

Asociación Mexicana de Agencias de Inteligencia de Mercado y Opinión

Nivel Socio Económico AMAI 2018

Nota Metodológica

Comité de Nivel Socioeconómico AMAI
Noviembre de 2017

CONTENIDO

INTRODUCCIÓN.....	3
PASOS PREVIOS AL MODELAJE	3
AJUSTE DEL MODELO	6
VALIDACIÓN DE RESULTADOS.....	12
CUESTIONARIO PROPUESTO.....	20
CONSIDERACIONES FINALES.....	23
ANEXOS.....	24

INTRODUCCIÓN

Más de 20 años han transcurrido desde la primera iteración del índice de Nivel Socioeconómico de AMAI. A lo largo de este tiempo, su practicidad y dinamismo le han ayudado a consolidarse como la herramienta de segmentación y clasificación más utilizada por la industria de investigación en el país. A lo largo de sus más de dos décadas de existencia, ha sido empleado por compañías de investigación, consultores y especialistas de mercadotecnia y comunicación, así como por agencias de publicidad, medios, marcas, empresas e instituciones oficiales.

Por su naturaleza de aplicabilidad es importante que el índice esté en constante actualización. Los componentes que le integran y las definiciones de los mismos, deben revisarse periódicamente para asegurar que estén brindando una adecuada interpretación de la dinámica del hogar mexicano, así como para valorar su pertinencia y otorgarle a cada uno un justo peso y una dimensión. Con base en este objetivo, el Comité de Nivel Socioeconómico se dio a la tarea de revisar el modelo vigente a Enero de 2017, la regla AMAI 8x7. Identificando las fortalezas y oportunidades de la regla, el comité emprendió la compleja tarea de efectuar una revisión detallada de la misma.

El presente ejercicio es el esfuerzo del Comité para documentar este proceso y, al mismo tiempo, hacer extensivo a todos los miembros AMAI y a la sociedad en general el conocimiento y aprendizajes generados durante esta etapa.

PASOS PREVIOS AL MODELAJE

Previo a comenzar el proceso de modelaje estadístico que llevaría a la actualización de la regla de Nivel Socioeconómico AMAI, el Comité sesionó numerosas ocasiones para definir una serie de estrategias para asegurar que el modelo generado cumpliera las expectativas del Comité.

Una de las primeras decisiones que se tomaron fue trabajar con fuentes oficiales disponibles y actuales. Al igual que la versión vigente de la regla AMAI, la nueva versión del modelo será generada a partir de los resultados oficiales de la Encuesta Nacional de Ingresos y Gastos de los Hogares¹, encuesta realizada por el Instituto Nacional de Estadística y Geografía mexicano (INEGI). Este estudio es una encuesta en hogares que realiza el INEGI cada dos años, con una muestra representativa a nivel nacional y por entidad federativa. Para la edición de 2014 la muestra estuvo conformada por más de 21,000 viviendas en el territorio nacional.

El estudio es especialmente robusto para entender los hábitos de consumo y gasto de los hogares mexicanos. Existe información relevante del gasto del hogar en cientos de categorías, que permiten describir a detalle el destino del dinero de las familias del país. Adicionalmente, La ENIGH contiene módulos en el cuestionario que registran no sólo información del gasto y el ingreso en los hogares,

¹ Encuesta Nacional de Ingresos y Gastos de los Hogares 2014. INEGI
<http://www.beta.inegi.org.mx/proyectos/enchogares/regulares/enigh/tradicional/2014/default.html>

sino también características de infraestructura de la vivienda que ocupan, características de los integrantes del hogar e incluso algunos datos interesantes acerca de sus hábitos de alimentación.

La ENIGH es la medición más confiable que se tiene en México del ingreso ya que cuenta con una metodología rigurosa para su medición, lo que la hace especialmente útil para estimaciones relacionadas con la pobreza y el bienestar social. Adicionalmente, la última edición de la encuesta, ENIGH 2016, sería publicada a mediados de 2017, así que los datos más recientes podrán ser utilizados para validar los resultados del modelo, como se muestra más adelante.

Habiendo establecido la fuente de datos a utilizar, una decisión importante del Comité fue definir la metodología a seguir para la generación del modelo. A lo largo de las distintas iteraciones de la regla AMAI, se han probado diversos tipos de modelos, incluyendo modelos de regresión, clasificación o árboles de clasificación. Después de estudiar diversas metodologías, se decidió continuar utilizando un modelo de regresión que permitiera generar un sistema de puntos. La decisión recayó en la simplicidad y practicidad asociada con el sistema de puntos, el cual no solo es sencillo de entender, sino también ágil de aplicar, incluso por personal en campo. El uso de una metodología similar a la utilizada para la creación de la regla AMAI 8x7 tendría además la ventaja de facilitar los ejercicios comparativos con el modelo vigente.

Otra decisión importante fue la de evaluar la propuesta de un modelo nuevo, independiente a la regla vigente. Esto implica que las variables que son parte de la regla AMAI 8x7 no iban a considerarse de forma predeterminada en esta nueva propuesta, sino que cada uno de los indicadores iba a evaluarse de forma individual, junto con otros componentes de la ENIGH. A la luz de los resultados, esta estrategia permitió validar y confirmar la regla AMAI 2018 e iteraciones anteriores.

Finalmente, el Comité tuvo que acordar uno de los temas principales del proceso de creación de la regla: la definición de la variable respuesta del modelo. Esta decisión no trivial requirió de un análisis descriptivo que permitiera identificar las variables en la ENIGH 2014 que presentaran una asociación importante con el concepto de Nivel Socioeconómico existente. Entre las variables analizadas se consideraron las Percepciones Totales, las Erogaciones Totales y el Ingreso Corriente.

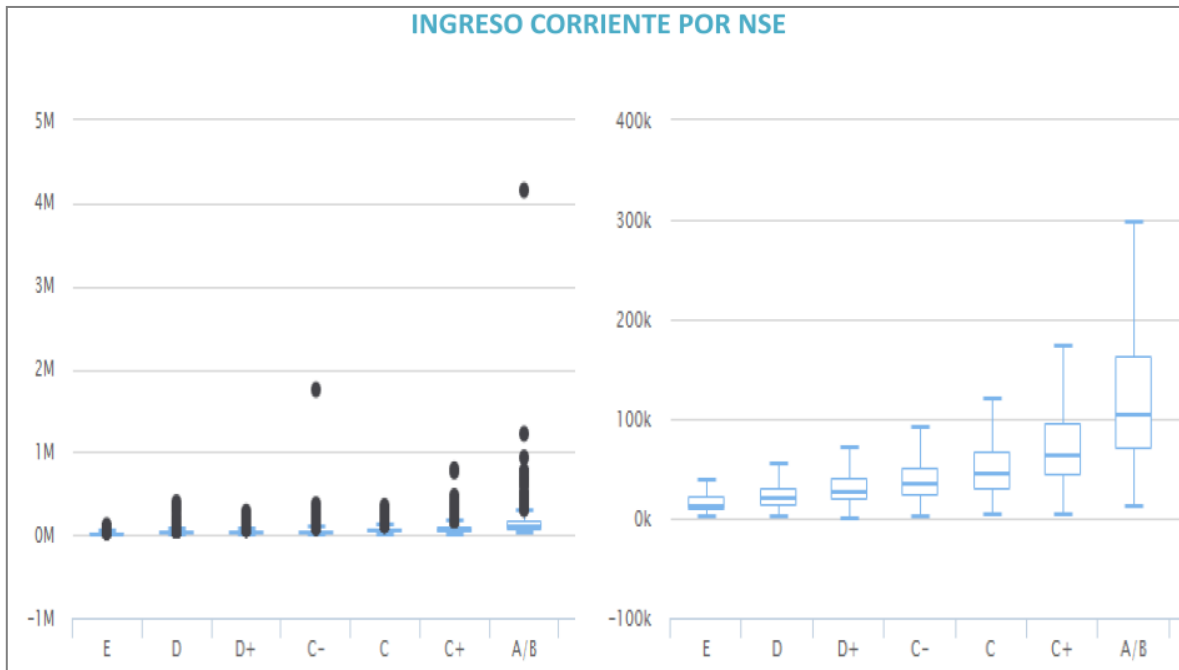


Figura 1. Asociación entre el Nivel SocioEconómico resultante de la regla AMAI 8x7 y el Ingreso Corriente Trimestral. El gráfico del lado derecho excluye los *outliers*.

De las variables analizadas, el Ingreso Corriente fue la que presentó la relación más fuerte con los Niveles Socioeconómicos. El Ingreso Corriente es definido por INEGI como la suma de los ingresos por trabajo, los provenientes de rentas, de transferencias, de estimación del alquiler y de otros ingresos. Se trata de un componente importante de los ingresos del hogar aunque no incluye la totalidad de los ingresos, sino que se refiere principalmente a los ingresos constantes de los integrantes de la familia.

