económica iniciada en 1982. En sus inicios este programa dependía de la Secretaría de Programación y Presupuesto y en 1992 fue transferido a la Secretaría de Desarrollo Social (SEDESOL), de reciente creación.

Desde su puesta en marcha, Solidaridad se propuso desarrollar en tres frentes el combate a la pobreza: el bienestar social, el apoyo a la producción y el desarrollo regional. Así pues PRONASOL constituyó el cuerpo central de la política social durante el sexenio de Salinas de Gortari y a pesar de que dicho programa fue objeto de polémica y de críticas en su momento, tuvo un gran efecto en las condiciones de vida de muchos mexicanos al mejorar el alcance y la calidad de los servicios a los que el público tenía acceso gracias a los proyectos de infraestructura como escuelas, clínicas rurales y electrificación que se llevaron a cabo (Favela Alejandro, Calvillo Minan, 2003).<sup>3</sup>

---

<sup>3</sup>Según el autor de “ El combate a la pobreza en México en el sexenio de Zedillo”, Alejandro Favela entre 1988 y 1992 fueron construidos o renovados 273 hospitales, 655 pueblos fueron electrificados y más de 123 mil km de camino recibieron mantenimiento, aunque reconoce que es difícil juzgar si las obras públicas hechas en el marco de Solidaridad habrían sido realizadas de cualquier manera en ausencia del programa.

## **Programas sociales de combate a la pobreza en el sexenio de Ernesto Zedillo (1994-2000)**

El PRONASOL se mantuvo un tramo del mandato presidencial zedillista (1994-2000), hasta que en 1997, Zedillo dio a conocer al sustituto de este programa que sería denominado como el Programa de Educación, Salud y Alimentación (PROGRESA) bajo el cual se reforzaron los programas dirigidos hacia grupos focalizados, restringiendo su atención únicamente a ciertas regiones rurales y reduciendo su intervención a tres aspectos: nutrición, salud y educación. Los apoyos en educación consistieron en becas para que los niños asistieran a escuelas primarias y secundarias, la salud se limitó a la atención médica básica de forma periódica y la alimentación a la distribución de suplementos nutricionales hacia infantes y mujeres embarazadas. Esto contrasta con los alcances del PRONASOL que incluía un rango mucho más amplio de acciones, se orientaba en un sentido operativo a la participación comunitaria y no de algunas familias y tampoco era exclusivo de las regiones más marginales del país.

## **Programas sociales de combate a la pobreza en el sexenio de Vicente Fox (2000-2006)**

En 2002 bajo la presidencia de Vicente Fox, PROGRESA se transformó en el programa Oportunidades que muy probablemente es el programa antipobreza más grande y más acertado de la historia de México (Juárez, 2008). Si bien hubo serie de modificaciones en este nuevo programa, estas no alteraron la esencia de su precursor PROGRESA, proveer la alimentación y salud, como también dar becas a los estudiantes, de los hogares seleccionados por el programa.

En general, Oportunidades es un programa que continúa los esfuerzos iniciados por PROGRESA, pero con mayor relevancia en términos del esfuerzo presupuestario, del número de familias y localidades cubiertas, de su penetración en el medio urbano, de su mecánica de operación, del contenido y frecuencia de las políticas de salud, de los años cubiertos por las becas educativas, de los incentivos para la formación de capital humano, y del monto total de los apoyos monetarios por familia.



Por ejemplo en el año 2003, con la intención de estimular a los jóvenes, Oportunidades incorporó un cuarto componente denominado “Jóvenes con Oportunidades” que consiste en el otorgamiento diferido de un apoyo monetario que se acumula entre el tercer grado de secundaria y el tercero de bachillerato y esté condicionado a la conclusión de la educación media superior. Cabe mencionar que los beneficiarios solo pueden destinar el dinero a ciertos fines: comenzar una carrera universitaria, iniciar un proyecto educativo, hacer mejoras a la vivienda, insertarse en un esquema de protección social a través del Seguro popular o IMSS, o bien abrir una cuenta bancaria y disponer del dinero dos años después. Otro componente es el denominado para adultos mayores, creado en 2006 y destinado a los miembros de hogares mayores de 70 años, el apoyo consiste en una transferencia monetaria condicionada a la asistencia del adulto mayor a consultas médicas semestrales. Todas estas modificaciones matizaron al programa Oportunidades.

Además de estos programas dentro del Poder Ejecutivo, el Congreso aprobó y promulgó el 20 de enero de 2004 la Ley General de Desarrollo Social (LGDS), lo cual representa un importante paso hacia la institucionalización de una política de estado, cuya finalidad es que toda la población tenga acceso al desarrollo social<sup>4</sup>.

Una de las aportaciones más relevantes de la LGDS son las pautas y criterios para el financiamiento de las políticas públicas sociales. Estas se definen de tal modo que los programas, fondos y recursos destinados al desarrollo social se consideren prioritarios y de interés público.

Pero encarar esta problemática requiere en la actualidad, además de un marco normativo, de la presencia de instituciones públicas en las cuales el estado pueda apoyarse para hacer frente a la pobreza. Bajo este contexto, surge el CONEVAL, que es un organismo público con autonomía técnica y de gestión, el cual se ha fijado como objetivo contribuir a normar y coordinar la evaluación de las políticas y programas de desarrollo social y de establecer los lineamientos y

---

<sup>4</sup>Josefina Vázquez Mota secretaria de SEDESOL señaló en 2002: “El desarrollo social es el proceso permanente mediante el cual se amplían las capacidades y las opciones de las personas y comunidades para que puedan ejercer plenamente sus libertades y derechos y realizar todo su potencial productivo y creativo, de acuerdo con sus aspiraciones, elecciones, intereses, convicciones y necesidades. (SEDESOL,2002) ”

criterios para la definición, identificación y medición de la pobreza.

De esta forma, la LGDS cumple su propósito a través del CONEVAL, quien establece los fundamentos en materia de definición y medición de la pobreza, al generar elementos que pueden ser utilizados en el análisis y evaluación de las políticas públicas.

### **Programas sociales de combate a la pobreza en el sexenio de Felipe Calderón Hinojosa (2006-2012)**

El Programa Sectorial de Desarrollo Social en el sexenio de Felipe Calderón Hinojosa, fue elaborado tomando como punto de partida la Visión México 2030 y el Plan Nacional de Desarrollo 2007- 2012. Los objetivos, las estrategias y las líneas de acción que propone este plan de desarrollo tienen como prioridad dar continuidad a los programas sociales existentes en nuestro país; así como atender a los grupos vulnerables que no habían sido incorporados a los beneficios de las políticas públicas.

El compromiso del gobierno federal establecido en El Programa Sectorial de Desarrollo Social propone como principio central el Desarrollo Humano Sustentable, el cual es entendido como el proceso permanente de ampliación de capacidades y libertades que permita a todos los mexicanos tener una vida digna sin comprometer el patrimonio de las generaciones futuras.(SEDESOL, 2012).

El Desarrollo Humano Sustentable se pretende lograr a través del desarrollo de capacidades básicas de educación, salud, nutrición, alimentación y vivienda que permitan una mayor igualdad de oportunidades, en especial para la población en condiciones de pobreza.

Para esto, el Gobierno Federal estableció acciones conjuntas con los responsables de las políticas de desarrollo social, como son educación, salud, infraestructura básica, fomento productivo y sustentabilidad. Asimismo buscó conectar más estrechamente la política económica con la política social ya que el crecimiento económico, la estabilidad macroeconómica y la generación

de empleos son elementos fundamentales para alcanzar un país sin pobreza. (SEDESOL, 2012).

## 1.2. Definición y medición de la pobreza actualmente en México

La pobreza es un fenómeno siempre presente, en mayor o menor medida en todas las sociedades, razón por la cual ha sido objeto de estudio al mismo tiempo que los gobiernos han buscado aplicar diversas políticas y leyes para enfrentarla.

En México el CONEVAL, menciona que la pobreza, en su acepción más amplia, está asociada a condiciones de vida que vulneran la dignidad de las personas, limitan sus derechos y libertades fundamentales, impiden la satisfacción de sus necesidades básicas e imposibilitan su plena integración social, lo cual implica que en nuestro país el CONEVAL ha reconocido múltiples dimensiones al definir la pobreza. En esta línea el CONEVAL bajo la perspectiva multidimensional definió en 2008 a la pobreza de la siguiente manera:

*“Una persona se encuentra en situación de pobreza multidimensional cuando no tiene garantizado el ejercicio de al menos uno de sus derechos para el desarrollo social, y si sus ingresos son insuficientes para adquirir los bienes y servicios que requiere para satisfacer sus necesidades.”*

Por otra parte, la LGDS en su artículo 36 establece que el CONEVAL debe emitir un conjunto de lineamientos y criterios para la definición, identificación y medición de la pobreza, en los cuales habrá de considerar, al menos, los ocho indicadores siguientes:

- Ingreso corriente per cápita.
  - Rezago educativo promedio en el hogar.
  - Acceso a los servicios de salud.
  - Acceso a la seguridad social.

- Calidad y espacios de la vivienda.
- Acceso a los servicios básicos en la vivienda.
- Acceso a la alimentación.
- Grado de cohesión social.

### **1.2.1. Enfoques para la medición de la pobreza en México**

De acuerdo a la LGDS, pueden distinguirse dos enfoques contrastantes para abordar la pobreza en México. Por un lado está el enfoque llamado de bienestar y por el otro el enfoque denominado de derechos.

Cabe mencionar que seleccionar uno u otro tiene repercusiones importantes en términos de las explicaciones de pobreza y en consecuencia, de las políticas públicas que se consideran apropiadas para disminuirla.

En la publicación presentada por el CONEVAL sobre la metodología para la medición multidimensional de la pobreza se limita al enfoque de bienestar como la aproximación de necesidades básicas insatisfechas; pues trata de identificar las condiciones de vida que brindan a los individuos las opciones de vida mínimamente aceptables en su sociedad, suponiendo que en base a las circunstancias y preferencias de cada individuo este define el abanico de opciones de vida que puede elegir y en caso de que estas opciones no le permitieran tener condiciones de vida aceptables, entonces se considera que el individuo es pobre.

Sin embargo, el progreso reciente en la comprensión de la pobreza, ofrece otra concepción de la misma basada en los derechos humanos. Según el CONEVAL el enfoque de derechos parte del reconocimiento de los derechos humanos como la expresión de las necesidades, valores, intereses y bienes que, por su urgencia e importancia, han sido considerados como fundamentales y comunes a todos los seres humanos.

La premisa bajo la cual se sustenta este enfoque consiste en que toda persona debe contar con una serie de garantías indispensables para la dignidad humana, toda persona, por el hecho de serlo, debe tener asegurados un conjunto de derechos sociales de manera irrenunciable e insustituible.(ONU, 2004).

Por otra parte, como ya se ha venido recalcando el concepto de pobreza comprende diversos componentes o dimensiones; así pues no basta con abordar unicamente el concepto bajo los dos enfoques ya propuestos bienestar y derechos; pues según el CONEVAL es también necesaria la incorporación del espacio de contexto territorial dentro de la medición de la pobreza.

La integración del espacio de contexto territorial dentro de la medición de la pobreza permite analizar el efecto de los fenómenos y problemáticas de las comunidades y localidades en las opciones de vida de toda persona, en especial la cohesión social y el disfrute de un medio ambiente sano. Esto permite vincular el concepto de pobreza con la creación de mecanismos de integración y cooperación social, así como con la búsqueda de un desarrollo sustentable y respetuoso del medio ambiente.

Por tanto la medición multidimensional de la pobreza adoptada por el CONEVAL concibe a la pobreza a partir de tres grandes espacios: el bienestar económico, los derechos sociales y el contexto territorial. Esto crea un cuerpo conceptual sólido para el análisis de la pobreza desde esta nueva perspectiva.

### **1.2.2. Método para la medición multidimensional de la pobreza considerando los enfoques establecidos por el CONEVAL**

La combinación de los espacios de bienestar, derechos y contexto territorial citados anteriormente son los elementos conceptuales bajo los cuales el CONEVAL propone establecer una metodología de medición de la pobreza multidimensional.

Para identificar el conjunto de la población en situación de pobreza, de acuerdo al CONEVAL es necesario identificar los individuos privados de uno o varios de los seis indicadores de desarrollo social: rezago educativo, acceso a los servicios de salud, acceso a la seguridad social, calidad y espacios de la vivienda, acceso a los servicios básicos en la vivienda y acceso a la alimentación, además de determinar si los ingresos de una persona son insuficientes para la satisfacción de sus necesidades.

El procedimiento establecido por el CONEVAL para determinar si los ingresos de una persona son insuficientes para la satisfacción de sus necesidades, es abordado a partir del ingreso que disponen las personas para la adquisición de bienes y servicios en el mercado. Cabe resaltar que el CONEVAL estipula que el ingreso al cual se refiere es el ingreso corriente el cual representa el flujo de entradas, no necesariamente monetarias (puede incluir los productos recibidos o disponibles en especie, entre otros).

Para determinar la suficiencia del nivel de ingresos de cada persona, se han establecido dos conceptos, el concepto de *línea de bienestar* y de *línea de bienestar mínima*, la línea de bienestar hace posible identificar a la población que no cuenta con los recursos suficientes para adquirir los bienes y servicios que requiere para satisfacer sus necesidades (alimentarias y no alimentarias) mientras que la línea de bienestar mínima permite identificar a la población que, aun al hacer uso de todo su ingreso en la compra de alimentos, no puede adquirir lo indispensable para tener una nutrición adecuada.

Por otra parte en lo referente al enfoque de derechos sociales, se ha construido un índice denominado índice de privación social cuya finalidad es identificar a la población con al menos una carencia en seis indicadores de derechos sociales. En primera instancia se lleva a cabo un proceso de identificación de carencias en cada indicador; es decir a través de una variable dicotómica se asigna un uno si el individuo presenta carencia en la dimensión respectiva y cero en caso contrario.

De esta forma, el índice de privación social se construye para cada persona a partir de la

suma de los seis indicadores asociados a las carencias sociales; así pues las personas que no presentan ninguna carencia tienen un índice de privación social igual a 0, mientras que aquéllos que presentan carencias en la totalidad de los indicadores tienen un índice igual a 6, como se muestra en la siguiente figura. Nótese que de manera implícita la construcción de este índice conlleva el supuesto de que cada una de las carencias tiene la misma importancia relativa.

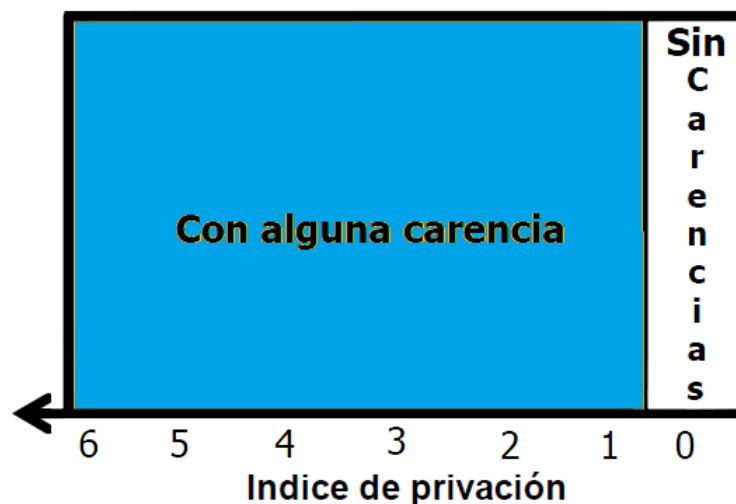


Figura 1.1: Índice de privación social

Fuente: [www.coneval.mx](http://www.coneval.mx)

Ahora bien, para diferenciar entre las personas con grandes carencias de los que tienen un número moderado de ellas se obtuvo un umbral de privación de derechos, el cual es un punto de corte ( $C=1$ ) del índice de privación social es decir; de acuerdo a la definición de pobreza multidimensional se considera que una persona experimenta carencias en el espacio de los derechos sociales cuando el valor del índice de privación social es mayor que cero o equivalentemente cuando el individuo padece al menos una de las seis carencias.

## Combinación del ingreso y el índice de privación social

Una de las grandes virtudes de la nueva metodología de medición reside en que, a partir de la combinación de las líneas de pobreza en las dimensiones de ingreso y derechos sociales es posible clasificar a cualquier persona en uno, y sólo uno, de los siguientes cuatro cuadrantes.(figura 1.2).

**I.-Pobres multidimensionales.** Población con ingreso inferior al valor de la línea de bienestar y que padece al menos una carencia social.

**II.-Vulnerables por carencias sociales.** Población que presenta una o más carencias sociales, pero cuyo ingreso es superior a la línea de bienestar.

**III.-Vulnerables por ingresos.** Población que no presenta carencias sociales y cuyo ingreso es inferior o igual a la línea de bienestar.

**IV.-No pobre multidimensional y no vulnerable.** Población cuyo ingreso es superior a la línea de bienestar y que no tiene carencia social alguna.

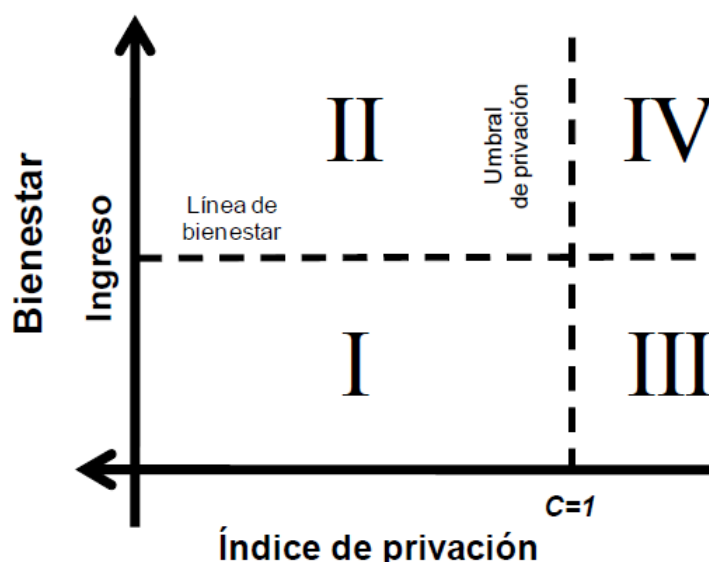


Figura 1.2: Población en situación de pobreza multidimensional

Fuente: [www.coneval.gob.mx](http://www.coneval.gob.mx)



### **1.3. Criterios para la definición de los indicadores de pobreza**

No toda la población pobre es igualmente pobre. Los indicadores de pobreza son una forma parcial e incompleta de capturar el fenómeno multidimensional de la pobreza.

Tanto el espacio de bienestar, el índice de privación social así como el contexto territorial, presentan características específicas del proceso de la metodología de medición que influyen en la construcción de los indicadores.

A continuación se detalla la construcción y fundamentación de los indicadores de carencia por rezago educativo, acceso a los servicios de salud, acceso a la seguridad social, calidad y espacios de la vivienda, servicios básicos en la vivienda y acceso a la alimentación y de otros indicadores considerados complementarios a estos últimos.

### **1.4. Medición del espacio de carencias sociales**

#### **1.4.1. Rezago educativo**

La educación constituye una de las claves del ascenso social, y un factor de justicia en la distribución de oportunidades, es decir tiene la capacidad de contribuir a la creación de una sociedad más justa, así como a la formación de individuos autónomos, creativos y participativos dentro de su sociedad.

Sin embargo, la educación por sí sola tiene poco impacto en el mejoramiento de las condiciones de vida de los grupos socialmente desfavorecidos; pues son los pobres quienes no acceden a la escuela, o bien en el mejor de los casos quienes acceden en condiciones inequitativas y quienes reciben educación irrelevante o ineficiente.

En nuestro país la función educativa, pública o privada, se desarrolla en el interior de un orden

normativo cuyos lineamientos deben ser cumplidos escrupulosamente; así pues dentro de los ordenamientos vigentes, relacionados directa o indirectamente con la educación pública mexicana, destacan el Artículo 3 de la Constitución Política de los Estados Unidos Mexicanos y los artículos 2, 3 y 4 de la Ley General de Educación (LGE) los cuales establecen que toda la población debe cursar la educación preescolar, primaria y secundaria, niveles que constituyen la educación básica obligatoria.

De acuerdo al CONEVAL en conjunción del Instituto Nacional para la Evaluación de la Educación (INEE) se considera las siguientes variables para determinar si alguna persona cuenta con carencia por rezago educativo:

- I. **Edad:** Número de años cumplidos por una persona al momento de la entrevista.
- II. **Año de nacimiento:** Diferencia entre el año en el que se realiza la medición y la edad reportada por el individuo.
- III. **Inasistencia a la escuela:** Condición de no asistencia de cada persona a una institución de enseñanza del Sistema Educativo Nacional (SEN).
- IV. **Nivel educativo:** Máximo nivel de estudios que reporta una persona haber cursado.

#### **1.4.2. Acceso a los servicios de salud**

El derecho a la salud ha sido consagrado en nuestro país a través del Artículo 4to. de la Constitución como uno de los derechos fundamentales de la población estableciendo que toda la población mexicana tiene derecho a la protección de la salud.

Por otro lado el estado de salud de las personas tiene diversas dimensiones y hay una considerable cantidad de indicadores que pueden ser utilizados para su medición. Sin embargo, a partir del marco normativo que se ha descrito ya, el CONEVAL considera que una persona se encuentra en situación de carencia por acceso a los servicios de salud cuando:

- No cuenta con adscripción o derecho a recibir servicios médicos de alguna institución que los presta, incluyendo el Seguro Popular, las instituciones públicas de seguridad social (IMSS, ISSSTE, PEMEX, ISSFAM) o los servicios médicos privados.

Para el cálculo del indicador, se identifica a las personas que cuentan con acceso a través de la siguiente variable:

- I. **Servicios de salud:** Identifica si una persona se encuentra afiliada o inscrita para recibir servicios de salud por parte de una institución pública o privada que ofrezca esos servicios.

Sin embargo, un indicador apropiado para representar la salud en las medidas de pobreza tendría que incluir además del acceso a los servicios de salud, información sobre otros elementos asociados al derecho a la salud, como la oportunidad y la efectividad de los servicios recibidos, pero contar con información como esta en un solo indicador derivaría en diversas restricciones conceptuales y metodológicas para analizarlos. Por ello estos aspectos serán considerados en otros indicadores complementarios.

