

Producto 3:

INFORME FINAL DE LA EVALUACIÓN

Proyecto sobre “EVALUACIÓN DEL PADRÓN DE BENEFICIARIOS DEL SISTEMA DE PROTECCIÓN SOCIAL EN SALUD PARA LA CORRECTA CLASIFICACIÓN DE BENEFICIARIOS” en el marco de la Licitación Pública LA-012000976-N35-2014 emitida por la *Comisión Nacional de Protección Social en Salud*.

México, D.F. a 29 de diciembre de 2014

TABLA DE CONTENIDO DEL INFORME FINAL

SECCIÓN	PÁGINA
I. Resumen Ejecutivo	
1.1 Objetivos	3
1.2 Descripción de la metodología	4
1.3 Principales resultados y conclusiones	5
II. Metodología y procesos conceptuales.	
2.1 Análisis de Contexto	15
2.2 Análisis conceptual y técnico-estadístico de las variables y de la medición que contiene la CECASOEH	18
2.3 Análisis conceptual	22
2.4 Análisis Técnico-Estadístico	30
2.5 Evaluación de la técnica estadística utilizada actualmente en la CECASOEH	39
2.6 Aplicación de la metodología (modelo, ponderadores y umbrales) y desarrollo de propuesta	51
2.7 Conclusiones y planteamiento propuesto	55
III. Metodología de piloteo en campo y validación del instrumento CECASOEH v 2.0	57

Se Anexa adicionalmente a este documento como parte del Informe Final:

- a) **Formato de CECASOEH v 2.0 desarrollado en el marco de este proyecto;**
- b) **Programa de cómputo con sistema de clasificación de la población de acuerdo a la CECASOEH v 2.0 desarrollado en el presente proyecto;**
- c) **Base de datos con 3,000 observaciones incluyendo los datos de los participantes en el Piloto de la CECASOEH v 2.0 (incluye ponderadores);**
- d) **Manual de Campo**
- e) **Bibliografía y referencias**

I. Resumen Ejecutivo

1.1 Objetivos del Estudio

Este documento presenta los resultados del proyecto sobre “EVALUACIÓN DEL PADRÓN DE BENEFICIARIOS DEL SISTEMA DE PROTECCIÓN SOCIAL EN SALUD PARA LA CORRECTA CLASIFICACIÓN DE BENEFICIARIOS” en el marco de la Licitación Pública LA-012000976-N35-2014 emitida por la *Comisión Nacional de Protección Social en Salud*.

El objetivo general del presente proyecto es realizar un estudio de validación de la Cédula de Características Socioeconómicas del Hogar (CECASOEH) y del modelo estadístico de clasificación utilizado actualmente por el Sistema de Protección Social en Salud.

Los objetivos específicos son:

1. Estimar la pertinencia de la medición obtenida por el instrumento CECASOEH en términos de su sensibilidad y pertinencia, mediante la comparación del instrumento que clasifica a los miembros del padrón del SPSS con un estándar con firmeza conceptual (que se denomina como de oro) que permita la construcción del nuevo régimen contributivo.
2. Revisar el modelo estadístico actual (análisis discriminante) utilizado por la CECASOEH, con el fin de analizar su pertinencia en cuanto a su confiabilidad y validez para mejorar el mismo o bien proponer una alternativa con el fin de que este nuevo tipo de análisis permita definir, con base en evidencia, los deciles socioeconómicos u otra alternativa de medición para la afiliación de familias.
3. Generar un marco conceptual y técnico para proponer el modelo estadístico para el instrumento de CECASOEH, basado en la investigación actual y relevante tanto aplicada como académica sobre los estudios de medición socioeconómica.
4. Proponer un modelo de actualización de las variables que forman parte del instrumento de captación para una CECASOEH versión 2.0 y del algoritmo del modelo estadístico de clasificación de las familias en deciles de ingreso u otro tipo de medida en función de resultados del desarrollo analítico y en campo.
5. Pilotear en campo, la efectividad del nuevo instrumento de captación de información (CECASOEH versión 2.0) o en su defecto, la propuesta de un nuevo instrumento de clasificación de familias en deciles de ingreso o la alternativa de medición.

6. Considerar que el modelo de medición y el instrumento depende de los procesos logísticos de incorporación de familias realizados en los Módulos de Afiliación y Orientación al SPSS.
7. Revisar el proceso de incorporación de familias al SPSS, con énfasis en el proceso de registro de información, con el fin de identificar áreas susceptibles de mejora.

Específicamente, se analizan tres dimensiones subyacentes al modelo de selección utilizado por la SPSS para clasificar a la población beneficiaria de acuerdo a los distintos esquemas de aportación (cuotas):

- a) Las variables a incorporar;
- b) La ponderación de cada dimensión;
- c) El umbral a partir del cual se considera que un hogar o individuo presenta la carencia de cada dimensión de bienestar

1.2 Descripción de la metodología

La identificación de las variables a incorporar así como su ponderación en el proceso de selección de beneficiarios, se llevó a cabo por medio de análisis discriminante que permite identificar los espacios que definen dos condiciones, la de pertenecer al grupo de potencial beneficiario de un programa social, y la de no serlo, a partir de arreglos multidimensionales que caracterizan estas condiciones, mediante valores típicos de las variables utilizadas. Partiendo de la información del ingreso y comparándolo con un punto de referencia establecido, se dispone de una base para caracterizar los perfiles que distinguen a los hogares con mayor y menor bienestar mediante un conjunto de indicadores.

La metodología de análisis discriminante es una herramienta que cuenta con los elementos y la robustez para constituirse como mecanismo de focalización para el SPSS y por eso se contempla como principal alternativa metodológica. En el caso del proyecto actual, el contar con un mecanismo de focalización es esencial para evitar errores de inclusión (es decir, catalogar a hogares que no pertenecen a los primeros cuatro deciles de medición del bienestar) o de exclusión (que consiste en clasificar erróneamente a hogares que pertenecen a los primeros cuatro deciles de medición del bienestar, como si pertenecieran a grupos de mayor nivel de bienestar en deciles superiores).

La importancia de evitar errores de inclusión y exclusión en el marco del SPSS radica en que un error de inclusión eximiría del pago del servicio a hogares que están en condición de realizar alguna

aportación, mientras que el error de exclusión llevaría a exigir una cuota de pago a hogares a los que no les correspondería realizar una aportación debido a que cuentan con las condiciones para clasificarse (bajo una metodología con mayor precisión) a los primeros cuatro deciles de nivel de bienestar.

1.3 Principales resultados y conclusiones

Las conclusiones del análisis del modelo utilizado actualmente por la SPSS para la clasificación de beneficiarios en la distribución del ingreso –lo cual se estima por medio de la aplicación del algoritmo empleado a los datos de la Encuesta Nacional de Ingreso Gasto de los Hogares (ENIGH) 2012 que la fuente más fidedigna con la que se cuenta para obtener información sobre el ingreso de toda la distribución de población del país- son que:

- En términos de las variables, la CECASOEH brinda un punto de partida correcto al considerar variables de distintas dimensiones que se complementan y logran caracterizar el estrato socio-económico al no contar con indicadores directos viables de ingreso.
- El modelo de clasificación actual del SPSS asigna correctamente al 14.4% de los hogares de acuerdo a su posición real en la distribución del ingreso.
- En general el modelo es un mejor predictor para los deciles bajos (del 1 al 4), reduciendo su precisión para el resto de los deciles, particularmente para el 9 y el 10, donde el margen de error es cercano a cero.
- Considerando estratos rural y urbano por separado, en lo rural el modelo predice con un alto grado de precisión para los primeros cuatro deciles rurales, pero un bajo nivel de precisión para el resto. Por otra parte, para lo urbano, el modelo es un buen predictor para los primeros cuatro deciles y los dos últimos, lo cual puede servir como referente.

Utilizando las mismas variables del modelo de clasificación del SPSS actual, se aplicó la metodología de análisis discriminante sobre los datos de la ENIGH para evaluar si mejora o no el grado de predicción en la clasificación respecto a la distribución por deciles de la ENIGH. La utilización de esta metodología implica el reestimar los ponderadores para identificar los que sean más idóneos para incrementar el poder de precisión del instrumento lo más posible. Los resultados dicha aplicación son:

- Cuando las variables originalmente incorporadas en la CECASOEH se procesan utilizando el análisis discriminante el poder de predicción de dichas variables se incrementa considerablemente. Con este modelo, se asigna atinadamente al 47.4% de los hogares, mejorando así el grado de precisión del modelo actual de clasificación del SPSS, que ubica adecuadamente solo al 14.4%.

Como resultado del análisis llevado a cabo en el presente estudio, se identifica un conjunto de variables que mantiene como base las variables planteadas por el modelo actual del SPSS complementándolas con un conjunto de indicadores que cuentan con mayor precisión para la captación de tenencia de activos y mejor caracterización de los deciles más altos. Las variables para el sector rural en dicha propuesta son las siguientes:

- Material del piso de la vivienda
- Si cuenta con excusado
- Educación del jefe del hogar
- Índice de dependencia
- Hacinamiento
- Seguro social
- Hijos que no asisten a la escuela
- Hijos que trabajan
- Tipo de servicios de salud
- Si cuenta con estufa
- Si cuenta con refrigerador
- Si cuenta con lavadora
- Número de prestaciones de todos los que trabajan en el hogar
- Si cuenta con vehículo 0, 1 o más de 2

Las variables urbanas son las siguientes:

- Tamaño de localidad
- Si cuenta con excusado
- Educación del jefe del hogar
- Índice de dependencia
- Hacinamiento
- Seguro social
- Hijos que no asisten a la escuela
- Hijos que trabajan
- Si cuenta con vehículo 0, 1 o más de 2
- Si cuenta con computadora 0, 1 o más de 2
- Número de prestaciones de todos los que trabajan en el hogar
- Número de integrantes que asisten a escuela pública

- Número de personas que laboran en el hogar entre el total de integrantes
- Tipo de empresa en la que labora (1. Tipo independiente, familiar o personal)
- Tenencia de la vivienda
- Número de miembros a los que les otorgaron beca escolar
- Si cuenta con celular
- Si cuenta con teléfono
- Tipo de servicios de salud
- Si cuenta con TV de paga
- Si cuenta con conexión a internet

Los resultados de la aplicación de este modelo, denominado CECASOEH v 2.0, utilizando análisis discriminante, y estimando su poder de decisión por medio de la ENIGH, son:

- Con la agregación en cuatro grupos y la combinación de modelos, se logra incrementar el poder de precisión a 56.6% de los hogares. Es decir, la metodología predice adecuadamente la ubicación de 56.6% de los hogares en el grupo socioeconómico que realmente le corresponde.
- Sin embargo, si se hace la agregación considerando únicamente tres grupos de deciles, el modelo combinado arroja un mayor grado de predicción para el primer grupo de deciles (1 al 4) y mejora significativamente el grado de predicción para el grupo de deciles intermedios (5 al 8), y para el grupo de deciles (9 y 10). Con este modelo, se clasifica atinadamente al 66.36% de los hogares, siendo este el grado de predicción más alto entre los ejercicios realizados.

Los análisis aquí planteados y su comparación permiten concluir lo siguiente:

- Es necesario evaluar, desde la perspectiva práctica de la implementación del Sistema, los objetivos que éste busca y en base al grado de precisión en la clasificación de los distintos grupos de deciles, si se requiere modificar los límites y condiciones de pertenencia para los regímenes contributivo y no contributivo. Por ejemplo, una alternativa es considerar un modelo que identifique con la mayor precisión posible a los últimos deciles de ingreso y establecer el régimen contributivo sólo para estos. Lo anterior requiere tomar en cuenta todas las consideraciones relevantes en cuanto al diseño del SPSS como política pública, la viabilidad operativa y el objetivo implícito de contar con la participación del afiliado en el co-financiamiento de su aseguramiento en salud con las consecuencias que esto conlleva.
- Dado el alto grado de predicción que el modelo actual del SPSS muestra para el primer grupo de deciles, aunado al incremento en el grado de precisión para los otros dos grupos

de deciles (particularmente los últimos deciles) producto del planteamiento de un modelo alternativo, la propuesta de este estudio radica en combinar ambas alternativas en un modelo que, comparado con el resto de los ejercicios realizados, permite asignar a un mayor porcentaje de los hogares atinadamente; o en otras palabras, muestra el menor margen de error.

- El cuadro siguiente resume los distintos modelos contemplados en el Análisis

Cuadro Resumen:

Comparación de análisis/ejercicios realizados respecto al % de hogares asignados atinadamente

Análisis/Ejercicio	% de hogares asignados atinadamente
Modelo de clasificación del SPSS actual, agregado en deciles	14.40%
Modelo de clasificación del SPSS actual, agregado en cuatro grupos de deciles	41.25%
Modelo de análisis discriminante con variables del modelo del SPSS actual, agregado en cuatro grupos de deciles	47.37%
Modelo combinado: algoritmo rural del modelo del SPSS actual + modelo de análisis discriminante con variables nuevas, agregado en cuatro grupos de deciles	56.64%
Modelo combinado: algoritmo rural del modelo del SPSS actual + modelo de análisis discriminante con variables nuevas, agregado en tres grupos de deciles	66.36%

Fuente: Elaboración propia con resultados de ejercicios realizados

- Asimismo, la clasificación en tres grupos o estratos coincide con la lógica actual de separación de regímenes y asignación adicional a estos por consideraciones particulares (contar con un menor de 5 años en el hogar o con una embarazada o ser beneficiarios de Prospera, entre otros).

En conclusión, se propone valorar la posibilidad de utilizar este modelo combinado alternativo con tres grupos de deciles o de clasificación como la mejor opción para lograr los objetivos propuestos.

Análisis de viabilidad de la propuesta por medio del piloteo de la CECASOEH v 2.0 en campo

Como resultado del análisis, investigación y recopilación de información, se diseñó una propuesta del instrumento de clasificación socioeconómica CECASOEH v 2.0 que incorpora las variables mencionadas en el apartado anterior.

El muestreo para la aplicación del instrumento se realizó mediante la selección de localidades representando estratos en zona urbana (78%) y rural (22%) de acuerdo con el Censo 2010 de INEGI. La muestra estará integrada por 3,000 hogares a nivel nacional. El grado de error para una muestra de tamaño 3,000 viviendas al 95% de confianza es de +1.79% considerando una muestra aleatoria simple (m.a.s.)

La prueba de campo consistió en las siguientes cuatro etapas principales:

- 1) Selección aleatoria de AGEBS proporcional al tamaño de la población.
- 2) En cada AGEBS, se seleccionaron las manzanas necesarias para realizar 10 entrevistas completas. Las manzanas se eligieron con un recorrido de acuerdo con las manecillas del reloj.
- 3) En campo, el encuestador se ubicó al frente de la manzana y, a partir de la esquina frontal izquierda, realizó un salto sistemático de 10 viviendas conforme a las manecillas del reloj, sin importar si ésta fue efectiva, rechazada, incompleta, etc. No se realizaron más de una entrevista por acera.
- 4) Selección de la persona de mayor edad con calidad de informante.

El operativo de campo se realizó en la Zona Metropolitana de la Ciudad de México, Xalapa, Tuxtla Gutiérrez, Puebla, León, Oaxaca. En total, las zonas cumplen con el criterio de selección y distribución urbano/rural, y conforman una muestra con representatividad de distintos niveles socioeconómicos. El levantamiento se realizó en la ciudad y la zona rural cercana a estas ciudades. El Público Objetivo fueron Hombres y Mujeres mayores de edad del hogar seleccionado.

Durante el levantamiento se aplicó la nueva versión de la Cédula de Características Socioeconómicas del Hogar (CECASOEH v 2.0) considerando las variables analizadas en la sección anterior.

La muestra quedó distribuida de la siguiente manera:

PLAZA	ZONA URBANA	LOCALIDAD RURAL	TOTAL
ZMCM	390	110	500
Xalapa	390	110	500
Tuxtla Gutiérrez	390	110	500
Puebla	390	110	500
León	390	110	500
Oaxaca	390	110	500
TOTAL	2340	660	3000

En cada AGEB y localidad se realizaron 10 entrevistas efectivas, de tal suerte que la selección de AGEB y localidades quedó distribuida de la siguiente manera:

PLAZA	AGEB URBANO	LOCALIDAD RURAL	TOTAL
ZMCM	39	11	50
Xalapa	39	11	50
Tuxtla Gutiérrez	39	11	50
Puebla	39	11	50
León	39	11	50
Oaxaca	39	11	50
TOTAL	234	66	300

Se incluyeron dentro de la muestra seleccionada 60 AGEB como posibles unidades de reemplazo, 10 en cada ciudad, para prevenir situaciones que imposibiliten la visita a alguna de las unidades seleccionadas. Las entrevistas se realizaron a papel y lápiz, cara a cara y utilizando un mismo cuestionario.