Esta variable no solo está fuertemente asociada con los niveles socioeconómicos entonces vigentes, sino también con los distintos indicadores que integran la regla AMAI 8x7. Hay una importante relación entre el ingreso corriente y el número de cuartos en la vivienda, el número de baños completos, el número de focos y también con el nivel educativo del jefe del hogar. Estas correlaciones, así como la asociación natural que se esperaría entre el nivel socioeconómico y parte de los ingresos del hogar, es lo que convirtió a esta variable en una candidata natural para el análisis de regresión que habría de generar la nueva regla AMAI.

Una vez que se estableció al ingreso corriente como la variable a modelar, se realizó un análisis de su distribución. Como suele ocurrir con las variables de ingreso, debido a la desigualdad de ingresos en la sociedad, la variable Ingreso Corriente presenta una distribución sesgada a la derecha. Aunque esto no es un problema inherente para el modelo, se realizó una transformación logarítmica para evitar problemas con la distribución de residuales y con otros supuestos. Adicionalmente, tres datos presentaban ingresos inusualmente altos en comparación con el resto de los hogares reportados por INEGI, los cuales no fueron considerados para el ajuste del modelo.

A diferencia de ejercicios anteriores, para esta regla **se consideraron todos los hogares** a nivel nacional, incluyendo los hogares en localidades rurales.

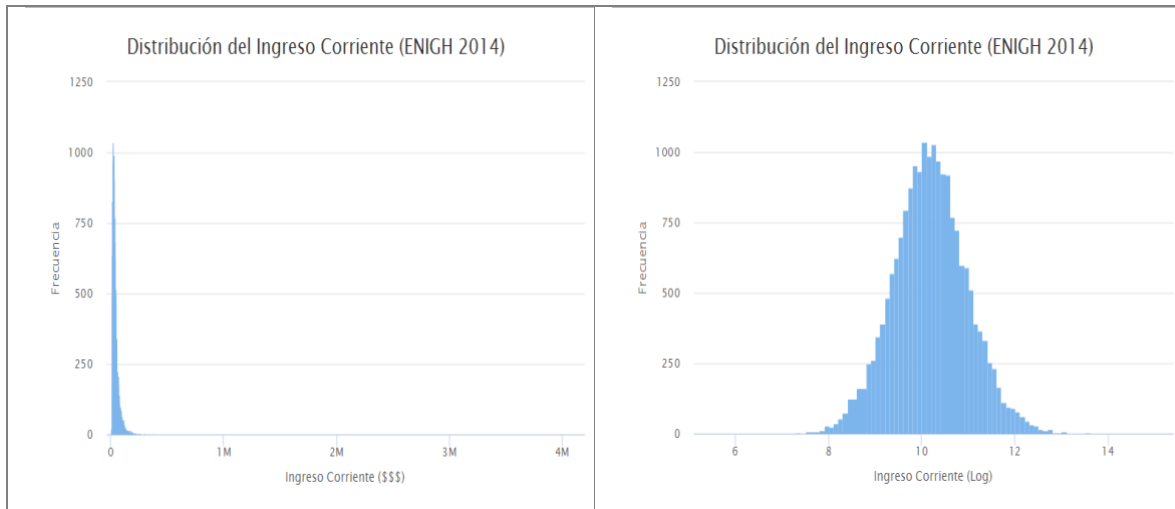


Figura 2. Distribución del Ingreso Corriente (Izquierda) y distribución del logaritmo natural del ingreso corriente (Derecha) reportado en ENIGH 2014.

AJUSTE DEL MODELO

Como se explicó, la principal fuente de información para el ajuste del modelo fue la base de datos de la Encuesta Nacional de Ingresos y Gastos de los Hogares 2014. Antes de iniciar con el proceso de ajuste del modelo, fue necesario realizar un procesamiento menor para llegar a la base de datos final.

La base de ENIGH 2014 disponible en el sitio web de INEGI se compone de distintas bases de datos, de acuerdo a los diferentes módulos del cuestionario aplicado. Para realizar el análisis, se realizaron pruebas de asociación del Ingreso Corriente con diversas variables contenidas en distintos módulos, por lo que fue necesario unir todas las bases de acuerdo a las llaves de identificación de los folios de hogar y/o vivienda. El modelo ajustado al final contiene variables de tres tablas distintas: la tabla VIVIENDAS, la tabla HOGARES y la tabla CONCENTRADOHOGAR. Cualquier persona interesada en replicar este resultado requeriría únicamente de juntar estas tres bases de datos con las llaves correspondientes.

Algunas variables utilizadas no existen en forma directa en la base de datos y se crearon a partir de variables existentes para el modelo. Por ejemplo: la variable "autos" se creó como la suma del total de vehículos en el hogar y se compone de la suma de las variables "num_auto", "num_van" y "num_pickup" dentro de la tabla de HOGARES. La variable "internet" se crea a partir de la variable "conex_inte" y únicamente convierte en dicotómica la variable original e invierte la escala, de forma que los hogares con internet tengan un valor más alto lo que permite que el coeficiente asociado sea positivo. La variable "edu_jefe" es "educa_jefe" pero definida como factor dentro del entorno, a fin de generar en el modelo un coeficiente por cada nivel de educación. Tal como se explicó en el capítulo anterior, se realizó una transformación al Ingreso Corriente, consistente en la aplicación del logaritmo natural de esta variable. Esta variable se definió como "ln_ing" y es la variable que se usa como respuesta en el modelo.

El modelo ajustado para la predicción del logaritmo natural de Ingreso Corriente es un modelo de regresión ajustado por Mínimos Cuadrados Ponderados, a fin de considerar el factor de expansión asociado a cada uno de los hogares en la ENIGH. Todos los errores estándar estimados fueron ajustados de acuerdo al diseño muestral.

Como primer ejercicio, se corrió un modelo ajustando el ingreso corriente transformado con las variables de la regla AMAI vigente 8x7. Este modelo conforma la base para hacer el comparativo y con ello tener una idea de la ganancia o pérdida de precisión, en términos de explicación del Ingreso Corriente, que tendríamos con el nuevo modelo.

Los resultados del modelo inicial y el código en R que lo genera se presentan en la sección de [Anexos](#) de este documento. Una evaluación visual del ajuste del modelo se muestra a continuación.

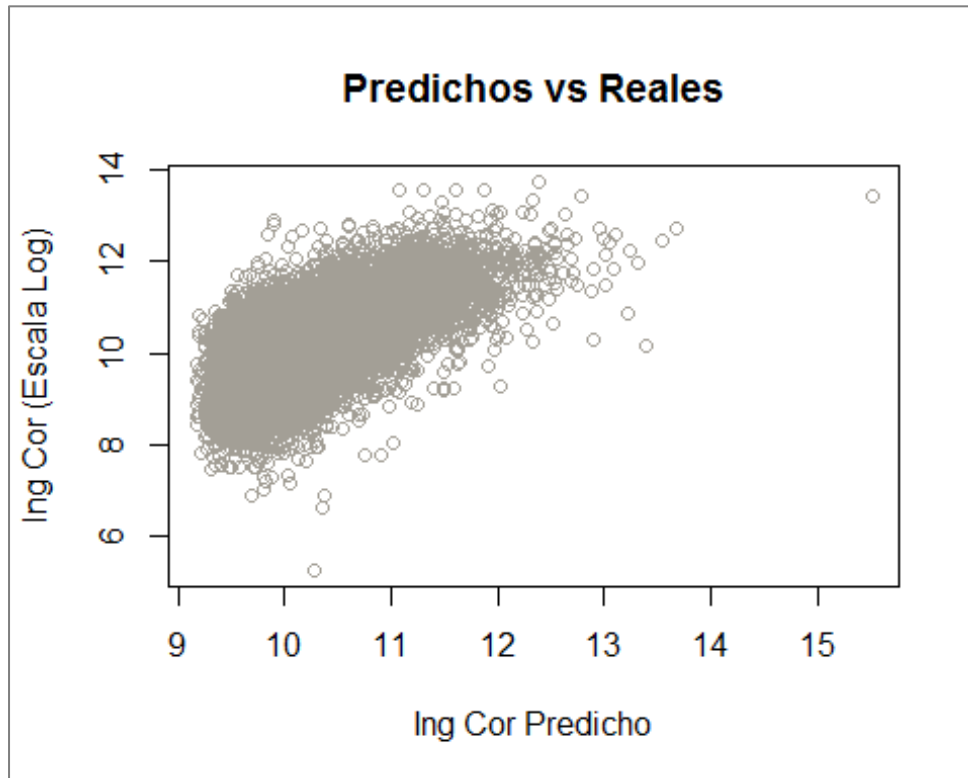


Figura 3. Asociación entre los valores predichos y el logaritmo natural del ingreso corriente reportado en ENIGH 2014. Resultado de ajustar un modelo predictivo con las variables de la regla AMAI 8x7

Los resultados de este modelo son la base comparativa que permite establecer el mínimo poder predictivo y ajuste esperado para la nueva regla. Cualquier nueva propuesta sólo sería aceptada si el nuevo modelo mejorara de forma significativa los criterios de bondad de ajuste presentados o bien si los igualaba utilizando un número menor de variables. Al final, ambos objetivos se alcanzarían.