### **1.4.3. Acceso a la seguridad social**

Según el artículo 2 de la Ley del Seguro Social, la seguridad social tiene como fin garantizar a los mexicanos el derecho a la salud, la asistencia médica, la protección de los medios de subsistencia y los servicios sociales para el bienestar individual y colectivo, así como el otorgamiento de una pensión que sea garantizada por el Estado.

Así mismo esta Ley establece dos regímenes para el acceso a la seguridad social: el obligatorio y el voluntario, así como un sub-régimen de inscripción voluntaria al régimen obligatorio.

De ahí que el indicador de seguridad social propuesto por el CONEVAL considera varios grupos: la Población Económicamente Inactiva, los trabajadores independientes sin servicios médicos (como los microempresarios), las personas sin pensión (contratados por honorarios) y

los adultos mayores de 65 años que no son beneficiarios de algún programa social de pensiones.

Bajo este contexto el CONEVAL identifica a la población con carencia por acceso a la seguridad social de acuerdo con los siguientes criterios:

- Para la población trabajadora asalariada, se considera que no tiene carencia en esta dimensión si dispone de las prestaciones de servicios médicos, incapacidad con goce de sueldo y Afore.
- Para la población trabajadora independiente se considera que no tiene carencia por acceso a la seguridad social cuando disponga de servicios médicos como prestación laboral o por contratación voluntaria al régimen obligatorio del IMSS, y además disponga de Afore.
- Para la población en general se considera que no tiene carencia cuando goce de alguna jubilación o pensión, o sea familiar de una persona dentro o fuera del hogar con acceso a la seguridad social. En el caso de la población en edad de jubilación (65 años y más), se considera que no tiene carencia por acceso a la seguridad social si es beneficiario de algún programa social de pensiones para adultos mayores.
- La población que no cumpla con al menos uno de los criterios mencionados anteriormente, se considera en situación de carencia por acceso a la seguridad social.

Con base en las siguientes variables se determina el indicador de carencia por acceso a la seguridad social:

- I. **Población Económicamente Activa:** Es el conjunto de la población de dieciseis años o más que declaró ser ocupada (es decir, durante el periodo de referencia realizó alguna actividad económica) o ser desocupada (personas que buscaron trabajo activamente en el período de referencia); además identifica a las personas que declararon realizar actividades consideradas no económicas (PNEA).
- II. **Acceso directo a la seguridad:** Integra variables que miden el tipo de trabajo de la población ocupada; es decir si realiza un trabajo subordinado o independiente, la incapacitación

dad con goce de sueldo de un trabajador, la afiliación de una persona a una institución que le proporciona servicios de salud como prestación laboral, si la persona cuenta con acceso a un sistema de jubilación, si la persona se encuentra afiliada a una institución que le proporciona servicios de salud por contratación voluntaria, si la persona cuenta con acceso a un sistema de jubilación o pensión para retiro por contratación voluntaria y finalmente personas que declaran ser jubiladas o pensionadas al momento de la entrevista.

III. **par:** Integrantes del hogar que, por su condición de parentesco, podrían tener acceso a la seguridad social por otros miembros de la familia.

IV. **jefss:** Indica si el/la jefe(a) del hogar tiene acceso directo a la seguridad social.

V. **conyss:** Indica si el o la cónyuge del jefe(a) tiene acceso directo a la seguridad social.

VI. **hijoss:** Indica si algún hijo(a) del jefe(a) del hogar tiene acceso directo a la seguridad social.

VII. **Servicios de Salud:** Mide si la persona declara contar con servicios médicos de alguna institución de seguridad social por algún familiar del hogar o de otro hogar, por muerte del asegurado o por contratación propia.

VIII. **Programa de Adultos Mayores:** Indica si la persona tiene 65 años o más y recibe algún programa de adultos mayores.

#### 1.4.4. Calidad y espacios de la vivienda

La calidad de vida de las personas se ve directamente impactada por el entorno físico en el que habitan, partiendo en primera instancia de la calidad de su vivienda.

Por otra parte el CONEVAL sostiene que una vivienda adecuada ayuda al desarrollo social y psicológico de las personas, así como a su adaptación al entorno sociocultural y económico del medio en donde se desenvuelven.

Se aborda pues, el concepto de calidad y espacios de la vivienda de acuerdo a los criterios formulados por la Comisión Nacional de Vivienda (CONAVI) y el CONEVAL desde dos perspectivas:

el material de construcción de la vivienda y sus espacios.

En el caso del material de construcción de la vivienda Coneval considera los pisos de tierra, el material del techo y de los muros mientras que en el caso de los espacios se evalúa el grado de hacinamiento (habitantes por cuarto).

De acuerdo con estos criterios, se considera como población en situación de carencia por calidad y espacios de la vivienda a las personas que residan en viviendas que presenten, al menos, una de las siguientes características:

- El material de los pisos de la vivienda es de tierra.
- El material del techo de la vivienda es de lámina de cartón o desechos.
- El material de los muros de la vivienda es de barro o bajareque; de carrizo, bambú o palma; lámina de cartón, metálica o asbesto; o material de desecho.
- La razón de personas por cuarto (hacinamiento) es mayor que 2.5.

La construcción del indicador se conforma a través de los siguientes subindicadores, las cuales se definen a continuación:

I. **icvpisos:** Indicador de carencia del material de piso de la vivienda.

II. **icvtechos:** Indicador de carencia del material de techos de la vivienda.

III. **icvmuros:** Indicador de carencia del material de muros de la vivienda.

IV. **icvhac:** Indicador de carencia por índice de hacinamiento de la vivienda.

Cabe mencionar que debido a que el indicador de carencia por calidad y espacios de la vivienda se construye a partir de los elementos mínimos indispensables de una vivienda digna, pueden existir otros factores que complementen a dicho indicador y que no hayan sido considerados en la elaboración de este y algunos de ellos serán incorporados dentro de los indicadores complementarios.

#### 1.4.5. Acceso a los servicios básicos en la vivienda

Al igual que el material de construcción de la vivienda y sus espacios, el acceso a servicios básicos es otra de las medidas que se utiliza para conocer el grado de bienestar del individuo, de esta forma se establece el mínimo de servicios a los cuales todo individuo junto con su grupo familiar debe de acceder.

La carencia por servicios básicos; según el CONAVI institución en la que el CONEVAL se apoya para establecer los lineamientos de este indicador se refiere a acceso al agua potable, disponibilidad de servicio de drenaje, servicio de electricidad y combustible para cocinar en la vivienda.

De acuerdo con los criterios propuestos por la CONAVI, se considera como población en situación de carencia por servicios básicos en la vivienda a las personas que residan en viviendas que presenten, al menos, una de las siguientes características:

- El agua se obtiene de un pozo, río, lago, arroyo, pipa; o bien, el agua entubada la obtienen por acarreo de otra vivienda, o de la llave pública o hidrante.
- No cuentan con servicio de drenaje, o el desagüe tiene conexión a una tubería que va a dar a un río, lago, mar, barranca o grieta.
- No disponen de energía eléctrica.
- El combustible que se usa para cocinar o calentar los alimentos es leña o carbón sin chimenea.

A partir de las variables y umbrales indicados por la CONAVI se construyen los subindicadores que conforman el indicador de Acceso a los servicios básicos en la vivienda, los cuales se muestran a continuación:

I. **isb agua:** Indicador de carencia de acceso al agua en la vivienda.

II. **isb dren:** Indicador de carencia de servicio de drenaje en la vivienda.

III. **isb luz:** Indicador de carencia de servicios de electricidad en la vivienda.

IV. **isb combus:** Indicador de carencia de servicio de combustible para cocinar en la vivienda.

Sin embargo, existen una serie de servicios que también son muy importantes para la calidad de vida del hombre, tales como pavimentación, alumbrado público, transporte público, recolección de basura, acceso a zonas de esparcimiento, acceso a servicios de salud incluso a servicios privados tales como los de telecomunicación principalmente a los telefónicos, pero para la construcción de este indicador solo se consideró aquellos que, indispensablemente, deben estar en funcionamiento al momento de ser ocupada la vivienda, los antes mencionados, de manera análoga al indicador anterior serán analizados dentro del conjunto de indicadores complementarios.

#### **1.4.6. Acceso a la alimentación**

La alimentación constituye un elemento esencial del bienestar humano y está determinada por factores sociales, culturales; así como por el grado de dependencia al autoconsumo y al mercado, pero principalmente por la capacidad adquisitiva las familias. Razón por la cual es posible contemplar el acceso a la alimentación como un indicador de la pobreza.

Al igual que en los indicadores anteriores, es necesario establecer una medida para evaluar el avance en el derecho a la alimentación y es aquí donde se ve involucrado el concepto de seguridad alimentaria, el cual ha ido evolucionado considerablemente desde 1945, al punto que la comunidad internacional lo ha consagrado la como uno de los derechos humanos fundamentales.

La seguridad alimentaria constituye, en principio, un impulso casi instintivo de los grupos humanos por asegurar su sobrevivencia frente a la escasez. Sin embargo la definición de seguridad alimentaria contemporánea se adoptó en la Declaración de Roma de 1996, resultado de la Cumbre Mundial de la Alimentación, como: la capacidad de un país para producir los alimentos que consume y la solvencia de la población para acceder a ellos.

Más recientemente, de acuerdo con la FAO (2006), la seguridad alimentaria existe cuando todas



las personas tienen en todo momento acceso físico y económico a suficientes alimentos inocuos y nutritivos para satisfacer sus necesidades alimenticias y sus preferencias en cuanto a los alimentos a fin de llevar una vida activa y sana.

Según el CONEVAL este concepto se encuentra estrechamente vinculado al acceso a la alimentación, por lo que se considera apropiado para la medición del indicador de carencia para esta dimensión. En este escenario surge la necesidad de evaluar de una manera sencilla práctica y rápida la Seguridad alimentaria en los hogares mexicanos, para ello el CONEVAL adopta una escala de de seguridad alimentaria basada en la propuesta de Pérez- Escamilla, Melgar-Quiñonez, Nord, Álvarez y Segall que reconoce cuatro posibles niveles de inseguridad alimentaria:

- I. **Inseguridad alimentaria severa:** Los hogares sólo con adultos que contestan afirmativamente de cinco a seis preguntas; los hogares con menores de edad que responden de ocho a doce preguntas de la escala.
- II. **Inseguridad alimentaria moderada:** Los hogares sólo con adultos que contestan afirmativamente de tres a cuatro preguntas; los hogares con menores de edad que responden de cuatro a siete preguntas de la escala.
- III. **Inseguridad alimentaria leve:** Los hogares conformados sólo por mayores de dieciocho años que contestan afirmativamente de una a dos preguntas de la escala.
- IV. **Seguridad alimentaria:** Los hogares constituidos sólo por adultos y los hogares con menores de edad que no responden de manera afirmativa a ninguna de las preguntas de la escala.

Dado lo anterior, el CONEVAL establece que se debe considerar en situación de carencia por acceso a la alimentación a los hogares que:

- Presenten un grado de inseguridad alimentaria moderado o severo.

### 1.4.7. Grado de cohesión social

Debido a los usos tan diversos que recibe el concepto de cohesión social, resulta difícil asociarle una definición unívoca que delimite con precisión sus componentes y alcances. Sin embargo, a pesar de que no existe una acepción clara del concepto, el CONEVAL plantea tres aspectos fundamentales relacionados con la cohesión social.

En primer lugar se plantea que la cohesión social sólo puede medirse como una característica de conjuntos de población, que en términos más sencillos se traduce como la propuesta de considerar que la unidad de análisis no sean los individuos, sino las comunidades o grupos sociales.

En segundo lugar, se recalca que cada matiz del concepto de pobreza impacta de manera distinta a la cohesión social, por lo que tratar de establecer si la cohesión social es parte intrínseca de la pobreza o no, no resulta tan evidente.

Finalmente y como punto tres, se sostiene la premisa de que una sociedad más equitativa puede generar mejores condiciones para desarrollar la cohesión entre sus miembros.

Como ya se señaló, la ambigüedad que frecuentemente caracteriza a la expresión cohesión social, condujo al CONEVAL a realizar la medición del grado de cohesión social en el espacio del territorio, es decir; se medirá a nivel municipal y estatal mediante cuatro indicadores:

- La desigualdad económica (coeficiente de Gini).
- La razón de ingreso de la población pobre multidimensional extrema respecto a la población no pobre multidimensional y no vulnerable.
- La polarización social.
- Las redes sociales; este último indicador se calcula solamente a nivel estatal.

## 1.5. Criterios para la definición de las líneas de bienestar y bienestar mínima

Para medir este espacio, la metodología del CONEVAL contempla dos etapas: la construcción de las líneas de bienestar y líneas de bienestar mínimo y la conformación del indicador de ingreso.

Para determinar el valor de las líneas de bienestar y de bienestar mínimo, el CONEVAL desarrolló una metodología con la cual construye la canasta alimentaria tomando en cuenta tres líneas básicas, que permiten analizar la ingesta energética y de otros nutrientes de la población:

- Los aportes nutricionales de cada alimento.
- Los requerimientos y recomendaciones de consumo de nutrientes.
- Los datos sobre consumo y frecuencia de consumo de alimentos de los hogares.

Para construir la canasta básica no alimentaria el CONEVAL la realizó a partir de la comparación de dos metodologías: el coeficiente de Engel o coeficiente de Orshansky y una propuesta del Dr. Enrique Hernández Laos.

La línea de bienestar equivalente a la suma de los costos de la canasta alimentaria y no alimentaria propuesta por el CONEVAL en 2008 para el área rural, es de: 1,202.80 pesos y la línea de bienestar mínimo equivalente al costo de la canasta alimentaria rural es de: 613.80 pesos.

Análogamente, la línea de bienestar equivalente a la suma de los costos de la canasta alimentaria y no alimentaria propuesta por el CONEVAL en 2008 para el área urbana, es de: 1,921.74 pesos y la línea de bienestar mínimo equivalente al costo de la canasta alimentaria urbana es de: 874.63 pesos.

### 1.5.1. Ingreso corriente

La definición de ingreso corriente para el CONEVAL y que se retoma en este estudio es:

*El flujo de entradas, no necesariamente monetarias (pues puede incluir los productos recibidos o disponibles en especie, entre otros), que le permiten a los hogares disponer de los satisfactores que requieren, sin disminuir los bienes o activos que poseen...<sup>5</sup>*

Como se sabe, las necesidades de las personas varían según su edad por lo que para poder hacer comparaciones entre los ingresos de hogares de diferente composición es necesario homogeneizar los costos. Esto puede ser aplicando un ponderador de economía de escala (un mayor número de personas pueden compartir ciertos factores, lo cual afecta el monto de los ingresos que requiere para satisfacer sus necesidades) y la transformación por adulto equivalente (distintas personas requieren de un monto distinto de recursos según su edad y otras características).

La definición de ingreso corriente de acuerdo a la metodología multidimensional de la pobreza es la siguiente:

$$\text{Ingreso corriente total} = \text{Ingreso corriente monetario} + \text{Ingreso corriente no monetario}$$

Donde:

**Ingreso corriente monetario** Remuneraciones por trabajo independiente (incluye autoconsumo), ingreso por renta de la propiedad, otros ingresos provenientes del trabajo y transferencias.

**Ingreso corriente no monetario** Pago en especie, transferencias en especie (regalos en especie, se excluyen las transferencias que se dan por primera vez) no incluye renta imputada.

---

<sup>5</sup>(CONEVAL, 2009)

# Capítulo 2

## Modelo de regresión logística

### 2.1. Nociones preliminares

Los métodos de regresión se han convertido en un componente principal de cualquier análisis de datos, cuya finalidad sea describir la relación entre una variable objetivo y una o más variables explicativas. Existen diferentes tipos de análisis de regresión que varían según el número, la naturaleza de las variables involucradas y la función matemática que describe la relación que las vincula. Si la función involucra a una sola variable independiente ( $x$ ) se dice que la regresión es simple, si hay varias, entonces se trata de una regresión múltiple. Por otro lado si la dependencia funcional de la variable dependiente ( $y$ ) respecto a las independientes, es expresable según una línea recta (o más generalmente, según un hiperplano) la regresión es lineal, de lo contrario, es no lineal.

Sin embargo, existen situaciones en las que resulta impropio usar la regresión simple o múltiple, lineal o no, para caracterizar la relación funcional entre cierta variable de respuesta y un conjunto de variables independientes. La más común de las circunstancias invalidantes es la que se da cuando la variable dependiente es categórica. Para este tipo de casos la regresión logística resulta ser una de las técnicas estadístico - inferenciales más empleadas en la producción científica contemporánea.

En términos generales en la regresión logística se desea estudiar la asociación entre una variable

categorica  $Y$  con  $k$  modalidades (variable por explicar) y un conjunto de variables explicativas (predictivas)  $x_1, x_2, \dots, x_k$  que pueden ser de cualquier tipo.

En un principio se analizara el caso donde  $k=2$  del modelo logístico binario.

## 2.2. Ajuste del modelo

Se dispone de una muestra de  $n$  individuos; la variable  $Y$  toma el valor de 1 si el individuo pertenece al grupo  $G_1$ , 0 si no, es decir, si pertenece al grupo  $G_0$ . Para cada individuo se conoce el valor de la variable por explicar  $Y$  y el valor del vector fila  $x = (x_1, \dots, x_k)$  por simplicidad analizaremos primeramente el caso que trata de evaluar el efecto de un solo factor, al que se representara mediante la variable  $x_i$  sobre el desenlace  $Y$ . Nótese que en este caso  $x = x_i$ .

Se busca construir un modelo que explique la asociación entre  $x_i$  y  $p_i = P(Y = 1 | x_i)$ . El primer enfoque es formular el siguiente modelo de regresión.

$$y = \beta_0 + \beta_1 * x + \epsilon \dots (1)$$

y estimar los parámetros por mínimos cuadrados de la forma habitual.

Dado que el propósito del análisis consiste en identificar qué variables son mejores para clasificar entre los dos grupos, la función lineal se encuentra con problemas de interpretación, tanto del modelo como de sus coeficientes estimados. Pues tomando esperanzas en (1) para  $x = x_i$  se tiene que:

$$E[y | x_i] = \beta_0 + \beta_1 * x$$

Por otra parte la esperanza de  $y$  es:

$$E[y | x_i] = P(y = 1 | x_i) * 1 + P(y = 0 | x_i) * 0 = p_i$$

En consecuencia, la predicción  $y_i$  estima la probabilidad de que un individuo con características definidas por  $x = x_i$  pertenezca a la población correspondiente a  $y = 1$ .

El inconveniente principal de esta formulación es que  $p_i \in (0, 1)$  y no se puede traducir esta asociación sobre la forma lineal debido a que no hay ninguna garantía de que la predicción,  $p_i = \beta_0 + \beta_1 * x_i$  verifique esta restricción pues toma valores dentro de  $\mathbb{R}$ .

### 2.3. Modelo logístico Logit

Si queremos que el modelo proporcione directamente la probabilidad de pertenecer a cada uno de los grupos, debemos transformar la variable respuesta de algún modo para garantizar que la respuesta prevista está entre cero y uno.

Si tomamos

$$p_i = F(\beta_0 + \beta_1 * x_i)$$

garantizaremos que  $p_i$  esté entre cero y uno, si exigimos que  $F$  tenga esa propiedad.

Una solución eventual consiste en tomar como  $F$  alguna función de distribución. Esto debido a que las funciones de distribución son funciones no decrecientes y acotadas entre cero y uno.

Habitualmente se toma como  $F$  la función de distribución logística estandarizada; también conocida con el nombre de Ley de Verhulst. Esto gracias a su simplicidad comparada con otras funciones utilizadas en la estimación de modelos de variable dependiente cualitativa tales como la función probit, que es la función inversa de la función de distribución normal centrada y reducida o bien la función log -log (modelo gompit) que es la función inversa de la función de distribución de la ley de Gompertz que tiene el inconveniente de ser asimétrica.

Así pues la función logística estandarizada (logit) está dada por:

$$y = \frac{1}{(1 + e)^{-bx}} \quad -\infty < x < \infty$$

Por otra parte, dado que el modelo que queremos estimar tiene la forma:

$$p_i = F(x_i\beta)$$

Si se adopta la función logística como representación del fenómeno estudiado, tendremos entonces

$$p_i = \frac{1}{(1 + e)^{-(\beta_0 + \beta_i x_i)}}$$

Y podremos escribir

$$1 - p_i = 1 - \frac{1}{(1 + e)^{-(\beta_0 + \beta_i x_i)}} = \frac{e^{-(\beta_0 + \beta_i x_i)}}{(1 + e)^{-(\beta_0 + \beta_i x_i)}}$$

O bien

$$G_i = \frac{p_i}{1 - p_i} = e^{(\beta_0 + \beta_i x_i)}$$

Y tomando logaritmos en ambos lados tenemos:

$$\ln \frac{p_i}{1 - p_i} = (\beta_0 + \beta_i x_i)$$

Con lo que nos encontramos de nuevo con un modelo lineal clásico. El logaritmo de la variable  $G_i$  representa la diferencia entre las probabilidades de pertenecer a ambas poblaciones en una escala logarítmica, y al ser una función lineal de las variables explicativas facilita la estimación y la interpretación del modelo.

## 2.4. Estimación de los parámetros

Sea  $\beta = (\beta_0, \dots, \beta_k)$  el vector de los coeficientes a estimar del modelo. El método más usado para la estimación de los coeficientes del modelo y de sus errores estándar es el de estimaciones de máxima verosimilitud es decir, estimaciones que maximicen la probabilidad de obtener los valores de la variable dependiente  $Y$  proporcionada por los datos de la muestra.