En cada plaza participaron dos equipos de trabajo, cada equipo se conformó por un supervisor y cuatro encuestadores.

Coordinación General:

- Gerente de Operaciones: 1
- Coordinador de campo: 1

Coordinación por plaza:

- Número de supervisores de campo por plaza: 2
- Número de entrevistadores por plaza: 8

Previamente al levantamiento de campo, el gerente de operaciones y el coordinador general de campo instruyeron al personal participante. La capacitación del personal se realizó en el Distrito Federal el 5 de diciembre del 2014. En dicha capacitación se utilizaron dos vías de comunicación:

- Presencial para el caso del personal que levantaría el campo en la ZMCM.
- Telefónica para el personal que levantaría el resto de las plazas.

Al finalizar la capacitación, todos los involucrados fueron registrados en un formato de lista de asistencia y se realizó una grabación en audio para su posterior consulta en caso de dudas. Adicionalmente se generó un manual de capacitación general para consulta del personal.

El supervisor de cada plaza fungió como responsable de la organización y supervisión de las actividades en campo, identificación de viviendas seleccionadas, asignación de las cargas de trabajo al equipo de entrevistadores y la verificación de la calidad de la información recolectada.

Al momento de la aplicación el supervisor se posicionó a una distancia prudente del equipo para poder vigilar el trabajo de todos los encuestadores a su cargo. Los encuestadores se encontraron en zonas de trabajo delimitadas por sus supervisores (a unas cuantas cuadras y siempre a la vista), de este modo los supervisores pudieron observar la aplicación de cuestionarios de varios encuestadores durante la misma jornada (moviéndose solo unas calles para atestiguar la realización).

El levantamiento de campo se realizó del 06 al 19 de diciembre del 2014.

Para lograr las 3000 entrevistas efectivas se realizaron 11,999 contactos, es decir el 25% de los casos se concretaron con éxito y cumplieron con los criterios de calidad. A continuación se especifica el tipo de incidencias registradas.

Tipo de Incidencias

TIPO DE INCIDENCIAS	Total	%
Entrevistas efectivas y validadas	3000	25.0%
No desea participar	2536	21.1%
Entrevista cortada / cancelada	301	2.5%
Contacto con menor de edad / no respondiente adecuado	590	4.9%
Hogar deshabilitado / abandonado	948	7.9%
Predio vacío	867	7.2%
Sin acceso / Restricción para ingreso	590	4.9%
No abrió	3061	25.5%
Otro	106	0.8%
TOTAL	11,999	100.0%

A continuación se describe cada uno de los tipos de incidencias encontradas en campo:

- **Entrevistas efectivas y validadas.** Cuando se obtiene información de todos los miembros de hogar y de todas las secciones del cuestionario de manera satisfactoria.
- **No desea participar.** Cuando se realiza algún contacto con el posible entrevistado y éste expresa que no le interesa o no quiere participar.
- **Entrevista cortada/ cancelada.** Cuando el entrevistado, una vez aceptada la entrevista, decide ya no continuar por alguna razón. Cabe mencionar que este tipo de incidencias representan menos del 2.5% del total de contactos realizados.
- **Contacto menor de edad / no es el respondiente adecuado.** Cuando la persona que se encuentra en el hogar aún no cumple la mayoría de edad y por la tanto no es candidato para la entrevista.
- **Hogar deshabitado / abandonado.** Se clasifican las viviendas aparentemente ocupadas pero que al momento de tocar no salió nadie. Se indaga con los vecinos e informan que en este momento nadie la habitaba.
- **Predio vacío.** Cuando de acuerdo al salto sistemático, el hogar que se debía tocar es un lote vacío.
- **Sin acceso / Restricción para ingreso.** Cuando el hogar se encuentra en alguna cerrada, edificio, etc. y que no se puede tener un contacto directo con el posible entrevistado.
- **No abrió.** Cuando después de insistir nadie abrió en el hogar.
- **Otros.** Cualquier caso no cubierto en los anteriores. Por ejemplo:
 - Se encontraron casos en los que el acceso a la zona de trabajo fue una barrera debido a que ésta era una zona considerada de alto riesgo.
 - En zonas rurales no se encontraron datos de identificación del lugar ni nomenclaturas o vías de acceso.

- Se mostró resistencia por parte de las autoridades locales quienes retuvieron a miembros del equipo de campo para indagar sobre la actividad realizada así como el fin de la misma. En ningún caso se registraron incidentes graves.

En general no se observa mayor dificultad para la aplicación del instrumento, únicamente en 301 casos de todos los contactos realizados (2.5%) el informante decidió dar por terminada la entrevista sin completar toda la información del cuestionario.

Como marco de referencia, en un levantamiento de campo regular observamos una tasa de abandono promedio del 30% durante los primeros 10 minutos de aplicación, esto quiere decir que 3 de cada 10 entrevistas que comienzan se dan por terminadas antes de que ésta sea finalizada completamente.

Con lo anterior se puede deducir que el instrumento se encuentra en una tasa de abandono muy por debajo de lo normal, lo que pronostica un desempeño muy positivo en términos de esfuerzo para el personal de campo, así como interés por parte del entrevistado y poca percepción de preguntas intrusivas.

Las dos principales secciones donde los entrevistados mostraron mayor renuencia a dar información fueron:

- Datos personales de los integrantes del hogar y parentesco con el jefe del hogar.
- Artículos con los que cuenta en su hogar.

En ambos casos, los informantes adjudicaron su abandono a razones de seguridad y falta de confianza, la labor del entrevistador para transmitir confianza y certidumbre a los respondentes acerca del operativo permitió que el cuestionario se completara hasta darlo por terminado.

Sin embargo, no se prevé que dichas razones afecten un proceso real de incorporación a los servicios ofrecidos por la SPSS debido a que la afiliación es voluntaria y requiere de la identificación plena de los miembros del hogar. Por este motivo, la conclusión es que el instrumento es adecuado para el propósito establecido.

Un aspecto importante a notar, es que en el conjunto de variables incorporadas en la versión ENCASEH v 2.0 se incluye a la participación en el sistema de seguridad social formal, el cual contribuye a incrementar el nivel de precisión del mecanismo de incorporación. La inclusión de dicha variable es por lo tanto con fines puramente estadísticos. La misma variable se utiliza adicionalmente con fines de establecer la elegibilidad a pertenecer al SPSS y se recomienda continuar con su utilización para este fin.

En cuanto a la implementación de las entrevistas se registra un tiempo promedio de aplicación de 08:29 minutos desde la aceptación a participar en la entrevista por parte del respondente y hasta la terminación del registro de todos los datos solicitados.

- Se registra una desviación estándar de .00190 minutos.
- El 67% de las entrevistas han durado entre 6 y 9 minutos.

Para la construcción de la base de datos, se realizó captura de los cuestionarios en formato electrónico utilizando una plataforma Online destinada para dicho propósito.

En cada plaza la captura inicia un día después del levantamiento, el coordinador de campo y el gerente de operaciones monitorean el avance y realizan una revisión y validación de la base de datos.

Se considera como factor de expansión el inverso de la probabilidad de selección de cada unidad final de muestreo con el propósito de hacer extensivas las conclusiones que del estudio se deriven al total de la población.

Los elementos en la muestra seleccionada no tienen la misma probabilidad de selección, para corregir el efecto del diseño de muestreo se pueden emplear los factores anexos. En cada una de las diferentes AGEb en la muestra se considera un diferente ponderador.

II. Metodología y Procesos Conceptuales

2.1 Análisis de Contexto

Los instrumentos de política pública para identificar el nivel socioeconómico de la población son relativamente recientes en México. Su utilización se hizo indispensable a partir de la transición en el enfoque de la política social pasando de una política basada en subsidios generalizados a una política estructurada con programas sociales focalizados a determinadas poblaciones que requerían ser identificadas con precisión por medio de instrumentos metodológicos robustos.

Por muchos años el país siguió una estrategia de construcción de una amplia red de subsidios al consumo abiertos a la población en general. A principios de los ochenta la estrategia tuvo que modificarse, pese a que en los años anteriores se registraron reducciones considerables en la pobreza y la desigualdad en el país. En 1982 la drástica caída en los precios del petróleo originó la crisis de la deuda que llevó al país a una severa recesión económica. En estos años se vivieron restricciones macroeconómicas sin precedentes, lo cual ocasionó que se considerara prohibitiva una amplia red de subsidios. Este es el inicio de la segunda fase.

Bajo las nuevas condiciones macroeconómicas de principios de la década de los ochenta, el gobierno mexicano redujo el gasto social considerablemente en aras de reducir el déficit público. En un contexto caracterizado por una creciente inflación, la devaluación del tipo de cambio y contracciones del Producto Interno Bruto (PIB), la prioridad de la política pública era estabilizar la economía.

A finales de la década de los ochenta, con la implementación de las primeras reformas estructurales, el gobierno mexicano reconoció explícitamente que los grupos más desfavorecidos de la sociedad eran quienes más habían sufrido con el estancamiento de la economía. La estrategia para hacer frente a esta situación fue la implementación del Programa Nacional de Solidaridad (Pronasol), el cual contaba con dos características que lo hacían distinto a las políticas anteriores. Por un lado, se basaba en el principio de la *focalización*. La idea central de este principio es que cuando se tienen recursos escasos una opción para maximizar el impacto de los programas es canalizar beneficios exclusivamente hacia la población objetivo. Esta estrategia constituyó la alternativa a la provisión de subsidios abiertos a la población en general, como se hizo durante la primera fase de sustitución de importaciones.

El Pronasol marcó un giro importante en la política social de México. Bajo los criterios de focalización, las políticas contra la pobreza fueron compensatorias, dirigidas a grupos específicos. Los métodos de focalización consisten en estimar la condición de susceptibilidad de recibir apoyos

de programas sociales para un hogar o individuo para garantizar que los apoyos lleguen a las familias que efectivamente los requieren, y con ello propiciar un uso transparente y eficiente de los recursos públicos asignados a dicha acción de política pública.

Se pueden identificar dos grandes áreas de focalización: la primera, dirige la inversión a ciertos gastos públicos que supuestamente benefician más a los pobres; la segunda, busca dirigir específicamente los recursos a los individuos más pobres, para lo que existen varias alternativas. Por una parte, puede considerarse la focalización geográfica, en la que se busca identificar regiones con alta concentración de población con carencias de satisfactores básicos, bajo la premisa de que dirigir los recursos a determinadas zonas beneficiará a un mayor número de personas pertenecientes a la población objetivo. Por otra parte, el indicador puede referirse al nivel de los individuos o sus hogares, buscando identificar aquéllos que viven en condiciones precarias.

La metodología de identificación de los programas sociales generalmente utiliza un proceso de direccionalidad en estos dos niveles. Como primer paso, se sigue una selección geográfica identificando localidades con grandes carencias socioeconómicas, con base en un índice de marginación bajo la premisa subyacente de que en ellas se concentra la mayor proporción de hogares con carencias básicas. El segundo paso consiste en evaluar la condición de bienestar de los hogares en las localidades seleccionadas, a partir de indicadores recolectados directamente en entrevista en los hogares. El método de estimación de la condición de bienestar establece los criterios que ponen en correspondencia una métrica del ingreso de las familias y una métrica de las características socioeconómicas de sus miembros, de las características de sus viviendas y de los bienes que poseen.

Los análisis que siguen un enfoque multidimensional para caracterizar el nivel socioeconómico familiar, requieren de tres tipos de decisiones:

- d) Las dimensiones a incorporar
- e) La ponderación de cada dimensión
- f) El umbral a partir del cual se considera que un hogar o individuo presenta la carencia de cada dimensión de bienestar

Para el presente proyecto, se plantea el uso de una de las herramientas metodológicas más utilizadas para identificar las dimensiones de bienestar relevantes a considerar para los propósitos del Sistema de Protección Social en Salud (SPSS) y en particular de un beneficio como el Seguro Popular. Una ventaja de dicha metodología es que además de permitir la identificación de variables, arroja también estimaciones sobre la ponderación de cada dimensión. En cuanto a los umbrales, la técnica deberá de hacer posible la clasificación de la población en 10 grupos, por

deciles de hogares. Esta constituye una clasificación relativa de umbrales comúnmente utilizada en programas sociales a lo largo de América Latina.

En el contexto del presente proyecto, la utilidad del análisis discriminante es que permite identificar los espacios que definen dos condiciones, la de pertenecer al grupo de potencial beneficiario de un programa social, y la de no serlo, a partir de arreglos multidimensionales que caracterizan estas condiciones, mediante valores típicos de las variables utilizadas. Partiendo de la información del ingreso y comparándolo con un punto de referencia establecido, se dispone de una base para caracterizar los perfiles que distinguen a los hogares con mayor y menor bienestar mediante un arreglo que incluye diversos indicadores. En el caso particular del SPSS la metodología permite por lo tanto ubicar a cada uno de los potenciales derecho habientes en un segmento específico de la distribución del ingreso, y por lo tanto, permite también establecer una cuota de cotización (la cual es igual a cero en los primeros cuatro deciles).

La metodología de análisis discriminante es una herramienta que cuenta con los elementos y la robustez para constituirse como mecanismo de focalización para el SPSS y por eso se contempla como principal alternativa metodológica. En el caso del proyecto actual, el contar con un mecanismo de focalización es esencial para evitar errores de inclusión (es decir, catalogar a hogares que no pertenecen a los primeros cuatro deciles de medición del bienestar) o de exclusión (que consiste en clasificar erróneamente a hogares que pertenecen a los primeros cuatro deciles de medición del bienestar, como si pertenecieran a grupos de mayor nivel de bienestar en deciles superiores).

La importancia de evitar errores de inclusión y exclusión en el marco del SPSS radica en que un error de inclusión eximiría del pago del servicio a hogares que están en condición de realizar alguna aportación, mientras que el error de exclusión llevaría a exigir una cuota de pago a hogares a los que no les correspondería realizar una aportación debido a que cuentan con las condiciones para clasificarse (bajo una metodología con mayor precisión) a los primeros cuatro deciles de nivel de bienestar.

2.2 Análisis conceptual y técnico-estadístico de las variables y de la medición que contiene la CECASOEH

El primer paso del presente proyecto consiste en evaluar a través de un análisis conceptual y técnico-estadístico el proceso de clasificación de los beneficiarios del Sistema de Protección Social en Salud (SPSS) con el fin de determinar si el proceso actual en cuanto a su instrumento y algoritmo es válido y confiable y, en su caso, proponer alternativas que permitan mejorar dicha clasificación.

Descripción general

Para realizar un análisis conceptual primero es necesario contar con un entendimiento general del objetivo de la clasificación de los beneficiarios del SPSS. Desde el momento de la concepción del Seguro Popular, se determinó que el financiamiento del mismo se basaría en un modelo de contribución tripartita, a manera de símil del esquema de financiamiento existente en la Seguridad Social. Las contribuciones serían a cargo del Gobierno Federal a través de la cuota social (CS) y la aportación solidaria federal (ASF); los gobiernos estatales con la aportación solidaria estatal (ASE); y las familias, a través de la cuota familiar (CF) que representa una cuota progresiva dependiendo de la clasificación de la familia a estrato o “decil” de ingresos o nivel socio-económico. Dicho decil es asignado en el momento de su afiliación al Seguro Popular por medio de la aplicación de un cuestionario conocido como la Cédula de Características Socioeconómicas del Hogar (CECASOEH) cuya información sirve de insumo a un algoritmo desarrollado para la asignación a deciles. (SSA, 2006).

La motivación implícita de incluir en el esquema de financiamiento una cuota familiar, radicaba en crear una cultura de la corresponsabilidad de modo que la población cuente con mayores elementos para exigir derechos y adquirir responsabilidad en las obligaciones; crear una cultura de pago anticipado como forma de reducir el gasto de bolsillo; y fomentar una cultura de atención preventiva para mantener la salud y reducir riesgos de enfermedades; en otras palabras, su objetivo es generar una corresponsabilidad de las familias en el cuidado y en la atención de su propia salud al ser partícipes del financiamiento de su aseguramiento en salud (SSA, 2006; Rubalcava, L, Vargas M, et al, 2011).