Como primer paso para encontrar la nueva propuesta del modelo de nivel socioeconómico, se realizó un análisis descriptivo de asociaciones, buscando identificar variables que presentan correlaciones interesantes que pudieran ser probadas en el modelo predictivo. Aunque no se presentan a detalle todos los resultados de este análisis, cabe mencionar su importancia dentro del proceso de ajuste de este y cualquier modelo.

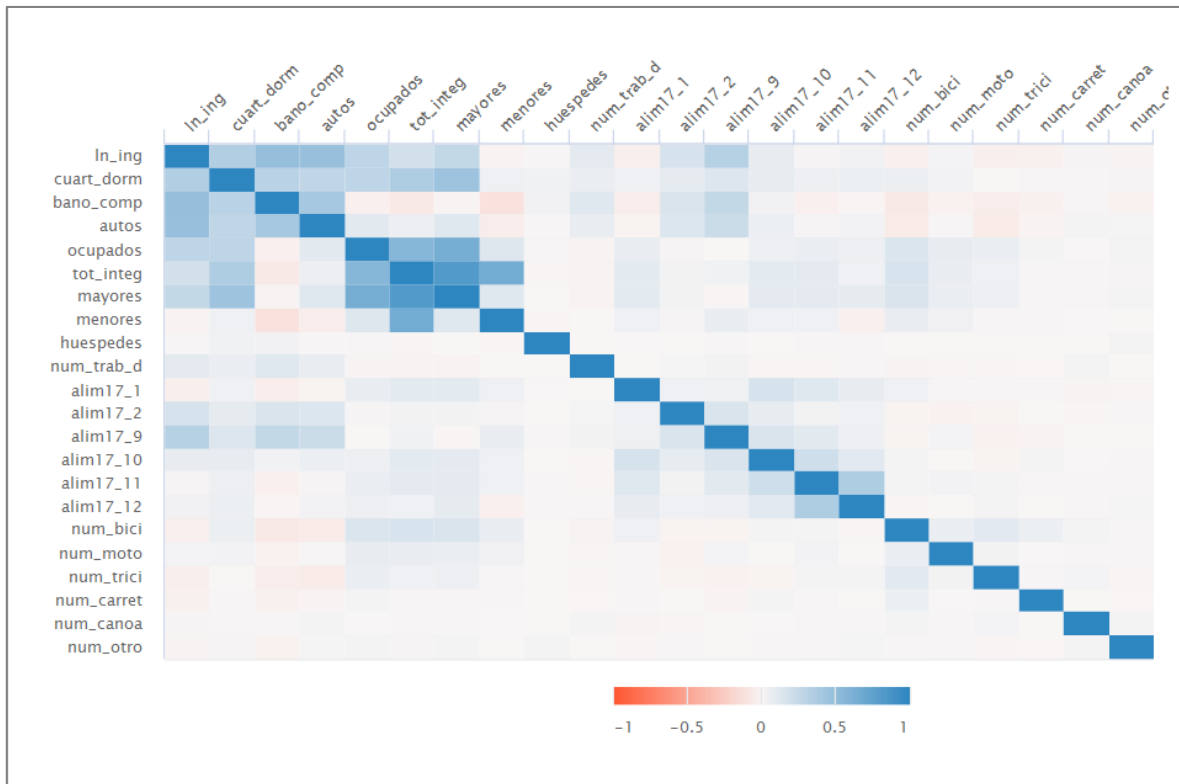


Figura 4. Representación Gráfica de una matriz de correlaciones entre el logaritmo natural del ingreso corriente y una serie de variables incluidas en el cuestionario de ENIGH 2014.

Entre las variables probadas se consideraron variables de la infraestructura básica de las viviendas como el material de las mismas, el número de habitaciones y dormitorios, la estructura del hogar, variables referentes a los hábitos alimenticios de los integrantes, la tenencia de distintos enseres como horno de microondas, refrigeradores o calentadores de gas, entre otros. En total, más de 50 posibles variables predictoras fueron evaluadas en algún momento.

Después del riguroso análisis descriptivo y tras varias iteraciones del modelo, **6 variables** resultaron presentar el mayor poder predictivo:

- Nivel educativo del jefe de hogar
- Número de baños completos en la vivienda
- Número de autos en el hogar (entendida como la suma de autos, vans y pick ups en el hogar)
- Tenencia de conexión a internet en el hogar
- Número de integrantes en el hogar mayores de 14 años que trabajan
- Número de dormitorios en la vivienda

Estas variables mostraban las asociaciones más claras con el ingreso corriente de los hogares. Las tres primeras son variables ya presentes en la regla AMAI 8x7

De estas variables, dos fueron particularmente causa de debate dentro del Comité. La primera fue la presencia de conexión a Internet; esta variable presenta una muy importante asociación con el ingreso corriente y, como más adelante se verá, su poder predictivo fue particularmente especial para discriminar los hogares de más bajo nivel socioeconómico, de forma similar a la discriminación que el material del piso de la vivienda lograba en versiones anteriores del modelo.

Esta variable podría estar sujeta a un cambio drástico en el futuro, puesto que la penetración de internet en el país ha tenido un incremento considerable en los últimos años. Este aspecto preocupaba debido a que su utilidad para el modelo de nivel socioeconómico podría estar limitada por esta posible volatilidad. Sin embargo, se decidió mantenerla debido a señales de estabilización en su comportamiento en las últimas mediciones y sobretodo porque uno de los objetivos del nuevo modelo es hacerlo más versátil con revisiones periódicas.

Otra variable que fue motivo de discusión fue la del número de integrantes ocupados. En este caso era una preocupación que el modelo clasificara como niveles socioeconómicos altos los hogares que presentaran un alto número de integrantes trabajando por el simple hecho de que se tratara de hogares más grandes, cuando, en general, hogares con muchos integrantes suelen tener condiciones de vida más difíciles.

Dado que la preocupación era válida, se probaron diversos mecanismos para evitar esta situación en el modelo final. Un primer acercamiento fue considerar no el número de integrantes ocupados sino la proporción de los mismos, es decir, el número de integrantes ocupados entre el total de integrantes del hogar. Sin embargo, esta alternativa reducía el nivel de ajuste del modelo, además de requerir de dos preguntas en el cuestionario para calcular el índice, lo que complica la aplicación en campo.

Una segunda estrategia para evitar esta afectación fue utilizar el total de integrantes como variable control en el modelo. El incluir el total de integrantes en el modelo, modificaría los coeficientes de las variables que estuvieran fuertemente asociadas con el total de habitantes del hogar, aislando el impacto de este factor y, en teoría, evitando que el modelo diera un peso exagerado a este componente de la nueva regla. Esta estrategia no sólo corrigió el coeficiente del número de integrantes ocupados, sino también el de número de cuartos, sin generar una pérdida del ajuste. Por lo anterior, se decidió utilizar el modelo que controla por el total de integrantes, aunque este indicador no sería usado para la regla final.

Después de numerosas iteraciones, se llegó a una propuesta final. Cabe mencionar que se probaron más de una docena de variaciones del modelo, incluyendo pruebas para buscar efectos no lineales e interacciones, sin embargo, se descartaron para la última versión puesto que no mejoraban de forma significativa el modelo. El modelo final ajustado y el código en R que lo genera se presentan en la sección de [Anexos](#) de este documento. Una inspección visual del ajuste se presenta a continuación.

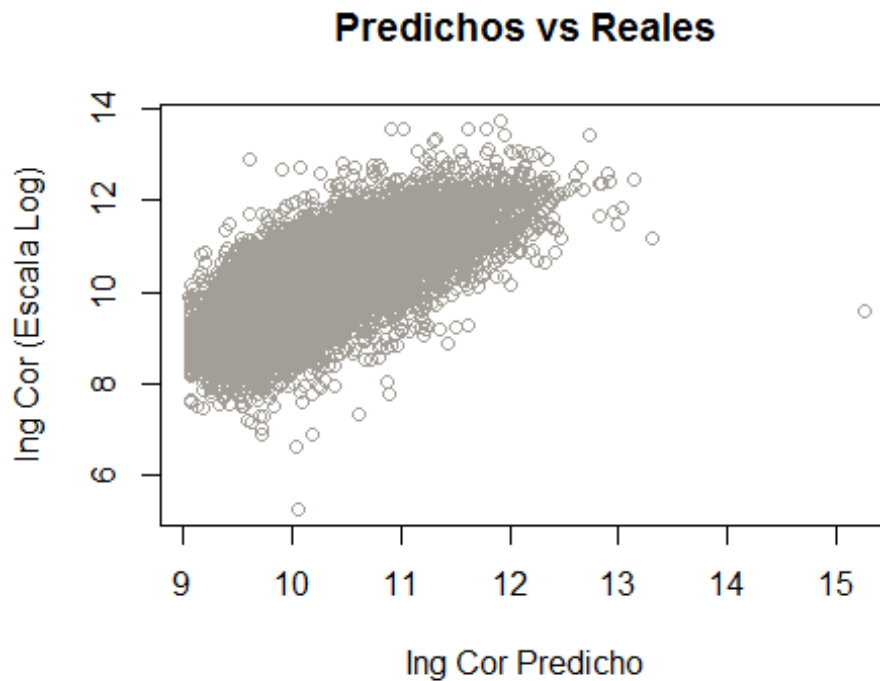


Figura 5. Asociación entre los valores predichos y el logaritmo natural del ingreso corriente reportado en ENIGH 2014. Resultado de ajustar un modelo predictivo con las variables de la regla AMAI 2018

Como se observa, el modelo identificado consta de menos variables y los indicadores de bondad de ajuste presentan un mejor valor. El AIC de la propuesta es significativamente menor que el obtenido con las variables originales, lo mismo que la *Deviance* del Modelo. Una estimación conservadora de la R cuadrada, de acuerdo a la relación de la *Deviance* del modelo ajustado en relación a la *Deviance* del modelo nulo, es también significativamente mayor para la propuesta AMAI 2018.