El método consiste en resolver  $k+1$  ecuaciones de verosimilitud que se obtienen derivando el logaritmo natural de la función de verosimilitud respecto a  $k+1$  coeficientes, como se muestra a continuación:

Sea  $\pi(x_m) = P(Y = 1 | x = x_m)$ , sea  $M$  el número de observaciones de  $x_m$  en la muestra, sea  $z_m$  la suma de los  $y_i$  para las observaciones, es decir el número de ellos para los cuales  $y_i = 1$ . Notando y el vector  $(y_1 \dots y_n)$  de los valores tomados por la variable por explicar, la verosimilitud de la muestra se escribe:

$$L(\beta, y) = \prod_{m=1}^M [(\pi(x_m))^{z_m} (1 - \pi(x_m))^{M-z_m}] \quad m \in 1, 2, 3, \dots M$$

y su logaritmo

$$\ln L(\beta, y) = \sum_{m=1}^M [z_m \ln \frac{\pi(x_m)}{1 - \pi(x_m)} + M * \ln(1 - \pi(x_m))]$$

de donde, utilizando la función logit y generando la ecuación (1) en el caso donde las interacciones se pueden tomar en cuenta, se obtiene:

$$\ln L(\beta, y) = \sum_{m=1}^M z_m \beta^T \bar{x}_m - \sum_{m=1}^M M * \ln(1 + \exp(\beta^T \bar{x}_m))$$

El estimador  $\hat{\beta}$  de máxima verosimilitud de  $\beta$  verifica:

$$\left\{ \frac{d \ln L(\beta, y)}{d \beta} \right\}_{\beta = \hat{\beta}} = 0$$

y es solución de la ecuación:

$$\sum_{m=1}^M z_m \bar{x}_m = \sum_{m=1}^M n(x_m) \bar{x}_m \frac{\exp(\beta^T \bar{x}_m)}{1 + \exp(\beta^T \bar{x}_m)}$$

Para resolver esta ecuación no lineal en  $\beta$  se utiliza un algoritmo iterativo del tipo Newton-Raphson.

## 2.5. ODDS Ratio y Riesgo Relativo

El *odds* asociado a cierto suceso se define como la razón entre la probabilidad de que dicho suceso ocurra y la probabilidad de que no ocurra; es decir, se trata de un número que expresa cuanto más probable es que se produzca a que no se produzca el hecho en cuestión.

Formalmente, si llamamos E a dicho suceso, P(E) a la probabilidad de que ocurra y O(E) al *odds* que le corresponde, entonces se tiene:

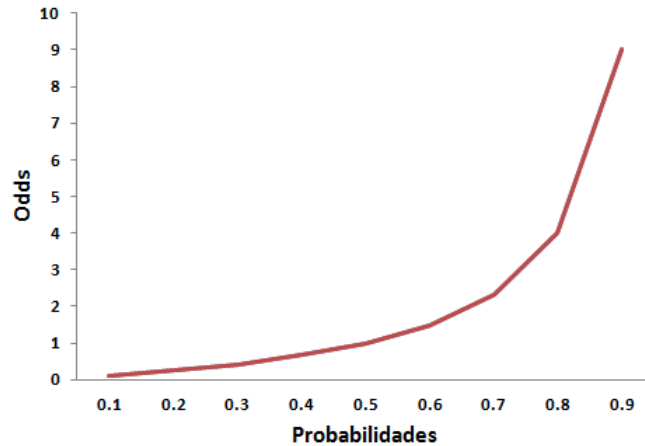
$$O(E) = \frac{P(E)}{1-P(E)}$$

Por otra parte, conocido el *odds*, se puede deducir la probabilidad. En general, si el *odds* de un suceso es O(E), entonces su probabilidad es:

$$P(E) = \frac{O(E)}{O(E)+1}$$

De modo que ambas informaciones son equivalentes y expresan la misma noción: cuantifican cuan probable es que algo ocurra (en particular, cuál es el riesgo de un acontecimiento).

Obviamente entre la probabilidad del suceso y el *odds* correspondiente hay una clara relación directa. Si aquella aumenta, este también lo hace, si P(E)=0, entonces O(E) también es nulo; pero en medida que P(E) tiende a la unidad, O(E) tiende a infinito. Lo anterior se ve reflejado en la siguiente figura:



Fuente: Silva Luis Carlos (2004). Regresión Logística.

El llamado *odds ratio* se define como la razón del *odds* correspondiente a un suceso bajo cierta condición entre el *odds* que le corresponde al mismo suceso bajo otra condición. Así pues:

$$OR = \frac{\frac{P_F(E)}{1-P_F(E)}}{\frac{P_{\hat{F}}(E)}{1-P_{\hat{F}}(E)}}$$

donde  $\hat{F}$  denota la no presencia del factor F y por tanto  $P_F(E)$  representa la probabilidad del evento E bajo la condición F y  $P_{\hat{F}}(E)$  representa la probabilidad del evento E bajo la no presencia del factor F.

## 2.6. Pruebas estadísticas al modelo logit

Usualmente en la estimación de un modelo de regresión, se efectúan pruebas estadísticas con la finalidad de determinar la calidad del ajuste global del modelo. A continuación se muestran algunas de las pruebas estadísticas comunmente utilizadas.

### 2.6.1. Deviance

Debido a que la función de verosimilitud mide la plausibilidad de un modelo de regresión logística, no debe sorprendernos que para valorar su capacidad predictiva sea central la consideración

de la verosimilitud.

Una forma de probar el ajuste global del modelo, es mediante la estadística llamada Devianza, representada por la función  $D(\beta) = -2L(\beta)$

$$D(\beta) = -2 \sum_{i=1}^n [y_i \log p_i + (1 - y_i) \log(1 - p_i)]$$

Por otra parte siempre que se ajusta un modelo, el algoritmo de la regresión logística computa dos lejanías: la que corresponde propiamente al modelo que se ha ajustado ( $D$ ) y la que corresponde al modelo nulo ( $D(\beta_0)$ ).

La lejanía del modelo nulo es más grande que la de cualquier modelo ampliado. Esto es razonable dado que se trata de un modelo mucho menos sofisticado y debe necesariamente tener una falta de ajuste mayor. La diferencia entre estas lejanías mide el aporte que hacen las variables incorporadas al modelo. Es decir para valorar dicho aporte se puede calcular:

$$D - D(\beta_0) = -2 \ln L(\beta_0) + 2 \ln L(\beta) = -2(\ln L(\beta_0) - \ln L(\beta)) = -2 \ln \frac{L(\beta_0)}{L(\beta)}$$

que se distribuye Ji-cuadrado con  $n-k-1$  grados de libertad donde  $k$  es el número de variables presentes en el modelo ampliado.

La diferencia entre estos dos valores prueba si la o las variables  $x_i$  son significativas.

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

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

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

### 2.6.2. Estadístico de Wald

En la regresión logística, una vía para probar si una variable independiente hace un aporte estadísticamente significativo a la explicación de la variable respuesta sin tener que usar la razón de verosimilitudes es la prueba de Wald. Esta prueba permite contrastar si un coeficiente  $\beta$  es significativamente diferente de cero.

El estadístico de prueba se obtiene dividiendo la estimación del coeficiente de interés por su error estándar:

$$Z_{Wald} = \frac{\hat{\beta}_i}{\sigma(\hat{\beta}_i)}$$

donde  $\sigma(\hat{\beta}_i)$  es el error estándar de  $\beta$ .

Las hipótesis estadísticas de la prueba son:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

por tanto la prueba Wald bajo  $H_0$  sigue una distribución T-student con  $n - p - 1$  grados de libertad y para muestras grandes se distribuye como una normal estándar. Se entiende que si  $Z_{Wald}$  es un valor alejado de cero entonces  $H_0$  sería falsa y la región crítica de la prueba estara dada por  $|Z_{Wald}| > t_{\frac{\alpha}{2}}$  para un nivel de significación adecuado.

### 2.6.3. Residuos de Pearson

Una vez ajustado un modelo de regresión logística podría ocurrir que no se cumpla el supuesto de linealidad entre el logit de la probabilidad del suceso de interes y las variables independientes o bien que la presencia de algunas observaciones extemas en el conjunto de datos que se está manejando perturbe la calidad del ajuste, o que otra función distinta a la logística fuese mejor para describir las observaciones.

Existen numerosos métodos para comprobar si el modelo ajustado es adecuado o no. Uno de ellos es el análisis de residuos en la regresión logística.

La medida que expresa la diferencia entre las respuestas observadas y las predichas por el modelo se conoce como residuo.

Los residuos crudos son las respuestas observadas menos las predichas por el modelo

$$y_i - n_i \hat{p}_i$$

donde  $n_i$  representa, en caso de datos agrupados, el número de individuos con un perfil de entrada específico,  $y_i$  es el número de estos que presentan la característica de interés y  $\hat{p}_i$  la probabilidad predicha para un individuo con tal perfil.

Los residuos crudos divididos por su error estándar son los llamados residuos crudos estandarizados o residuos de Pearson. Para el caso de datos no agrupados (es decir  $n_i = 1$ ), los residuos anteriores toman respectivamente las siguientes formas:

$$y_i - p_i y$$
$$e_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}}$$

Si el modelo es adecuado  $e_i$  tiene media cero y varianza uno, de aquí se construye el estadístico de prueba que se distribuye asintóticamente como:

$$\chi^2 = \sum_{i=1}^n e_i^2$$

que sigue una distribución  $\chi^2$  con  $n - (k + 1)$  grados de libertad con  $k$  variables.

#### **2.6.4. Variables DUMMY en regresión logística**

Cuando alguna de las variables independientes es de índole cualitativa o nominal, su inclusión entre las covariables consideradas en el modelo de regresión logística exige un tratamiento especial.

Puesto que el entorno en que se inscribe el problema de la regresión logística tal y como ocurre en otros tipos de regresión es cuantitativo, resulta inaceptable hacer intervenir de manera directa a las variables cualitativas o nominales entre las que integran el perfil de entrada. Sin embargo, es obvio que rasgos tales como la raza, la religión o el grupo sanguíneo que se tiene, podrían ser modificadores de la probabilidad de ocurrencia de un suceso. Por tanto resultan ser con frecuencia variables cuya incorporación en el análisis resulta conveniente.

Así pues; si una de las variables explicativas es categórica, con  $c$  valores posibles, se crean  $c-1$  variables dicotómicas como variables explicativas: las llamadas variables dummy a las que denotaremos  $z_1, z_2, \dots, z_{c-1}$ . A cada categoría o clase de la variable nominal le corresponde un conjunto de valores de los  $z_i$  con el cual se identifica dicha clase.

La manera más usual de definir estas variables es la siguiente: si el sujeto pertenece a la primera categoría, entonces las  $c-1$  variables dummy valen 0: se tiene  $z_1 = z_2 = \dots = z_{c-1} = 0$ ; si el sujeto se halla en la segunda categoría, entonces  $z_1 = 1$  y las restantes valen cero;  $z_2$  vale 1 sólo para aquellos individuos que están en la tercera categoría, en cuyo caso las otras variables asumen el valor de cero, y así sucesivamente hasta llegar a la última categoría, para la cual  $z_{c-1}$  es la única que vale 1.

En cualquier caso, si se ajusta un modelo que incluya una variable nominal con  $c$  clases, ésta debe ser sustituida por las  $c-1$  variables dummy y a cada una de ellas correspondería su respectivo coeficiente.

## Capítulo 3

# Análisis y reducción de variables bajo la perspectiva de los modelos de score

El capítulo anterior estuvo enfocado en la estimación e interpretación de los coeficientes en un modelo de regresión logística. El modelo de regresión que se ajusta en esta tesis, consta de aproximadamente de 17 variables que pueden ser potencialmente incluidas en el modelo de regresión. Este capítulo se plantean los métodos de análisis y reducción de variables.

El análisis de variables consiste básicamente en investigar la relación entre las variables que se pretenden incluir, con la finalidad de precisar el aporte en términos de predicción de dichas variables al modelo de regresión al mismo tiempo que se busca minimizar el número de variables.

Acorde a un artículo publicado por Bolton Christine (2009), algunas de las características que deben ser tomadas en cuenta en el proceso de selección de variables son:

- Las variables deben ser predictivas.
- Las variables deben ser lógicas y lo más simples posible, esto con la finalidad de hacer fácilmente la interpretación de los resultados.
- Las variables deben estar disponibles y ser lo más estable posibles en el tiempo.
- Las variables que se incluyan, no deben propiciar problemas de multicolinealidad.



- Las variables que se excluyan, no deben representar una pérdida significativa de información.

### **3.1. Análisis Bivariado de los datos: Tablas de Contingencia**

El estudio de la influencia de una variable (variable independiente) sobre la forma en la que se modifica otra (variable dependiente) es conocido como análisis bivariado.

Analizar la distribución de una variable con relación a otras es un paso fundamental previo al ajuste de un modelo de regresión, pues permite al investigador elegir las variables adecuadas para su análisis estadístico y minimizar el riesgo de tener exceso de información o bien información redundante en el modelo.

Por otra parte Bolton Christine(2009), en el documento titulado “Logistic Regresion and its application in credit scoring”, el proceso de selección y análisis de variables previo al ajuste de un modelo de regresión debe ser realizado cuidadosamente para cada una de las variables candidatas.

Las técnicas existentes para realizar este tipo de análisis varían según la naturaleza de las variables. Cabe mencionar que la mayoría de las variables disponibles para la construcción del modelo de regresión de esta tesis, son de tipo cualitativo.

El primer paso en el estudio de la relación entre dos variables cualitativas (o categóricas) es construir la tabla de la distribución de frecuencias conjunta, es decir las frecuencias de los cruces de todas las clases de una variable por todas las clases de la otra, conocida con el nombre de Tabla de Contingencia.

La tabla de contingencia, es una tabla de doble entrada, donde en cada casilla figura el número

de casos o individuos que poseen ambos niveles o características de los factores analizados.

Consideremos una población y particiones exhaustivas y mutuamente excluyentes. Si X y Y son dos variables categóricas, la tabla de contingencia se reduce a una tabla como la que se muestra a continuación:

Cuadro 3.1: Estructura: Tabla de Contingencia

x/y	1	2	Total
1	$n_{11}$	$n_{12}$	$n_{1\cdot}$
2	$n_{21}$	$n_{22}$	$n_{2\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	$n$

donde:

- $n_{ij}$  = número de observaciones que tienen el atributo i y j.
- $n_{i\cdot}$  = número de individuos que tienen el atributo i (marginal i).
- $n_{\cdot j}$  = número de individuos que tienen el atributo j (marginal j).

La tabla de contingencia se define por el número de atributos o variables que se analizan conjuntamente y el número de modalidades o niveles de los mismos. El ejemplo propuesto es una tabla de contingencia 2x2, ya que tiene dos atributos (X,Y) y cada uno de ellos tiene dos niveles.

Por medio de la tabla de contingencia se pueden resolver distintas cuestiones por ejemplo, a través de dicha tabla se podrá determinar si existe una relación estadísticamente significativa entre las variables estudiadas. En segundo lugar nos será posible cuantificar dicha relación y estudiar su relevancia; dado que si dos variables son independientes significa que los valores de una de ellas no están influidos por la modalidad o el nivel que adopte la otra, lo cual a su vez nos daría la certeza de no estar introduciendo información fuertemente correlacionada al modelo de regresión.

## 3.2. Prueba Chi-cuadrada

La prueba Chi-cuadrada es empleada para identificar dependencia entre las variables cualitativas del modelo. La hipótesis nula de la prueba postula la independencia entre los factores, siendo la hipótesis alternativa la de dependencia entre los factores.

Para realizar este contraste se disponen los datos en una tabla de frecuencias. Para cada valor o intervalo de valores se indica la frecuencia absoluta observada o empírica ( $O_{ij}$ ). A continuación, y suponiendo condiciones de independencia, es posible calcular para cada valor o intervalo de valores la frecuencia esperada:

$$E_{ij} = \frac{n_{i.} * n_{.j}}{N}$$

El estadístico de prueba se basa en las diferencias entre la  $O_{ij}$  y  $E_{ij}$  y se define como:

$$\hat{\chi}^2 = \sum_{i=1}^h \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

El valor de  $\hat{\chi}^2$  calculado se compara con el valor tabulado de una para un nivel de confianza determinado y  $(n-1)(k-1)$  grados de libertad, pudiéndose tener las siguientes regiones de rechazo para la prueba.

Si  $\hat{\chi}^2 > \hat{\chi}^2_{(n-1)(k-1)}$  se rechaza  $H_0$  (dependencia entre las variables)

Si  $\hat{\chi}^2 < \hat{\chi}^2_{(n-1)(k-1)}$  se acepta  $H_0$  (independencia entre las variables)

## 3.3. Pesos de Evidencia (WOE)

El estadístico Pesos de Evidencia o WOE por sus siglas en inglés, se basa en una evaluación cuantitativa de los datos. El poder de predicción en cada atributo o grupo de atributos en una variable se calcula a través de este estadístico. Su cálculo consiste en medir la diferencia de las proporciones de los dos estados de la variable dependiente (en este caso pobre y no pobre) en cada atributo de las variables independientes.

$$WOE_i = \log \frac{\frac{N_i}{\sum_{i=1}^n N_i}}{\frac{P_i}{\sum_{i=1}^n P_i}}$$

donde P es el número de ocurrencias, N es el número de no ocurrencias del evento en observación y el índice  $i$  representa el atributo que está siendo evaluado. La condición es ningún valor en cero para todas las  $N_i$  y todas las  $P_i$ . Guoping Zeng (2013)

El evento de interés de esta tesis es la medición de la pobreza, por lo que la definición de WOE considerando tal evento queda expresada de la siguiente manera:

$$WOE_i = 100 * \log \frac{(Distribución de no pobres en el atributo i)}{(Distribución de pobres en el atributo i)}$$

Cuando se obtienen valores negativos del WOE significa que se tienen proporciones altas de Pobres sobre los No Pobres, lo cual nos da una medida acerca de si un grupo en particular está asociado con altos o bajos niveles de riesgo, entendiendo la palabra riesgo como que tan propenso es que se sea Pobre.

Para que el WOE está definido, ninguna de las clases debe estar formada por pobres o por no pobres únicamente.

### 3.4. Valor de la información (IV)

El valor de la información (IV) es una medida del poder de predicción global de la característica, y se define como:

$$IV_{group} = \frac{N_i}{\sum_{i=1}^n N_i} - \frac{P_i}{\sum_{i=1}^n P_i} * WOE_i$$

donde N es el número de no ocurrencias, P es el número de ocurrencias, WOE es el peso de la evidencia e  $i$  representa el atributo que está siendo evaluado.

Para calcular el IV de una variable (la cual cubre el total de atributos o posibles valores de la variable), el IV de cada atributo debe ser sumado.

$$IV_{group} = \sum_{i=1}^n \left[ \frac{N_i}{\sum_{i=1}^n N_i} - \frac{P_i}{\sum_{i=1}^n P_i} * WOE_i \right]$$

donde  $n$  es el número total de atributos. El IV es siempre positivo, grandes valores del IV indican que la variable es predictiva, valores pequeños indican que no. Guoping Zeng (2013)

Siddiqi (2006) considera que una característica con un IV:

- menor a 0.02 es tiene nulo valor predictivo.
- entre 0.02 y 0.1 es de predicción débil.
- entre 0.1 y 0.3 es de predicción media.
- mas de 0.3 es de predicción fuerte.

Siddiqi (2006) aconseja que las características con  $IV \leq 2\%$  deben excluirse del modelo.

### 3.5. Coeficientes de correlación

Los coeficientes de correlación son medidas que indican a través de una expresión numérica el grado de relación existente entre dos variables y nos proporcionan una idea sobre la medida en que estas se relacionan. Los límites de los coeficientes de correlación van de  $-1$  a  $+1$  pasando por el cero, una correlación de  $+1$  significa que existe una relación lineal perfecta (positiva), una correlación de  $-1$  significa que existe una relación lineal inversa perfecta; mientras que una correlación de cero se interpreta como la no existencia de una relación lineal entre las variables estudiadas.

Algunos de los coeficientes de correlación más empleados son:

- El coeficiente de correlación de Pearson (1896). Pensado para variables cuantitativas (escala mínima de intervalo), es un índice que mide el grado de covariación entre distintas variables relacionadas linealmente.

La fórmula de este coeficiente suele ser expresada de la siguiente manera:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

En algunas situaciones, resulta difícil precisar a partir de qué valor de  $r_{xy}$  podemos considerar que existe una correlación lineal entre dos variables. Siempre se debe tener en cuenta para la interpretación el tipo de variables a las que se les aplica. Sin embargo para tener un referente, consideremos la siguiente tabla de correlaciones y su interpretación:

Cuadro 3.2: Escala de Correlaciones

Correlación	Descripción
$r = 1$	Correlación Perfecta.
$0.8 < r < 1$	Correlación muy alta.
$0.6 < r < 0.8$	Correlación alta.
$0.4 < r < 0.6$	Correlación moderada.
$0.2 < r < 0.4$	Correlación baja.
$0 < r < 0.2$	Correlación muy baja.
$r = 0$	Correlación nula.
$-1 < r < 0$	Aplica la misma escala que los valores positivos, reflejando una correlación inversa.

Fuente : <http://www.mcgraw-hill-educacion.com>

- El coeficiente de correlación rho de Spearman (1904) es el coeficiente de correlación de Pearson, pero aplicado después de transformar las puntuaciones originales en rangos. También toma valores entre  $-1$  y  $+1$ , y se interpreta exactamente igual que el coeficiente de correlación de Pearson. El coeficiente de correlación de Spearman es calculado de la siguiente manera:

$$\rho = 1 - \frac{6 \sum(x_R - y_R)^2}{n^3 - n}$$

donde  $(x_R - y_R)$  hace referencia a los respectivos rangos de la variable independiente (x) y la variable dependiente (y) para la misma observación y n se refiere al número total de casos que han sido ranqueados.

## 3.6. Análisis de Cluster

El análisis de cluster es una técnica multivariada, cuya idea básica es agrupar un conjunto de observaciones en grupos. Este agrupamiento se basa en la idea de distancia o similitud entre las observaciones; es decir agrupa elementos (o variables) tratando de lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre los grupos. Cada cluster se describe como la clase a la que sus miembros pertenecen.

Esta técnica es de gran utilidad en el proceso de selección y reducción de variables. Podemos encontrarnos dos tipos fundamentales de métodos de clasificación: Jerárquicos y No Jerárquicos. En los métodos jerárquicos no se conoce previamente el número de grupos a formar, mientras que en los no jerárquicos, se parte de un número fijo de Clusters (F. Hair Joseph, Anderson E. Rolph, 1999).

### 3.6.1. Medidas de Similitud

Una vez establecidas las variables y los objetos a clasificar el siguiente paso consiste en establecer una medida de proximidad o de distancia entre ellos que cuantifique el grado de similaridad entre cada par de objetos.