Asimismo, aquellas familias asignadas a los primeros dos deciles de ingreso, con base en el instrumento y metodología utilizados, no debían pagar ninguna cuota siendo considerados dentro de un “régimen no contributivo”, mientras que aquellas familias de los deciles 3 al 10 pagaban una

cuota progresiva dependiendo del decil, y con base en un modelo de asignación de cuotas, siendo incluidos así dentro del “régimen contributivo” (SSA, 2006).

A partir del diseño inicial del Seguro Popular y con la evolución e implementación de éste durante más de 10 años, se han presentado algunos cambios relevantes en este proceso. A partir de la implementación del programa Seguro Médico para una Nueva Generación en 2007 (ahora Seguro Médico Siglo XXI) y posteriormente de la Estrategia de Embarazo Saludable, se han ido modificando los límites entre los regímenes no contributivo y contributivo. Dentro del primero, entran automáticamente también aquellas familias que cuenten con algún niño menor de 5 años o con una mujer embarazada aun cuando se encuentren entre los deciles 4 al 7 así como los beneficiarios de los programas de combate a la pobreza extrema del Gobierno Federal (Prospera principalmente), residentes en localidades rurales, los beneficiarios que residan en localidades de muy alta marginación con menos de doscientos cincuenta habitantes y, las que determine la Comisión Nacional de Protección Social en Salud (Reglamento de la Ley General de Salud en materia de Protección Social en Salud, artículo 127). Es así que en la actualidad y a partir de junio de 2010, el régimen no contributivo incluye a todos aquellos individuos clasificados entre los deciles 1 y 4 más aquellos que entran en los supuestos anteriormente mencionados; mientras que el régimen contributivo lo constituyen aquellos clasificados entre los deciles 5 y 10 con las excepciones ya mencionadas (CNPSS, SSA, 2014). Parte de la justificación de la ampliación del régimen no contributivo radica en la observación de las características de las familias afiliadas al SPSS y sobre todo por los costos implícitos en el proceso de cobro de las cuotas familiares por parte de los REPSS para aquellos del régimen contributivo (Rubalcava, L, Vargas M, et al, 2011).

Por otro lado, cabe destacar que la bolsa de recursos del SPSS está financiada en 5/6 por el Gobierno Federal (a través de la CS y la ASF) y 1/6 por los gobiernos estatales (a través de la ASE). La cuota familiar no se contempla dentro de la bolsa sino que es recolectada directamente por los Regímenes de Protección Social en Salud (REPSS) en el momento de afiliación y administrado directamente por ellos para los gastos que determinen las disposiciones reglamentarias de la ley y que consideren sean necesarios para el SPSS; si bien sigue manteniendo el objetivo original e implícito de ser una herramienta de corresponsabilidad del usuario en el financiamiento de su aseguramiento en salud y por ende en la atención y cuidado de su salud (Rubalcava, L, Vargas M, et al, 2011).

Al cierre de junio de 2014, el SPSS contaba con 55,601,952 afiliados, los cuales están clasificados en los deciles de ingreso de la siguiente forma:

Cuadro 1. Afiliados al SPSS por decil de ingreso (a junio de 2014, nacional)

Decil de ingreso	Número total de afiliados	% de afiliados en el decil respecto al total
1	34,151,759	61.4%
2	16,916,144	30.4%
3	3,472,177	6.2%
4	882,430	1.6%
5 a 10	179,442	0.3%
TOTAL	55,601,952	100%

Fuente: CNPSS, Secretaría de Salud. Junio 2014.

El esquema de cuotas actual es el siguiente:

Cuadro 2. Cuota Familiar según decil de ingreso (2014)

Decil de ingreso	Cuota anual por familia (pesos)
1	0
2	0
3	0
4	0
5	2,075.0
6	2,833.6
7	3,647.9
8	5,650.4
9	7,519.0
10	11,378.9

Fuente: CNPSS, Secretaría de Salud.

El instrumento de recolección de información sobre los beneficiarios en el momento de la afiliación en el Módulo de Afiliación y Orientación (MAO) sigue siendo la CECASOEH y el algoritmo original para el uso de dicha información con el fin de clasificar a los beneficiarios en deciles y por ende solicitarles el pago de cuotas específicas (o la exención del pago), sigue siendo el mismo y se cuenta con poca información que permita documentar cómo fueron construidos exactamente dicho instrumento y particularmente dicho algoritmo. Igualmente cabe destacar, que aun cuando la unidad de afiliación del SPSS siguió siendo la familia, el financiamiento cambió en el 2010 para tomar como referencia al individuo. La cuota familiar y el modelo de clasificación asociado a ésta, sin embargo, persistió como tal con las reglas originales (Lineamientos para la afiliación, operación, integración del Padrón Nacional de Beneficiarios y determinación de la

Cuota Familiar del Sistema de Protección Social en Salud, CNPSS, SSA, 2005), cabe destacar que como consecuencia, el modelo de identificación de nivel socioeconómico permite identificar a cada individuo en el hogar correspondiente, y es el hogar el que se ubica en la distribución del ingreso.

El hecho de no contar con un entendimiento claro de por qué se utilizan determinadas variables y no otras y, en particular, qué metodología utiliza el algoritmo diseñado e implementado para ponderar dichas variables y combinarlas a manera de generar una clasificación por deciles (sólo se tienen los puntajes); aunado a los resultados de evaluaciones recientes sobre el padrón de beneficiarios; a la discrepancia que existe con otras encuestas respecto a los resultados de la asignación de la población a deciles de ingreso; al hecho de que el 99.6% de los beneficiarios del SPSS pertenecen ya al régimen no contributivo (92.4% clasificados en los deciles 1 y 2); y a la existencia de retos operativos que conlleva la aplicación del proceso y en particular el cobro de cuotas por parte de los REPSS; justifican la necesidad de evaluar el esquema actual de asignación y, tomando en cuenta el contexto actual del SPSS, proponer instrumentos y/o métodos alternativos para la clasificación de sus beneficiarios a partir de lo encontrado en el análisis.

En este sentido, cabe destacar también cuáles son los objetivos primordiales de realizar un levantamiento de información de los afiliados en el momento de su afiliación y re-afiliación así como del proceso de clasificación de los mismos. Si bien el objetivo principal sigue siendo contar con una herramienta confiable y viable de clasificación de los beneficiarios a categorías o estratos de nivel socio-económico (deciles) con la finalidad de establecer a qué régimen deben pertenecer y por ende cuál debe ser su nivel de corresponsabilidad en el financiamiento de este esquema de aseguramiento ligado a su capacidad de pago (como en cualquier esquema de aseguramiento público o privado); este ejercicio también brinda una oportunidad única de contar con información objetiva, válida y confiable que permita caracterizar correctamente el perfil de los beneficiarios del SPSS para mejorar la implementación del mismo y optimizar la toma de decisiones continua, entre otros. Por ello, resulta de gran relevancia el análisis y, en su caso, el mejoramiento del instrumento y/o de la metodología actual.

2.3 Análisis conceptual

Descripción del instrumento y algoritmo

La Cédula de Características Socio-económicas de los Hogares (CECASOEH) es el instrumento de recolección de información para la posterior clasificación de los beneficiarios a estratos socio-económicos. Oficialmente está definida como el “instrumento estandarizado elaborado por la Secretaría de Salud que es aplicado por los Regímenes Estatales de Protección Social en Salud para la determinación del nivel socioeconómico de la familia que solicita su afiliación al Sistema de Protección Social en Salud y sirve para la identificación del núcleo familiar y sus integrantes, así como para la determinación de la cuota familiar correspondiente” (Lineamientos para la Afiliación, Operación, Integración del Padrón Nacional de Beneficiarios y Determinación de la Cuota Familiar del Sistema de Protección Social en Salud, CNPSS, SSA).

En cuanto a su aplicación ésta se realiza in situ en el MAO durante el proceso de afiliación y reafiliación sin posterior verificación a nivel del hogar del individuo. La ventaja de este esquema de implementación es que es sencillo y costo-efectivo en recursos y en tiempo de aplicación; la posible desventaja es que depende de información auto-reportada por el individuo (quien tiene incentivos a sub-reportar su estatus socio-económico para entrar en el régimen no contributivo) que no es verificada con un trabajo de campo en los hogares (como se realizaría con otros instrumentos utilizados para la clasificación de población, por ejemplo, la ENCASEH de Prospera; es importante destacar sin embargo, que en el caso de la ENCASEH existe un procedimiento de selección y focalización en dos etapas, en las que en primera instancia se pre seleccionan áreas geográficas con mayor concentración de población objetivo potencial, lo cual se esperaría que incrementara su grado de acierto. Esta deficiencia, sin embargo, puede tener efectos mínimos sobre el resultado de la clasificación si el instrumento de recolección de información utilizado está adecuadamente diseñado en cuanto a las dimensiones o conceptos a incluir, las variables para caracterizar éstas y sobre todo la forma de preguntar y obtener la información específica sobre dichas variables.

De igual manera, el uso de una metodología robusta para la ponderación de las variables y su clasificación, puede minimizar este efecto. Adicionalmente, cabe destacar que en el caso de los beneficios de salud a los que se accede por medio de la afiliación o re-afiliación, puede existir el incentivo adicional a aportar información veraz ante el riesgo de que quedar excluidos del servicio en caso de verificarse el reporte incorrecto de datos.

En cuanto a su composición, la CECASOEH recaba la siguiente información directa sobre el hogar:

- Composición del hogar
- Características de cada miembro del hogar: edad, estado civil, entidad de origen, sexo, si hay discapacidad, escolaridad, estatus laboral, si hay embarazo, entre otros
- Beneficios de programas sociales (e.g. si es beneficiario de Prospera)
- Condiciones de la vivienda
- Activos en la vivienda

Como resultado, a partir de la CECASOEH se obtienen el siguiente conjunto de variables que se utilizan para el proceso de asignación de nivel socio-económico o decil para los beneficiarios:

Cuadro 3. Variables del algoritmo de la CECASOEH

Variable	Descripción	Medición
Rural	Tipo de localidad de residencia	1 rural 0 urbano
Piso	Material del piso de la vivienda	1 tiene piso de tierra 0 algún otro material
Baño	Tenencia de servicio sanitario en la vivienda	1 no cuenta con servicio sanitario 0 tiene excusado, sanitario, retrete, fosa, letrina u hoyo negro
Escolar 1	Jefe del hogar sin educación formal	1 0 años de escolaridad 0
Escolar 2	Jefe del hogar sin primaria completa	1 1-5 años de escolaridad 0
Escolar 3	Jefe del hogar sin secundaria completa	1 6-8 años de escolaridad 0
Edad	Edad del Jefe del Hogar	0 a 97 años de edad
Depdem	Índice de dependencia demográfica	(Personas entre 0 y 14 años + personas de 65 años y más)/(resto de las personas en el hogar)
Hacina	Índice de hacinamiento	(Total de personas)/(número de cuartos sin contar la cocina)
VHSDVD	Tenencia de videocassetera o DVD	1 No tiene 0 Si tiene
Estufa	Tenencia de estufa de gas	1 No tiene 0 Si tiene
Refri	Tenencia de Refrigerador	1 No tiene 0 Si tiene
Lava	Tenencia de Lavadora	1 No tiene 0 Si tiene
Vehi	Tenencia de Vehículo	1 No tiene 0 Si tiene
Sexoj	Sexo del jefe del hogar	1 Mujer 0 Hombre
SS	Algún miembro en la vivienda cuenta con seguridad social	1 No cuenta con seguridad social 0 Cuenta con seguridad social
Noasisten	Niños entre 5 y 15 años que no asisten a la escuela	Total de niños
Trabajan	Niños entre 12 y 15 años que trabajan	Total de niños
Niños	Niños menores de 12 años	Total de niños

Fuente: Elaboración propia con información proporcionada por la CNPSS

En cuanto al modelo, se cuenta con un algoritmo que utiliza los siguientes puntajes o ponderadores para cada variable separando por estrato urbano y rural, en donde el criterio de clasificación de la población por área geográfica se toma del INEGI:

Cuadro 4. Ponderaciones de variables por estrato

Variable	URBANO	RURAL
	Puntaje	Puntaje
Localidad de residencia	* 0.000	*0.687
Material de piso	* 0.81	* 0.454
Servicio Sanitario	* 0.587	* 0.067
Escuela 1 Jefe del hogar	* 0.621	* 0.46
Escuela Primaria	* 0.302	* 0.266
Escuela Secundaria	* 0.000	* 0.005
Edad del Jefe del hogar	* 0.018	* 0.012
Dependencia demográfica	* 0.238	* 0.15
Hacinamiento	* 0.179	* 0.119
Video casetera o DVD	* 0.151	* 0.091
Estufa	* 0.741	* 0.741
Refrigerador	* 0.948	* 0.625
Lavadora	* 0.096	* 0.141
Vehículo	* 0.248	* 0.246
Sexo del Jefe de hogar	* 0.126	* 0.121
Seguridad Social	* 0.447	* 0.388
No asistencia	* 0.49	* 0.295
Trabajan niños	* 0.39	* 0.009
Total de niños	* 0.32	* 0.268

Fuente: Elaboración propia con información proporcionada por la CNPSS

Comparación conceptual con estándares

Para analizar conceptualmente el esquema actual en términos de sus dimensiones, variables utilizadas y la ponderación de éstas (su metodología) es necesario referirse a otros estándares o estrategias de clasificación de individuos a estratos socio-económicos específicos a manera de referencia (o estándar de oro) comparativa. Se utilizarán como referentes el Índice de Marginación del Consejo Nacional de Población (CONAPO); la Metodología para la Medición de la Pobreza Multidimensional en México (CONEVAL); y el Índice de Desarrollo Humano (ONU).

Índice de Marginación

El indicador más comúnmente utilizado en México para medir el bienestar ha sido el Índice de Marginación (IM). El Índice de Marginación ha sido ampliamente utilizado como herramienta de diagnóstico y como insumo para el diseño de políticas y programas específicos (e.g. la estrategia de Microrregiones utilizada en su momento por la Secretaría de Desarrollo Social).

El IM es una medida resumen que permite estratificar jerárquicamente unidades territoriales, como son las entidades federativas, los municipios y las localidades del país, según el impacto global de distintas carencias que enfrenta la población.

El índice de marginación considera cuatro dimensiones de la marginación mediante la identificación de nueve factores de exclusión o carencia de oportunidades. La agregación de estos factores en una variable resumen se realizó mediante el Análisis de Componentes Principales, que permite transformar un conjunto de variables o indicadores en uno nuevo y facilita una interpretación más sencilla del fenómeno original al reducir el análisis a un menor número de variables. Con este método se proyecta el espacio definido por los nueve indicadores sobre un espacio unidimensional.¹ Para el cálculo del IM se realizaron pruebas sobre los componentes principales y se concluyó la pertinencia de sólo tomar en cuenta el primero de ellos para conformar el índice de marginación.²

Una vez que se estimaron los coeficientes que ponderan cada una de las variables estandarizadas para obtener la primera componente principal, se obtuvo el índice de marginación como una combinación lineal de los indicadores estandarizados. Este índice conlleva a una ordenación robusta de las unidades en estudio (estados o municipios), ya que se construye en un intervalo, cualidad que a su vez permite llevar a cabo una estratificación de las unidades objeto de estudio en conjuntos homogéneos de acuerdo con la síntesis de los nueve factores de exclusión social. La técnica empleada para llevar a cabo la estratificación fue la desarrollada por Dalenius y Hodges, y se dividió el rango del Índice de Marginación en 5 subconjuntos.

¹ La base del análisis por componentes principales es el coeficiente de correlación lineal o la covarianza. A partir de los valores que toman estas medidas de asociación cuando se consideran todas las parejas de indicadores involucrados en el estudio es posible definir nuevos indicadores resumen denominados componentes principales. La primera componente principal es el indicador resumen que explica la mayor heterogeneidad entre los casos; es decir, tiene la mayor varianza. La segunda componente es el indicador resumen que ocupa el segundo lugar en heterogeneidad y que no está correlacionado con el primero y así sucesivamente.