Considerando que los diversos criterios de bondad de ajuste coinciden en que, hablando en términos del Ingreso Corriente, la nueva propuesta del modelo brinda un ajuste superior para los hogares mexicanos en 2014 que las variables vigentes en la regla AMAI 8x7.

Una vez verificado que el modelo realmente mejora el ajuste en relación al que incluía solamente las variables de la regla AMAI 8x7, se procedió a utilizar estos resultados para generar los puntajes que serían la base de la nueva regla AMAI 2018.

Al igual que como se hizo para la definición del modelo, en la determinación de los puntajes se optó por usar un procedimiento independiente al usado en versiones anteriores de la regla. El procedimiento utilizado para asignar los puntos a los distintos componentes de la regla fue el siguiente:

1. Se definió un número total de puntos máximo, valor que podría ser arbitrario. Se acordó utilizar un puntaje máximo de **300** puntos, dado que es un número redondo cercano al puntaje máximo de la regla AMAI 8x7.

2. Este puntaje máximo se repartió de forma **proporcional** a las categorías más altas de cada variable utilizada en la regla, definiendo esta categoría más alta no como el máximo sino de acuerdo al **percentil 95** de la distribución de cada una de las variables en la base de ENIGH 2014. Esto se decidió para evitar que los datos atípicos presentes en la distribución de las variables incorporadas al modelo pudiese sesgar los resultados.
3. Habiendo asignado proporcionalmente el puntaje máximo a cada variable, los puntos se repartieron también **proporcionalmente**, entre las categorías inferiores de dicha variable.

Baños Completos		Número de dormitorios		Número de ocupados	
RESPUESTA	PUNTOS	RESPUESTA	PUNTOS	RESPUESTA	PUNTOS
0	0	0	0	0	0
1	24	1	6	1	15
2 ó más	47	2	12	2	31
		3	17	3	46
		4 ó más	23	4 ó más	61
Número de Autos		Internet		Educa Jefe	
RESPUESTA	PUNTOS	RESPUESTA	PUNTOS	RESPUESTA	PUNTOS
0	0	No tiene	0	No estudió	0
1	18	Si tiene	31	Primaria incompleta	10
2 ó más	37			Primaria completa	22
				Secundaria incompleta	23
				Secundaria completa	31
				Carrera comercial	35
				Carrera técnica	35
				Preparatoria incompleta	35
				Preparatoria completa	43
				Licenciatura incompleta	59
				Licenciatura completa	73
				Diplomado o Maestría	101
				Doctorado	101

Figura 6. Puntajes Asociados a cada variable dentro del modelo AMAI 2018.

4. De esta forma, se tienen los puntajes asignados a cada una de las variables y a sus distintos niveles. Con estos puntajes, se obtuvo la suma de puntos para todos los hogares dentro de la muestra de ENIGH 2014. Posteriormente, se utilizó el procedimiento de estratificación univariado de **Dalenius-Hodges**, (*Cumulative root frequency method*) para obtener los puntos de corte que minimizan la variabilidad intra-grupos.
5. Se probaron distintas soluciones con diversos números de grupos. Aunque el análisis presentó evidencia como para sugerir la partición del nivel D en dos subgrupos, se definió continuar

utilizando un corte de **7 niveles socioeconómicos** dado que la partición de este nivel implica la definición conceptual de un nuevo grupo.

Puntos de corte del NSE	
A/B	205+
C+	166 a 204
C	136 a 165
C-	112 a 135
D+	90 a 111
D	48 a 89
E	0 a 47

Figura 7. Puntos de corte definidos para la clasificación de Niveles Socioeconómicos de la regla AMAI 2018

De esta manera, se definieron los puntos que serían la base del cálculo para la nueva regla AMAI 2018. Al igual que en la última versión, de acuerdo a las respuestas del entrevistado, se asignará una serie de puntos que al final se sumará. Dicha suma será contrastada con los puntos de corte definidos para asignar el respectivo hogar a su nivel socioeconómico correspondiente.

VALIDACIÓN DE RESULTADOS

Durante todo el proceso de modelaje, análisis de variables y cálculo de puntajes, se realizaron numerosas validaciones de la consistencia y confiabilidad de los resultados. Ya se describió cómo los criterios de bondad de ajuste se utilizaron para definir si existía una mejora importante en el nivel predictivo del modelo: adicionalmente, se realizaron otro tipo de validaciones, tanto cualitativas como cuantitativas, para verificar que la nueva propuesta brindara resultados consistentes y coherentes al aplicarse en la práctica.

Una de las validaciones básicas fue verificar que todas las variables del modelo se pudieran clasificar dentro de las dimensiones que considera AMAI bajo su concepto de nivel socioeconómico. Es importante recordar que la AMAI define al Nivel Socio Económico como el nivel de bienestar que tiene un hogar y que todos sus miembros comparten, y que permite determinar qué tan cubiertas están las necesidades de espacio, sanidad, practicidad, entretenimiento, comunicación y planeación y futuro en un hogar. De esta manera, un primer punto era verificar que los distintos indicadores estuvieran relacionados con estos aspectos, algo que se cumple pues la nueva regla considera estas necesidades, incluyendo las de entretenimiento, pues en la actualidad la disponibilidad de internet se asocia fuertemente con este concepto.

Ya se comentó cómo el Comité debatió intensamente la relevancia y validez de las distintas variables a incorporar en la nueva regla AMAI 2018. Adicionalmente, se realizaron una serie de validaciones cuantitativas para asegurar que los resultados del modelo fueran coherentes con la realidad de los hogares mexicanos. Aprovechando que la última versión de la Encuesta Nacional de Ingreso y Gasto

en Hogares (ENIGH 2016) estaría disponible poco tiempo después de finalizar las tareas de modelaje, se decidió utilizar la nueva regla para generar distribuciones de niveles socioeconómicos usando esta nueva información. Cabe hacer énfasis en que los datos de la ENIGH 2016 usados para validar la información de ninguna manera fueron usados para generar la regla, por lo que las distribuciones calculadas son un buen indicio de la fiabilidad de los resultados del modelo.

Una vez disponible la base de datos de la ENIGH 2016, un primer ejercicio fue comparar las distribuciones de los Niveles Socioeconómicos generados bajo la regla AMAI 2018 con las distribuciones obtenidas con la ENIGH 2014 con la regla 8X7. Dicho comparativo, a nivel nacional, se muestra en la Figura 8

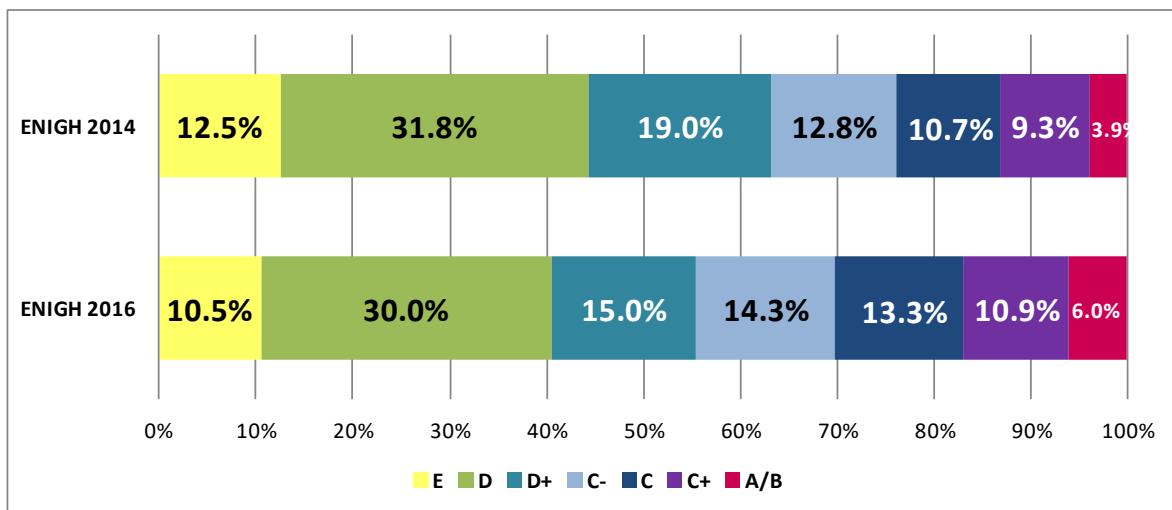


Figura 8. Comparativo de la distribución nacional de niveles socioeconómicos. Regla AMAI 2018 y regla AMAI 8x7.