Se pueden agrupar observaciones según la similitud expresada en términos de una distancia. Dados dos vectores  $x_i, x_j$  pertenecientes a  $R^k$ , diremos que hemos establecido una distancia entre ellos si definimos una función  $d$  con las propiedades siguientes:

- I.  $d : R^k \times R^k \rightarrow R^+$ , es decir  $d(x_i, x_j) \geq 0$ .
- II.  $d(x_i, x_j) = 0 \forall i$ , la distancia entre un elemento y sí mismo es cero.
- III.  $d(x_i, x_j) = d(x_j, x_i)$  la distancia es simétrica.
- IV.  $d(x_i, x_j) \leq d(x_i, x_p) + d(x_p, x_j)$  la distancia verifica la propiedad triangular.

### 3.6.2. Métodos Cluster

Dentro de los métodos Cluster, podemos distinguir los jerárquicos de los no jerárquicos.

- Jerárquicos : La pertenencia a un grupo (Cluster) en un nivel de jerarquía condiciona a la pertenencia en grupos de un nivel superior (matriz de distancia o similaridad).
  - I. Métodos jerárquicos aglomerativos: se comienza con los objetos o individuos de modo individual; de este modo, se tienen tantos clusters iniciales como objetos. Luego se van agrupando de modo que los primeros en hacerlo son los más similares y al final, todos los subgrupos se unen en un único cluster.
  - II. Métodos jerárquicos divisivos: se actúa al contrario. Se parte de un grupo único con todas las observaciones y se van dividiendo según lo lejanos que están.

Entre los algoritmos jerárquicos destacan los siguientes métodos:

- Método de las distancias mínimas (Vecino más cercano): se busca la mayor semejanza entre los elementos o grupos más cercanos.
- Método de las distancias máximas (Vecino más lejano): se calcula la mínima distancia entre los elementos más alejados.
- Método de las distancias medias: se calcula la media de las distancias entre elementos.

En cualquier caso, de ambos métodos se deriva un dendograma, que es un gráfico que ilustra cómo se van haciendo las subdivisiones o los agrupamientos, etapa a etapa.

Consideremos un ejemplo sencillo con solo 5 objetos. Sea la matriz de distancias entre estos objetos la dada por:

$$D = [d_{ik}]_{ik} = \begin{pmatrix} 0 & & & & \\ 9 & 0 & & & \\ 3 & 7 & 0 & & \\ 6 & 5 & 9 & 0 & \\ 11 & 10 & 2 & 8 & 0 \end{pmatrix}$$



Cada uno de los objetos comienza siendo un cluster. Como  $\min_{i,k} d_{ik} = d_{53} = 2$  los objetos 3 y 5 se unen para formar el cluster (35). Para construir el siguiente nivel, se calcula la distancia entre el cluster (35) y los restantes objetos 1, 2 y 4. Así, usando el método de la distancia mínima:

$$d_{(35),1} = \min\{d_{31}, d_{51}\} = \min\{3, 11\} = 3$$

$$d_{(35),2} = \min\{d_{32}, d_{52}\} = \min\{7, 10\} = 7$$

$$d_{(35),4} = \min\{d_{34}, d_{54}\} = \min\{9, 8\} = 8$$

Reconstruyendo la matriz de distancias con las distancias anteriores:

$$D = [d_{ik}]_{ik} = \begin{pmatrix} 0 & & & & \\ 3 & 0 & & & \\ 7 & 9 & 0 & & \\ 8 & 6 & 5 & 0 & \end{pmatrix}$$

De la matriz anterior, seleccionamos nuevamente la menor distancia,  $d_{(35),1} = 3$  y junto, así, el cluster (35) con el 1.

Calculo ahora las distancias del nuevo cluster a los dos elementos que quedan:

$$d_{(351),2} = \min\{d_{(35),2}, d_{12}\} = \min\{7, 9\} = 7$$

$$d_{(351),4} = \min\{d_{(35),4}, d_{14}\} = \min\{8, 6\} = 6$$

La matriz de distancias queda como:

$$D = [d_{ik}]_{ik} = \begin{pmatrix} 0 & & & & \\ 7 & 0 & & & \\ 6 & 5 & 0 & & \end{pmatrix}$$

La mínima distancia se alcanza entre los clusters 2 y 4  $d_{24} = 5$ . Se obtienen así dos clusters: (351) y (24). La distancia que los separa es:

$$d_{(351),(24)} = \min\{d_{(351),2}, d_{(351),4}\} = \min\{7, 6\} = 6$$

Así, la matriz de distancias queda como:

$$D = [d_{ik}]_{ik} = \begin{pmatrix} 0 & \\ 6 & 0 \end{pmatrix}$$

Cuando la distancia es iguala 6, todos los objetos se unen en un único cluster.

- No Jerárquicos : Obtienen una única partición de los datos mediante la optimización de algun método iterativo. Se usan para agrupar objetos, pero no variables, en un conjunto de k clusters ya predeterminado. No se tiene que especificar una matriz de distancias ni se tienen que almacenar las iteraciones.

Se parte de un conjunto inicial de clusters elegidos al azar, que son los representantes de todos ellos; luego se van cambiando de modo iterativo. Se usa habitualmente el método iterativo de las k-medias.

El método de las k-medias es un método que permite asignar a cada observación el cluster que se encuentra más próximo en términos del centroide (media). En general, la distancia empleada es la euclídea.

Los pasos que lleva a cabo este algoritmo son :

- I. Se toman al azar k clusters iniciales.

II. Para el conjunto de observaciones, se vuelve a calcular las distancias a los centroides de los clusters y se resignan a los que están más próximos. Se vuelven a recalcular los centroides de los  $k$  clusters después de las reasignaciones de los elementos.

III. Se repiten los dos pasos anteriores hasta que no se produzca ninguna reasignación, es decir, hasta que los elementos se estabilicen en algún grupo. Usualmente, se especifican  $k$  centroides iniciales y se procede al paso (2) y, en la práctica, se observan la mayor parte de reasignaciones en las primeras iteraciones.

**Ejemplo:** Supongamos dos variables  $x_1$  y  $x_2$  y 4 elementos: A, B, C, D, con la siguiente tabla de valores:

	$x_1$	$x_2$
A	5	3
B	-1	1
C	1	-2
D	-3	-2

Se quiere dividir estos elementos en dos grupos ( $k = 2$ ). De modo arbitrario, se dividen los elementos en dos clusters (AB) y (CD) y se calculan los centroides de los dos clusters.

Cluster (AB):

$\bar{x}_1$	$\bar{x}_2$
$\frac{5-1}{2} = 2$	$\frac{3+1}{2} = 2$

Cluster (CD) :

$\bar{x}_1$	$\bar{x}_2$
$\frac{1-3}{2} = -1$	$\frac{-2-2}{2} = -2$

En el paso (2), calculamos las distancias euclídeas de cada observación al grupo de centroides y reasignamos cada una al grupo más próximo. Si alguna observación se mueve de grupo, hay que volver a calcular los centroides de los grupos. Así, las distancias son:

$$d^2(A, (AB)) = (5 - 2)^2 + (3 - 2)^2 = 10$$

$$d^2(A, (CD)) = (5 + 1)^2 + (3 + 2)^2 = 61$$

Como A está más próximo al cluster (AB) que al cluster (CD), no se reasigna. Se hace lo mismo para el elemento B:

$$d^2(B, (AB)) = (-1 - 2)^2 + (1 - 2)^2 = 10$$

$$d^2(B, (CD)) = (-1 + 1)^2 + (1 + 2)^2 = 9$$

Por lo cual, el elemento B se reasigna al cluster (CD) dando lugar al cluster (BCD). A continuación, se vuelven a calcular los centroides:

Cluster	$\bar{x}_1$	$\bar{x}_2$
A	5	3
(BCD)	-1	-1

Nuevamente, se vuelven a calcular las distancias para cada observación para ver si se producen cambios con respecto a los nuevos centroides:

	A	(BCD)
A	0	52
B	40	4
C	41	5
D	89	5

Como no se producen cambios, entonces la solución para  $k = 2$  clusters es: A y (BCD). Si se quiere comprobar la estabilidad de los grupos, es conveniente volver a correr el algoritmo con otros clusters iniciales (una nueva partición inicial).

### 3.7. Kolmogorov Smirnov

Una vez que se ha calculado el score con la regresión logística, se pretende determinar si los puntajes asignados están acorde con el grupo al que pertenecen ; es decir se espera que los individuos clasificados como No Pobres reciban menor puntuación y viceversa los Pobres mayor puntuación.

Una de las pruebas no paramétricas , para determinar el poder de discriminación del modelo es la prueba de Kolmogorov- Smirnov. Esta prueba se basa en una medida de separación y consiste en medir cuán distintas son las funciones de de distribución, en este caso de Pobres y No Pobres para cada rango percentil del puntaje score Nieto Soraida (2010).

El estadístico de prueba se calcula como la máxima diferencia absoluta entre sus distribuciones, entonces busca detectar la discrepancias existentes entre las frecuencias relativas acumuladas de las dos muestras de estudio. La prueba se construye sobre las hipótesis nula y alternativa como sigue:

$H_0$  : *Las distribuciones poblacionales son iguales.*

$H_1$  : *Las distribuciones poblacionales son diferentes.*

Para esta prueba se requiere tener dos muestras de una variable continua, o al menos una escala ordinal. Los valores de la variable se agrupan en  $k$  categorías o intervalos según sea conveniente. Para cada categoría se calculan las frecuencias relativas acumuladas. El estadístico esta dado como la máxima diferencia de las distribuciones de frecuencias relativas acumuladas.

$$KS = \max | \%Distribucion\ acum.\ No\ Pobres - \%Distribucion\ acum.\ Pobres |$$

El estadístico KS es un número entre 0 y 1. Entre más grande sea dicha diferencia, más fuerte será el poder de discriminación del modelo.

### 3.8. Índice de Gini

En 1960 se pensó medir la desigualdad en los índices de salud con la curva de Lorenz, el índice de Gini se deriva de ésta. El índice de Gini es uno de los instrumentos más utilizados para medir la desigualdad entre dos poblaciones. En este caso se utiliza para medir la desigualdad de las poblaciones de individuos clasificados como pobres y no pobres.

Sea  $F(x)$  y  $G(x)$  las funciones de distribución de los individuos clasificados como pobre y no pobre respectivamente, con  $x$  el puntaje del score, la curva de Lorenz es el subconjunto:

$$L(F, G) = \{(u, v) | u = F(x), v = G(x), \text{con}, x \in R\}$$

Cuando el score discrimina bien a los pobres de los no pobres, el puntaje de score para los no pobres es menor que el puntaje score para los pobres, entonces la curva de Lorenz de  $F$  y  $G$  es cóncava hacia arriba como se muestra en la siguiente figura.

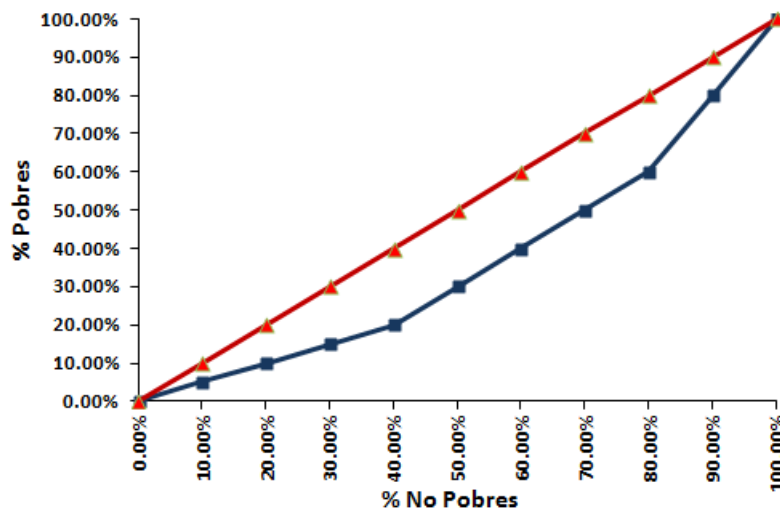


Figura 3.1: Gráfico Índice de Gini

Fuente : Bolton Christine (2009)

Cuando  $F(x) = G(x)$  la curva de Lorenz  $L(F;G)$  describe la recta  $u = v$  con  $u, v \in (0, 1)$ . Por lo tanto mientras  $L$  se separe más de la recta  $v = u$ , mayor será la diferencia entre  $F(x)$  y  $G(x)$ .

El coeficiente de Gini puede ser calculado de la siguiente manera:

Área bajo la curva azul (por la regla del trapecio):

$$A = \sum_{i=1}^n \frac{1}{2} (b_i - b_{i-1}) * (a_{i-1} + a_i)$$

Área bajo la identidad (triángulo):

$$\frac{1}{2} * 1 * 1 = 0.5$$

$$\Rightarrow Gini = \frac{1/2 - A}{1/2}$$

# Capítulo 4

## Un modelo de Score para medir Pobreza en México

De acuerdo con el CONEVAL, la pobreza corresponde a múltiples factores. La definición, identificación y medición de la pobreza contempla dos enfoques: bienestar económico y de derechos sociales. El primero es medido operativamente por el ingreso corriente per cápita y el segundo mediante seis indicadores de carencias sociales (rezago educativo, acceso a los servicios de salud, acceso a la seguridad social, acceso a la alimentación, calidad y espacios de la vivienda).

Debido a que la variable dependiente a estudiar en esta tesis es dicotómica (Pobre o no Pobre) en esta sección se estima un modelo de regresión logística, con el propósito de reproducir con un menor número de variables la clasificación de individuos en situación de pobreza hecha por el CONEVAL. Una vez obtenido el modelo de regresión, se asigna un puntaje a cada uno de los individuos el cuál ordena a estos de acuerdo a su grado de pobreza. Los puntajes más altos corresponden a las personas más pobres y en sentido contrario los puntajes más bajos corresponden a las personas menos pobres. El puntaje obtenido, es propuesto como una métrica para hacer evaluaciones de pobreza en México, que puede ser usada para realizar estrategias de combate a la pobreza.

La base de datos empleada en la estimación del modelo se construye de la siguiente manera. Utilizamos la base de datos del MCS 2008 (Módulo de Condiciones Socioeconómicas de la ENIGH



2008), la cual contiene las variables necesarias para la medición multidimensional de la pobreza, establecidas en la LGDS.

Dicha base de datos, incorpora información sobre ingresos, salud, educación, seguridad social, calidad y espacios de la vivienda, servicios básicos, alimentación y cohesión social, recolectados del 21 de agosto al 27 de noviembre de 2008. El MCS 2008 ofrece resultados a nivel nacional con corte urbano y rural para cada una de las entidades federativas. Es importante mencionar que la base de datos que se está empleando se encuentra a nivel persona, por lo que el valor de los indicadores de carencia de las dimensiones con información a nivel del hogar se imputa a todas las personas del mismo.

Para definir la población objetivo del modelo de regresión logística, es necesario definir en primera instancia el concepto de individuo en situación de pobreza. El CONEVAL define los individuos en situación de pobreza como aquellos cuyos ingresos son inferiores al valor de la línea de bienestar y que padecen al menos una carencia social (Cuadrante I). Los individuos que cuyo ingreso es superior a la línea de bienestar y que no tienen carencia social alguna son clasificados como No Pobres (Cuadrantes II, III y IV).

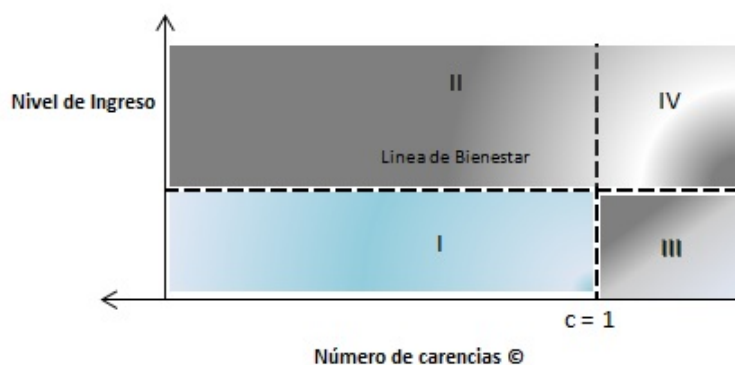


Figura 4.1: Población en situación de pobreza multidimensional

Fuente: [www.coneval.mx](http://www.coneval.mx)

Por construcción, las definiciones de pobreza están directamente relacionadas con:

- I. El nivel de ingresos.

## II. El número de carencias.

Por otra parte, el nivel de ingresos se mide a través de la línea de bienestar, la cual tiene un valor en la zona urbana de \$1,921 mensuales y en la zona rural de \$1,202.8 mensuales. El número de carencias se mide a través de los seis indicadores de pobreza establecidos por el CONEVAL, que a su vez han sido construidos en base a 17 variables.

Como ya se ha mencionado anteriormente, uno de los objetivos de esta tesis es replicar la clasificación de pobreza hecha por el CONEVAL con un menor número de variables de las que actualmente se emplean. Sin embargo, a diferencia de las 17 variables usadas en la construcción de los indicadores de pobreza, la variable Ingreso forma parte directa de la definición de pobreza del CONEVAL; es decir todos aquellos que tengan un ingreso mensual mayor a la línea de bienestar (rural y/o urbana) son prácticamente clasificados como NO POBRES de acuerdo al CONEVAL.

Debido a que la variable Ingreso forma parte directa de la definición de pobreza del CONEVAL no resulta conveniente usarla en el modelo de regresión; pues como se muestra a continuación por definición a partir de cierto umbral clasifica perfectamente a la población de Pobres y no Pobres. La variable ingresos, en este caso se emplea para decidir sobre que población resulta conveniente realizar el modelo de regresión.

En la Tabla 4.1 se ha distribuido a la población en 10 rangos de ingresos de igual tamaño (10% de la población en cada uno), se observa que aproximadamente el 90% de la población de dicha base de datos, tiene ingresos menores a \$5,000. De ese 90% de la población el 43.8% ha sido clasificado como Pobre multidimensional, mientras que el 46.2% de la población ha sido clasificada como No Pobre multidimensional. El umbral promedio a partir del cual es conveniente realizar el modelo de regresión es sobre individuos con ingresos menores a \$2,191, registros de la base de datos con ingresos mayores a este son clasificados automáticamente como No Pobres.

Cuadro 4.1: Distribución de la población por ingreso

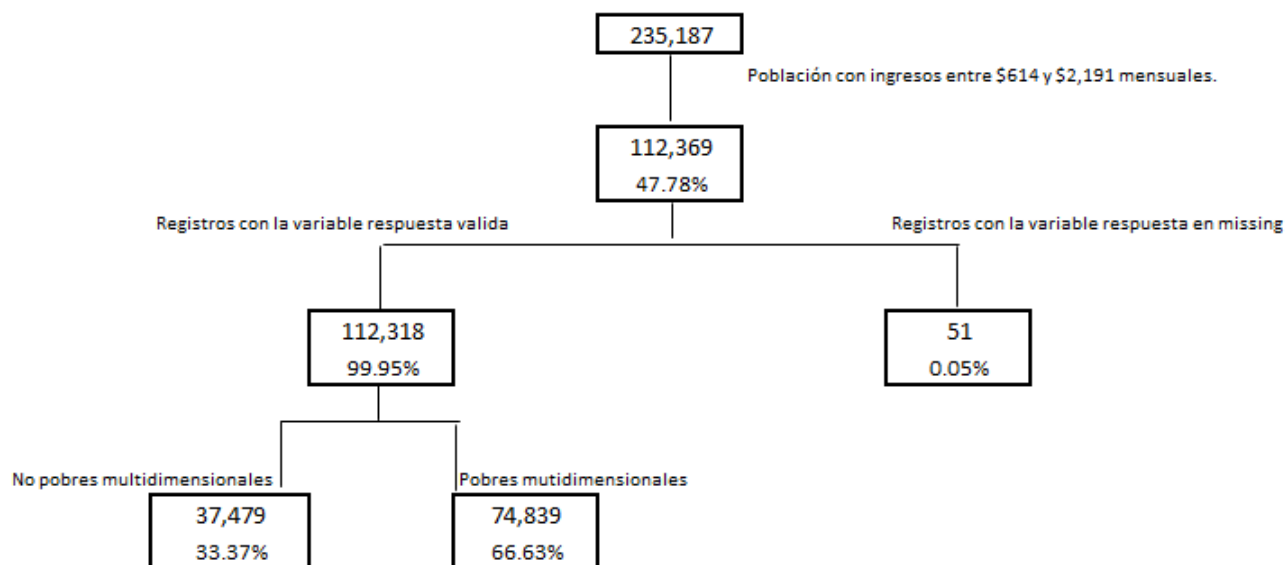
Rangos de ingreso mensual MNX	No pobres	Pobres	Total
$\geq 539$	449	23,056	23,505
% del Total	0.2 %	9.8 %	10 %
(539 a 851]	986	22,518	23,504
% del Total	0.4 %	9.6 %	10 %
(851 a 1,132]	2,018	21,479	23,497
% del Total	0.9 %	9.1 %	10 %
(1,132 a 1,421]	7,360	15,765	23,125
% del Total	3.1 %	6.7 %	10 %
(1,421 a 1,772]	9,215	14,679	23,894
% del Total	3.9 %	6.2 %	10 %
(1,772 a 2,191]	18,039	5,460	23,499
% del Total	7.7 %	2.3 %	10 %
(2,191 a 2,779]	23,501	0	23,501
% del Total	10 %	0 %	10 %
(2,779 a 3,700]	23,506	0	23,506
% del Total	10 %	0 %	10 %
(3,700 a 5,693]	23,500	0	23,500
% del Total	10 %	0 %	10 %
$\leq 5,693$	23,510	0	23,510
% del Total	10 %	0 %	10 %
Total	132,084	102,957	100 %

Fuente : Cálculos elaborados con base en ENIGH 2008

Con base a lo anterior, se han considerado los siguientes umbrales para segmentar a la población sobre la cual se hace el modelo de regresión:

- Población cuyo ingreso corriente per cápita se encuentre entre la línea de bienestar mínima (\$614 ) y \$2,191.

De la base original MCS (Módulo de Condiciones Socioeconómicas de la ENIGH 2008) se extrajo a la población que se encuentra entre los límites establecidos por el umbral de ingresos previamente definido, la distribución final de la base de datos con respecto a la variable de pobreza con la que se trabajó se muestra a continuación:



Cuadro 4.2: Diagrama distribución base de datos respecto a la variable pobreza.