² El primer criterio empleado fue el de seleccionar aquellos componentes principales cuyo valor propio sea superior al promedio de la varianza de los datos estandarizados. El segundo criterio fue la elaboración de una gráfica de sedimentación.

Los nueve indicadores de exclusión considerados dentro del índice de marginación son:

- % Población analfabeta de 15 años o más,
- % Población sin primaria completa de 15 años o más,
- % Ocupantes en viviendas sin drenaje ni servicio sanitario exclusivo,
- % Ocupantes en viviendas sin energía eléctrica,
- % Ocupantes en viviendas sin agua entubada,
- % Viviendas con algún nivel de hacinamiento,
- % Ocupantes en viviendas con piso de tierra,
- % Población en localidades con menos de 5 000 habitantes,
- % Población ocupada con ingreso de hasta 2 salarios mínimos.

El índice de marginación ha sido calculado para estados, municipios y localidades utilizando los datos tanto de los Censos como de los Conteos de Población. Es importante subrayar que el ordenamiento de las unidades de análisis se ha establecido con relación al conjunto de las mismas. Esto significa que un municipio que puede tener mayores rezagos al interior de su entidad federativa, puede no ser clasificado en niveles de muy alta marginación cuando se le compara con municipios mucho más rezagados de otros estados del país.

El IM funge como un buen parámetro de comparación para el ejercicio de asignación de estrato socio-económico del SPSS dado que considera a toda la población (toda la distribución, no solo aquellos debajo de un estándar o línea como en el caso de la pobreza o programas sociales asociados a la misma); y se centra prioritariamente en variables observables y básicas para caracterizar el estrato socio-económico de toda la población. Si bien su ámbito es de orden territorial.

En cuanto a las variables, en términos comparativos, si bien el IM es más exhaustivo en el uso de variables que caracterizan la vivienda y sí utiliza una variable de clasificación de ingresos como es el salario mínimo; las dimensiones contempladas y tipo de variables para captar información socio-económica son relativamente similares y adhoc a los objetivos que busca cada instrumento. En cuanto a las metodologías, asumimos que son distintas dado el desconocimiento que se tiene de aquella utilizada en el proceso de clasificación del SPSS. En este sentido, el IM puede aportar algunos elementos a considerar en un proceso alternativo de clasificación para el SPSS.

Metodología para la Medición de la Pobreza Multidimensional

A partir de 2008, la medición de la pobreza en México se realiza con un modelo multidimensional. Este enfoque combina la clasificación de individuos como pobres en base a su ingreso respecto a

una línea de bienestar de referencia con la clasificación de individuos como pobres en base a carencias en otras dimensiones asociadas a derechos sociales como son: educación, salud, seguridad social, vivienda, servicios básicos y alimentación. Este enfoque permite considerar una gama más amplia de dimensiones (no necesariamente estimables con criterios monetarios) – no sólo el ingreso- caracterizables a través de distintas variables y a las que se les asigna igual peso ponderado en la construcción de un índice (índice de privación) o conteo de carencias general. Asimismo, en el orden territorial considera también las dimensiones o conceptos de cohesión social y acceso a infraestructura social básica.

En este sentido, aun cuando la metodología está enfocada en la clasificación de pobres monetarios y de pobres en carencias sociales, permite clasificar a toda la población en las distintas dimensiones y ampliar el espectro de las variables disponibles, obtenibles a través de instrumentos de recolección básicos, para clasificación de individuos en estratos. Es por ello, que funge como referente, en términos de variables y metodológicamente, para el proceso de clasificación del SPSS. Si bien el proceso del SPSS ya considera un conjunto importante y base de estas variables, podría tomar en cuenta otras variables aquí utilizadas que junto con una metodología de ponderación distinta pudieran ser una alternativa para mejorar la clasificación del SPSS.

Las variables utilizadas para la medición multidimensional de la pobreza en México incluyen:

- Ingreso corriente per cápita
- Población de 3 a 15 años que no cuenta con la educación secundaria terminada y no asiste a la escuela.
- Población nacida hasta 1981 que no cuenta con la educación primaria terminada.
- Población nacida a partir de 1982 que no cuenta con la educación secundaria terminada.
- Población no afiliada o inscrita para recibir servicios de salud por parte del Seguro Popular, y que tampoco es derechohabiente de alguna institución de seguridad social.
- Población ocupada y asalariada que no recibe por parte de su trabajo las prestaciones de servicios médicos, incapacidad con goce de sueldo y SAR o Afore.
- Población ocupada y no asalariada que no recibe como prestación laboral o por contratación propia servicios médicos por parte de una institución pública de Seguridad Social y SAR o Afore.
- Población sin acceso a la seguridad social por alguno de los primeros dos criterios, que no goce de alguna jubilación o pensión, ni sea familiar directo de una persona dentro o fuera del hogar con acceso a la seguridad social.
- Población de 65 años o más que no dispone de acceso a la seguridad social por alguno de los criterios anteriores ni es beneficiario de algún programa social de pensiones para adultos mayores.
- Población que reside en viviendas donde el material de la mayor parte de los pisos es tierra.

- Población que reside en viviendas donde el material de la mayor parte del techo es lámina de cartón o desechos.
- Población que reside en viviendas donde el material de la mayor parte de los muros es barro o bajareque; carrizo, bambú o palma; lámina de cartón, metálica o de asbesto; o material de desecho.
- Población que reside en viviendas con hacinamiento.
- Población que reside en viviendas donde 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.
- Población que reside en viviendas que no disponen de drenaje, o el desagüe tiene conexión a una tubería que va a dar a un río, lago, mar, barranca o grieta.
- Población que reside en viviendas que no disponen de energía eléctrica.
- Población en hogares con un grado de inseguridad alimentaria moderado o severo.
- Medidas de polarización social, redes sociales o de desigualdad económica.
- Población en localidades que se encuentran a una distancia mayor a tres kilómetros de un camino pavimentado.
- Población en localidades sin acceso a alcantarillado y electrificación pública.
- Población en localidades sin acceso a rellenos sanitarios públicos o algún servicio público de eliminación de basura.

Índice de Desarrollo Humano

El Índice de Desarrollo Humano es un indicador del nivel de desarrollo de una población. Aun cuando no es un indicador que permita clasificar a la población en distintos grupos o caracterizarla de manera detallada, sí permite obtener una medición agregada multidimensional que es además sencillamente comparable.

El IDH puede ser usado como referencia en base a la pertinencia de su enfoque: selecciona dimensiones específicas (nivel de vida decoroso, vida larga y saludable y conocimiento); identifica las variables para caracterizar cada dimensión (e.g. PIB per cápita, esperanza de vida, tasa de inscripción, tasa de alfabetismo de adultos, años de escolaridad, etc); y las pondera con igual peso relativo para obtener un índice final comparable.

De igual manera este enfoque de selección de dimensiones, variables y ponderación, se puede extender para la construcción de otros índices agregados similares como: el Índice de Desarrollo Humano Ampliado (IDHA), el Índice de Pobreza Humana (IPH), el Índice de Desarrollo Relativo al Género (IDG), entre otros.

De estos índices se pueden obtener más variables de referencia para el SPSS que favorezcan la clasificación de la población en estratos socio-económicos considerando distintas dimensiones desde el ingreso hasta derechos sociales básicos.

Entre las variables utilizadas para la estimación de estos índices, destacan:

- Esperanza de vida al nacer
- Tasa de alfabetización de adultos (se considera a las personas mayores a 15 años de edad)
- Tasa de matriculación escolar (se considera a las personas entre 6 y 24 años de edad)
- PIB per cápita real en dólares PPC
- Porcentaje de personas que se estima no sobrevivirán hasta la edad de 40 años
- Porcentaje de adultos analfabetas
- Porcentaje de personas sin acceso a agua potable
- Porcentaje de niños menores de 5 años de edad con peso insuficiente
- Porcentaje de personas sin acceso a servicios de salud
- Salarios no agrícolas
- Porcentaje de población económicamente activa

Para el caso de México, existen distintos ejercicios para la estimación del IDH a nivel nacional, estatal y municipal.

2.4 Análisis Técnico-Estadístico

Con el fin de realizar un análisis técnico-estadístico que permita emitir conclusiones sobre la precisión y confiabilidad del instrumento y metodología de clasificación actual de beneficiarios del SPSS, se realizaron varios ejercicios que se describen a continuación.

Análisis con la ENIGH

La Encuesta Nacional de Ingreso Gasto de los Hogares (ENIGH) es la fuente más fidedigna con la que se cuenta para obtener información sobre el ingreso de toda la distribución de población y por ende, poder clasificarla en estratos –deciles. Los datos de la ENIGH 2012 muestran de manera general lo siguiente:

Cuadro 5. Distribución de hogares por estrato rural-urbano

DISTRIBUCION DE HOGARES POR ESTRATO		
Urbano	24,487,689	78.09
Rural	6,871,873	21.91
TOTAL	31,359,562	100

Fuente: Elaboración propia con datos de la ENIGH 2012

Cuadro 6. Distribución de hogares por decil

DISTRIBUCION DE HOGARES POR DECIL			
Decil	Urbano	Rural	Total
1	5.9	24.6	10
2	7.6	18.5	10
3	8.9	14.0	10
4	9.6	11.3	10
5	10.4	8.7	10
6	10.8	7.2	10
7	11.3	5.4	10
8	11.6	4.3	10
9	11.8	3.6	10
10	12.2	2.3	10
Total	100.0	100.0	100

Fuente: Elaboración propia con datos de la ENIGH 2012

El 78.09% de la población se clasifica como urbana, mientras que el 21.91% como rural. Para cada estrato (rural-urbano) se muestra la distribución por deciles (e.g. el 24.6% de la población rural pertenece al primer decil; el 12.2% de la población urbana pertenece al décimo decil).

De acuerdo con el contexto del SPSS y los límites establecidos en los regímenes no contributivo y contributivo, se presenta también la información agrupándola de la siguiente forma: deciles 1 al 4, deciles 5 y 6, deciles 7 y 8, deciles 9 y 10. Esto permite tener un panorama de análisis masadhoc al contexto clasificatorio actual y los resultados esperados de éste (donde se espera que se clasifique con mayor precisión a los primeros deciles en comparación con los últimos). Los datos para la ENIGH por estrato agrupados de esta forma muestran la siguiente distribución:

**Cuadro 7. Distribución de hogares por decil
(porcentajes)**

Decil	Urbano	Rural	Total
1-4	32.0	68.5	40
5-6	21.1	15.9	20
7-8	22.9	9.7	20
9-10	24.0	5.9	20
Total	100.0	100.0	100

Fuente: Elaboración propia con datos de la ENIGH 2012

Aplicación de metodología actual de clasificación del SPSS con datos de la ENIGH

Con el objetivo de comparar que tan preciso es el modelo de clasificación actual del SPSS, con respecto a la clasificación por deciles que hace la ENIGH, se realizó el siguiente ejercicio:

Se utilizó como referencia la clasificación por deciles de la ENIGH y los datos de la encuesta, separando inicialmente los estratos rural y urbano y posteriormente obteniendo estimaciones nacionales. Con los datos de la ENIGH se estimaron las variables que utiliza el modelo actual del SPSS y se procedió a clasificar a los individuos utilizando también los puntajes/ponderadores especificados por el algoritmo del SPSS. Finalmente, se compara la predicción del modelo de clasificación del SPSS con la ubicación del hogar de acuerdo a la ENIGH. Esto permite evaluar la precisión de asignación del modelo del SPSS tomando como referente la asignación por deciles de la ENIGH, según se muestra en el siguiente cuadro:

Cuadro 8. Comparación entre distribución por deciles de la ENIGH y distribución por deciles al aplicar el modelo del SPSS sobre la ENIGH, nacional

Deciles ENIGH	Deciles PUNTAJE									
	1	2	3	4	5	6	7	8	9	10
1	25.7	28.3	17.1	13.3	6.5	4.4	2.3	1.6	0.6	0.3
2	15.9	28.7	21.1	16.6	7.5	5.4	2.2	2.1	0.5	0.1
3	11.3	24.8	22.2	17.9	9.3	7.3	3.4	3.0	0.6	0.2
4	8.0	20.8	22.5	18.4	11.3	9.7	4.8	3.1	1.1	0.3
5	6.0	16.4	21.4	19.4	12.6	11.2	6.1	4.7	1.6	0.6
6	4.2	13.5	19.6	19.6	16.3	12.6	7.0	4.8	1.9	0.5
7	2.7	10.5	17.0	19.1	16.5	13.9	9.0	7.3	2.9	1.0
8	1.5	7.2	15.9	17.8	16.8	16.2	10.4	9.0	4.0	1.2
9	0.7	5.4	12.3	17.4	17.8	18.6	9.9	10.6	5.3	2.0
10	0.3	3.3	8.4	13.0	15.7	19.5	14.0	14.5	7.7	3.7
Total	7.6	15.9	17.8	17.3	13.0	11.9	0.0	6.1	2.6	1.0

Fuente: Elaboración propia con datos de la ENIGH 2012 y modelo de clasificación del SPSS

El cuadro muestra lo siguiente: de aquellos hogares clasificados por la ENIGH en el decil 1, el 25.7% son también clasificados por el modelo del SPSS en el decil 1, mientras que el 28.3% son clasificados en decil 2, el 17.1% en el decil 3 y así sucesivamente.

La diagonal muestra entonces la coincidencia entre la clasificación de la ENIGH y la del modelo del SPSS usando las variables y algoritmos del modelo y aplicándolo a datos de la ENIGH. Como se observa, el modelo es mejor predictor para los primeros deciles que para los últimos.

Dado que por definición cada decil incluye a un 10% de la población, el 25.7% de predicción correcta del modelo de SPSS para el decil 1 implica que se asigna correctamente a este decil a un 2.57% de la población; al decil 2 a un 2.83% de la población; al decil 3 a un 1.71% de la población y así sucesivamente. Si sumamos estos porcentajes para todos los deciles concluimos que el modelo de clasificación del SPSS asigna correctamente al 14.4% de los hogares. Esta medida sirve como referencia general de la precisión del modelo para predecir o asignar deciles.

Si utilizamos la clasificación agregada, adhoc al contexto del SPSS, se observa lo siguiente:

Cuadro 9. Comparación entre distribución por grupos de deciles de la ENIGH y distribución por grupos de deciles al aplicar el modelo del SPSS sobre la ENIGH, nacional

Deciles ENIGH	Deciles PUNTAJE			
	1-4	5-6	7-8	9-10
1-4	78.1	15.2	5.7	1.0
5-6	60.1	25.6	11.8	2.5
7-8	48.7	30.7	16.5	4.2
9-10	34.3	34.7	23.0	8.0

Fuente: Elaboración propia con datos de la ENIGH 2012 y modelo de clasificación del SPSS

Según muestra en el cuadro 9, la coincidencia para el grupo de los primero cuatro deciles es de 78%, lo que confirma que el modelo del SPSS es un buen predictor para la parte baja de la distribución y los que tiene clasificados dentro del régimen no contributivo. Sin embargo, para los deciles intermedios baja el grado de coincidencia y para los últimos deciles, la predicción es bastante errada. En general, este modelo logra asignar con precisión al 41.25% de los hogares.

Estos ejercicios se pueden analizar también separando por estrato rural³ y urbano para alcanzar un mejor entendimiento de la efectividad del modelo.

Cuadro 10. Comparación entre distribución por grupos de deciles de la ENIGH y distribución por grupos de deciles al aplicar el modelo del SPSS sobre la ENIGH, rural

Deciles ENIGH	Deciles PUNTAJE			
	1	2	3	4
1-4	93.9	5.94	0.15	
5-6	88.84	11.0	0.2	
7-8	82	17	0.8	
9-10	72.54	24.39	3.07	

Fuente: Elaboración propia con datos de la ENIGH 2012 y modelo de clasificación del SPSS

³Cabe destacar, que en el listado de puntajes del algoritmo de clasificación de SPSS para el estrato rural se le asigna un puntaje a la variable localidad de residencia. Se requeriría aclarar por qué se presenta esto en el diseño del modelo.