La distribución de los niveles socioeconómicos de la regla AMAI 2018 es consistente con la distribución generada con ENIGH 2014, las principales diferencias se presentan en los niveles socioeconómicos medios y medios altos donde hay una redistribución de los hogares mexicanos.

Este comportamiento se presenta también al generar las distribuciones de acuerdo a tamaños de localidad. Cabe mencionar que, dado que el modelo considera también hogares rurales, es posible obtener estimaciones para las localidades de menor tamaño.

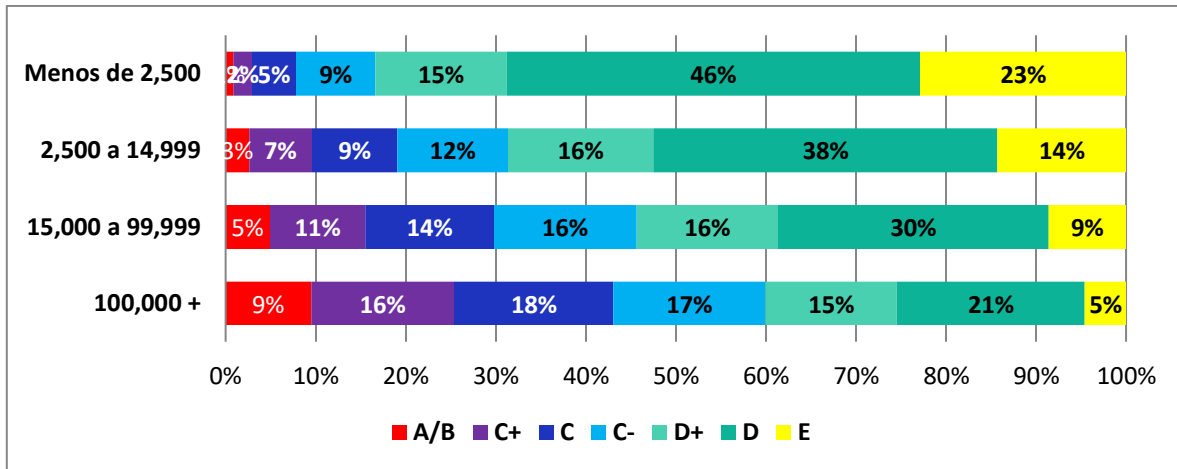


Figura 9. Distribución de niveles socioeconómicos de la regla AMAI 2018 por tamaño de localidad.
Estimación a partir de la ENIGH 2016

Los resultados por tamaño de localidad son consistentes con el comportamiento esperado de acuerdo a la naturaleza de los hogares. En las zonas rurales los niveles socioeconómicos A/B y C+ son mucho menos usuales que en las zonas urbanas.

Otro comparativo importante era evaluar las distribuciones de los niveles socioeconómicos generados por la nueva regla en las distintas entidades federativas del país. Se esperaba ver diferencias marcadas por estado, con algunas zonas presentando números más altos de algunos niveles socioeconómicos. Los resultados de esta validación, se muestran a continuación.

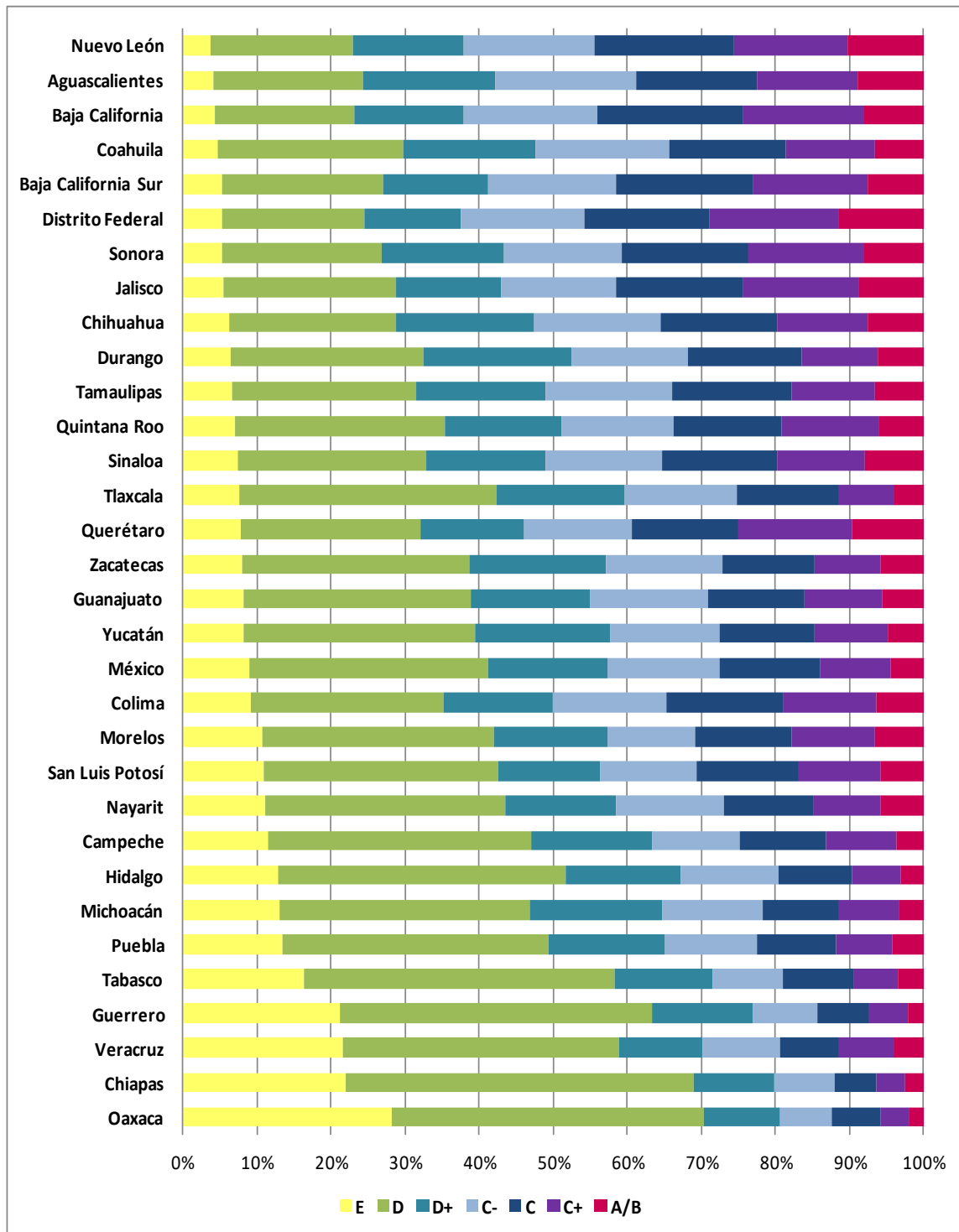


Figura 10. Distribución de niveles socioeconómicos de la regla AMAI 2018 por entidad. Estimación a partir de la ENIGH 2016

Otro aspecto relevante a validar es la distribución de los niveles socioeconómicos generados por la regla AMAI 2018 en las principales zonas metropolitanas del país, definidas de acuerdo a los criterios establecidos por el CONAPO.

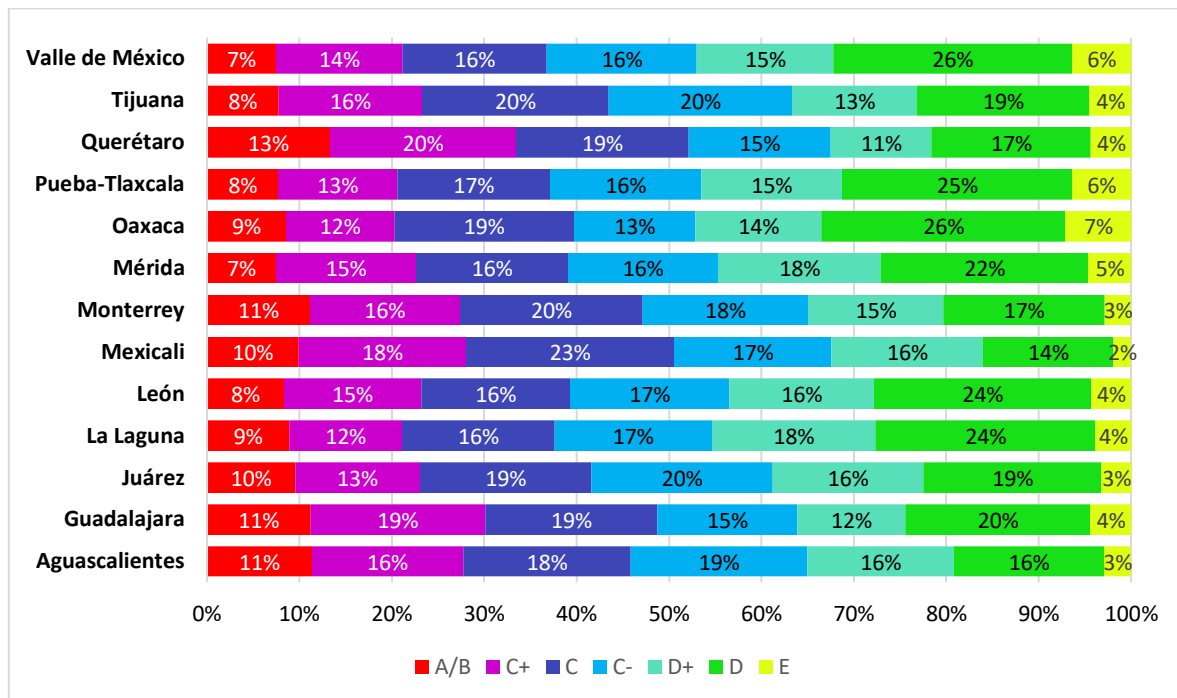


Figura 11. Distribución de niveles socioeconómicos de la regla AMAI 2018 para un grupo de zonas metropolitanas del país. Estimación a partir de la ENIGH 2016