Una consecuencia de emplear la variable ingresos como parte de la segmentación de la población, es que dicha variable ya no puede ser usada en el modelo de regresión para replicar la clasificación de pobreza hecha por el CONEVAL, por lo cual únicamente trataremos de replicar la definición de pobreza del CONEVAL seleccionando las mejores variables del conjunto de 17 variables usadas en la construcción de los indicadores de pobreza del CONEVAL:

Cuadro 4.3: Lista de variables empleadas en la construcción por Indicador de Pobreza

Indicador de Pobreza	Variables que conforman cada indicador
Indicador de rezago educativo	Edad Inasistencia a escuela Nivel educativo
Indicador de salud	Servicios de Salud
Indicador de acceso a la seguridad social	Acceso directo a la seguridad social Acceso directo a la seg. soc. de la jefatura del hogar Acceso directo a la seg. soc. de cónyuge de la jefatura del hogar Acceso directo a la seg. soc. de hijos(as) de la jefatura del hogar Servicios médicos por otros núcleos familiares o por contratación propia Programa de adultos mayores
Indicador de calidad de la vivienda	Indicador de carencia del material de piso de la vivienda Indicador de carencia del material de techos de la vivienda Indicador de carencia del material de muros de la vivienda
Indicador de servicios básicos en la vivienda	Indicador de carencia de acceso al agua Indicador de carencia de servicio de drenaje Indicador de carencia de servicios de electricidad
Indicador del grado de inseguridad alimentaria	Grado de inseguridad alimentaria

## 4.1. Análisis exploratorio de los datos

El análisis preliminar de los datos, es un proceso necesario en la comprensión de los rasgos generales de la estructura de estos y de las relaciones que puedan existir entre las variables.

Como un primer paso, se asegurará que los datos de la base estuvieran bien estructurados; es decir se validó que la distribución de la población por categoría en cada variable respecto a la variable regresora no fuera nula.

Lo anterior debido a que cuando los datos no están bien estructurados o hay muy pocos, se puede presentar el problema de la separación en regresión logística, el cual ocasiona que el proceso de estimación de los estimadores por máxima verosimilitud de los parámetros, no converja.

Con la finalidad de descartar problemas en la estructura de los datos se llevó a cabo un análisis de las variables a través de la elaboración de Tablas de Contingencia. Lo anterior debido a que el conjunto de variables independientes con las que se construirá el modelo de regresión, en su

mayoría son de tipo categórico.

No se detectó ninguna variable con frecuencias nulas en sus categorías; sin embargo algunas de ellas presentan frecuencias menores al 4 % de la población, situación que debe tenerse en cuenta al momento de la selección de variables para la construcción del modelo.

Las tablas de contingencia para todas las variables candidatas a ser consideradas en el modelo de regresión, se presenta en el Apéndice A de esta tesis.

#### **4.1.1. Selección de variables**

Después de haber obtenido la muestra sobre la que se desarrollara el modelo, el siguiente paso en el proceso de construcción del modelo de regresión es la elaboración del análisis bivariado sobre las variables independientes.

Las siguientes evaluaciones fueron hechas para cada una de las variables:

- Tabla de Contingencia (Test Chi Cuadrado).
- Correlación de Pearson.
- Pesos de Evidencia (WOE).
- Valor de la Información (IV).

Para el cálculo del WOE e IV, se requiere que todas las variables sean de tipo categórico. La mayoría de las variables a introducir al modelo lo son; sin embargo la variable edad es de tipo continua.

Con la finalidad de calcular el WOE e IV para la variable edad, se hizo una categorización para dicha variable. Cada una de las categorías concentrará como mínimo aproximadamente un 5 % de la población. Esto es únicamente para poder calcular el IV para esta variable y tener una idea del poder de predicción global de esta; sin embargo la variable edad seguirá siendo considerada

de tipo continuo.

A continuación se muestra la categorización para la variable edad:

Si edad  $\leq 20$  entonces edad1=1.

Si edad  $> 20$  y edad  $\leq 40$  entonces edad1=2.

Si edad  $> 40$  y edad  $\leq 60$  entonces edad1=3.

Si edad  $> 60$  y edad  $\leq 80$  entonces edad1=4.

Si edad  $> 80$  entonces edad1=5.

#### **4.1.2. Análisis Bivariado de los datos : Tablas de Contingencia**

Algunas tablas de contingencia ya han sido construidas en el apartado 3.2 para el análisis exploratorio de los datos. Sin embargo en esta sección el enfoque del análisis de las variables, a través de tablas de contingencia consistirá en tratar de medir la asociación entre las variables que conformen cada tabla y valorar si dichas variables son estadísticamente significativas para el modelo.

Todas las variables que conformarán el modelo de regresión son de tipo categórico, por lo que la correlación entre ellas será medida a través de la elaboración de Tablas de Contingencia.

Consideremos las variables Inasistencia a la escuela y Nivel educativo, dos variables categóricas candidatas para el modelo de regresión. La tabla de contingencia resultante entre estas dos variables se presenta a continuación:

Cuadro 4.4: Tabla de Contingencia: Inasistencia a la escuela, Nivel Educativo

	Nivel educativo				
Inasistencia a la escuela	Primaria incompleta	Secundaria incompleta	Sec.o mayor nivel	NA	Total
Si asiste	21,488	7,656	7,188	0	36,332
%within Inasistencia a la escuela	59.1 %	21.1 %	19.8 %	.0 %	100 %
%within Niv edu	47.9 %	30.01 %	20.03 %	.0 %	32.3 %
%of Total	19.1 %	6.8 %	6.4 %	.0 %	32.3 %
No asiste	23,362	17,704	28,235	0	69,391
%within Inasistencia a la escuela	33.7 %	25.6 %	40.7 %	.0 %	100 %
%within Niv edu	52.1 %	69.9 %	79.7 %	.0 %	61.8 %
%of Total	20.8 %	15.8 %	25.1 %	.0 %	61.8 %
NA	0	0	0	6,597	6,597
%within Inasistencia a la escuela	.0 %	.0 %	.0 %	100 %	100 %
%within Niv edu	.0 %	.0 %	.0 %	100 %	5.9 %
%of Total	.0 %	.0 %	.0 %	5.9 %	5.9 %
TOTAL	44,850	25,450	35,423	6,597	112,320
%within Inasistencia a la escuela	39.9 %	22.7 %	31.5 %	5.9 %	100 %
%within Niv edu	100 %	100 %	100 %	100 %	100 %
%of Total	39.9 %	22.7 %	31.5 %	5.9 %	100 %

Cuadro 4.5: Pearson Chi Square

	Value	df	Asym Sig 2 sided
Pearson Chi Square	1.197E5 <sup>a</sup>	6	.000
Likelihood Ratio	5.729E4	6	.000
Linear by Linear Assotiation	2.441E4	1	.000
N of valid Cases	112,320		

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis

c. Based on normal approximation.



Cuadro 4.6: Symmetric Measures

	Value	Asymp Std Error	Approx. Tb.	Approx. Sig.
Interval by Interval Pearson's R	.466	.003	176.611	.000c
Ordinal by Ordinal Spearman Correlation	.408	.003	149.712	.000c
N of valid Cases	112,320			

El valor de la prueba  $\chi^2$  correspondiente a esta tabla es 119,720 , dado que el valor calculado de la  $\hat{\chi}^2_{(n-1)(k-1)}$  para un nivel de confianza del 95 % ( 5 % nivel de significación) es mayor que el valor de las tablas, se rechaza la hipótesis nula de independencia entre las variables, aceptando por tanto que la inasistencia a la escuela influye en el nivel educativo. El ejercicio anterior se realizó para todas las variables categóricas que se introducirán al modelo:

- Edad.
- Inasistencia a la escuela.
- Nivel educativo.
- Servicios de salud.
- Acceso directo a la seguridad social de la jefatura del hogar.
- Acceso directo a la seguridad social del cónyuge del jefe(a) del hogar.
- Acceso directo a la seguridad social de los hijos del jefe(a) del hogar.
- Servicios médicos por otros núcleos familiares o por contratación propia.
- Programa de adultos mayores.
- Acceso directo a la seguridad social.
- Indicador de carencia del material de piso, techo y muros de la vivienda.
- Indicador de carencia de acceso al servicio de agua, drenaje y electricidad.
- Grado de inseguridad alimentaria.

En términos generales se observa que no hay independencia entre las variables; sin embargo, antes de decidir si se deben o no incluir en el modelo es importante revisar el grado de asociación entre estas a través del cálculo de los coeficientes de correlación.

### 4.1.3. Coeficientes de correlación

Las Tablas de Contingencia y los resultados de la prueba  $\chi^2_{(n-1)(k-1)}$  en la sección anterior muestran que no hay independencia entre las variables. Por ello antes de introducirnos en el ajuste del modelo de regresión, analizaremos de manera más puntual la magnitud y el sentido (positivo o negativo) de la correlación entre las variables.

Los coeficientes de correlación son medidas que indican a través de una expresión numérica el grado de relación existente entre dos variables y nos proporcionan una idea sobre la medida en que estas se relacionan. El estadístico utilizado para medir la correlación será el estadístico de Pearson, descrito en secciones anteriores.

Enseguida, se presentan los coeficientes de correlación para todas las variables candidatas a ser introducidas al modelo de regresión:

- Var1 : Edad.
- Var2 : Inasistencia a la escuela.
- Var3 : Nivel educativo.
- Var4 : Programa adultos mayores.
- Var5 : Servicios de salud.
- Var6 : Acceso directo a la seguridad social.
- Var7 : Acceso directo a la seguridad social de la jefatura del hogar.

Cuadro 4.7: Coeficientes de Correlación de Pearson.

Matriz de correlaciones																	
	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11	Var12	Var13	Var14	Var15	Var16	Var17
Var1	1																
Var2	.326**	1															
Var3	-.143**	.466**	1														
Var4	-.580**	-.123**	.202**	1													
Var5	.060**	-.037**	.049**	-.073**	1												
Var6	.218**	.118**	.095**	-.137**	.328**	1											
Var7	.033**	-.013**	.066**	-.060**	.404**	.386**	1										
Var8	-.006*	-.009**	.035**	.008**	.133**	.143**	.051**	1									
Var9	.035**	.030**	.046**	-.010**	.120**	.200**	.064**	.010**	1								
Var10	.167**	-.029**	.034**	-.123**	.457**	-.047**	.285**	.054**	.058**	1							
Var11	0.004	.008**	-.086**	-0.006	-.092**	-.054**	-.097**	-.033**	-.044**	-.072**	1						
Var12	-.017**	0.003	-.040**	.009**	-.048**	-.027**	-.054**	-.018**	-.019**	-.037**	.228**	1					
Var13	0.004	.006**	-.043**	-0.003	-.045**	-.031**	-.055**	-.017**	-.028**	-.040**	.227**	.203**	1				
Var14	-.020**	0.004	-.065**	.013**	-.093**	-.061**	-.100**	-.042**	-.063**	-.071**	.166**	.067**	.105**	1			
Var15	-.015**	-.012**	-.090**	.010**	-.088**	-.066**	-.114**	-.042**	-.065**	-.087**	.215**	.085**	.129**	.256**	1		
Var16	.009**	.009**	-.045**	-.012**	-.037**	-.021**	-.040**	-.017**	-.018**	-.029**	.148**	.065**	.097**	.172**	.156**	1	
Var17	-.015**	-.012**	-.092**	.014**	-.092**	-.053**	-.095**	-.046**	-.010**	-.070**	.144**	.087**	.078**	.085**	.111**	.060**	1

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

- Var8 : Acceso directo a la seguridad social del cónyuge del jefe(a) del hogar.
- Var9 : Acceso directo a la seguridad social de los hijos del jefe(a) del hogar.
- Var10 : Servicios médicos por otros núcleos familiares o por contrataci?n propia.
- Var11 : Indicador de carencia del material de piso de la vivienda.
- Var12 : Indicador de carencia del material de techo de la vivienda.
- Var13 : Indicador de carencia del material de muros de la vivienda.
- Var14 : Indicador de carencia de acceso al servicio de agua.
- Var15 : Indicador de carencia de acceso al servicio de drenaje.
- Var16 : Indicador de carencia de acceso al servicio de electricidad.
- Var17 : Grado de inseguridad alimentaria.

Como se ha indicado el coeficiente de correlación de Pearson es un índice cuyos valores absolutos oscilan entre 0 y 1. Cuanto más cerca de 1 mayor es la correlación, y menor cuanto más cerca

de cero.

En la tabla anterior es posible observar que las correlaciones para casi todas las variables oscilan entre valores de  $-0.02$  a  $0.4$ , lo cual de acuerdo al cuadro 3.2, significa que en su mayoría las variables están guardando correlaciones bajas entre si.

Sin embargo algunas variables están presentando correlaciones moderadas; esto es una correlación por encima del  $\pm 0.4$ . Tal es el caso de la variable Inasistencia a la escuela contra la variable Nivel Educativo, las cuales presentan una correlación positiva moderada de  $0.466$ . La explicación de esto resulta natural; pues a mayor Nivel Educativo la asistencia a la escuela se vuelve casi un hecho de ahí la correlación positiva entre estas variables.

Debido a que no se disponen de muchas variables para construir el modelo de regresión y a que las correlaciones en su mayoría son bajas ; se ha decidido continuar con la construcción del modelo sin remover ninguna de las variables del cuadro anterior.

#### **4.1.4. Análisis de Cluster**

El análisis de cluster es una técnica cuya idea básica es agrupar un conjunto de observaciones y/o variables en un número dado de clusters o grupos. Este agrupamiento se basa en la idea de distancia o similitud entre las observaciones o variables, de forma que el grado de asociación/similitud entre miembros del mismo cluster sea más fuerte que el grado de asociación/similitud entre miembros de diferentes clusters.

Dicha técnica nos permitió agrupar las variables más semejantes y en combinación con el IV que es una medida que indica el poder de predicción de cada variable, seleccionaremos las variables más representativas de cada cluster, las cuales serán las candidatas a ser introducidas al modelo de regresión logística.

Se considerarán las diecisiete variables que conforman los indicadores de pobreza definidos por el

CONEVAL, trece de ellas son variables de tipo binario como se muestra en el siguiente cuadro, por lo que habrá de seleccionar una distancia apropiada para dicha situación.

Cuadro 4.8: Lista de variables empleadas en la construcción por Indicador de Pobreza

Indicador de Pobreza	Variables que conforman cada indicador	Tipo de variable
Indicador de rezago educativo	Edad	Catagórica
	Inasistencia a escuela	Catagórica Binaria
	Nivel educativo	Catagórica
Indicador de salud	Servicios de Salud	Catagórica
Indicador de acceso a la seguridad social	Acceso directo a la seguridad social	Catagórica Binaria
	Acceso directo a la seg. soc. de la jefatura del hogar	Catagórica Binaria
	Acceso directo a la seg. soc. de cónyuge	Catagórica Binaria
	Acceso directo a la seg. soc. de hijos(as)	Catagórica Binaria
	Servicios médicos por otros núcleos familiares	Catagórica Binaria
	Programa de adultos mayores	Catagórica Binaria
Indicador de calidad de la vivienda	Indicador de carencia del material de piso de la vivienda	Catagórica Binaria
	Indicador de carencia del material de techos de la vivienda	Catagórica Binaria
	Indicador de carencia del material de muros de la vivienda	Catagórica Binaria
Indicador de servicios básicos en la vivienda	Indicador de carencia de acceso al agua	Catagórica Binaria
	Indicador de carencia de servicio de drenaje	Catagórica Binaria
	Indicador de carencia de servicios de electricidad	Catagórica Binaria
Indicador del grado de inseguridad alimentaria	Grado de inseguridad alimentaria	Catagórica

Uno de los índices más habituales, tomado como distancia cuando se consideran variables dicotómicas es el índice de Jaccard, que se difine de la siguiente manera:

Sea

$X_{Aj}$  = Valor del individuo A en la variable  $j$ -ésima  $\in \{1, 0\}$ .

$X_{Bj}$  = Valor del individuo B en la variable  $j$ -ésima  $\in \{1, 0\}$ .

$V = \sum_j X_{Aj}(1 - X_{Bj})$  No. de atributos donde A es 1 y B es 0.

$R = \sum_j X_{Aj}X_{Bj}$  No. de atributos donde A y B son 1.

$S = \sum_j (1 - X_{Aj})(1 - X_{Bj})$  No. de atributos donde  $A$  y  $B$  son 0.

$T = \sum_j (1 - X_{Aj})X_{Bj}$  No. de atributos donde  $A$  es 0 y  $B$  es 1.

$U = R + S + T + V$  No. total de atributos o variables.

el coeficiente de Jaccard se define como:

- Índice de Jaccard .  $C = \frac{R}{R+T+V}$

La distancia que hemos seleccionado, es apropiada únicamente para las trece variables dicotómicas que se pretenden incluir al modelo de regresión, las cuatro variables categóricas restantes serán analizadas más adelante.

Con la finalidad de verificar la existencia de grupos entre las variables que conforman los indicadores realizamos un análisis de conglomerados jerárquicos con el método del vecino más cercano (single linkage) y la distancia de Jaccard, obteniendo el siguiente dendograma para las trece variables dicotómicas que fueron consideradas en este análisis.

En el dendograma podemos ubicar cinco grupos como se muestra a continuación:

Cuadro 4.9: Tabla de cluster

Cluster	Variables	IV
Cluster1	Acceso directo a la seguridad social	0.0194
	Acceso directo a la seguridad social de la jefatura del hogar	0.3500
	Inasistencia a la escuela	0.0046
	Servicios de salud	0.2547
	Programa adultos mayores	0.000072
Cluster2	Indicador de carencia de acceso al servicio de agua	0.0049
	Indicador de carencia de acceso al servicio de drenaje	0.0054
	Indicador de carencia del material de piso de la vivienda	0.0246
	Indicador de carencia del material de muros de la vivienda	0.0108
Cluster3	Acceso directo a la seguridad social de la jefatura del hogar	0.0342
	Indicador de carencia del material de techos de la vivienda	0.0132
Cluster4	Indicador de carencia de acceso al servicio de luz	0.0007
Cluster5	Acceso directo a la seguridad social del conyugue del hogar	0.0619

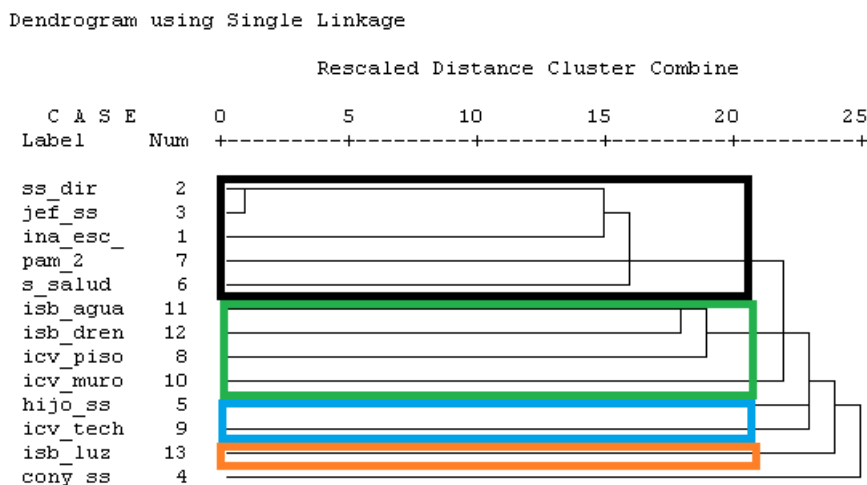


Figura 4.2: Dendrograma variables empleadas en la construcción por Indicador de Pobreza

El Cluster número uno, se formó principalmente de las variables referentes a la seguridad social y los servicios de salud, el cluster número dos esta conformado por las variables relacionadas con los indicadores de carencia en los servicios y materiales de la vivienda; sin embargo no todas las variables fueron agrupadas dentro de estos dos clusters. El cluster tres contiene una variable referente a la seguridad social y otra referente al material de la viviente. La variable indicadora de carencia al acceso de servicio de luz y la variable acceso directo a la seguridad del cónyugue

fueron agrupadas de manera individual en dos clusters.

Seleccionaremos a las variables de cada cluster que tengan un IV mayor en cada grupo, debido a que el IV es una medida que nos indica el poder de predicción global de cada variable. Del grupo de variables que conforman el cluster número uno, la variable Acceso directo a la seguridad social presenta un IV mayor 0.3500, lo que significa que tiene un poder de predicción fuerte, del cluster número dos la mejor variable de acuerdo a su IV es la variable Indicador de carencia del material de piso de la vivienda y finalmente del Cluster 3 la variable Acceso directo a la seguridad social de la jefatura del hogar, es la variable con el mejor IV; sin embargo es importante decir que en estos dos últimos casos el poder de predicción de estas variables es catalogado como débil. El cluster 4 y 5 solo constan de una variable por lo que se consideraran ambas variables en el análisis.

Si realizamos una nueva agrupación a través del análisis de Clusters de las variables que seleccionamos en base a su IV en el análisis anterior, obtenemos el siguiente dendrograma:

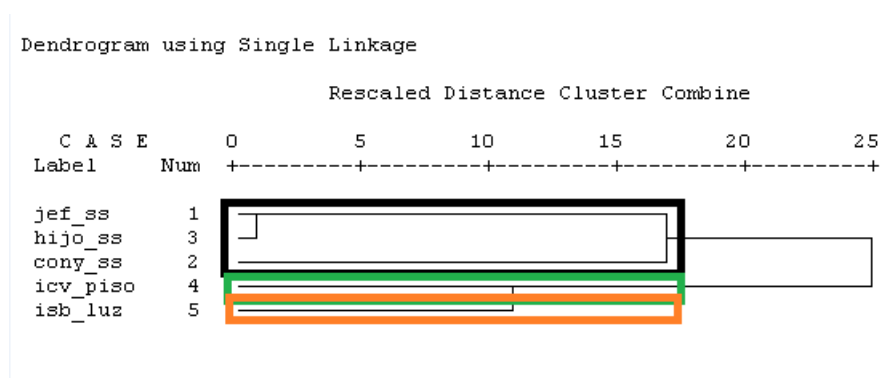


Figura 4.3: Dendrograma variables empleadas en la construcción por Indicador de Pobreza después de selección por IV.