Cuadro 11. Comparación entre distribución por grupos de deciles de la ENIGH y distribución por grupos de deciles al aplicar el modelo del SPSS sobre la ENIGH, urbano

Deciles ENIGH	Deciles PUNTAJE			
	1-4	2	3	4
1-4	66.8	10.3	12.56	10.39
5-6	43.98	13.8	20.8	21.41
7-8	30	12	27.2	31
9-10	12.31	6.24	18.5	63.0

Fuente: elaboración propia con datos de la ENIGH 2012 y modelo de clasificación del SPSS

En el caso rural, el modelo predice con mucha exactitud la clasificación a los primeros cuatro deciles de ingreso pero es muy deficiente en la clasificación del resto. Por su parte, en lo urbano, el modelo del SPSS es un mejor predictor en los extremos: deciles 1 al 4 y para los deciles 9 y 10, siendo más impreciso para los deciles del 5 al 8.

Con el fin de poder comparar este análisis con otras alternativas y proponer una estrategia de clasificación más precisa para todos los estratos se realizaron varios ejercicios adicionales:

- 1) Utilizando las mismas variables del modelo de clasificación del SPSS, se aplicó la metodología de análisis discriminante sobre los datos de la ENIGH para evaluar si mejora o no el grado de predicción en la clasificación respecto a la distribución por deciles de la ENIGH.
- 2) Utilizando algunas variantes en el conjunto de variables utilizadas y aplicando análisis discriminante sobre los datos de la ENIGH se plantea un modelo alternativo que pretende clasificar con mucha mayor precisión a los últimos deciles e incrementar el nivel de predicción general del modelo, es decir, asignar atinadamente a un porcentaje mayor de hogares.
- 3) Con este mismo fin, se evalúa también la aplicación combinada del modelo actual con el modelo alternativo como una opción viable para elevar el porcentaje de hogares asignados correctamente y por ende, el grado de predicción y precisión del modelo.

La metodología de análisis discriminante se describirá con detalle en el apartado 2 de este documento. Por su parte, los resultados de estos ejercicios se presentarán en el apartado 3 con el fin de plantear una propuesta de clasificación alternativa que combinando distintos modelos permita obtener la mejor clasificación posible para todos los grupos de deciles.

Análisis con otras fuentes

Además de la ENIGH, se intentó procesar un análisis similar con datos provenientes de otras fuentes. Para ello, se realizó el ejercicio de utilizar las variables y el algoritmo del modelo de clasificación del SPSS con datos de la Encuesta Nacional de Salud y Nutrición (ENSANUT) 2012. Sin embargo, cabe destacar que el ingreso del hogar no se puede obtener a partir de la ENSANUT, por lo que los responsables de esta encuesta realizaron un ejercicio de imputación para poder obtener la distribución con esta encuesta. De esta manera, la clasificación por deciles de la ENSANUT que se usará para la comparación no se obtuvo directamente, sino corresponde a la indicada en la encuesta. De igual manera, la variable que se refiere a la tenencia de videocasetera o DVD no se puede obtener por medio de la ENSANUT.

Cuadro 12. Distribución de hogares con base en la ENSANUT

Decil	Distribución ingreso per cápita de ENSANUT	Distribución aplicando modelo de SPSS	Porcentaje de coincidencia
1	3,916,074	3,537,739	-9.7%
2	2,697,283	3,486,531	29.3%
3	2,649,491	3,468,713	30.9%
4	2,611,208	3,455,785	32.3%
5	2,596,074	3,288,538	26.7%
6	2,895,501	3,153,983	8.9%
7	3,022,531	2,720,014	-10.0%
8	3,136,021	2,261,157	-27.9%
9	3,438,399	2,028,433	-41.0%
10	2,465,346	2,027,035	-17.8%
Total	29,427,928	29,427,928	-

Fuente: Elaboración propia con datos de la ENSANUT 2012 y modelo de clasificación del SPSS

Como se puede observar en el Cuadro 12, el primer decil y los cuatro últimos presentan un margen de error negativo, esto indica que con la distribución que se obtiene a partir del modelo de SPSS, un menor número de hogares se encuentra en estos deciles en comparación con el número de hogares que “deberían” de pertenecer a estos deciles. Por otro lado, a partir del segundo decil hasta el 6, el número de hogares es mayor con respecto a la distribución que se obtiene con el ingreso.

Al igual que con la ENIGH, se puede observar en general que el modelo de SPSS CECASOEH sobreestima para los menores deciles de ingreso, con excepción del decil 1, y subestima para los deciles más altos.

Asimismo, se consideró realizar este mismo ejercicio utilizando como fuente el Censo de Población y Vivienda; sin embargo, no es posible debido a que sólo se reporta el ingreso laboral y por ello no sería un parámetro válido.

Aun cuando este ejercicio adicional aporta a la discusión; se tomará como referente principal el ejercicio derivado de la ENIGH dada la utilidad, precisión en la información y pertinencia de esta encuesta para los objetivos de clasificación de beneficiarios a estratos socio-económicos.

Conclusiones generales del análisis

En resumen, a partir del análisis conceptual y técnico-estadístico de la CECASOEH se puede concluir lo siguiente:

- Al aplicar el modelo de clasificación del SPSS con datos de otras fuentes como la ENIGH y compararlo con la distribución que se obtiene a partir de esta encuesta, se puede concluir que el modelo es en general buen predictor para los deciles bajos (del 1 al 4), reduciendo su precisión para el resto de los deciles, particularmente para el 9 y el 10, donde el margen de error es muy amplio. Esto implica que el porcentaje de hogares asignado atinadamente sea menor al 50%.
- Considerando estratos rural y urbano por separado, en lo rural esta conclusión es todavía más acertada, dado que se predice con un alto grado de precisión para los primeros cuatro deciles rurales y sólo para estos. Por otra parte, para lo urbano, el modelo es un buen predictor para los primeros cuatro deciles y los dos últimos, lo cual puede servir como referente.
- En este sentido, el modelo no es descartable sino que combinado con un modelo alternativo puede seguir caracterizando correctamente a la parte baja de la distribución (deciles 1 al 4) aumentando también el grado de predicción para los deciles medios (5 al 8) y últimos (9 y 10). Esto permitiría aumentar el porcentaje de hogares asignados atinadamente significativamente para contar con un mejor modelo de predicción.
- Los análisis que se mostrarán en el apartado 3 de este documento debieran arrojar más luz sobre el entendimiento del modelo actual y la viabilidad de otras alternativas para optimizar el nivel de precisión del modelo.
- En términos de las variables, la CECASOEH brinda un punto de partida correcto al considerar variables de distintas dimensiones que se complementan y logran caracterizar el estrato socio-económico al no contar con indicadores directos viables de ingreso.

- En base a los análisis del apartado 3, se podrá concluir con mayor certeza si el conjunto de variables utilizado actualmente debe permanecer así o si se requiere incorporar algunas alternativas de variables adicionales que mejoren el nivel de precisión del modelo en general. En este caso, se podrían utilizar como referencia las variables usadas en otros índices como el IM, el IDH y la medición multidimensional de la pobreza y ya existentes y probadas en instrumentos como la ENIGH.
- En cuanto a la metodología se considera fundamental contar con la documentación de su diseño y aplicación, y se considera además que la clasificación del modelo podría mejorar, particularmente para los deciles medios y altos, si se utiliza un modelo robusto y probado como el análisis discriminante (según se describirá en los apartados 2 y 3).
- Aun cuando la recolección de información es in situ, auto-reportada y no verificada en el hogar (siendo así más costo-efectiva), el contar con una selección de variables adecuada (que podría ser la actual con variantes) junto con una metodología robusta como el análisis discriminante, podrían minimizar el efecto de este factor sobre los resultados de clasificación.
- Cualquier planteamiento propuesto debe tomar en cuenta que un objetivo explícito de este proceso, debe ser la obtención de información confiable y de calidad que además de permitir la clasificación efectiva de beneficiarios en estratos para el cobro de la cuota familiar, permita caracterizar de manera valiosa y útil a los beneficiarios del SPSS (generar un padrón confiable y de calidad) para la toma de decisiones y mejora en la implementación del SPSS.
- De igual manera es necesario documentar con toda precisión el proceso de elaboración del instrumento, recolección de información y sobretodo de construcción y aplicación de una metodología de clasificación como parte de los procedimientos elementales del SPSS.

2.5 Evaluación de la técnica estadística utilizada actualmente en la CECASOEH y propuesta metodológica

Generalmente, para la identificación de hogares susceptibles de recibir apoyos, la información sobre las características socioeconómicas y demográficas de los hogares, así como sobre el equipamiento de sus viviendas y los bienes que poseen, se obtiene directamente de cada vivienda. Uno de los antecedentes de mayor importancia en México de este tipo de instrumentos son las Encuestas de Características Socioeconómicas de los Hogares (ENCASEH), que se aplicaron a manera de censo dentro de un conjunto de localidades seleccionadas por primera vez en los años 1997-1998 para su incorporación al Programa de Educación, Salud y Alimentación (PROGRESA). La unidad de análisis considerada para la focalización es el hogar y no sus miembros de manera aislada, bajo la consideración de que la familia es la unidad fundamental para el desarrollo de los individuos. La CECASOEH sigue principios similares para la clasificación de la población para los fines de la SPSS.

Para identificar a la población susceptible de acceder a beneficios sociales en este contexto, se requiere definir al menos tres aspectos centrales: las variables que caracterizan el nivel socioeconómico, la ponderación de cada una de dichas variables, y la determinación del nivel de dichas variables que hacen susceptible de atención a un individuo u hogar. Este enfoque multidimensional es el marco conceptual básico, y señalado por la literatura, para la elaboración de un modelo de clasificación.

Una vez establecida la condición inicial de bienestar como punto de referencia, es común aplicar técnicas de estadística multivariada, como el análisis discriminante, para evaluar las condiciones de vida de los hogares y la caracterización de los grupos susceptibles a ser beneficiarios de una acción determinada. Existen otras alternativas metodológicas como el modelo de Regresión Logístico, el cual estima la probabilidad de pertenecer a un grupo determinado con base en un conjunto de características. Para los propósitos del presente estudio se opta por el análisis discriminante, debido a que éste es el método más ampliamente utilizado en la literatura sobre el tema y en la instrumentación de programas de beneficio social. Incluso, es el método empleado por el CONEVAL para el desarrollo de la metodología multidimensional de la pobreza, así como del programa Prospera (antes Oportunidades). El uso de este método por lo tanto permitirá mantener la consistencia en el conjunto del sector social del Gobierno.

Otro aspecto relevante es que la regla de clasificación del análisis discriminante es en gran medida endógena al procedimiento. En otros procedimientos, esta regla es por lo general exógena, como sucede al fijar un conjunto de necesidades básicas. Ello facilita caracterizar distintas parametrizaciones entre los espacios de ingreso y de otras variables socioeconómicas según distintos contextos. Así, el peso que puede tener en la función discriminante una característica

particular, por ejemplo el nivel de escolaridad del jefe del hogar, no es el mismo en un contexto de alta escolaridad generalizada frente a otro de escasa escolaridad.

Adicionalmente, el análisis discriminante se presta a un aspecto práctico de suma importancia en la identificación de la población susceptible de recibir un apoyo. Por lo general, captar adecuadamente el ingreso o el consumo requiere de trabajos de campo sumamente meticulosos y muchas veces con elevados costos, siendo bastante más sencillo recabar información sobre aspectos como la escolaridad, las condiciones de la vivienda, o la posesión de bienes y enseres domésticos.

Generalmente, para su implementación, se parte de una clasificación inicial comparando el ingreso contra el valor del umbral de bienestar establecido, y se incorpora la información disponible sobre las características de los miembros del hogar y de su vivienda: la composición y tamaño de los hogares; la edad, el uso de lengua indígena, el analfabetismo, la escolaridad, la participación laboral y el tipo de ocupación de los miembros del hogar; la presencia de personas discapacitadas; el equipamiento de las viviendas y la posesión de bienes y enseres domésticos; el acceso y uso de superficie agrícola, así como la posesión de animales de trabajo y/o cría.

El proceso de discriminación y clasificación está relacionado con la separación de grupos de observaciones y la asignación de nuevas observaciones a grupos previamente definidos. Mediante el análisis discriminante se busca una función matemática para clasificar a los hogares en cada grupo de acuerdo al perfil que los caracteriza.

Conceptualmente existen dos tipos de análisis vinculados a la separación (cuando se trata de desagregar las observaciones en grupos) y clasificación (cuando se desea determinar el grupo al que pertenecen nuevas observaciones que son similares a otras observaciones de grupos previamente establecidos). En este caso, el procedimiento consiste en la clasificación en dos grupos: hogares susceptibles de recibir los apoyos y hogares no susceptibles.

Específicamente, el análisis discriminante logístico es un método parcialmente paramétrico. El modelo de regresión multinomial logístico es la base teórica del modelo de análisis discriminante logístico (*discrim logistic*, STATA), y también es una generalización del modelo de regresión logístico binomial (Mc Cullagh, 1989) en el que se desea estimar la probabilidad de que el individuo presente o no a un aspecto en particular, dado un conjunto de variables que explican características de los individuos.

Consideremos una variable de respuesta politómica Y con más de dos categorías de respuesta que denotaremos por Y_1, Y_2, \dots, Y_k .

Se pretende explicar la probabilidad de cada categoría de respuesta en función de un conjunto de covariables $X = \{x_1, x_2, \dots, x_n\}$ observadas. Es decir, ajustar un modelo de la forma:

$$p_j = P[Y = Y_j | X = x] = f_j(x) \quad \forall j = 1, \dots, k$$

para cada vector de valores observados de las variables explicativas X .

La idea básica detrás del modelo *logit multinomial* es comparar dos respuestas a la vez. Esta comparación podría ser entre cualesquiera dos categorías, sin embargo, suele hacerse contra la primera o última categoría de respuesta.

Para construir el modelo de respuesta multinomial se consideran $(k - 1)$ transformaciones *logit*, definidas respecto a la categoría de referencia, las cuales se definen como:

$$L_j(x) = \ln \left[\frac{p_j(x)}{p_k(x)} \right] \quad \forall j = 1, \dots, k - 1$$

Esta expresión indica la relación de la probabilidad de respuesta Y_j con respecto a la Y_k , condicionado a las observaciones de las variables explicativas.

El modelo lineal se define de la siguiente forma:

$$L_j(x) = \sum_{i=0}^n \beta_{ij} x_i = x' \beta_j \quad \forall j = 1, \dots, k - 1$$

Para cada vector de valores observados de las variables explicativas $x = (x_0, x_1, \dots, x_n)$ donde $x_0 = 1$ (término constante) y $\beta_j = (\beta_{0j}, \beta_{1j}, \dots, \beta_{nj})$ el vector de parámetros asociados a la categoría Y_j .

Las probabilidades asociadas a una determinada categoría tienen la siguiente expresión:

$$p_j(x) = \frac{\exp(\sum_{i=0}^n \beta_{ij} x_i)}{1 + \sum_{j=1}^{k-1} \exp(\sum_{i=0}^n \beta_{ij} x_i)} \quad \forall j = 1, \dots, k - 1$$

Si la categoría de referencia es la última la probabilidad asociada es:

$$p_k(x) = \frac{1}{1 + \sum_{j=1}^{k-1} \exp(\sum_{i=0}^n \beta_{ij} x_i)} \quad \forall j = 1, \dots, k - 1$$

En este caso, el objetivo es clasificar a los sujetos por *decil de ingreso* de acuerdo a ciertas características socio demográfico, entonces la variable respuesta es *decil de ingreso*. La Encuesta de Ingresos y Gastos de los Hogares (ENIGH) es la fuente mediante la cual se obtuvieron las estimaciones, pues ENIGH permite establecer un panorama estadístico del comportamiento de los ingresos y gastos de los hogares; adicionalmente ofrece información sobre las características ocupacionales y socio-demográficas de los integrantes del hogar, así como las características de la infraestructura de la vivienda y el equipamiento de los hogares, además que la representatividad es nacional y para los ámbitos rural y urbano.

Debido a que existen fuertes diferencias socio-demográficas entre la población rural y urbana, se optó por realizar dos modelos uno para cada población. En el modelo urbano la categoría de referencia es la última y en el modelo rural la primera.