Una vez que se determinó que los resultados de la distribución de los niveles socioeconómicos era coherente y consistente con la realidad del país, incluso en niveles desagregados, validaciones adicionales se realizaron para verificar que distintas variables, tanto consideradas en el modelo como no consideradas, tuvieran un comportamiento congruente con las definiciones y características de los niveles socioeconómicos.

Como era de esperarse, hay una correlación importante entre el nivel socio económico estimado y el ingreso corriente de los hogares, en una relación que tiene sentido con la definición de calidad de vida de los distintos niveles socioeconómicos.

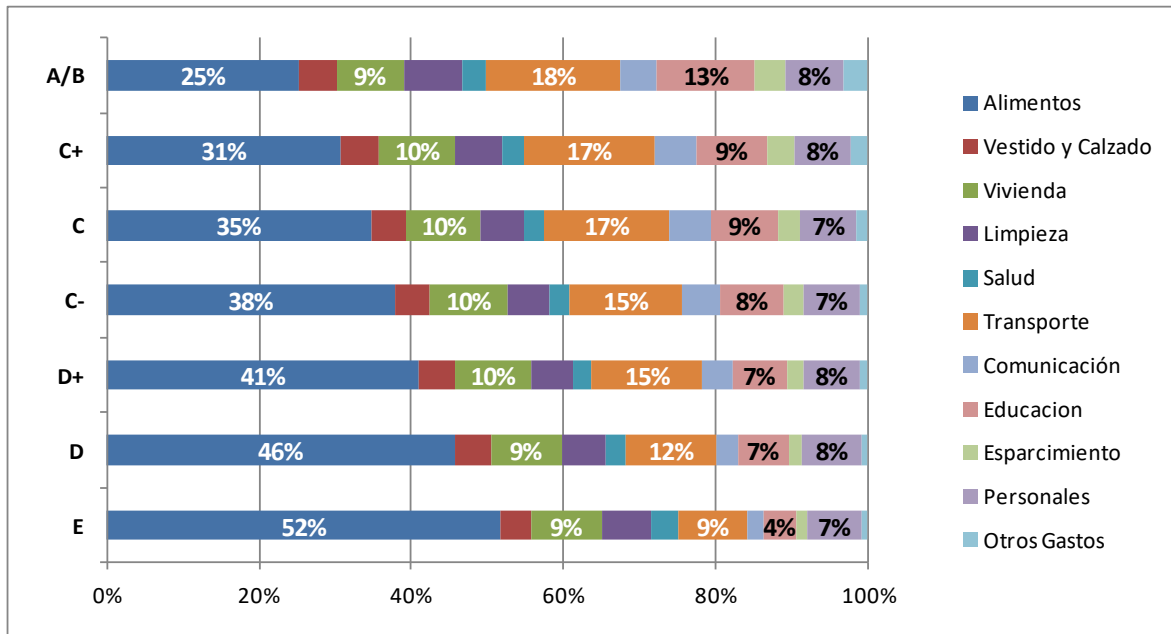


Figura 12. Distribución de gasto de acuerdo a los niveles socioeconómicos de la regla AMAI 2018. Estimación a partir de la ENIGH 2016

De igual forma en la distribución de gasto se identificó un comportamiento consistente con la definición de niveles socioeconómicos.

Otra validación a realizar es el comparativo de las distribuciones de diversos indicadores de la ENIGH en los distintos niveles socio económicos. En primer lugar se comparó el comportamiento de los indicadores que integran la nueva regla AMAI.

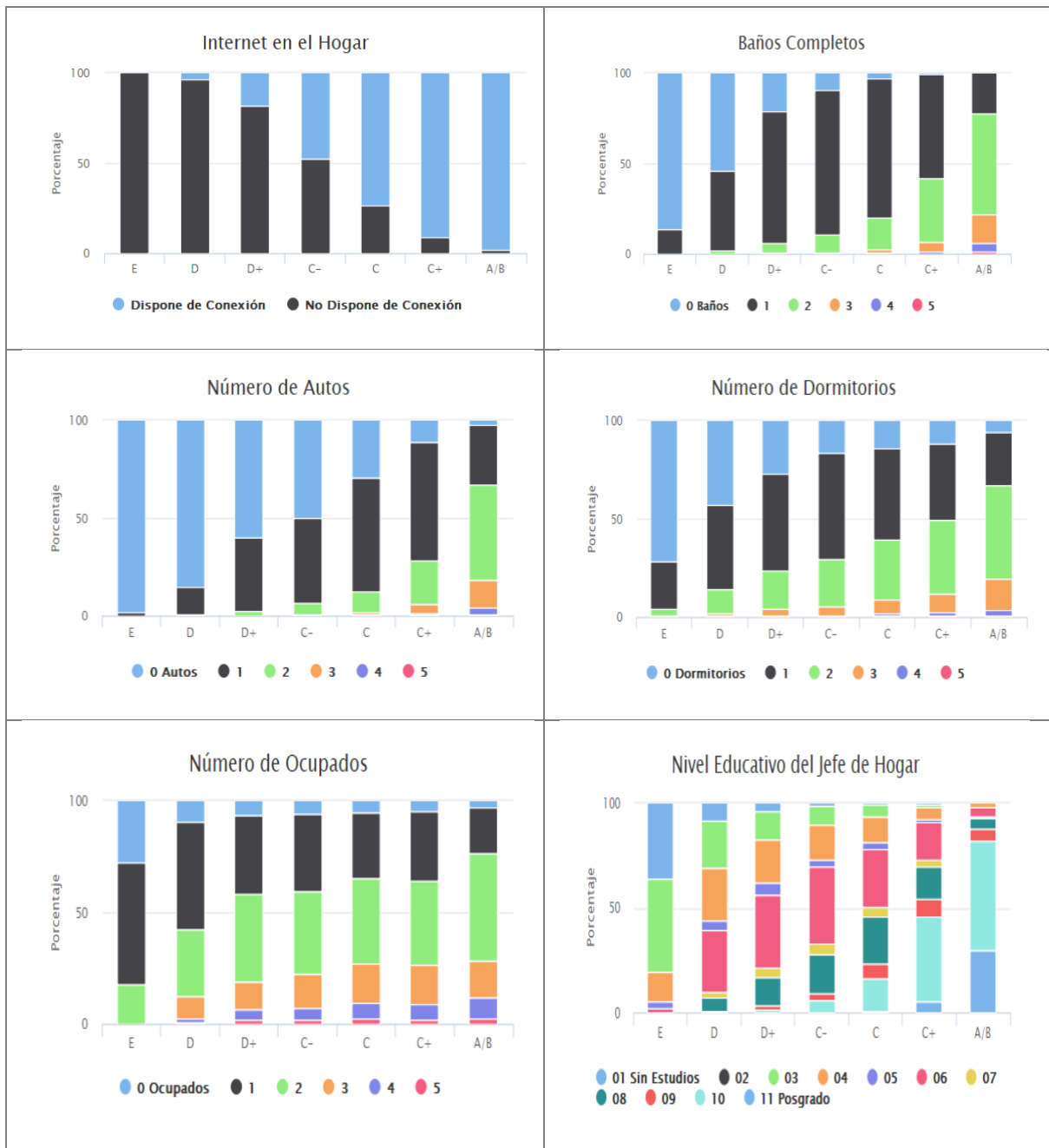


Figura 13. Distribución de las variables del modelo de acuerdo a los niveles socioeconómicos de la regla AMAI 2018. Estimación a partir de la ENIGH 2016

La distribución de los diferentes indicadores utilizados en la regla AMAI 2018 refleja una correlación fuerte entre estas variables y los niveles socioeconómicos, como habría de esperarse. Destaca que el número de integrantes ocupados no tiene un impacto tan marcado en la definición de los NSE, debido al ajuste que se realizó en el modelo, mediante el uso de una variable control, como se describió en la sección anterior.