Es posible observar que en esta nueva agrupación se forman tres clusters, el primero está conformado por variables referentes a la seguridad social siendo la variable más representativa por su IV, la variable Acceso directo a la seguridad social de la jefatura del hogar 0.3500, el segundo cluster hace referencia a la calidad del material de la vivienda y únicamente consta de la variable



Indicador de carencia del material de piso de la vivienda, el tercer cluster hace referencia a la calidad de los servicios básicos en la vivienda y se conforma únicamente de la variable Indicador de carencia de acceso al servicio de luz. En conclusión, de este análisis de conglomerados conformado por trece variables dicotómicas candidatas a ser empleadas en el modelo de regresión, se han seleccionado tres variables:

- Acceso directo a la seguridad social de la jefatura del hogar.
- Indicador de carencia del material de piso de la vivienda.
- Indicador de carencia de acceso al servicio de luz.

Por otra parte, cuatro variables no fueron incluidas en el análisis anterior debido a que son variables con más de una categoría y el índice de Jaccard trabaja con datos de presencia/ausencia; es decir es un índice binario.

Las cuatro variables que no fueron incluidas en el análisis anterior son:

- Edad IV = 0.0133
- Nivel educativo IV = 0.0826
- Servicios de Salud IV = 0.5341
- Grado de inseguridad alimentaria IV = 0.1264

Estas variables en su mayoría, de acuerdo a su IV tienen un poder de predicción medio y fuerte, además son un complemento de información respecto a las variables que anteriormente seleccionamos. Con la finalidad de tener para todos los indicadores de pobreza construidos por el CONEVAL al menos una variable de cada uno de estos, se ha decidido conservar en adición a las tres variables que se seleccionaron por medio del análisis de Cluster e IV, a estas cuatro variables. El grupo de variables candidatas a ser empleadas en el modelo de regresión se redujo 7 variables como se muestra en el siguiente cuadro:

Cuadro 4.10: Lista de variables empleadas en la construcción por Indicador de Pobreza

Indicador de Pobreza	VARIABLES que conforman cada indicador	IV
Indicador de rezago educativo	Edad	0.0133
	Nivel educativo	0.0826
Indicador de salud	Servicios de Salud	0.5341
Indicador de acceso a la seguridad social	Acceso directo a la seg. soc. de la jefatura del hogar	0.0342
Indicador de calidad de la vivienda	Indicador de carencia del material de piso de la vivienda	0.0246
Indicador de servicios básicos en la vivienda	Indicador de carencia de acceso al servicio de luz	0.0007
Indicador del grado de inseguridad alimentaria	Grado de inseguridad alimentaria	0.1264

## 4.2. Definición de la variable respuesta y variables explicativas

En el caso de la medición de la Pobreza en México, el evento no deseado es la posibilidad de que una persona sea catalogada como Pobre Multidimensional, de acuerdo al CONEVAL esto ocurre cuando la persona presenta al menos una carencia social y cuyo ingreso corriente per cápita es inferior al valor de la línea de bienestar. De este modo, la variable respuesta binaria según la definición anterior, toma el valor de 1 si la persona en cuestión es considerada pobre multidimensional y en caso contrario toma el valor de cero.

Se busca calcular la probabilidad de que la variable respuesta tome el valor de uno. Usaremos el modelo de regresión logística binaria con la variable dependiente dicotómica: Pobreza multidimensional y las 7 variables explicativas que se seleccionaron en el apartado anterior.

Para cada variable fue necesario seleccionar una categoría de referencia, excepto para la variable edad la cual es de tipo continuo. Para las variables categóricas de tipo binario la categoría codificada con el valor uno será la categoría de referencia, por tanto se obtiene la probabilidad de la categoría 1 frente a la categoría 0. Por ejemplo para la variable Acceso directo a la seguridad social de la jefatura del hogar, codificada de la siguiente manera:

- No cuenta con acceso directo a la seguridad social de la jefatura del hogar = 0.
- Cuenta con acceso directo a la seguridad social de la jefatura del hogar = 1.

Cuadro 4.11: Categoría de referencia para cada variable

Variable	Valor	Categoría	Ref
Nivel educativo	0	Primaria incompleta o menos	x
	1	Primaria completa o secundaria incompleta	
	2	Secundaria completa o mayor nivel educativo	
	3	No aplica	
Servicios de Salud	0	No cuenta con servicios médicos	x
	1	Cuenta con seguro popular	
	2	Cuenta con IMSS	
	3	Cuenta con ISSTE	
	4	Cuenta con servicios de Salud PEMEX	
	5	Otros	
Servicios de Salud	0	No cuenta con servicios de salud	x
	1	Cuenta con servicios de salud	
Acceso directo a la seg. soc. de la jefatura	0	No cuenta con servicios de salud	x
	1	Cuenta con servicios de salud	
Indicador de carencia del material de piso	0	Sin carencia	x
	1	Con carencia	
Indicador de carencia de acceso al servicio de luz	0	Sin carencia	x
	1	Con carencia	
Grado de inseguridad alimentaria	0	Seguridad alimentaria	x
	1	Inseguridad alimentaria leve	
	2	Inseguridad alimentaria moderada	
	3	Inseguridad alimentaria severa	

Estaríamos obteniendo la probabilidad de ser catalogado como pobre multidimensional cuando se cuenta con acceso directo a la seguridad social de la jefatura del hogar frente a la opción de no contar con dicha característica.

Para las variables con más de dos categorías, es necesario construir variables dummy. En cada caso se optó por seleccionar aquella categoría que tuviera el cociente del logaritmo de odds más cercano a cero (neutral) como la categoría pivote para la construcción de las variables dummy; debido a que esta categoría cuenta con una proporción similar de personas catalogadas como pobres y no pobres. El cuadro 4.11 muestra la categoría de referencia para cada una de las variables.

El modelo se desarrolló usando el paquete estadístico SPSS en su versión 16.0. En las siguientes secciones se exponen los resultados obtenidos.

### 4.3. Ajuste del modelo de regresión logística

El ajuste del modelo de regresión logística, se basó en el logaritmo de la verosimilitud. Dicho ajuste fue realizado con el método Enter, el método que por default nos da el programa SPSS.

Para el modelo que contiene sólo el término constante,  $-2LogLikelihood$ , es igual a 143, 035.916. El proceso ha necesitado de tres ciclos para estimar correctamente el término constante, porque la variación de  $-2LogLikelihood$  entre el primero y el segundo bucle ha cambiado en menos del criterio fijado por el programa 0.001. El valor calculado del parámetro es  $\beta_0 = 0.692$ .

Cuadro 4.12: Log likelihood. Primer ajuste al modelo de Regresión

Iteración	-2 Log likelihood	Constant
Step 1	143,053.247	0.665
2	143,035.917	0.691
3	143,035.916	0.692

La siguiente tabla permite evaluar el ajuste del modelo de regresión (hasta ahora con un parámetro en la ecuación) comparando los valores predichos con los observados. Se ha empleado un punto de corte de la probabilidad de Y para clasificar a los individuos de 0.5. Lo anterior implica que para aquellos sujetos para los que la ecuación (únicamente con el parámetro  $\beta_0$ ) calcule una probabilidad  $< 0.5$  se clasificarán como no Pobres, mientras que si la probabilidad resultante es  $\geq 0.5$  se clasificarán como Pobres.

En este primer paso el modelo ha clasificado correctamente a un 66.6% de los casos y ningún sujeto No Pobre multidimensional ha sido clasificado correctamente.

Cuadro 4.13: Variables en la Ecuación

	B	S.E.	Wald.	df.	Sig.	Exp $\beta$
Step 0 Constant	.692	.006	1.194E4	1	.000	1.997

Cuadro 4.14: Log likelihood. Ajuste Final al modelo de Regresión

Iteración	-2 Log likelihood	Constant
Step 1	124,467.291	0.831
2	123,638.384	0.892
3	123,630.411	0.892
4	123,630.410	0.892

a.Method Enter

b.Constant is included in the model

c.Initial  $-2$  Log Likelihood 143,035.916

d.Estimation terminated it iterartion number 4 becасue parameter estimates changed by less than .001

Observed	Predicted		Percentage Correct
	No pobre multidimensional	Pobre multidimensional	
No pobre multidimensional	0	37, 478	0 %
Pobre multidimensional	0	74, 738	100 %
Overall Percentage			66.6 %

a.Constant is included in the model .

b.The cut value is .500

El siguiente cuadro muestra el parámetro estimado  $\beta_0$ , su error estándar (E.T.) y su significación estadística con la prueba de Wald, el cual nos permite valorar si el parámetro estimado  $\beta_0$  es significativamente diferente de cero. En este caso el índice de Wald tiene una significación estadística de 0.000 por lo que se rechaza la hipótesis de que el parámetro es cero.

A continuación se inicia un segundo paso (BLOQUE 1) en el cuál de manera automática el programa a través del método ENTER basado en el criterio de verosimilitud contrasta las nuevas variables para introducirlas al modelo.

Nótese como disminuye el  $-2\text{LogLikelihood}$  respecto al paso anterior (el modelo con una sola constante tenía un  $-2\text{LogLikelihood}$  de 143, 035.916, mientras que ahora se reduce a

Cuadro 4.15: Prueba Chi -square sobre los coeficientes del modelo

Step	Chi- square	df	Sig
Step	19,417.175	15	.000
Block	19,417.175	15	.000
Model	19,417.175	15	.000

a.Estimation terminated it iterartion number 4.

123,618.741, el proceso termina con 4 iteraciones.

En la tabla 4.15 se muestra una prueba  $\chi^2$  que evalúa la hipótesis nula de que los coeficientes  $\beta$  de todos los términos (excepto la constante) incluidos en el modelo son cero. El estadístico  $\chi^2$  para este contraste es la diferencia entre el valor de  $-2\text{LogLikelihood}$  para el modelo sólo con la constante y el valor de  $-2\text{LogLikelihood}$  para el modelo actual.

El programa SPSS ofrece tres entradas: Step, Bloque y Modelo.

- La primera fila (Step) es la correspondiente al cambio de verosimilitud (de  $-2\text{LogLikelihood}$ ) entre pasos sucesivos en la construcción del modelo, contrastando la  $H_0$  de que los coeficientes de las variables añadidas en el último paso son cero.
- La segunda fila (Block) es el cambio en  $-2\text{LogLikelihood}$  entre bloques de entrada sucesivos durante la construcción del modelo.
- La tercera fila (Modelo) es la diferencia entre el valor de  $-2\text{LogLikelihood}$  para el modelo sólo con la contaste y el valor de  $-2\text{LogLikelihood}$  para el modelo actual. <sup>1</sup>

En este caso al haber introducido un único bloque de variables en un solo paso coinciden los tres valores.

---

<sup>1</sup>Cómo hacer una Regresión Logística con SPSS, paso a paso.(I), Aguayo Canela Mariano

A continuación se observa el resumen del modelo aporta tres medidas complementarias a la anterior para validar el ajuste global del modelo. La primera es el valor del  $-2\text{LogLikelihood}$  y las otras dos son los Coeficientes de Determinación  $R^2$ , parecidos a los que se obtienen en el ajuste de una regresión lineal. Un modelo perfecto tendría un valor de  $-2\text{LogLikelihood}$  muy pequeño y un  $R^2$  cercano a uno.

- $-2\text{LogLikelihood}$  mide hasta que punto un modelo se ajusta bien a los datos . El resultado de esta medición recibe también el nombre de *desviación* . Cuanto más pequeño sea el valor, mejor será el ajuste.
- La R cuadrada de Cox y Snell es un coeficiente de determinación generalizado que se utiliza para estimar la proporción de varianza de la variable dependiente explicada por las variables predictoras (independientes). Sus valores oscilan entre 0 y 1. En este caso el valor de la R cuadrada de Cox y Snell indica que el 67.7% de la variación de la variable dependiente es explicada por las variables incluidas en el modelo, un ajuste bastante bueno. <sup>2</sup>
- La R cuadrada de Nagelkerke es una versión corregida de la R cuadrada de Cox y Snell.PENDIENTE

Finalmente una forma de evaluar la ecuación de regresión y el modelo obtenido es construir una tabla de  $2 \times 2$  clasificando a todos los individuos de la muestra según la concordancia de los valores observados con los predichos o estimados por el modelo. La tabla de clasificación para nuestro modelo es la siguiente:

---

<sup>2</sup>Cómo hacer una Regresión Logística con SPSS, paso a paso.(I), Aguayo Canela Mariano

Observed	Predicted		Percentage Correct
	No pobre multidimensional	Pobre multidimensional	
No pobre multidimensional	17,782	19,696	47.4%
Pobre multidimensional	7,936	66,902	89.4%
Overall Percentage			75.4%

a.Constant is included in the model.

b.The cut value is .500

En la tabla de clasificación es posible comprobar que el modelo clasifica en un 47.4% a los individuos No Pobres multidimensionales y en 89.4% a los individuos Pobres multidimensionales, proporcionando un ajuste global del 75.4% un porcentaje adecuado, del que se deduce que el modelo tiene una buena capacidad predictiva.

En el siguiente cuadro se muestran los resultados que se obtienen al hacer la regresión logística.

	B	S.E.	Wald	df	Sig	Exp $\beta$
jef ss	-.721	0.018	1.565E3	1	0.000	0.486
icv pisos	.050	0.031	2.623	1	0.105	1.051
isb luz	-.461	0.070	42.769	1	0.000	0.631
dumserv sal1	.242	0.018	180.700	1	0.000	1.273
dumserv sal2	-.978	0.020	2.347E3	1	0.000	0.376
dumserv sal3	-1.425	0.040	1.295E3	1	0.000	0.240
dumserv sal4	-.975	0.093	109.398	1	0.000	0.377
dumserv sal5	-.199	0.071	7.833	1	0.005	0.819
duminiv edu1	.462	0.018	631.256	1	0.000	1.588
duminiv edu2	-.033	0.019	3.167	1	0.075	0.967
duminiv edu3	.437	0.035	159.282	1	0.000	1.548
dum ali1	-.259	0.016	258.692	1	0.000	0.772
dum ali2	.710	0.024	865.269	1	0.000	2.034
dum ali3	.774	0.028	747.337	1	0.000	2.169
edad	.000	0.000	6.071	1	0.014	.999
Constant	.932	0.024	1.556E3	1	0.000	2.540

a.Variables entered on step 1 : jef ss, icv pisos, isb luz, dumserv sal1, dumserv sal2, dumserv sal3, dumserv sal4, dumserv sal5, duminiv edu1, duminiv edu3, dum ali1, dum ali2, dum ali3, edad.

Para determinar si las variables introducidas al modelo son estadísticamente significativas se analizará el *p-vale* asociado al estadístico de Wald para evaluar la hipótesis nula  $\beta_i = 0$ .



Obsérvese que a un nivel de significancia de 0.05 la hipótesis nula  $\beta_i = 0$  se aceptaría para las una de las variables dummy referente al nivel educativo y para la variable que mide la calidad del material de los pisos de la vivienda, para el resto de las variables la hipótesis se rechaza.

Con un nivel de significancia de 0.01 la hipótesis nula  $\beta_i = 0$  se aceptaría para la variable edad. El modelo se volverá a ajustar removiendo las variables para las cuales la hipótesis nula fue aceptada tanto al nivel de significancia 0.05 como 0.01. El ajuste se muestra a continuación:

	B	S.E.	Wald	df	Sig	Exp $\beta$
jef ss	-.722	0.018	1.572E3	1	0.000	0.486
isb luz	-.444	0.070	40.592	1	0.000	0.641
dumserv sal1	.239	0.018	177.324	1	0.000	1.270
dumserv sal2	-.985	0.020	2.413E3	1	0.000	0.373
dumserv sal3	-1.439	0.039	1.332E3	1	0.000	0.237
dumserv sal4	-.981	0.093	110.819	1	0.000	0.375
dumserv sal5	-.205	0.071	8.330	1	0.004	0.814
duminiv edu1	-.485	0.015	1.076E3	1	0.000	1.623
duminiv edu3	.483	0.031	242.250	1	0.000	1.620
dum ali1	-.261	0.016	262.764	1	0.000	0.771
dum ali2	.713	0.024	874.096	1	0.000	2.040
dum ali3	.779	0.028	760.422	1	0.000	2.179
Constant	.892	0.018	2.402E3	1	0.000	2.441

a. Variables entered on step 1: jef ss, isb luz, dumserv sal1, dumserv sal2, dumserv sal3, dumserv sal4, dumserv sal5, duminiv edu1, duminiv edu3, dum ali1, dum ali2, dum ali3.

En este ajuste del modelo el parámetro estimado  $\beta_0$  así como el resto de los parámetros, de acuerdo con la prueba de Wald, son significativamente diferentes de cero.

La clasificación de todos los individuos de la muestra según la concordancia de los valores observados con los predichos o estimados por el modelo, muestra que el modelo clasifica nuevamente en un 47.5% a los individuos No Pobres multidimensionales y en 89.4% a los individuos Pobres multidimensionales, proporcionando un ajuste global del 75.4% un porcentaje bastante alto, del que se deduce que el modelo tiene una alta capacidad predictiva.

Observed	Predicted		Percentage Correct
	No pobre multidimensional	Pobre multidimensional	
No pobre multidimensional	17,801	19,677	47.5%
Pobre multidimensional	7,958	66,880	89.4%
Overall Percentage			75.4%

a.Constant is included in the model.

b.The cut value is .500

## 4.4. Construcción del Score

Los métodos o modelos de credit scoring, son algoritmos que de manera automática evalúan el riesgo de crédito de un solicitante de financiamiento de alguna entidad bancaria. El resultado de la evaluación se refleja en la asignación de un puntaje o score que permite comparar y ordenar a los evaluados en función de su riesgo.

En esta sección la técnica del credit scoring usualmente empleada en instituciones financieras, será implementada para asignar un score o calificación de acuerdo al grado o nivel de pobreza de los individuos con los cuales se desarrolló el modelo de regresión logística en la sección anterior.

Los puntajes del score son resultado de un re-escalamiento y una traslación del modelo logístico y queda determinado con la siguiente ecuación:

$$Score = Offset + Factor * \log(odds)$$

donde Offset es un término de traslación (o compensación) y Factor es un término de re-escalamiento y son calculados resolviendo el siguiente sistema de ecuaciones:

$$Score = Offset + Factor * \ln(odds)$$

$$Score + pdo = Offset + Factor * \ln(2 * odds)$$

El término Offset y Factor deben satisfacer la condición de que en la relación de buenos y malos de 1:1 es tal que el Odds se duplica cada determinado número de puntos (pdo).

Resolviendo las ecuaciones anteriores se obtiene:

$$pdo = Factor * \ln(2) \text{ por tanto}$$
$$Factor = pdo / \ln(2) \quad Offset = Score - Factor * \ln(Odds)$$

Para tener una mejor lectura del puntaje asociado a cada individuo de acuerdo a su grado de pobreza, se procede a reescalar los valores del modelo logístico estimado en la sección anterior de la siguiente manera:

Se tomó una escala de puntajes tal que el valor de 600 corresponde a una relación de No Pobres/Pobres de 60:1 y un incremento en el puntaje de 15 unidades coincide con el doble de la relación No Pobres/Pobres.

$$Factor = 15 / \ln(2) = 21.64 \quad Offset = 600 - 21.64 * \ln(60) = 511.40$$

Por lo que la ecuación para calcular el score de pobreza a nivel individuo es:

$$Score = 511.40 + 21.64 * \log(odds)$$

## 4.5. Validación del Score

Una vez elaborado el modelo estadístico es necesario verificar su poder de discriminación y la correcta ordenación de los puntajes. En la sección anterior se evaluó a través de una tabla de clasificación la capacidad predictiva del modelo, la cual fue del 75.4% lo que indica un alto poder de discriminación.

Sin embargo, en el diseño de modelos scoring existen algunos índices adicionales que son de gran importancia y de frecuente uso:

- Kolmogorov-Smimov.

- Índice de Gini.
- La curva ROC.

los cuales se basan en la función de distribución o probabilidades acumuladas y se describen con mayor detalle en el Capítulo 3 de esta tesis. El índice de Gini es únicamente una medida global de calidad del modelo, mientras que en el coeficiente K-S, aparte de medir la calidad de ajuste, identifica el valor del score para el cual se maximiza dicho coeficiente.

Después de calcular los puntajes de cada uno de los individuos involucrados en la construcción del modelo de regresión, se procede a crear cuatro rangos para hacer la validación del score del modelo en base a las métricas anteriormente mencionadas.

La siguiente tabla, muestra la distribución de los individuos Pobres y No Pobres a lo largo de cuatro rangos de score. Es posible observar que el modelo ordena correctamente a la población por riesgo, en los rangos de score más bajos se concentra la población clasificada como No Pobre, mientras que en los rangos con puntajes más altos se concentra la mayoría de la población Pobre. En consecuencia, la tasa de Pobres en cada rango score sigue una tendencia creciente, lo cual confirma la correcta separación y ordenación del riesgo de Pobreza proporcionada por el modelo.

Cuadro 4.16: Distribución de individuos Pobres y no Pobres por Rango de Puntaje

Rango de score	Total	Total %	Dist. Acum No Pobres	% No Pobres	Dist. Acum Pobres	% Pobres
$\leq 420$	26,194	23.32 %	17,954	47.91 %	8,240	11.09 %
(420, 472]	27,410	24.40 %	25,328	67.58 %	28,276	37.78 %
(472, 565]	31,826	28.34 %	32,529	86.79 %	52,901	70.69 %
(565, 603]	26,885	23.94 %	37,478	100.00 %	74,837	100.00 %
<i>TOTAL</i>	112,315	100 %				

El estadístico de Kolmogorv Smirnov, puede ser calculado a través de la tabla anterior tomando el valor absoluto de la máxima diferencia en la distribución acumulada de pobres y no pobres por rango de score (columna P). En este caso la máxima diferencia de todos los rangos de score (KS) se ubica en el rango número uno y le corresponde un valor de 36.89 % .

En términos gráficos la KS es el largo de la línea representada por una flecha en el rango score que maximiza la separación entre las dos funciones de distribución, como se muestra a continuación.