El modelo urbano tendría la siguiente expresión, el cual está formado por 21 variables explicativas, las cuales se presentan en el cuadro 1 y nueve funciones.

$$L_{decil\ j}(x) = \sum_{i=0}^{21} \beta_{ij}x_i = x' \beta_j \quad \forall j = 1, \dots, 9$$

Los coeficientes de la funciones del modelo urbano se observan en el cuadro 1. Por ejemplo, la función 1 y 9 se escribe de la siguiente forma:

$$\begin{aligned} L_{decil1} = & 0.6x_1 + 0.7x_2 + 0.3x_3 - 1.3x_4 - 0.4x_5 + 1.7x_6 - 0.3x_7 + 0.4x_8 - 2.2x_9 - 1.2x_{10} \\ & + 1.2x_{11} - 0.5x_{12} + 1.2x_{13} - 0.3x_{14} + 0.7x_{15} + 0.9x_{16} - x_{17} - 0.5x_{18} \\ & + 1.1x_{19} + 0.9x_{20} - 3.2x_{21} - 1.7 \end{aligned}$$

⋮

$$\begin{aligned} L_{decil9} = & 0.02x_1 - 0.12x_2 + 0.11x_3 + 0.2x_4 - 0.13x_5 + 0.65x_6 - 0.09x_7 + 0.18x_8 - 0.11x_9 \\ & - 0.22x_{10} - 0.04x_{11} - 0.25x_{12} + 0.25x_{13} - 0.05x_{14} + 0.02x_{15} + 0.16x_{16} \\ & - 0.47x_{17} - 0.04x_{18} + 0.38x_{19} + 0.23x_{20} - 0.53x_{21} + 0.33 \end{aligned}$$

Una vez evaluadas las características de un individuo en las nueve funciones. El siguiente paso es obtener la probabilidad de pertenecer a determinada categoría mediante las funciones calculadas.

Para obtener la probabilidad de que el individuo pertenezca a la categoría 1, es decir, al primer decil, se tendría que evaluar la siguiente función y de manera sucesiva hasta la categoría 9.

$$p_1(x) = \frac{\exp(L_{decil1})}{1 + \sum_{j=1}^9 \exp(\sum_{i=0}^{21} \beta_{ij} x_i)} \quad \forall j = 1, \dots, 9$$

$$\vdots$$

$$p_9(x) = \frac{\exp(L_{decil9})}{1 + \sum_{j=1}^9 \exp(\sum_{i=0}^{21} \beta_{ij} x_i)} \quad \forall j = 1, \dots, 9$$

Como en el modelo urbano la categoría de referencia fue la última la probabilidad asociada al décimo decil de ingreso es:

$$p_{10}(x) = \frac{1}{1 + \sum_{j=1}^9 \exp(\sum_{i=0}^{21} \beta_{ij} x_i)} \quad \forall j = 1, \dots, 9$$

Una vez obtenidas la probabilidad de que el individuo pertenezca a los diferentes grupos se clasifica al individuo en la categoría cuya probabilidad asociada es la mayor de todas.

Cuadro 2.5.1. Funciones del modelo urbano

Variables		FUNCIONES								
		F1	F2	F3	F4	F5	F6	F7	F8	F9
x_1	Tamaño de localidad	0.6	0.4	0.4	0.2	0.1	0.1	0.0	0.0	0.0
x_2	Servicio sanitario	0.7	0.6	0.4	0.2	0.0	0.5	0.3	0.4	-0.1
x_3	Tenencia de la vivienda	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.2	0.1
x_4	Recibe un pago por su trabajo	-1.3	-0.6	-0.3	-0.2	0.0	0.2	0.2	0.2	0.2
x_5	Educación del jefe del hogar	-0.4	-0.4	-0.3	-0.3	-0.3	-0.3	-0.3	-0.2	-0.1
x_6	Hacinamiento	1.7	1.6	1.5	1.5	1.4	1.3	1.1	1.0	0.7
x_7	Tipo de servicio de salud	-0.3	-0.3	-0.3	-0.3	-0.2	-0.2	-0.1	-0.1	-0.1
x_8	Dependencia demográfica	0.4	0.3	0.3	0.2	0.1	0.1	0.1	0.1	0.2
x_9	Seguridad Social	-2.2	-1.2	-0.9	-0.7	-0.6	-0.5	-0.4	-0.3	-0.1
x_{10}	Número de computadoras	-1.2	-1.0	-0.9	-0.9	-0.8	-0.6	-0.5	-0.4	-0.2
x_{11}	No asisten a la escuela	1.2	0.9	0.5	0.6	0.4	0.4	0.1	0.4	0.0
x_{12}	Hijos que trabajan	-0.5	-0.5	-0.5	-0.4	-0.5	-0.6	-0.2	-0.4	-0.3
x_{13}	Televisión de paga	1.2	1.0	0.9	0.7	0.7	0.5	0.5	0.4	0.3
x_{14}	Conexión a internet	-0.3	0.2	0.3	0.2	0.2	0.2	0.1	0.0	0.0
x_{15}	Teléfono	0.7	0.3	0.3	0.3	0.2	0.1	0.1	0.1	0.0
x_{16}	Celular	0.9	0.6	0.6	0.4	0.3	0.3	0.1	0.3	0.2
x_{17}	Número de vehículos	-1.0	-1.1	-1.1	-1.0	-0.9	-0.9	-0.8	-0.7	-0.5
x_{18}	Número de prestaciones	-0.5	-0.3	-0.3	-0.2	-0.1	-0.1	-0.1	-0.1	0.0
x_{19}	Escuela pública	1.1	1.2	1.1	1.1	1.0	0.9	0.7	0.6	0.4
x_{20}	La empresa es negocio propio	0.9	1.0	1.0	0.9	0.8	0.7	0.6	0.4	0.2
x_{21}	Número de ocupados entre total de integrantes	-3.2	-3.3	-3.4	-3.2	-2.6	-2.4	-2.0	-1.1	-0.5
	Constante	-1.7	-0.9	-1.2	-0.2	0.4	0.4	1.1	0.3	0.3

Elaboración propia EASE 2014.

El procedimiento para llevar a cabo la clasificación de la población rural es similar en la definición de la variable dicotómica.

Cuadro 2.5.2. Funciones del modelo urbano

Variables		FUNCIONES								
		F1	F2	F3	F4	F5	F6	F7	F8	F9
x_1	Material del piso	-0.13	-0.25	0.03	-0.04	-0.04	-0.63	-0.27	0.04	-0.19
x_2	Servicio sanitario	-0.06	-0.09	-0.09	-0.17	-0.17	0.05	0.10	-0.16	-0.49
x_3	Educación del jefe del hogar	0.00	0.02	0.06	0.06	0.06	0.04	0.13	0.15	0.23
x_4	Hacinamiento	-0.09	-0.14	-0.18	-0.30	-0.30	-0.41	-0.57	-1.11	-1.33
x_5	Dependencia demográfica	-0.03	-0.14	-0.29	-0.57	-0.57	-0.74	-0.91	-0.77	-0.95
x_6	Seguridad Social	0.98	2.00	2.22	1.92	1.92	2.34	2.12	2.29	2.49
x_7	Tipo de servicio de salud	0.13	0.03	0.02	0.24	0.24	0.18	0.28	0.34	0.17
x_8	Estufa	-0.42	-0.59	-0.87	-1.20	-1.20	-1.12	-0.76	-0.94	-0.70
x_9	Refrigerador	-0.09	-0.25	-0.16	-0.14	-0.14	-0.13	-0.10	-0.15	-0.32
x_{10}	Lavadora	-0.18	-0.16	-0.29	-0.11	-0.11	-0.12	-0.39	-0.57	-0.34
x_{11}	No asisten a la escuela	-0.16	-0.19	-0.06	-0.33	-0.33	-0.47	-0.02	-0.46	-0.61
x_{12}	Hijos que trabajan	0.08	0.17	0.06	0.37	0.37	0.51	-0.05	0.52	0.00
x_{13}	Numero de prestaciones	0.84	1.04	1.15	1.27	1.27	1.28	1.34	1.39	1.39
x_{14}	Vehículo	-0.10	0.16	0.11	0.17	0.17	0.20	0.23	0.51	0.69
	Constante	0.16	0.02	-0.13	-0.31	-0.31	-0.14	-0.62	-0.59	-0.78

Elaboración propia EASE 2014.

Finalmente, cuando la población está clasificada en las 10 categorías de acuerdo a esta metodología, se reclasifica la población del primer al cuarto decil de ingreso, del quinto al octavo decil y del noveno al décimo decil de ingreso.

Descripción de variables

Material del piso

<i>mat_pisos</i>	1	si la vivienda tiene piso de tierra
	0	si la vivienda tiene piso de cemento o firme
	0	si la vivienda tiene piso linóleum, congóleum o firme
	0	si la vivienda tiene piso laminado
	0	si la vivienda tiene mosaico, mármol o vitropiso
	0	si la vivienda tiene piso de madera, duela o parquet

Servicio de salud

<i>serv_sal</i>	1	si la persona cuenta con Seguro Popular
	2	si la persona cuenta con servicios médicos del IMSS
	3	si la persona cuenta con servicios médicos del ISSSTE o ISSSTE estatal
	4	si la persona cuenta con servicios médicos de Pemex, Defensa o la Marina
	5	si la persona cuenta con otros servicios médicos
	0	si la persona no cuenta con ninguno de los servicios médicos referidos

Educación del jefe del hogar

<i>educa_jefe</i>	1	Sin instrucción
	2	Preescolar
	3	Primaria incompleta (menos de 6 años)
	4	Primaria completa
	5	Secundaria incompleta (menos de 3 años)
	6	Secundaria completa
	7	Preparatoria incompleta (menos de 3 años)
	8	Preparatoria completa
	9	Superior incompleta (menos de 3 años)
	10	Superior completa
	11	Posgrado

Hacinamiento

Número de cuartos sin contar la cocina, si la cocina es dormitorio entonces si se toma en cuenta.

$$= \frac{\text{Total de integrantes del hogar}}{\text{Número de cuartos}}$$

Dependencia demográfica

$$= \frac{\text{Integrantes entre 0 y 14 años + integrantes mayores de 65 años}}{\text{Integrantes entre 15 y 64 años}}$$

Índice de ocupados

$$= \frac{\text{Numero de ocupados en el hogar}}{\text{Total de integrantes del hogar}}$$

Número de personas que reciben un pago por su trabajo

recibe: Número de personas que reciben un pago por su trabajo

Número de prestaciones

presta: Número total de prestaciones del hogar (suma de todas las prestaciones de las personas que trabajan en el hogar)

No asisten a la escuela

hijos5_15: Número de integrantes del hogar entre 5 y 15 años que no asisten a la escuela.

Hijos que trabajan

hijos12_15: Número de integrantes del hogar entre 12 y 15 años que no trabajan.

Escuela pública

publica: Número de integrantes del hogar que asisten a escuela pública

La empresa es negocio propio

<i>c_l</i>	1	El tipo de empresa en la que labora es independiente, personal o familiar
	0	en otro caso

Seguridad social

<i>segsoc</i>	1	si trabaja y cuenta con incapacidad y acceso a jubilación o fondo para el retiro y cuenta con servicio médico en su trabajo
	1	si cuenta con jubilación
	0	en otro caso

Servicio sanitario

<i>excusado</i>	1	si el hogar cuenta excusado, retrete, sanitario, letrina u hoyo negro
	0	en otro caso

Estufa

<i>n_estufa</i>	1	si el hogar cuenta con estufa
	0	en otro caso

Refrigerador

<i>n_refri</i>	1	si el hogar cuenta con refrigerador
	0	en otro caso

Lavadora

<i>n_lava</i>	1	si el hogar cuenta con lavadora
	0	en otro caso

Vehículo

<i>n_vehi1</i>	1	si el hogar cuenta con un auto o pick-up o van
	2	si el hogar cuenta con dos o más autos (auto, pick-up o van)
	0	en otro caso

Computadora

<i>n_compu</i>	1	si el hogar cuenta con computadora
	2	si el hogar cuenta con dos o más computadoras
	0	en otro caso

Conexión a internet

<i>conex_inte</i>	1	si el hogar cuenta con conexión a internet
	0	en otro caso

Televisión de paga

<i>tv_paga</i>	1	si el hogar cuenta con televisión de paga
	0	en otro caso

Teléfono⁴

<i>teléfono</i>	1	si el hogar cuenta con línea telefónica
	0	en otro caso

Celular

<i>celular</i>	1	si en el hogar alguien cuenta con celular
	0	en otro caso

Tenencia

<i>Tenencia</i>	1	Rentada
	2	Prestada
	3	Propia pero se está pagando
	4	Propia
	5	Esta intestada o en litigio
	6	Otra situación

Elaboración de un marco conceptual y técnico para sustentar el proceso de clasificación de los ingresos

De acuerdo con las conclusiones generales obtenidas del análisis de aplicación del modelo actual de clasificación que utiliza el SPSS (conjunto de variables y algoritmo: puntajes) sobre los datos de la ENIGH y al comparar la distribución por deciles que arroja este modelo versus la distribución por deciles que arroja la propia encuesta, se sabe que el modelo es buen predictor para los primeros cuatro deciles, siendo impreciso para el resto, particularmente los últimos deciles.

Dado este resultado y considerando las recomendaciones de la literatura e intervenciones-referencias de modelos de clasificación de beneficiarios que apuntan hacia el uso de enfoques multidimensionales con métodos como el análisis discriminante, se realizaron los siguientes ejercicios técnico-estadísticos.

⁴⁴ En este caso por cuestiones metodológicas tanto en la variable teléfono como en la de celular únicamente se incluyen las opciones binarias de 0 en el caso de que en el hogar nadie cuente con celular o teléfono; y 1 para cuando si exista celular o teléfono en el hogar, independientemente de que pueda ser más de 1 celular y/o teléfono.

2.6 Aplicación de la metodología (modelo, ponderadores y umbrales) propuesto

Aplicación de análisis discriminante con variables del modelo de clasificación del SPSS

Utilizando las mismas variables del modelo de clasificación del SPSS actual, se aplicó la metodología de análisis discriminante sobre los datos de la ENIGH para evaluar si mejora o no el grado de predicción en la clasificación respecto a la distribución por deciles de la ENIGH.

A continuación se presentan los resultados para cada decil y también los resultados agregando los deciles en grupos acorde al contexto del SPSS, según se describió anteriormente:

Cuadro 13. Comparación entre distribución por deciles de la ENIGH y distribución por deciles con variables del modelo de SPSS y aplicación de análisis discriminante, nacional

Deciles ENIGH	Deciles PUNTAJE									
	1	2	3	4	5	6	7	8	9	10
1	35.7	14.5	4.9	7.4	8.0	7.7	3.9	6.9	5.8	5.2
2	20.6	16.5	7.7	11.3	10.1	8.5	4.5	9.6	5.5	5.6
3	12.2	11.3	8.1	12.0	10.6	11.8	6.2	12.3	7.0	8.6
4	7.6	9.6	6.1	12.1	12.5	10.8	7.9	12.5	7.6	13.4
5	4.6	6.4	5.2	8.3	12.1	10.7	9.2	14.7	8.9	20.0
6	3.7	4.2	2.9	7.8	10.8	10.4	8.2	18.2	11.2	22.7
7	3.2	2.2	2.4	5.1	8.1	7.5	8.3	18.2	13.2	31.9
8	1.5	1.2	1.3	2.9	4.4	5.7	6.6	16.2	11.5	48.7
9	1.1	0.8	0.8	1.6	2.5	3.8	4.9	12.6	9.6	62.4
10	0.3	0.2	0.3	0.7	1.0	1.2	2.2	6.1	5.0	83.1

Fuente: Elaboración propia con datos de la ENIGH 2012 y variables de clasificación del SPSS

Cuadro 14. Comparación entre distribución por grupos de deciles de la ENIGH y distribución por grupos de deciles con variables del modelo de SPSS y aplicación de análisis discriminante, nacional

Deciles ENIGH	Deciles PUNTAJE			
	1-4	5-6	7-8	9-10
1-4	65.7	13.1	11.7	9.6
5-6	42.1	16.5	20.6	20.8
7-8	29.4	13.4	26.4	30.8
9-10	12.5	6.7	18.4	62.4

Fuente: Elaboración propia con datos de la ENIGH 2012 y variables de clasificación del SPSS

Según se observa, este modelo es un buen predictor para el grupo de primeros cuatro deciles y mejora importantemente la predicción para el último grupo de deciles (9 y 10). Para el grupo de deciles intermedios (5 al 8), el porcentaje es ligeramente menor y mayor para los deciles 5-6 y 7-8, respectivamente, comparado con los resultados del modelo de clasificación actual del SPSS (ver cuadro 7), por lo que sigue siendo un mal predictor. Con este modelo, se asigna atinadamente al 47.4% de los hogares, mejorando así ligeramente el grado de precisión del modelo actual de clasificación del SPSS como se describió en el apartado 1.