Se realizó también una validación similar, considerando en esta ocasión indicadores que no se usaron en el modelo.

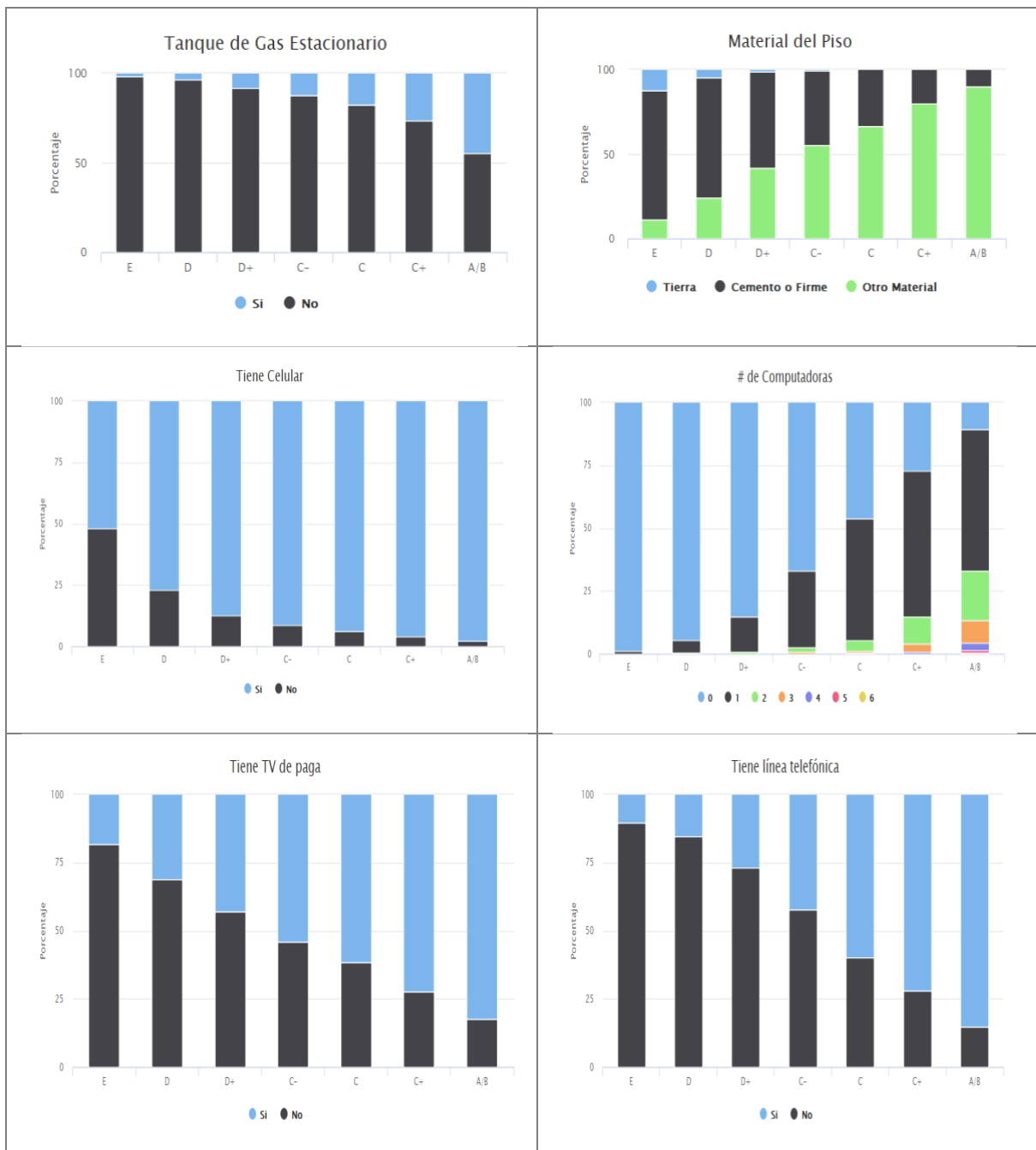


Figura 14. Distribución de algunas variables ajenas al modelo de acuerdo a los niveles socioeconómicos de la regla AMAI 2018. Estimación a partir de la ENIGH 2016

Se aprecia una importante relación entre la tenencia de distintos bienes y los niveles socioeconómicos estimados por la regla AMAI 2018. Estos resultados también ayudan a validar que la estratificación generada es consistente con el estilo de vida y el nivel de comodidades establecido dentro de la definición de los niveles socioeconómicos AMAI.

Adicional a estos ejercicios, se contrastaron las asociaciones de los niveles socioeconómicos con muchos otros indicadores presentes en la ENIGH 2016. Todos los resultados mostraron una relación consistente entre los niveles socioeconómicos de la regla AMAI 2018 y los estándares de calidad de vida que los definen.

CUESTIONARIO PROPUESTO

Antes de cerrar la nota técnica de la creación del modelo de nivel socioeconómico AMAI 2018, vale la pena explicar a detalle las definiciones de las variables consideradas en la nueva regla. Como se detalló, son 6 variables las que integran el nuevo modelo:

- Nivel educativo del jefe de hogar
- Número de baños completos en la vivienda
- Número de autos en el hogar (entendida como la suma de autos, vans y pick ups en el hogar)
- Tenencia de conexión a internet en el hogar
- Número de integrantes en el hogar de 14 años o más que trabajan
- Número de dormitorios en la vivienda

Las definiciones que INEGI presenta de cada una de ellas se muestran a continuación:

Variable	Definición de ENIGH 2014	Pregunta Textual en el cuestionario de ENIGH 2014
Nivel educativo del jefe de hogar	Educación formal del jefe del hogar	¿Hasta qué año o grado aprobó (NOMBRE) en la escuela?
Número de baños completos en la vivienda	Número de baños con excusado y regadera.	¿Cuántos baños tiene esta vivienda con excusado y regadera?
Número de autos en el hogar	Número de automóviles/camionetas/vans con los que cuenta el hogar	¿Este hogar cuenta con... ...automóvil? ...camioneta cerrada o con cabina? ...camioneta de caja? ¿Cuántos tiene?
Tenencia de conexión a internet en el hogar	Hogares que cuentan con el servicio de internet	¿Este hogar cuenta con internet?
Número de integrantes en el hogar de 14 años o más que trabajan	Número de personas que tienen trabajo y 14 años de edad o más.	¿Cuántos trabajos tuvo (NOMBRE) el mes pasado?
Número de dormitorios en la vivienda	Número de cuartos de la vivienda que son usados habitualmente para dormir, aunque también tengan otros usos.	¿Cuántos cuartos se usan para dormir, sin contar pasillos ni baños?

Buscando preservar en la medida de lo posible la redacción precisa de acuerdo al cuestionario utilizado para la ENIGH 2014, pero haciendo las adecuaciones pertinentes para su aplicación en campo, a continuación se presenta la propuesta de cuestionario a aplicar para producir una estimación del Nivel Socioeconómico de los hogares mexicanos:

1. Pensando en el jefe o jefa de hogar, ¿cuál fue el último año de estudios que aprobó en la escuela?

RESPUESTA	PUNTOS
Sin Instrucción	0
Preescolar	0
Primaria Incompleta	10
Primaria Completa	22
Secundaria Incompleta	23
Secundaria Completa	31
Preparatoria Incompleta	35
Preparatoria Completa	43
Licenciatura Incompleta	59
Licenciatura Completa	73
Posgrado	101

2. ¿Cuántos baños completos con regadera y W.C. (excusado) hay en esta vivienda?

RESPUESTA	PUNTOS
0	0
1	24
2 ó más	47

3. ¿Cuántos automóviles o camionetas tienen en su hogar, incluyendo camionetas cerradas, o con cabina o caja?

RESPUESTA	PUNTOS
0	0
1	18
2 ó más	37

4. Sin tomar en cuenta la conexión móvil que pudiera tener desde algún celular ¿este hogar cuenta con internet?

RESPUESTA	PUNTOS
NO TIENE	0
SÍ TIENE	31

5. De todas las personas de 14 años o más que viven en el hogar, ¿cuántas trabajaron en el último mes?

RESPUESTA	PUNTOS
-----------	--------

0	0
1	15
2	31
3	46
4 ó más	61

6. En esta vivienda, ¿cuántos cuartos se usan para dormir, sin contar pasillos ni baños?

RESPUESTA PUNTOS

0	0
1	6
2	12
3	17
4 ó más	23

CONSIDERACIONES FINALES

La regla AMAI 2018, descrita en este documento, se trata de un esfuerzo realizado por parte del Comité de Niveles Socioeconómicos para actualizar la regla AMAI 8x7 y solventar las áreas de oportunidad que tenía el modelo anterior. En opinión del Comité, la nueva propuesta consolida la definición de los estratos de calidad de vida en país y facilita la aplicación de la regla en campo, sin sacrificar la precisión de la clasificación.