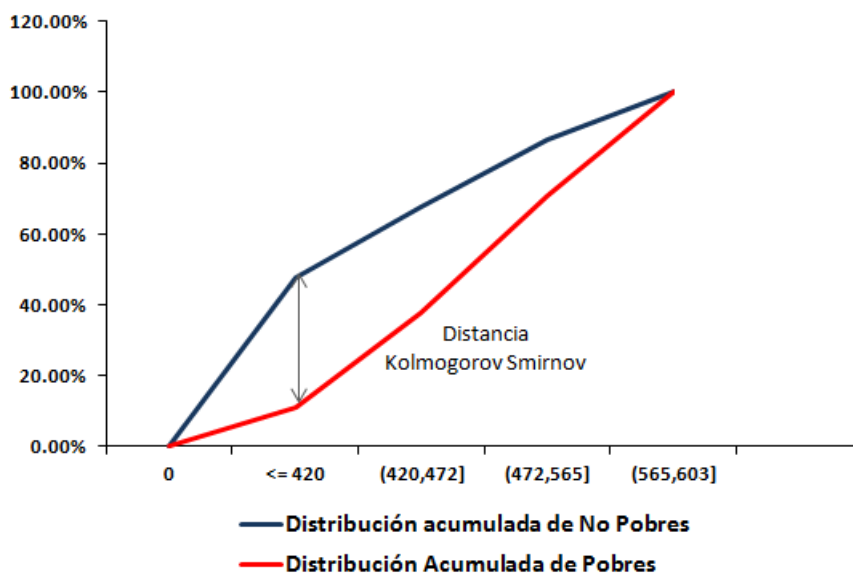


Figura 4.4: KS.

La curva ROC a diferencia de la medida Kolmogorov-Smirnov muestra la información en una sola curva graficando la distribución acumulada de Pobres contra la distribución acumulada de No Pobres para cada rango percentil del puntaje score. La medida que recoge el área bajo la curva se conoce como el coeficiente de Gini. Realmente este índice se define como dos veces el área formada por la diagonal y la curva. Este índice tendrá la propiedad de que en el caso de una perfecta clasificación su valor será de 1, mientras que una clasificación aleatoria con una curva ROC sobrepuesta a la diagonal tendrá un valor de índice igual a 0.

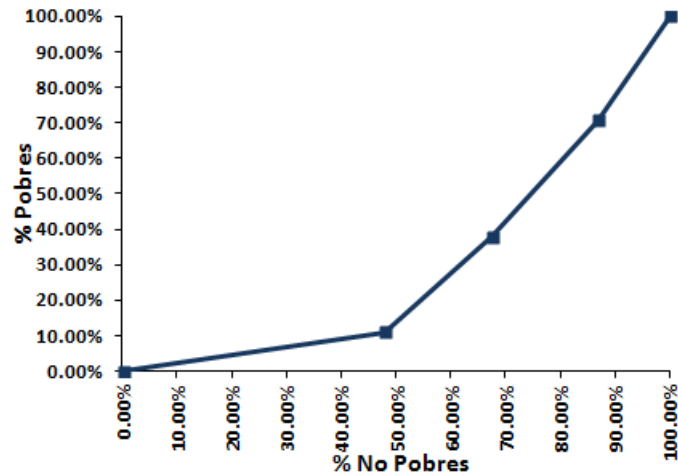


Figura 4.5: Curva ROC.

En este caso el índice de Gini, tiene un valor del 0.58, lo cual indica un alto poder de discriminación del modelo.

## 4.6. Aplicaciones del modelo de puntaje estadístico

Una de las ventajas de diseñar evaluaciones que combinen métodos cuantitativos y cualitativos de investigación es que estos permiten reportar datos de forma transparente y objetiva.

En este caso el puntaje estadístico propuesto es una medida consistente que produce resultados comparables y proporciona puntos de referencia para su evaluación a través del tiempo, que pueden ser considerados inclusive para ejecutar modelos de combate a la pobreza.

Julio Boltvinik menciona que los modelos de combate a la pobreza desde la perspectiva de focalización ya han cobrado fuerza en México, inicialmente con el programa Progresá continuado como Oportunidades. En este tipo de programas solo aquellos que prueban no tener medios para resolver sus necesidades, reciben transferencias del gobierno. Conviene por tanto, revisar a detalle en qué consiste este enfoque, pues se ha proclamado como la mejor solución al problema

porque es más eficiente y no hay dispendio por que no se otorgan apoyos a quienes no los necesitan.

Por otra parte Fernando Cortes (2014) menciona el gasto social realizado por los gobiernos mexicanos desde 1997 hasta 2012 no se ha traducido en disminuciones en la proporción de la población o el número de personas en situación de pobreza. Sobre la base de la percepción de que los programas sociales son inefectivos suele derivarse la conclusión de que los programas sociales destinados a ayudar a los pobres deben modificarse, o bien ser sustituidos.

El gasto en desarrollo social ha aumentado en las últimas dos décadas aunque el crecimiento no ha sido sostenido.

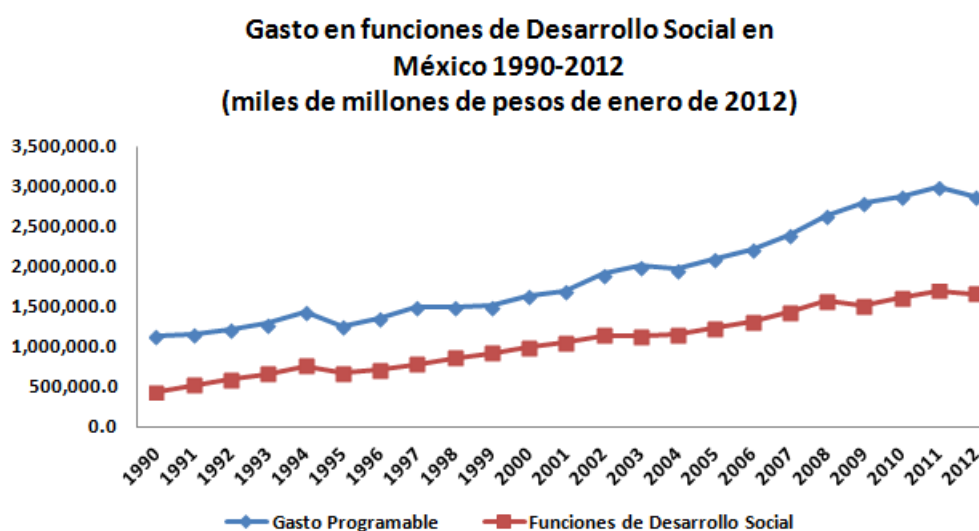


Figura 4.6: Gasto en funciones de Desarrollo Social en México 1990 - 2012

Fuente: 1990 a 1994, cálculos del Centro de Estudios de las Finanzas Públicas de la Cámara de Diputados.

A pesar del volumen creciente de recursos económicos dedicados al gasto social, el porcentaje de población en situación de pobreza no ha disminuido a la par del aumento del gasto. Las

incidencias de las pobreza alimentaria, de capacidades y de patrimonio presentan en 2014 un incremento respecto a 2012 y los años anteriores.

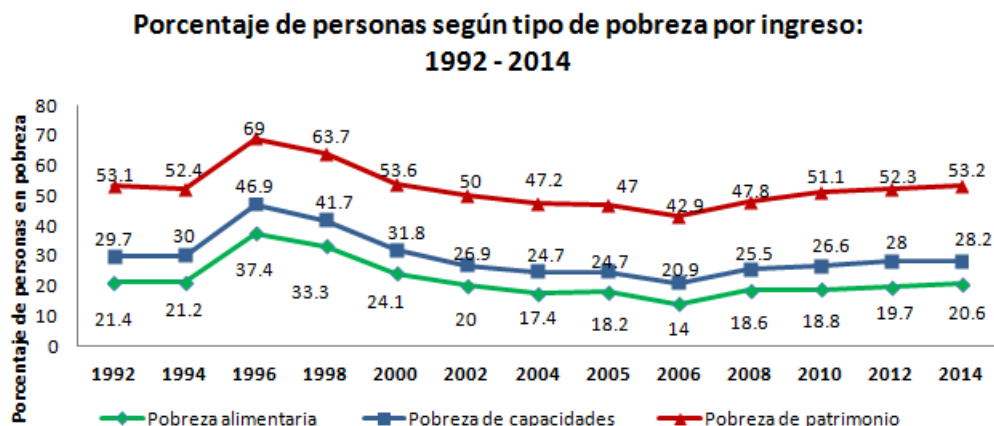


Figura 4.7: Porcentaje de personas según tipo de pobreza por ingreso: 1992 - 2014

Fuente: estimaciones del CONEVAL con base en las ENIGH de 1992 a 2012

Acorde al análisis realizado por Fernando Cortes(2014), con base a las dos gráficas mostradas anteriormente se observa que el gobierno mexicano ha realizado un esfuerzo financiero sostenido para combatir la pobreza a lo largo de los últimos 15 años, pero su incidencia no cede. Fernando Cortes(2014) sugiere que esto se puede deber a que los apoyos no llegan a los pobres sino a otros sectores sociales (problemas de focalización), a que los recursos se enredan en la jungla burocrática quedando como remanente volúmenes reducidos para apoyar a los pobres, que las cantidades entregadas son insuficientes para que la población salga de esa condición, o que no son suficientes para que lo hagan por sus propios medios y se puedan autosostener.

En lo que respecta a la hipótesis sostenida por Fernando Cortes(2014) respecto a los problemas de focalización, es importante mencionar que todos los programas dirigidos a los pobres cometen inevitablemente dos errores conocidos como tipo I y II. El error tipo I o error de exclusión, consiste en descartar de los beneficios a los individuos realmente pobres. En el error tipo II, o



de inclusión, se admiten no pobres entre los beneficiarios.

Adicional a lo anterior, cualquier programa de focalización implica que algunas personas, típicamente funcionarios de gobierno, juzguen las solicitudes de los aspirantes a beneficiarios.

La primera ventaja que proporciona el modelo de score propuesto es que a pesar de que el puntaje estadístico técnicamente implica cierto nivel de complejidad, su uso en la práctica es automatizado, emplear un modelo de score implicaría que la selección y evaluación de las solicitudes de los aspirantes a beneficiarios fueran más objetivas, el puntaje estadístico es una medida consistente, dos personas con las mismas características son idénticamente evaluadas.

Por otra parte el uso del modelo de puntaje propuesto daría la pauta para el desarrollo de modelos de combate a la pobreza focalizada más certeros, en el que sería posible afinar las características de la población objetivo y diseñar políticas de combate acorde a sus necesidades, como se muestra a continuación.

Se ha ejecutado el modelo de puntaje desarrollado en el capítulo anterior a la base de datos del MCS (Módulo de Condiciones Socioeconómicas de la ENIGH) 2008, 2010, 2012 y 2014 respectivamente, la cual contiene las variables necesarias para la medición multidimensional de la pobreza, establecidas en la LGDS.

En el siguiente cuadro se muestra la distribución por rango de score de la población entre las muestras.

RANGO		2008		2010		2012		2014	
1.-	<= 488.11	30,098	12.80%	28,063	11.92%	23,507	11.05%	25,359	11.73%
2.-	(488.11 , 498.60]	20,379	8.67%	18,090	7.69%	14,336	6.74%	15,313	7.08%
3.-	(498.60 , 509.19]	21,868	9.30%	20,577	8.74%	18,022	8.47%	19,069	8.82%
4.-	(509.19 , 519.68]	17,992	7.65%	18,230	7.74%	16,170	7.60%	17,361	8.03%
5.-	(519.68 , 530.21]	22,181	9.44%	29,187	12.40%	33,272	15.64%	35,886	16.60%
6.-	(530.21 , 530.7 ]	24,021	10.22%	20,942	8.90%	13,429	6.31%	12,875	5.96%
7.-	(530.7 , 535.87]	29,246	12.44%	32,129	13.65%	33,179	15.60%	32,743	15.15%
8.-	(535.87 , 541.19]	23,588	10.03%	20,723	8.80%	17,954	8.44%	16,732	7.74%
9.-	(541.19 , 551.30]	22,428	9.54%	21,917	9.31%	21,048	9.90%	20,787	9.62%
10.-	> 551.30	23,263	9.90%	25,529	10.85%	21,757	10.23%	20,041	9.27%
		235,064	100.00%	235,387	100.00%	212,674	100.00%	216,166	100.00%

Cuadro 4.17: Diagrama distribución base de datos respecto a la variable pobreza.

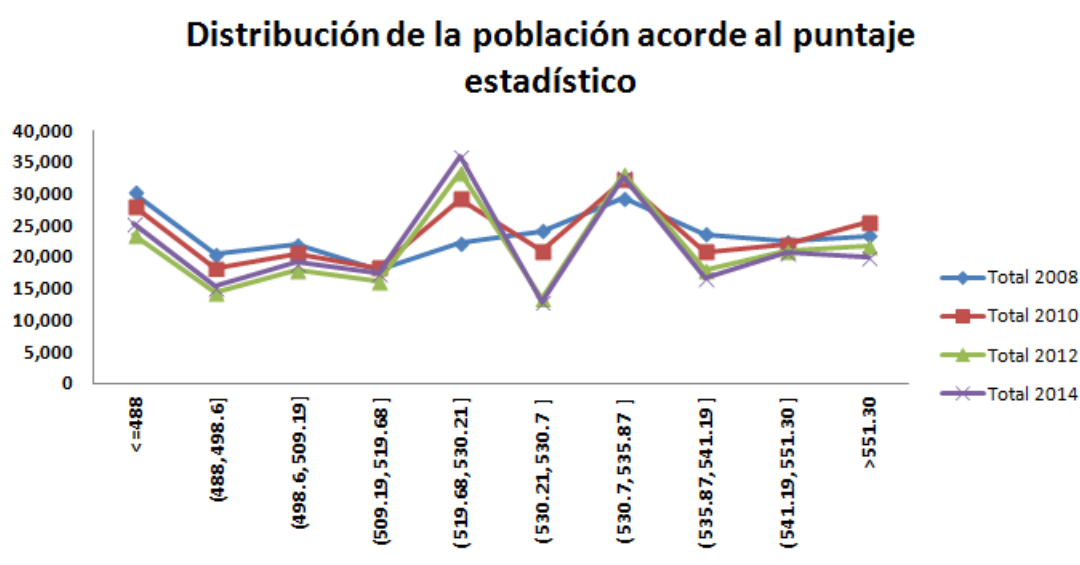


Figura 4.8: Distribución de la población acorde al puntaje estadístico

Fuente: estimaciones con información del CONEVAL con base en las ENIGH de 2008 a 2012

A través del uso del puntaje construido en la sección anterior, es posible observar que la población ha registrado movimientos principalmente en el rango 5 y 7. El número de personas que está siendo evaluadas con puntajes más altos se ha incrementado en dichos rangos respecto a 2008 lo que significa que la proporción de personas en situación de pobreza es mayor que en 2008 para ese sector de la población.

Con dicha información es posible generar análisis que permitan diseñar políticas de combate a

la pobreza focalizadas con mayor precisión, por ejemplo, en el siguiente gráfico se muestra la distribución de la población del rango 5 del 2008 al 2014, en dicho gráfico es posible observar que Guanajuato, Guerrero, Morelos, Puebla, Quintana Roo, San Luis Potosí, Tlaxcala y Zacatecas son estados de la república mexicana en los que la población evaluada en este rango de puntaje, se ha incrementado con el paso del tiempo.

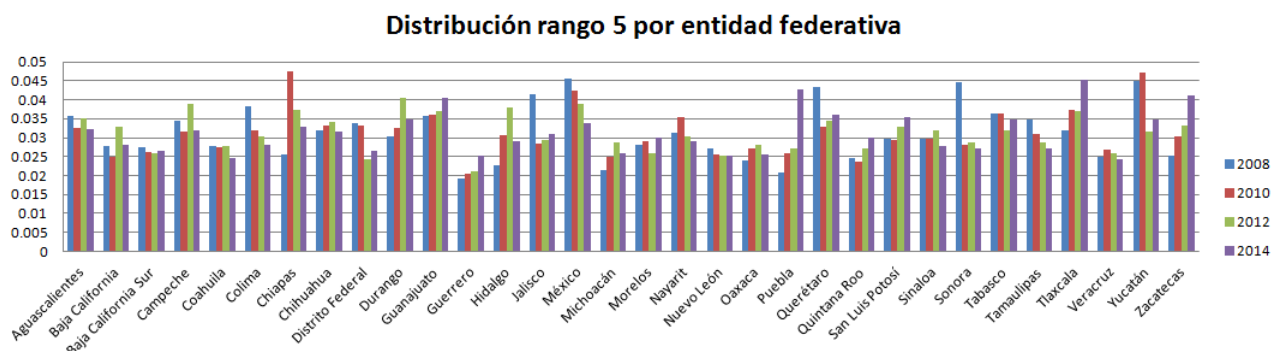


Figura 4.9: Distribución rango 5 por entidad federativa.

Fuente: estimaciones con información del CONEVAL con base en las ENIGH de 2008 a 2012

El estado de Puebla en 2014 particularmente registra un aumento de la población en el rango 5 de un 70 % en 2014 respecto a 2012. A través del puntaje asignado, es posible generar un análisis que permita explicar mejor el comportamiento de la población en dicho rango de puntaje para el estado de Puebla y diseñar un análisis que sirva de soporte para la creación de una estrategia de combate a la pobreza para el mismo.

En primer lugar se analiza cada una de las variables del modelo y su respectiva distribución. Las variables Servicio de Salud e Indicador de carencia alimentaria, son las variables que presentan cambios en su distribución a lo largo del tiempo en algunas de sus categorías.

La variable Servicio de Salud, muestra un incremento en la población catalogada como pobre en la categoría Seguro Popular y una disminución en la categoría IMSS. El programa de Seguro Popular fue implementado por la Secretaria de Salud en 2002, este programa se define como instrumento de protección financiera para la población no derechohabiente, misma que a través

de un pago denominado prepago recibe por un período determinado servicios de salud, medicamentos etc. para mantener y preservar la salud. Por otra parte recordemos que únicamente aquellos que gozan de un trabajo formal son derechohabientes del IMSS, el incremento en el número de personas catalogadas como pobres podría indicar que a pesar de los esfuerzos que realiza el gobierno con este programa para garantizar el acceso universal a los servicios de salud, siempre está presente un número significativo de ciudadanos que enfrentan una barrera financiera para acceder a los mismos.

El 14 de agosto el Instituto Nacional de Estadística y Geografía (INEGI) presentó los resultados de la Encuesta Nacional de Ocupación y Empleo. El documento agrega que la tasa de desocupación de 3.2 por ciento de la entidad poblana fue inferior al 4.3 por ciento, registrada a nivel nacional durante los meses de abril, mayo y junio de 2015. La informalidad laboral en Puebla registró 72.4 por ciento, es decir, este porcentaje es de aquellos trabajadores que no gozan de prestaciones y no están contratados de acuerdo con lo que marca la Ley Federal del Trabajo vigente.

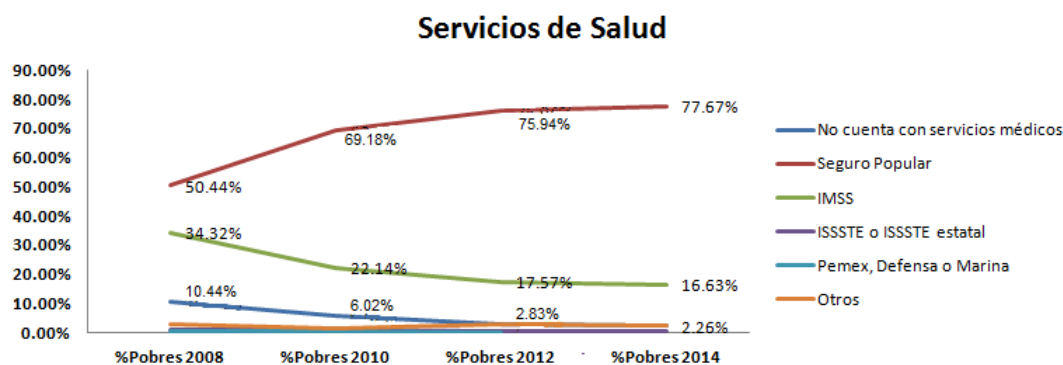


Figura 4.10: Servicios de Salud.

Fuente: estimaciones con información del CONEVAL con base en las ENIGH de 2008 a 2012

Es posible agregar al análisis anterior más variables que puedan segmentar y definir con más precisión las características de la población sobre las que técnicamente es más recomendable focalizar una estrategia de combate a la pobreza.

Por ejemplo al análisis anterior, podemos agregarle la variable edad y sexo:

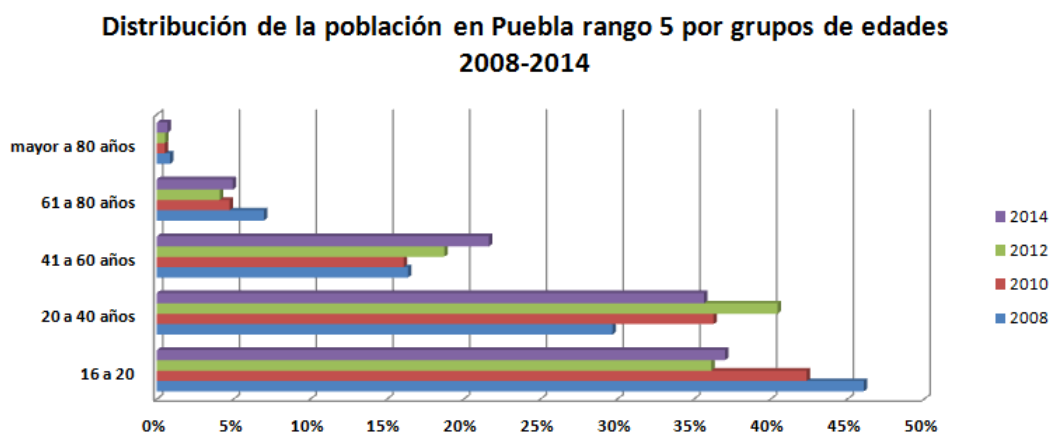


Figura 4.11: Distribución de la población en Puebla rango 5 por grupos de edades 2008- 2014

Fuente: estimaciones con información del CONEVAL con base en las ENIGH de 2008 a 2012

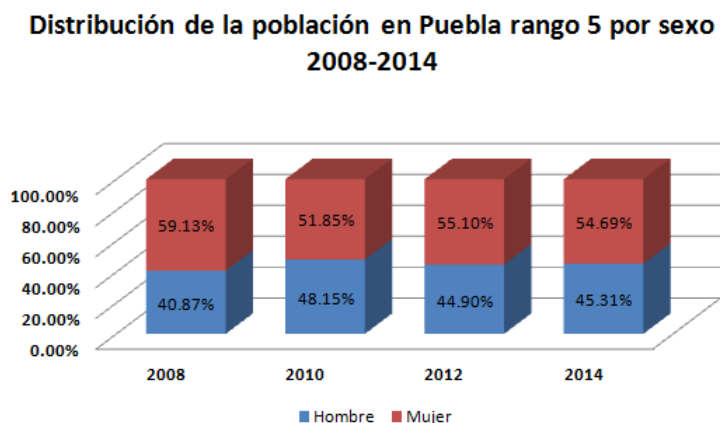


Figura 4.12: Distribución de la población en Puebla rango 5 por sexo 2008 - 2014

Fuente: estimaciones con información del CONEVAL con base en las ENIGH de 2008 a 2012

En el primer gráfico se observa que desde 2008 el 40.40 % de la población que muestra cambios

en el rango 5 para el estado de Puebla, es menor a 20 años y en promedio el 55.19 % son mujeres.