Combinación de modelos: algoritmo rural actual del modelo de SPSS más modelo con aplicación de análisis discriminante sobre un conjunto alternativo de variables

Con el objetivo de incrementar la precisión en la clasificación de los grupos de últimos deciles (9 y 10) y deciles intermedios (5 al 8), manteniendo la predicción del primer grupo de deciles (1 al 4), se propone un modelo combinado donde se utiliza el algoritmo rural del modelo de clasificación del SPSS actual y un modelo alternativo que aplica análisis discriminante sobre un conjunto de variables distinto.

Este conjunto de variables mantiene como base las variables planteadas por el modelo actual del SPSS complementándolas con varias más precisas para la captación de tenencia de activos y mejor caracterización de los deciles más altos, derivadas del análisis discriminante presentado en la sección anterior. Las variables rurales son las siguientes:

- Material del piso de la vivienda
- Si cuenta con excusado
- Educación del jefe del hogar
- Índice de dependencia
- Hacinamiento

- Seguro social
- Hijos que no asisten a la escuela
- Hijos que trabajan
- Tipo de servicios de salud
- Si cuenta con estufa
- Si cuenta con refrigerador
- Si cuenta con lavadora
- Número de prestaciones de todos los que trabajan en el hogar
- Si cuenta con vehículo 0, 1 o más de 2

Las variables urbanas son las siguientes:

- Tamaño de localidad
- Si cuenta con excusado
- Educación del jefe del hogar
- Índice de dependencia
- Hacinamiento
- Seguro social
- Hijos que no asisten a la escuela
- Hijos que trabajan
- Si cuenta con vehículo 0, 1 o más de 2
- Si cuenta con computadora 0, 1 o más de 2
- Número de prestaciones de todos los que trabajan en el hogar
- Número de integrantes que asisten a escuela pública
- Número de personas que laboran en el hogar entre el total de integrantes
- Tipo de empresa en la que labora (1. Tipo independiente, familiar o personal)
- Tenencia de la vivienda
- Número de miembros a los que les otorgaron beca escolar
- Si cuenta con celular
- Si cuenta con teléfono
- Tipo de servicios de salud
- Si cuenta con TV de paga
- Si cuenta con conexión a internet

El siguiente cuadro muestra los resultados de la aplicación de este modelo sobre los datos de la ENIGH:

Cuadro 15. Comparación entre distribución por grupos de deciles de la ENIGH y distribución por grupos de deciles con combinación de modelos propuesto, nacional

Deciles ENIGH	Deciles PUNTAJE			
	1-4	5-6	7-8	9-10
1-4	82.4	9.6	5.9	2.1
5-6	51.3	19.6	19.1	10.0
7-8	26.1	15.5	29.9	28.5
9-10	9.9	6.5	14.5	69.1

Fuente: Elaboración propia con datos de la ENIGH 2012 y modelo propuesto

Con la agregación en cuatro grupos y la combinación de modelos, se logra incrementar la coincidencia entre la asignación de esta propuesta y la asignación de la ENIGH para todos los grupos. Destaca en particular, el alto nivel de predicción para el primer grupo de deciles y para el último. Con este modelo, se clasifica atinadamente al 56.6% de los hogares.

Si se hace la agregación considerando únicamente tres grupos de deciles, el modelo combinado arroja los siguientes resultados:

Cuadro 16. Comparación entre distribución por tres grupos de deciles de la ENIGH y distribución por tres grupos de deciles con combinación de modelos propuesto, nacional

Deciles ENIGH	Deciles PUNTAJE		
	1-4	5-8	9-10
1-4	77.5	19.6	2.9
5-8	28.3	51.8	19.9
9-10	6.0	20.9	73.2

Fuente: Elaboración propia con datos de la ENIGH 2012 y modelo propuesto

Este modelo combinado propuesto permite mantener el alto grado de predicción para el primer grupo de deciles (1 al 4) y mejorar significativamente el grado de predicción para el grupo de deciles intermedios (5 al 8) y para el grupo de últimos deciles (9 y 10). Con este modelo, se clasifica atinadamente al 66.36% de los hogares, siendo este el grado de predicción más alto entre los ejercicios realizados.

2.7 Conclusiones y planteamiento propuesto

En resumen, lo que se pretende es presentar una propuesta que permita cumplir con los siguientes objetivos:

- Mejorar en la medida de lo posible la clasificación del primer grupo de deciles de ingreso (1 al 4) que constituyen el régimen no contributivo y que parecen caracterizar al grueso de la población beneficiaria del SPSS para minimizar errores de exclusión.
- Mejorar significativamente la clasificación a los deciles que constituyen el régimen contributivo, en particular, los deciles 9 y 10 que tienen mayor capacidad de financiar su aseguramiento en salud. Esto permitiría minimizar los errores de inclusión.
- Contar con información de calidad para la constitución del padrón que además de servir para el proceso de pago de la cuota familiar, provea información relevante para la toma de decisiones del sistema.
- Documentar con precisión todo el proceso de diseño metodológico.

Los análisis aquí planteados y su comparación permiten concluir lo siguiente:

- Es necesario evaluar, desde la perspectiva práctica de la implementación del Sistema, los objetivos que éste busca y en base al grado de precisión en la clasificación de los distintos grupos de deciles, si se requiere modificar los límites y condiciones de pertenencia para los regímenes contributivo y no contributivo. Por ejemplo, una alternativa es considerar un modelo que identifique con la mayor precisión posible a los últimos deciles de ingreso y establecer el régimen contributivo sólo para estos. Lo anterior requiere tomar en cuenta todas las consideraciones relevantes en cuanto al diseño del SPSS como política pública, la viabilidad operativa y el objetivo implícito de contar con la participación del afiliado en el co-financiamiento de su aseguramiento en salud con las consecuencias que esto conlleva.
- Dado el alto grado de predicción que el modelo actual del SPSS muestra para el primer grupo de deciles, aunado al incremento en el grado de precisión para los otros dos grupos de deciles (particularmente los últimos deciles) producto del planteamiento de un modelo alternativo, la propuesta de este estudio radica en combinar ambas alternativas en un modelo que, comparado con el resto de los ejercicios realizados, permite asignar a un mayor porcentaje de los hogares atinadamente; o en otras palabras, muestra el menor margen de error.

Cuadro 17. Comparación de análisis/ejercicios realizados respecto al % de hogares asignados atinadamente

Análisis/Ejercicio	% de hogares asignados atinadamente
Modelo de clasificación del SPSS actual, agregado en deciles	14.40%
Modelo de clasificación del SPSS actual, agregado en cuatro grupos de deciles	41.25%
Modelo de análisis discriminante con variables del modelo del SPSS actual, agregado en cuatro grupos de deciles	47.37%
Modelo combinado: algoritmo rural del modelo del SPSS actual + modelo de análisis discriminante con variables nuevas, agregado en cuatro grupos de deciles	56.64%
Modelo combinado: algoritmo rural del modelo del SPSS actual + modelo de análisis discriminante con variables nuevas, agregado en tres grupos de deciles	66.36%

Fuente: Elaboración propia con resultados de ejercicios realizados

- Asimismo, la clasificación en tres grupos o estratos coincide con la lógica actual de separación de regímenes y asignación adicional a estos por consideraciones particulares (contar con un menor de 5 años en el hogar o con una embarazada o ser beneficiarios de Prospera, entre otros).

En conclusión, se propone valorar la utilización de este modelo combinado alternativo con tres grupos de deciles o de clasificación como la mejor opción para lograr los objetivos propuestos.

III. Piloteo en campo y validación del instrumento CECASOEH v 2.0

El objetivo de la presente Sección es realizar una descripción detallada del Piloto de la “Cédula de Características Socioeconómicas del Hogar (CECASOEH v 2.0)” cuyo levantamiento de campo se realizó del 06 al 19 de diciembre del 2014.

a) Antecedentes.

El instrumento que permite clasificar a las familias que buscan afiliarse al SPSS es la Cédula de Características Socioeconómicas del Hogar (CECASOEH), que sirve para la determinación del nivel socioeconómico de la familia que solicita su afiliación al SPSS y sirve para la identificación del núcleo familiar y sus integrantes, así como para la determinación de la cuota familiar.

El proceso de clasificación actualmente utilizado por el SPSS ubica en los cuatro primeros deciles a aproximadamente 52 millones de personas que representan el 99.6% del total de afiliados al mes de diciembre de 2013; es decir, están clasificados dentro del régimen no contributivo.

El presente proyecto busca validar la consistencia y confiabilidad de la información proporcionada por la CECASOEH y registrada en el Padrón Nacional de Beneficiarios para fortalecer su integración y cumplimiento con la normatividad que permitan la creación de competencias institucionales de un Sistema de Protección Social en Salud.

b) Objetivos del Piloto.

Realizar un piloto en campo para probar y validar la Cédula de Características Socioeconómicas del Hogar (CECASOEH versión 2.0) con representación del Padrón Nacional de Beneficiarios a nivel nacional y con significancia estadística.

c) Metodología.

Se realizaron aplicaciones de la CECASOEH versión 2.0 cara a cara a papel y lápiz en hogares dentro de seis entidades federativas que representan el 53.5% de la población que actualmente se encuentra en el régimen contributivo decil I al V.

d) Cobertura geográfica.

El levantamiento de campo se realizó dentro de 6 ciudades de las entidades federativas que representan el 53.5% del total de población que actualmente entra en el régimen contributivo del Padrón Nacional de Beneficiarios del Sistema de Protección Social en Salud (decil I al V):

- Zona Metropolitana de la Ciudad de México
- Xalapa
- Tuxtla Gutiérrez
- Puebla
- León
- Oaxaca

En cada una de estas ciudades, los elementos de interés se encuentran concentrados en dos diferentes estratos: zona urbana y zona rural. De acuerdo al Censo Poblacional 2010 de INEGI, en la zona urbana se concentra el 78% de la población objetivo, el restante 22% lo ubicamos en la zona rural.

e) Público Objetivo.

Hombres y Mujeres de 18 años y más, jefes del hogar seleccionado.

f) Diseño muestral.

Se realizó un muestreo aleatorio con selección de unidades primarias de muestreo (AGEB y localidades) en las plazas seleccionadas con una proporción de 78% en zonas urbanas y 22% en localidades rurales.

- Primera Etapa. Selección aleatoria simple de AGEB y localidades dentro de las seis ciudades previamente mencionadas. El número de AGEB y localidades determinado mediante el procedimiento antes mencionado, será asignado, en cada ciudad, en proporción de 78% en área urbana y 22% en área rural.
- Segunda Etapa. Selección aleatoria simple de 10 hogares en cada una de las AGEB elegidas en la primera etapa. La selección en esta etapa depende de las condiciones geográficas de cada AGEB, ya sea por elección aleatoria de hogares por manzana con salto sistemático dentro de una AGEB (zonas urbanas), o bien, mediante selección aleatoria desde un punto de referencia reconocible dentro de la AGEB (zonas rurales).

- Tercera etapa: Selección del informante en cada hogar seleccionado quien debe ser el jefe del hogar.

En AGEB urbanas la selección de los hogares se determina mediante ubicando al encuestador al frente de la manzana, comenzando desde un punto central, continúa la numeración en sentido contrario a las manecillas del reloj. Se realiza un salto sistemático de 10 viviendas por manzana.

- Se tiene que obtener 10 entrevistas completas dentro de cada AGEB.
- No se pueden realizar más de una entrevista por acera.

En localidades rurales se eligen los hogares desde un punto de referencia reconocible en la zona (Iglesia, plaza principal, etc.), se deberá mover desde el punto de referencia, en sentido contrario a las manecillas del reloj, contando hogares sobre la región, sistemáticamente, cada “r” hogares. El valor de “r” se determina de acuerdo al total de casas estimadas sobre la región, (cada 10 hogares, cada 5, cada 3) según lo permita la situación particular en cada sitio.

g) Tamaño de la Muestra.

El elemento último de muestreo son hombres y mujeres mayores de edad, los cuales están distribuidos en zonas geográficas denominadas áreas geo-estadísticas básicas (AGEB).

El tamaño de muestra se determina considerando un margen de error de ± 0.03 en la estimación de proporciones con una confianza del 95%

$$n = \frac{Z^2 p(1-p)}{d^2} = \frac{(1.96)^2 0.5(1-0.5)}{0.03^2} = 1067.1$$

El valor $p = 0.5$, considera el tamaño de muestra del peor de los casos, que es sería suficiente para lograr la precisión antes mencionada, independientemente del valor real de p .

Si se considera un factor de tamaño de muestra por el efecto del diseño de muestreo de 2.5, lo cual determina un tamaño de muestra de:

$$n = 1067.1 \times 2.5 = 2667.7$$

El tamaño final de muestreo se eleva hasta 3000 para evitar una posible reducción en la precisión de las estimaciones por cuestiones no controlables.

Estas 3000 aplicaciones de la CECASOEH se distribuyeron de manera equitativa por plaza (500 por plaza) considerando un 78% en zonas urbanas y 22% en localidades rurales.

La muestra quedó distribuida de la siguiente manera:

PLAZA	ZONA URBANA	LOCALIDAD RURAL	TOTAL
ZMCM	390	110	500
Xalapa	390	110	500
Tuxtla Gutiérrez	390	110	500
Puebla	390	110	500
León	390	110	500
Oaxaca	390	110	500
TOTAL	2340	660	3000

Como se mencionó anteriormente, para determinar las zonas de levantamiento se realizó una selección aleatoria simple de AGEB (áreas geo-estadísticas básicas) dentro de las zonas urbanas y de igual forma se seleccionaron localidades en zonas rurales de las mismas plazas.

En cada AGEB y localidad se realizaron 10 entrevistas efectivas, de tal suerte que la selección de AGEB y localidades quedó distribuida de la siguiente manera:

PLAZA	AGEB URBANO	LOCALIDAD RURAL	TOTAL
ZMCM	39	11	50
Xalapa	39	11	50
Tuxtla Gutiérrez	39	11	50
Puebla	39	11	50
León	39	11	50
Oaxaca	39	11	50
TOTAL	234	66	300

(Ver documentos: AGEB y Localidades seleccionadas).

h) Reemplazo de AGEB.

Se incluyeron dentro de la muestra seleccionada 60 AGEB como posibles unidades de reemplazo, 10 en cada ciudad, para prevenir situaciones que imposibiliten la visita a alguna de las unidades seleccionadas.

i) Levantamiento de campo.

Las entrevistas se realizaron a papel y lápiz, cara a cara y utilizando un mismo cuestionario (Ver documento Anexo Cuestionario CECASOEH 2.0).

El procedimiento del levantamiento de campo en zona urbana se realiza de la siguiente manera:

- El equipo de campo se posiciona en el AGEB seleccionado con el mapa correspondiente (Ver Imagen 1. Ejemplo de AGEB).
- Dentro del AGEB se ubicó la manzana seleccionada.
- En la manzana, se ubica al encuestador al frente de la manzana, comenzando desde un punto central, continúa la numeración en sentido contrario a las manecillas del reloj en la esquina frontal izquierda, se realiza un salto sistemático de 10 viviendas.