Sin embargo, es importante acentuar el hecho de que el modelo de nivel socioeconómico busca brindar un panorama general del nivel de bienestar de los hogares del país, pero que puede ser insuficiente para analizar a detalle algún subsegmento de la distribución. Es decir, aunque el modelo es adecuado para representar a los distintos sectores del país, podría haber un grupo interesado en analizar con detalle algún sector particular de población, como podría ser la base de la pirámide o en los niveles socioeconómicos altos, por lo que se podría pensar en definir una regla especializada para estratificar mejor estos segmentos.

Precisamente con estas consideraciones en mente se ha desarrollado la regla AMAI 2018 y por ello es que se ha decidido hacer de conocimiento público la metodología detallada usada para su generación. Parte de los objetivos del Comité es promover la generación de nuevos estudios y metodologías que permitan entrar en un esquema de mejora continua de la regla, y con ello contar con aplicaciones más específicas que generen sub-segmentos dentro de algunos estratos específicos o bien, que puedan implementar preguntas adicionales que mejoren la segmentación de acuerdo a las necesidades específicas del estudio. El sistema de puntos elegido por el Comité permite tener un modelo adaptable y versátil que puede ser ajustado y mejorado de acuerdo a las necesidades específicas del usuario.

Finalmente, la regla AMAI 2018 comienza también un nuevo modelo de evaluación y ratificación de la regla de Nivel socioeconómico. A partir de esta iteración, el Comité se compromete a realizar revisiones regulares de la regla conforme nuevas ediciones de la Encuesta Nacional de Ingreso y Gasto en Hogares sean publicadas. Esto no implica que la regla cambie con mayor frecuencia, simplemente indica que el Comité tendrá ya fechas pre-establecidas para validar si los cambios sociales, económicos y culturales del país han alterado el estilo de vida de los hogares mexicanos al grado de que la regla vigente deje de ser válida. Por supuesto, si los cambios no reflejan esto, el Comité podrá ratificar la regla vigente, brindando a la industria la confianza de que la segmentación propuesta sigue siendo vigente.

ANEXOS

Códigos de Análisis en R

Modelo ajustando el ingreso corriente transformado con las variables de la regla AMAI vigente 8x7

```

library(foreign)
library(survey)
#Se define la ruta de la BD
setwd("~/Work/NSE/Modelo")
#Leemos La BD
BASE <- read.dta("BASE_Full.dta", convert.factors=T, convert.dates=T, warn.missing.labels=T)
#Se crean algunas variables necesarias para aplicar la regla AMAI 8x7
BASE$regaderas=BASE$bano_regad+BASE$bano_comp
BASE$focos=BASE$focos_inca+BASE$focos_ahor
#Se define el diseño de muestra complejo para incorporarlo en el ajuste del modelo
design1<-svydesign(id=~upm, strata=~est_dis, weights=~factor_hog, data = BASE)
#Ajuste del modelo
modelo.original<-svyglm(ln_ing~num_cuarto+bano_comp+regaderas+focos+material_pisos+num_au
to+edu_jefe, design1, subset=ing_cor<1000000)
#Coeficientes y principales resultados del Modelo
summary(modelo.original)
## Call:
## svyglm(formula = ln_ing ~ num_cuarto + bano_comp + regaderas + focos +
##   material_pisos + num_auto + edu_jefe, design1, subset = ing_cor < 1e+06)
##
## Survey design:
## svydesign(id = ~upm, strata = ~est_dis, weights = ~factor_hog,
##   data = BASE)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.973914   0.033634  266.808 < 2e-16 ***
## num_cuarto    0.058979   0.005732   10.289 < 2e-16 ***
## bano_comp     0.109210   0.026606    4.105 4.18e-05 ***
## regaderas     0.034213   0.026682    1.282  0.1999
## focos         0.023144   0.002922    7.921 3.57e-15 ***
## material_pisos1 0.115739   0.041564    2.785  0.0054 **
## material_pisos2 0.251190   0.016622   15.112 < 2e-16 ***
## material_pisos3 0.430770   0.017987   23.948 < 2e-16 ***
## num_auto      0.276709   0.012467   22.195 < 2e-16 ***
## edu_jefe02    -0.063050   0.130901   -0.482  0.6301
## edu_jefe03     0.120340   0.024994    4.815 1.57e-06 ***
## edu_jefe04     0.247802   0.025352    9.775 < 2e-16 ***
## edu_jefe05     0.223141   0.034342    6.498 9.88e-11 ***
## edu_jefe06     0.325201   0.024210   13.433 < 2e-16 ***
## edu_jefe07     0.341270   0.032488   10.504 < 2e-16 ***
## edu_jefe08     0.424491   0.027820   15.258 < 2e-16 ***
## edu_jefe09     0.535849   0.041249   12.991 < 2e-16 ***
## edu_jefe10     0.688845   0.029651   23.232 < 2e-16 ***
## edu_jefe11     0.947524   0.046290   20.469 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```



```
## (Dispersion parameter for gaussian family taken to be 0.3539119)
##
## Number of Fisher Scoring iterations: 2
#Akaike's Information Criteria
modelo.original$aic
# [1] 39660.88
#Valores Predichos contra Reales
plot(fitted(modelo.original),modelo.original$y,col="#a39f96",main="Predichos vs Reales",x
lab="Ing Cor Predicho",ylab="Ing Cor (Escala Log)")
#Correlacion entre Predichos y Reales
cor(fitted(modelo.original),modelo.original$y)
## [1] 0.652611
#Deviance del Modelo
modelo.original$deviance
## [1] 6648.235
#R Cuadrada
print(R2 <- 1-modelo.original$deviance/modelo.original$null.deviance)
## [1] 0.4659177
```

Modelo ajustando el ingreso corriente transformado con las variables que integran la regla AMAI 2018

```
library(foreign)
library(survey)
#Se define la ruta de la BD
setwd("~/Work/NSE/Modelo")
#Leemos La BD
BASE <- read.dta("BASE_Full.dta",convert.factors=T,convert.dates=T,warn.missing.labels=T)
#Creamos algunas variables para usar en el modelo final
BASE$autos<-BASE$num_auto+BASE$num_van+BASE$num_pickup
BASE <- within(BASE, internet <- relevel(internet, ref = 2))
#Se define el diseño de muestra compleja utilizado en La ENIGH
design1<-svydesign(id=~upm,strata=~est_dis,weights=~factor_hog,data = BASE)
#Ajuste del modelo
modelo.AMAI.2018<-svyglm(ln_ing~cuart_dorm+bano_comp+autos+internet+ocupados+edu_jefe+tot
_integ,design1,subset=(ing_cor<1000000))
#Coeficientes del Modelo
summary(modelo.AMAI.2018)
## Call:
## svyglm(formula = ln_ing ~ cuart_dorm + bano_comp + autos + internet +
## ocupados + edu_jefe + tot_integ, design1, subset = (ing_cor <1e+06))
##
## Survey design:
## svydesign(id = ~upm, strata = ~est_dis, weights = ~factor_hog, data = BASE)
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.986310 0.025518 352.158 < 2e-16 ***
## cuart_dorm 0.060989 0.008662 7.041 2.48e-12 ***
## bano_comp 0.250480 0.010675 23.463 < 2e-16 ***
## autos 0.194344 0.009665 20.109 < 2e-16 ***
## internet1 0.322823 0.013874 23.268 < 2e-16 ***
## ocupados 0.162173 0.006536 24.812 < 2e-16 ***
## edu_jefe02 -0.025323 0.142407 -0.178 0.859
## edu_jefe03 0.106901 0.024046 4.446 9.15e-06 ***
## edu_jefe04 0.233745 0.023910 9.776 < 2e-16 ***
## edu_jefe05 0.240938 0.031983 7.533 6.96e-14 ***
## edu_jefe06 0.331525 0.023599 14.049 < 2e-16 ***
## edu_jefe07 0.366085 0.029733 12.313 < 2e-16 ***
```

```

## edu_jefe08  0.455238  0.026532  17.158 < 2e-16 ***
## edu_jefe09  0.617822  0.039479  15.649 < 2e-16 ***
## edu_jefe10  0.770849  0.028743  26.818 < 2e-16 ***
## edu_jefe11  1.062254  0.050718  20.944 < 2e-16 ***
## tot_integ   0.024253  0.003766   6.440 1.44e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.3080391)
##
## Number of Fisher Scoring iterations: 2
#Akaike's Information Criteria
modelo.AMAI.2018$aic
## [1] 37270.79
#Valores Predichos contra Reales
plot(fitted(modelo.AMAI.2018),modelo.AMAI.2018$y,col="#a39f96",main="Predichos vs Reales"
,xlab="Ing Cor Predicho",ylab="Ing Cor (Escala Log)")
#Correlacion entre Predichos y Reales
cor(fitted(modelo.AMAI.2018),modelo.AMAI.2018$y)
## [1] 0.7147783
#Deviance del Modelo
modelo.AMAI.2018$deviance
## [1] 5820.707
#R Cuadrada
print(R2 <- 1-modelo.AMAI.2018$deviance/modelo.AMAI.2018>null.deviance)
## [1] 0.5386235

```