Como se ha podido observar, el enfoque cuantitativo a través del uso del puntaje estadístico produce métodos de evaluación que explican el fenómeno de la pobreza de manera más completa y permite proponer y elaborar estrategias más eficaces. En este caso se ha detectado que el estado de Puebla para la población evaluada con puntajes de 519 a 530, principalmente mujeres menores de 20 años se están produciendo cambios en el comportamiento de la variable Servicios de Salud, provocando que más personas sean evaluadas como pobres al presentar un incremento en la población catalogada como pobre en la categoría Seguro Popular y una disminución en la categoría IMSS. Lo anterior puede ser explicado por el incremento de la informalidad laboral en Puebla que registró 72.4 por ciento en 2015, es decir, este porcentaje es de aquellos trabajadores que no gozan de prestaciones y no están contratados de acuerdo con lo que marca la Ley Federal del Trabajo vigente.

Diversos estudios demuestran que en países como el nuestro, una parte del presupuesto familiar es destinado al pago de bienes y servicios necesarios para atender la salud de sus integrantes, mismo que la mayoría de las veces afecta económicamente a las familias más pobres y vulnerables, y a pesar de que en cierta medida se están cubriendo los problemas básicos de salud de este sector de la población con el programa del Seguro Popular, valdría la pena evaluar los umbrales de cobertura de dicho programa.

# Capítulo 5

## Conclusiones

El número de carencias se mide a través de los seis indicadores de pobreza establecidos por el CONEVAL, que a su vez han sido contruidos con base a 17 variables, se desarrolló un modelo de regresión que permitió clasificar a cada uno de los individuos pertenecientes a la muestra del MCS 2008 (Módulo de Condiciones Socioeconómicas de la ENIGH 2008) acorde a su grado de pobreza empleando un menor número de variables de las que fueron consideradas por el CONEVAL en la construcción de cada indicador de pobreza.

El modelo de regresión empleado en esta tesis considera solo 7 variables explicativas y la clasificación de todos los individuos de la muestra según la concordancia de los valores observados con los predichos o estimados por el modelo, mostró que el modelo clasifica adecuadamente un 47.5% a los individuos No Pobres multidimensionales y en 89.4% a los individuos Pobres multidimensionales, proporcionando un ajuste global del 75.4% un porcentaje bastante alto, del que se deduce que el modelo tiene una alta capacidad predictiva.

Con un menor número de variables y un 75.4% de la muestra ajustada correctamente, se construyó un puntaje o calificación estadística para cada individuo con el objetivo de comparar y ordenar a los individuos evaluados como pobres o no pobres.

No se sugiere cambiar la metodología empleada actualmente por el CONEVAL para evaluar pobreza; sin embargo se mostró que hoy en día el estudio del cambio de los índices de pobreza

en el tiempo en relación con el gasto social y variación en los porcentajes de pobres es compleja. Motivo por el cual se propone el modelo de puntaje desarrollado en esta tesis como un soporte para el desarrollo de estrategias de combate a la pobreza, que permita afinar las características de la población objetivo y diseñar políticas de combate acorde a sus necesidades siendo una métrica viable para determinar la magnitud de los efectos de los programas de combate a la pobreza a través del tiempo.

Considerando lo anterior, se ejecutó el modelo de puntaje desarrollado en este trabajo con los datos de la base del MCS (Módulo de Condiciones Socioeconómicas de la ENIGH ) 2008, 2010, 2012 y 2014 respectivamente, la cual contiene las variables necesarias para la medición multi-dimensional de la pobreza, establecidas en la Ley de Desarrollo Social brindando las siguientes ventajas:

- La primera ventaja que proporciona el modelo de puntaje estadístico propuesto es que a pesar de que dicho puntaje técnicamente implica cierto nivel de complejidad, su uso en la práctica es automatizado, emplear un modelo de score implicaría que la selección y evaluación de las solicitudes de los aspirantes a beneficiarios fueran más objetivas, el puntaje estadístico es una medida consistente, dos personas con las mismas características son idénticamente evaluadas.
- Se generó un análisis con resultados comparables y se proporcionaron puntos de referencia para su evaluación a través del tiempo. En dicho análisis, se identificó que los estados de la distribución de la población del rango 5 y 7 del 2008 al 2014, ha registrado movimientos a la alza a lo largo del tiempo, es decir; el número de personas que está siendo evaluada con un puntajes más altos se ha incrementado en dichos rangos respecto a 2008 lo que significa que la proporción de personas en situación de pobreza es mayor que en 2008 para ese sector de la población.
- A través del puntaje estadístico se pueden generar el desarrollo de modelos de combate a



la pobreza focalizados más certeros, en el que sería posible afinar las características de la población objetivo y diseñar políticas de combate acorde a sus necesidades. Los estados de Guanajuato, Guerrero, Morelos, Puebla, Quintana Roo, San Luis Potosí, Tlaxcala y Zacatecas son estados de la república mexicana en los que la población evaluada en el rango 5 de puntaje, se ha incrementado con el paso del tiempo. El estado de Puebla en 2014 particularmente registra un aumento de la población en el rango 5 de un 70 % en 2014 respecto a 2012. Se analizó cada una de las variables del modelo y su respectiva distribución. Las variables Servicio de Salud e Indicador de carencia alimentaria, son las variables que presentaron cambios en su distribución a lo largo del tiempo en algunas de sus categorías.

- La variable Servicio de Salud, mostró un incremento en la población catalogada como pobre en la categoría Seguro Popular y una disminución en la categoría IMSS. El programa de Seguro Popular fue implementado por la Secretaría de Salud en 2002, este programa se define como instrumento de protección financiera para la población no derechohabiente, misma que a través de un pago denominado prepago recibe por un período determinado servicios de salud, medicamentos etc. para mantener y preservar la salud. Por otra parte recordemos que únicamente aquellos que gozan de un trabajo formal son derechohabientes del IMSS, el incremento en el número de personas catalogadas como pobres podría indicar que a pesar de los esfuerzos que realiza el gobierno con este programa para garantizar el acceso universal a los servicios de salud, siempre está presente un número significativo de ciudadanos que enfrentan una barrera financiera para acceder a los mismos. Lo anterior puede ser explicado por el incremento de la informalidad laboral en Puebla que registró 72.4 por ciento en 2015, es decir, este porcentaje es de aquellos trabajadores que no gozan de prestaciones y no están contratados de acuerdo con lo que marca la Ley Federal del Trabajo vigente.

La importancia de la evaluación de los programas sociales no radica únicamente en determinar el número de beneficiarios directos e indirectos que cada programa establezca, mucho menos

en justificar decisiones alineadas a los intereses del gobierno en turno; sino en la capacidad de los gobiernos de poder diseñar nuevos enfoques que permitan atender de manera focalizada y eficiente las necesidades de la población que más lo requiere.

El enfoque cuantitativo produce métodos de evaluación que solo van a calcular aquello que pueda ser medido de este modo, como consecuencia las metas y los objetivos que se plantean programas y planes. Se necesita integrar a la medición de la pobreza y de las carencias sociales un enfoque cualitativo que explique el fenómeno de la pobreza de manera más completa y permita proponer y elaborar estrategias más eficaces.

Entre las dimensiones de la pobreza que hoy se evalúan es necesario tener en cuenta que en tiempos futuros, estas deberían ajustarse o expandirse a nuevos elementos para la elaboración de un análisis más exhaustivo.

# Referencias

- I. Amarya Sen, *Desarrollo y Libertad*, Buenos Aires, Editorial Planeta S.A, 2000. Recuperado de: <http://www.cdh.uchile.cl/media/publicaciones/pdf/5/260.pdf>
- II. Boltvinik Julio, Enrique Hernández Laos, (2001). *Pobreza y Distribución de Ingreso en México*. Siglo Veintiuno Editores. México, D.F. Tercera Edición.
- III. Banco Mundial de México, *La pobreza en México: una evaluación de las condiciones, tendencias y estrategia de gobierno*, México: Banco Mundial, 2004, Traducción Marcela Pimentel Lusarreta, página 2-4.
- IV. Bolton Christine (2009), *Logistic regression and its application in credit scoring*. Recuperado de: <http://repository.up.ac.za/bitstream/handle/2263/27333/dissertation.pdf>
- V. CONEVAL (2015), *Dirección de Información y Comunicación social prensa No 005*. Recuperado de: [http://www.coneval.gob.mx/Documents/Comunicado005 Medici?n pobreza 2014.pdf](http://www.coneval.gob.mx/Documents/Comunicado005%20Medici%27n%20pobreza%202014.pdf).
- VI. CONEVAL (2015), *Metodología de Medición Multidimensional de la pobreza en México*. Recuperado de: [http://www.coneval.gob.mx/SalaPrensa/Documents/Comunicado005 Medición pobreza 2014.pdf](http://www.coneval.gob.mx/SalaPrensa/Documents/Comunicado005%20Medici%27n%20pobreza%202014.pdf).

- VII. Cortes Fernández, Escobar Agustín et al. (2008). *Método científico y Político Social a propósito de las evaluaciones cualitativas de programas sociales*. El Colegio de México. México, D.F. Primera Edición.
- VIII. Cortes Fernández, (2014), *El gasto social y la pobreza*. Recuperado de:  
<http://www.pued.unam.mx/archivos/opinion/009.pdf>
- IX. Favela Alejandro, Calvillo Minan, et al. *El combate a la pobreza en México en el sexenio de Zedillo.*, Plata y Valdés Editores, México Primera edición, 2003.
- X. F. Hair Joseph, Anderson E. Rolph et al, *Análisis Multivariante*, 5ta edición, Pearson Prentice Hall Iberia; Madrid, 1999.
- XI. Guoping Zeng (2013), *Metric Divergence Measures and Information Value in Credit Scoring*. Journal of Mathematics, Volume 2013, Article ID 848271, página 10. Recuperado de:  
<http://www.hindawi.com/journals/jmath/2013/848271/>
- XII. Juárez Bolaños Diego, et al. *Combate a la pobreza en México: balance y perspectivas del Programa Oportunidades*, primera edición, 2008 Universidad Autónoma de Nuevo León, página 7.
- XIII. Julio Boltivnik, *Políticas focalizadas de combate a la pobreza en México*  
Recuperado de : <http://www.colmex.mx/academicos>
- XIV. Naem Siddiqi. (2006). *Developing and Implementing Intelligent Credit Scoring*. Jhon Wiley and Sons, Inc., Hoboken, New Jersey.

- XV. Nieto Soraida (2010). *Credito al Consumo: La Estadística aplicada a un problema de Riesgo Crediticio*. (Tesis Licenciatura). Universidad Autónoma Metropolitana.
- XVI. Programa Sectorial de Desarrollo Social 2007-2012 SEDESOL <http://www.sedesol.gob.mx>
- XVII. Silva Luis Carlos, Barroso Isabel Maria. (2004). *Regresión Logística*. Editorial Hespírides S.L. Madrid . Primera Edición.

# Apéndice A

## Análisis Exploratorio de los datos

El análisis preliminar de los datos, es un proceso necesario en la comprensión de los rasgos generales de la estructura de estos y de las relaciones que puedan existir entre las variables.

Como un primer paso, se aseguró que los datos de la base estuvieran bien estructurados; es decir se validó que la distribución de la población por categoría en cada variable respecto a la variable regresora no fuera nula.

Lo anterior debido a que cuando los datos no están bien estructurados o hay muy pocos, se puede presentar el problema de la separación en regresión logística, el cual ocasiona que el proceso de estimación de los estimadores por máxima verosimilitud de los parámetros, no converja (el algoritmo de Newton-Raphson crece infinitamente).

Con la finalidad de descartar problemas en la estructura de los datos se llevó a cabo un análisis de las variables a través de la elaboración de Tablas de Contingencia. Lo anterior debido a que el conjunto de variables independientes con las que se construirá el modelo de regresión, en su mayoría son de tipo categórico.

No se detectó ninguna variable con frecuencias nulas en sus categorías; sin embargo algunas de ellas presentan frecuencias menores al 4% de la población, situación que debe tenerse en cuenta al momento de la selección de variables para la construcción del modelo.

Adicional a la revisión de la correlación entre las variables a introducir al modelo de regresión, en esta sección se revisa la correlación entre las variables independientes y la variable pobreza.

Las tablas de contingencia para todas las variables candidatas a ser consideradas en el modelo de regresión, se presenta a continuación.

Cuadro A.1: Tabla de Contingencia edad vs pobreza.

Edad Rangos	No Pobre multidimensional	Pobreza multidimensional	Total
1 Count	16,044 14.3 %	35,885 31.9 %	51,929 46.2 %
2 Count	11,076 9.9 %	20,919 18.6 %	31,995 28.5 %
3 Count	6,999 6.2 %	11,566 10.3 %	18,565 16.5 %
4 Count	2,916 2.6 %	5,468 4.9 %	8,384 7.5 %
5 Count	444 0.4 %	1,001 0.9 %	1,445 1.3 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.2: Tabla de Contingencia inasistencia a la escuela vs pobreza.

Inasistencia a la escuela	No Pobre multidimensional	Pobreza multidimensional	Total
Si Asiste	11,939 10.6 %	24,393 21.7 %	36,332 32.3 %
No Asiste	23,696 21.1 %	45,693 40.7 %	69,389 61.8 %
NA	1,844 1.6 %	4,753 4.2 %	6,597 5.9 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.3: Tabla de Contingencia nivel educativo vs pobreza.

Nivel Educativo	No Pobre multidimensional	Pobreza multidimensional	Total
Cuenta con primaria incompleta o menos	11,922 10.6 %	32,925 29.3 %	44,847 39.9 %
Cuenta con primaria o sec. incompleta	9,154 8.2 %	16,296 14.5 %	25,450 22.7 %
Cuenta con sec. o mayor nivel educativo	14,558 13 %	20,864 18.6 %	35,422 31.5 %
NA	1,844 1.6 %	4,753 4.2 %	6,597 5.9 %
Total	37,478 33.4 %	74,838 66.6 %	112,316 100 %

Cuadro A.4: Tabla de Contingencia servicio de salud vs pobreza.

Servicios de Salud	No Pobre multidimensional	Pobreza multidimensional	Total
No Cuenta con servicios medicos	9,840 8.8 %	36,048 32.1 %	45,888 40.9 %
Seguro Popular	7,456 6.6 %	23,623 21.0 %	31,089 27.7 %
IMSS	16,985 15.1 %	1,328 11.6 %	30,013 26.7 %
ISSSTE	2,595 2.3 %	1,213 1.1 %	3,808 3.4 %
Pemex o defensa marina	276 .2 %	242 .2 %	518 .5 %
Otros	317 .3 %	685 .6 %	1,002 .9 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.5: Tabla de Contingencia seguridad social vs pobreza.

Acceso directo seguridad social	No Pobre multidimensional	Pobreza multidimensional	Total
Sin acceso	31,266 27.8 %	71,807 63.9 %	103,073 91.8 %
Con acceso	6,213 5.5 %	3,032 2.7 %	9,245 8.2 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %



Cuadro A.6: Tabla de Contingencia acceso directo a la seguridad social a través del jefe del hogar vs pobreza.

Acceso directo seguridad social jefe del hogar	No Pobre multidimensional	Pobreza multidimensional	Total
No cuenta	23,174 20.6 %	64,964 57.8 %	88,138 78.5 %
Si cuenta	14,305 12.7 %	9,875 8.8 %	24,180 21.5 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.7: Tabla de Contingencia acceso directo a la seguridad social a través del cónyuge del jefe del hogar vs pobreza.

Acceso directo seguridad social a través del cony.	No Pobre multidimensional	Pobreza multidimensional	Total
No cuenta	35,286 31.4 %	73,728 65.6 %	109,014 97.1 %
Si cuenta	2,193 2 %	1,111 1 %	3,304 2.9 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.8: Tabla de Contingencia acceso directo a la seguridad social a través de los hijos vs pobreza.

Acceso directo seguridad social a través de los hijos.	No Pobre multidimensional	Pobreza multidimensional	Total
No cuenta	33,134 29.5 %	70,070 62.4 %	103,204 91.9 %
Si cuenta	4,345 3.9 %	4,769 4.2 %	9,114 8.1 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.9: Tabla de Contingencia programa adulto mayores vs pobreza.

Programa adultos mayores.	No Pobre multidimensional	Pobreza multidimensional	Total
No cuenta	1,843 1.6 %	3,727 3.3 %	5,570 5 %
Si cuenta	722 .6 %	1,360 1.2 %	2,082 1.19 %
NA	39,914 31.1 %	69,752 62.1 %	104,666 93.2 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.10: Tabla de Contingencia servicio médicos por otros núcleos familiares vs pobreza.

servicio m?dicos por otros núcleos familiares	No Pobre multidimensional	Pobreza multidimensional	Total
No cuenta	28,522 25.4 %	69,841 62.2 %	98,363 87.6 %
Si cuenta	8,957 8.0 %	4,998 4.4 %	13,965 12.4 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.11: Tabla de Contingencia indicador de carencia del material de piso de la vivienda vs pobreza.

Indicador de carencia del material de piso de la vivienda	No Pobre multidimensional	Pobreza multidimensional	Total
Sin Carencia	35,800 31.9 %	68,658 61.1 %	104,458 93 %
Con Carencia	1,679 1.5 %	6,181 5.5 %	7,860 7 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.12: Tabla de Contingencia indicador de carencia del material del techo de la vivienda vs pobreza.

Indicador de carencia del material de piso de la vivienda	No Pobre multidimensional	Pobreza multidimensional	Total
Sin Carencia	36,825 32.8 %	72,179 64.3 %	109,004 97 %
Con Carencia	654 .6 %	2660 2.4 %	3,314 3 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.13: Tabla de Contingencia indicador de carencia del material del muros de la vivienda vs pobreza.

Indicador de carencia del material del muro de la vivienda	No Pobre multidimensional	Pobreza multidimensional	Total
Sin Carencia	36,869 32.8 %	72,451 64.5 %	109,320 97.3 %
Con Carencia	610 .5 %	2,388 2.1 %	2,998 2.7 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.14: Tabla de Contingencia carencia de acceso al agua vs pobreza.

Carencia de acceso al agua	No Pobre multidimensional	Pobreza multidimensional	Total
Sin Carencia	33,913 30.2 %	66,110 58.9 %	100,023 89.1 %
Con Carencia	3,566 3.2 %	8,729 7.8 %	12,295 10.9 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.15: Tabla de Contingencia carencia de servicio de drenaje vs pobreza.

Carencia de servicio de drenaje	No Pobre multidimensional	Pobreza multidimensional	Total
Sin Carencia	33,348 29.7 %	64,783 57.7 %	9,8131 87.4 %
Con Carencia	4,131 3.7 %	10,056 9 %	14,187 12.6 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.16: Tabla de Contingencia carencia de servicio de electricidad vs pobreza.

Carencia de servicio de electricidad	No Pobre multidimensional	Pobreza multidimensional	Total
Sin Carencia	37,158 33.1 %	74,001 65.9 %	111,159 99 %
Con Carencia	321 .3 %	838 .7 %	1,159 1 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.17: Tabla de Contingencia grado de inseguridad alimentaria vs pobreza.

Grado de inseguridad alimentaria	No Pobre multidimensional	Pobreza multidimensional	Total
Seguridad alimentaria	21,463 19.1 %	29,596 26.4 %	51,059 45.5 %
Inseguridad alimentaria leve	10,511 9.4 %	21,816 19.4 %	3,2327 28.8 %
Inseguridad alimentaria moderada	3,352 3 %	13,659 12.2 %	17,011 15.1 %
Inseguridad alimentaria severa	2,153 1.9 %	9,768 8.7 %	11,921 10.6 %
Total	37,479 33.4 %	74,839 66.6 %	112,318 100 %

Cuadro A.18: Correlación variable independiente vs la variable Pobreza

	Pobreza Multidimensional	
Nivel Educativo	Pearson Correlation	-0.098
	Sig (2 tailed)	0.000
	N	112,318
Inasistencia a la escuela	Pearson Correlation	0.006
	Sig (2 tailed)	0.050
	N	112,318
Servicio de Salud	Pearson Correlation	-0.290
	Sig (2 tailed)	0.000
	N	112,318
Acceso a la seguridad social	Pearson Correlation	-0.215
	Sig (2 tailed)	0.000
	N	112,318
Acceso a la s.s. a través del jefe	Pearson Correlation	-0.286
	Sig (2 tailed)	0.000
	N	112,318
Acceso a la s.s. a través del cónyuge	Pearson Correlation	-0.122
	Sig (2 tailed)	0.000
	N	112,318
Acceso a la s.s. a través de los hijos	Pearson Correlation	-0.090
	Sig (2 tailed)	0.000
	N	112,318
Programa adultos mayores	Pearson Correlation	0.000
	Sig (2 tailed)	0.955
	N	112,318
Servicios médicos	Pearson Correlation	-0.246
	Sig (2 tailed)	0.000
	N	112,318
Carencia del material del piso de la vivienda	Pearson Correlation	0.070
	Sig (2 tailed)	0.000
	N	112,318
Carencia del material del techo de la vivienda	Pearson Correlation	0.050
	Sig (2 tailed)	0.000
	N	112,318
Carencia del material del muro de la vivienda	Pearson Correlation	0.046
	Sig (2 tailed)	0.000
	N	112,318
Carencia de acceso al agua	Pearson Correlation	0.032
	Sig (2 tailed)	0.000
	N	112,318

	Pobreza Multidimensional	
Carencia servicios de drenaje	Pearson Correlation Sig (2 tailed) N	0.034 0.000 112, 318
Carencia servicio de electricidad	Pearson Correlation Sig (2 tailed) N	0.012 0.000 112, 318
Grado de inseguridad alimentaria	Pearson Correlation Sig (2 tailed) N	0.194 0.000 112, 318