Imagen 1. Ejemplo de AGEB

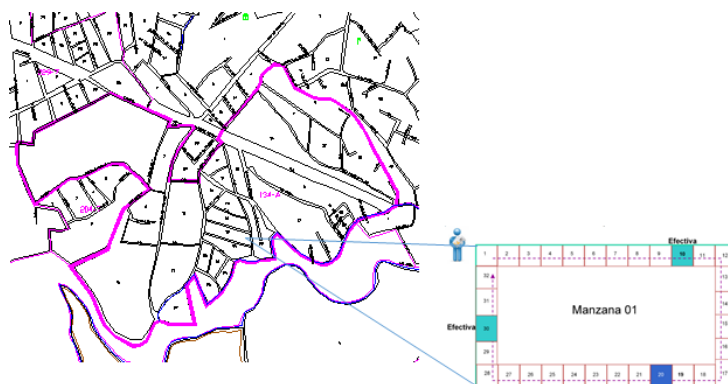
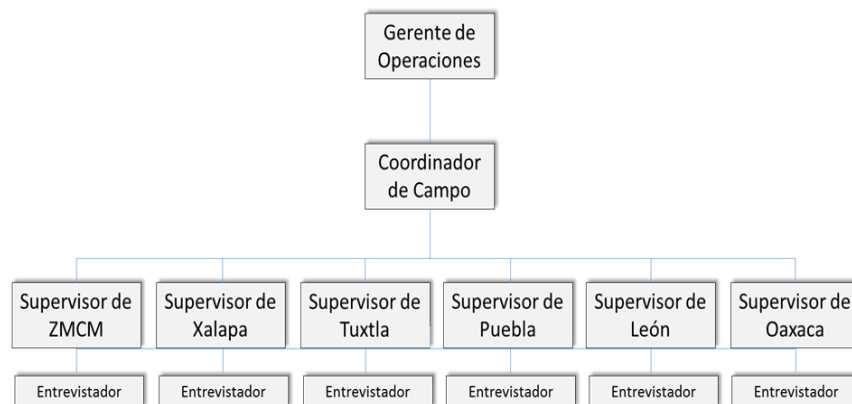


Imagen 2. Ejemplo de localidad



j) Constitución del Personal de campo.

La estructura general del personal de campo se integra de la siguiente forma:



En cada plaza participaron dos equipos de trabajo, cada equipo se conformó por un supervisor y cuatro encuestadores.

Coordinación General:

- Gerente de Operaciones: 1
- Coordinador de campo: 1

Coordinación por plaza:

- Número de supervisores de campo por plaza: 2
- Número de entrevistadores por plaza: 8

k) Capacitación del personal de campo.

Previamente al levantamiento de campo, el gerente de operaciones y el coordinador general de campo instruyeron al personal participante.

La capacitación del personal se realizó en el Distrito Federal el 5 de diciembre del 2014. En dicha capacitación se utilizaron dos vías de comunicación:

- Presencial para el caso del personal que levantaría el campo en la ZMCM.
- Telefónica para el personal que levantaría el resto de las plazas.

Al finalizar la capacitación, todos los involucrados fueron registrados en un formato de lista de asistencia (Ver formato de lista de asistencia) y se realizó una grabación en audio para su posterior consulta en caso de dudas.

Adicionalmente se generó un manual de capacitación general para consulta del personal (Ver manual de capacitación).

l) Supervisión del trabajo de campo.

El supervisor de cada plaza fungió como responsable de la organización y supervisión de las actividades en campo, identificación de viviendas seleccionadas, asignación de las cargas de trabajo al equipo de entrevistadores y la verificación de la calidad de la información recolectada.

Al momento de la aplicación el supervisor se posicionó a una distancia prudente del equipo para poder vigilar el trabajo de todos los encuestadores a su cargo.

Los encuestadores se encuentran en zonas de trabajo delimitadas por sus supervisores (a unas cuantas cuadras y siempre a la vista), de este modo los supervisores pueden observar la aplicación de cuestionarios de varios encuestadores durante la misma jornada (moviéndose solo unas calles para atestiguar la realización).

Tipo de supervisión:

- Directa o coincidental. A través de acompañamientos de los entrevistadores para verificar la adecuada localización de las áreas, correcta selección del respondente y aplicación de los instrumentos de campo.
- Posterior. Durante todo el trabajo de campo, el supervisor seleccionó al azar cuestionarios de cada uno de los entrevistadores y realizó visitas a las viviendas para verificar que efectivamente se haya realizado la entrevista, aplicando el cuestionario de manera correcta y que se haya seleccionado al respondente adecuado.
- Telefónica. Este tipo de supervisión se realizó en escritorio, en donde se tomaron de manera aleatoria cuestionarios con información telefónica del respondente y se aplicaron preguntas de validación.

Tabla III.1. Tipo de Supervisión

TIPO DE SUPERVISIÓN	%
Directa	20%
Posterior	55%
Telefónica	5%
Sin supervisión	20%
TOTAL	100%

m) Reporte de incidencias de campo

El levantamiento de campo se realizó del 06 al 19 de diciembre del 2014.

Para lograr las 3000 entrevistas efectivas se realizaron 11,999 contactos, es decir el 25% de los casos se concretaron con éxito y cumplieron con los criterios de calidad.

Tabla III. 2. Tipo de Incidencias

TIPO DE INCIDENCIAS	Total	%
Entrevistas efectivas y validadas	3000	25.0%
No desea participar	2536	21.1%
Entrevista cortada / cancelada	301	2.5%
Contacto con menor de edad / no respondiente adecuado	590	4.9%
Hogar deshabilitado / abandonado	948	7.9%
Predio vacío	867	7.2%
Sin acceso / Restricción para ingreso	590	4.9%
No abrió	3061	25.5%
Otro	106	0.8%
TOTAL	11,999	100.0%

A continuación se describe cada uno de los tipos de incidencias encontradas en campo:

- Entrevistas efectivas y validadas. Cuando se obtiene información de todos los miembros de hogar y de todas las secciones del cuestionario de manera satisfactoria.
- No desea participar. Cuando se realiza algún contacto con el posible entrevistado y éste expresa que no le interesa o no quiere participar.
- Entrevista cortada/ cancelada. Cuando el entrevistado, una vez aceptada la entrevista, decide ya no continuar por alguna razón. Cabe mencionar que este tipo de incidencias representan menos del 2.5% del total de contactos realizados.
- Contacto menor de edad / no es el respondiente adecuado. Cuando la persona que se encuentra en el hogar aún no cumple la mayoría de edad y por lo tanto no es candidato para la entrevista.
- Hogar deshabilitado / abandonado. Se clasifican las viviendas aparentemente ocupadas pero que al momento de tocar no salió nadie. Se indaga con los vecinos e informan que en este momento nadie la habitaba.
- Predio vacío. Cuando de acuerdo al salto sistemático, el hogar que se debía tocar es un lote vacío.
- Sin acceso / Restricción para ingreso. Cuando el hogar se encuentra en alguna cerrada, edificio, etc. y que no se puede tener un contacto directo con el posible entrevistado.
- No abrió. Cuando después de insistir nadie abrió en el hogar.
- Otros. Cualquier caso no cubierto en los anteriores. Por ejemplo:
 - Se encontraron casos en los que el acceso a la zona de trabajo fue una barrera debido a que ésta era una zona considerada de alto riesgo.

- En zonas rurales no se encontraron datos de identificación del lugar ni nomenclaturas o vías de acceso.
- Se mostró resistencia por parte de las autoridades locales quienes retuvieron a miembros del equipo de campo para indagar sobre la actividad realizada así como el fin de la misma. En ningún caso se registraron incidentes graves.

n) Incidencias en la aplicación del cuestionario

En general no se observa mayor dificultad para la aplicación del instrumento, únicamente en 301 casos de todos los contactos realizados (2.5%) el informante decidió dar por terminada la entrevista sin completar toda la información del cuestionario.

Como marco de referencia, en un levantamiento de campo regular observamos una tasa de abandono promedio del 30% durante los primeros 10 minutos de aplicación, esto quiere decir que 3 de cada 10 entrevistas que comienzan se dan por terminadas antes de que ésta sea finalizada completamente.

Con lo anterior se puede deducir que el instrumento se encuentra en una tasa de abandono muy por debajo de lo normal, lo que pronostica un desempeño muy positivo en términos de esfuerzo para el personal de campo, así como interés por parte del entrevistado y poca percepción de preguntas intrusivas.

En todo caso, a continuación se enumeran las dos principales secciones donde los entrevistados mostraron mayor renuencia a dar información:

- Datos personales de los integrantes del hogar (Ver Sección V. Características Sociodemográficas de los Integrantes del Hogar), particularmente en la pregunta 1 (Nombre de los integrantes del hogar) y pregunta 2 (parentesco con el jefe del hogar).
- De igual forma, la pregunta H4 que solicita proporcionar los artículos con los que cuenta en su hogar (Ver Sección IV. Datos del Hogar).

En ambos casos, los informantes adjudicaron su abandono a razones de seguridad y falta de confianza, la labor del entrevistador para transmitir confianza y certidumbre a los respondentes acerca del operativo permitió que el cuestionario se completara hasta darlo por terminado.

Sección V. Características Sociodemográficas de los Integrantes del Hogar

V. CARACTERÍSTICAS SOCIODEMOGRÁFICAS DE LOS INTEGRANTES DEL HOGAR												
DATOS GENERALES							EDUCACIÓN					
R E N G L Ó N	1. ¿Cuál es el nombre de los integrantes de este hogar, empezando por el jefe o la jefa? (Incluya a los niños chiquitos, a los ancianos, a los huéspedes y también a los trabajadores domésticos, así como a los familiares de éstos que duermen aquí).	2. ¿Qué parentesco tiene (NOMBRE) con el jefe(a) del hogar REGISTRAR 99 SI ES EL JEFE DEL HOGAR	3. ¿Cuántos años cumplidos tiene (NOMBRE)?	4. (NOMBRE) es...		5. ¿(NOMBRE) asiste actualmente a la escuela?		6. ¿La escuela a la que asiste (NOMBRE) es...?		7. ¿Le otorgaron beca de algún programa público a (NOMBRE) para este año escolar?		
				1. Hombre	2. Mujer	1. SÍ CONTINUAR	2. NO PASAR A P.8	1. Pública o de gobierno	2. Privada o de paga	1. SÍ	2. NO	
NO.	NOMBRE	PARENTEZCO	EDAD	GÉNERO		ESCUELA		TIPO DE ESCUELA		BECA		
1				1	2	1	2	1	2	1	2	
2				1	2	1	2	1	2	1	2	
3				1	2	1	2	1	2	1	2	
4				1	2	1	2	1	2	1	2	
5				1	2	1	2	1	2	1	2	
6				1	2	1	2	1	2	1	2	
7				1	2	1	2	1	2	1	2	
8				1	2	1	2	1	2	1	2	
9				1	2	1	2	1	2	1	2	
10				1	2	1	2	1	2	1	2	

Sección IV. Datos del Hogar

IV. DATOS DEL HOGAR			
PREGUNTA	CÓDIGOS DE RESPUESTAS		
H1. ¿Esta vivienda...?	01	Es rentada?	
	02	Es prestada?	
	03	Es propia pero la están pagando?	
RESPUESTA ÚNICA	04	Es propia?	
	05	Está intestada o en litigio?	
	06	Otra Situación Especificar: _____	
H2. ¿Qué tipo de baño o escusado tiene su vivienda?	01	Con conexión de agua	
	02	Le echan agua	
	03	Sin admisión de agua (letrina seca o húmeda)	
RESPUESTA ÚNICA	04	Pozo u hoyo negro	
	05	No tiene SALTAR A PREGUNTA H4	
H3. ¿El baño o escusado es para uso exclusivo de los habitantes de su hogar?	01	SÍ	
RESPUESTA ÚNICA	02	NO	
H4. ¿En su hogar tienen y sirve...?	1. SÍ	2. NO	
	1	2	Refrigerador
	1	2	Lavadora
	1	2	Reproductor VHS, DVD o Blu-Ray
RESPUESTA ÚNICA POR FILA	1	2	Automóvil u otro vehículo propio H4.1 ¿Cuántos automóviles tienen? <input type="text"/>
	1	2	Teléfono fijo
	1	2	Horno de microondas
	1	2	Computadora H4.2 ¿Cuántas computadoras tienen? <input type="text"/>
	1	2	Estufa de gas o eléctrica
	1	2	Calentador de agua (bolíer)
	1	2	Conexión a Internet
	1	2	Televisión de Paga
	1	2	Teléfonos Celulares H4.3 ¿Cuántos celulares tienen? <input type="text"/>

o) Tiempo de aplicación del cuestionario.

Se registra un tiempo promedio de aplicación de 08:29 minutos desde la aceptación a participar en la entrevista por parte del respondente y hasta la terminación del registro de todos los datos solicitados.

- Se registra una desviación estándar de .00190 minutos.
- El 67% de las entrevistas han durado entre 6 y 9 minutos.

p) Captura de entrevistas y construcción de base de datos

Para la construcción de la base de datos, se realizó captura de los cuestionarios en formato electrónico utilizando una plataforma Online destinada para dicho propósito.

En cada plaza la captura inicia un día después del levantamiento, el coordinador de campo y el gerente de operaciones monitorean el avance y realizan una revisión y validación de la base de datos.

q) Factores de expansión.

Se considera como factor de expansión el inverso de la probabilidad de selección de cada unidad final de muestreo con el propósito de hacer extensivas las conclusiones que del estudio se deriven al total de la población.

Los elementos en la muestra seleccionada no tienen la misma probabilidad de selección, para corregir el efecto del diseño de muestreo se pueden emplear los factores anexos. En cada una de las diferentes AGEB en la muestra se considera un diferente ponderador.

Los ponderadores están calculados con base a lo sugerido por el estimador de Hansen-Hurwitz:

$$\hat{t}_{\pi}$$

Para

un total poblacional τ de valores asociados con las unidades en la población dado por:

$$\hat{t}_{\pi} = \sum_{i=1}^v \frac{y_i}{\pi_i}$$

Donde la suma se toma sólo sobre los diferentes v valores de muestra y π_i denota la probabilidad de selección del i -ésimo elemento de la muestra. El estimador de Hansen-Hurwitz aplica con cualquier diseño de muestreo, esto es, con o sin sustitución.

El inverso de la probabilidad de selección π_i corrige efectos no deseados (sesgo) en la estimación del total τ .

Con base a lo anterior, se consideran como ponderadores a valores asociados con elementos en la muestra, los inversos multiplicativos de la probabilidad de selección de cada elemento en la misma.

Los procedimientos de selección son similares, tanto en la AGEB urbana, como en la localidad rural. En localidades rurales, en cada una de las seis ciudades previamente determinadas, se eligen 11 al azar (22% de un total de 50), y después se eligen familias, al azar, dentro de cada localidad elegida. En las urbanas, se eligen primero, aleatoriamente 39 AGEB en cada ciudad, y después, se eligen familias al azar dentro de la AGEB correspondiente.

En la elección de las familias dentro de cada AGEB, en zona rural y urbana, aunque no es estrictamente un muestreo aleatorio simple, se puede decir, que la probabilidad de selección es casi igual al inverso multiplicativo del total de la población de la AGEB.

De acuerdo a lo anterior calculamos los factores de expansión (probabilidades de selección) con base al procedimiento descrito a continuación, en AGEB urbana y localidad rural.

- Para localidad rural

En caso de la localidad rural, por ejemplo, dentro del Distrito Federal y su área conurbada, existen 276 AGEB rurales con población de al menos 20 habitantes (se descartan las de menor tamaño para evitar problemas con la tasa de no respuesta).

Entonces la probabilidad de estar en la muestra mencionada de 11, para cada una de estas 276 AGEB está dada por:

$$C(1, 1) \times C(275, 10) / C(276, 11) = 11/276$$

Luego, por ley multiplicativa de la probabilidad, en la AGEB del municipio de Tecámac en la localidad de la base con población de 33 (incluida en la muestra) la probabilidad de selección en cada uno de sus elementos es:

$$(11/276) \times (1/33) = 0.0012$$

Finalmente, el factor de expansión, considerado como el inverso de la probabilidad de selección, es igual a $1/0.0012 = 828$ (incluyendo todos los decimales en las operaciones anteriores).

- Para AGEB urbana

En zona urbana, por ejemplo, en Tuxtla Gutiérrez, existen 198 AGEB con población de al menos 20, (se descartan las de menor tamaño para evitar problemas con las no respuestas) luego, la probabilidad de estar en la mencionada muestra de 39, para cada una de estas 198 AGEB está dada por:

$$C(1, 1) \times C(197, 38) / C(198, 39) = 39/198$$

Luego, por ley multiplicativa de la probabilidad, en la AGEB 2637, con población de 340 (incluida en la muestra) la probabilidad de selección en cada uno de sus elementos es:

$$(39/198) \times (1/340) = 0.000579$$

Finalmente, el factor de expansión, considerado como el inverso de la probabilidad de selección, es igual a $1/0.000579 = 1726.15$.

Los demás ponderadores se calculan con el mismo procedimiento y se resumen en el archivo adjunto (Ver documentos: Factores de ponderación AGEB y Factores de ponderación Localidades).