

Construcción de un índice socioeconómico para medir desigualdades educativas en el Valle de México a partir de la encuesta COMIPEMS

Wilfrido A. Gómez-Arias

Resumen

Se propone el método de correspondencias múltiples (MCM) para construir un índice socioeconómico (ISE) que permita analizar las desigualdades educativas entre los estudiantes que presentan el examen de ingreso al nivel medio superior en la Zona Metropolitana del Valle de México (ZMVM). Para este análisis, se emplea la encuesta de la Comisión Metropolitana de Instituciones Públicas de Educación Media Superior (COMIPEMS 2021), que recopila información sobre las características demográficas, académicas y socioeconómicas de los aspirantes. El MCM se prefiere porque las variables disponibles son categóricas y este método utiliza las frecuencias relativas de cada categoría, lo que lo hace especialmente adecuado para este tipo de datos. El enfoque propuesto ofrece un instrumento único y replicable para analizar en retrospectiva la brecha socioeconómica que subyace en las trayectorias escolares de la ZMVM. Los resultados muestran que los estudiantes con un ISE muy alto —interpretado a partir del primer eje factorial del MCM como una escala continua de estatus socioeconómico, agrupada en cinco niveles: muy bajo, bajo, medio, alto y muy alto— acceden aproximadamente cinco veces más a las instituciones más demandadas que aquellos con un índice muy bajo, lo que confirma que el origen socioeconómico sigue siendo determinante en las oportunidades de admisión.

Palabras clave: índice socio económico, método de correspondencias múltiples, desigualdades educativas, comipems

Introducción

Diversos estudios cuantitativos han documentado las brechas socioeconómicas entre quienes presentan el examen de la Comisión Metropolitana de Instituciones Públicas de Educación Media Superior (COMIPEMS) y los resultados que obtienen. La evidencia señala que las y los estudiantes provenientes de hogares con mayores niveles socioeconómicos logran sistemáticamente puntajes más altos y, en consecuencia, tienen mayores probabilidades de ingresar a las opciones de su preferencia. Por el contrario, quienes viven en contextos menos favorecidos alcanzan puntajes más bajos y ven reducidas sus oportunidades de acceder a dichas opciones (Solís et al., 2013; Blanco, 2014; Rodríguez, 2015; García, 2016). El efecto agregado es una distribución desigual de oportunidades: el puntaje termina por alinearse con la posición socioeconómica de cada estudiante. Así, aun cuando en términos formales todas y todos obtienen un lugar, no todos los lugares son equivalentes. El concurso opera, en la práctica, como un filtro que canaliza a los alumnos con mejores condiciones socioeconómicas hacia las escuelas más demandadas (como las preparatorias de la UNAM y el IPN), mientras que concentra a muchos jóvenes de entornos desfavorecidos en planteles con menores recursos y prestigio. Por ejemplo, Estrada (2017) encontró que los estudiantes de alto nivel socioeconómico tenían casi el doble de probabilidades de ingresar a una preparatoria de la UNAM o el IPN en comparación con estudiantes de bajo nivel socioeconómico pero con rendimiento académico similar. Incluso

dentro de la propia UNAM persiste una segregación interna: la Escuela Nacional Preparatoria agrupa sobre todo a los deciles socioeconómicos altos, mientras que el Colegio de Ciencias y Humanidades concentra a los deciles bajos (Mata y Suárez, 2016).

Aunque estas investigaciones han sido fundamentales para visibilizar la desigualdad educativa, la mayoría han recurrido a diversas estrategias para medir un índice socioeconómico de los estudiantes. Algunos estudios han utilizado encuestas específicas y han construido índices integrados, como el Índice de Orígenes Sociales (IOS), que busca medir la posición socioeconómica de la familia de origen (Solís et al., 2013). Otros han aprovechado los datos de la propia encuesta COMIPEMS, construyendo índices socioeconómicos mediante técnicas como el Análisis Factorial por Componentes Principales (PCA) para resumir variables como el ingreso familiar, la escolaridad de los padres y el acceso a ciertos bienes (Rodríguez, 2015), o han empleado modelos como el de Crédito Parcial de Masters considerando variables como número de libros, computadoras y escolaridad parental para analizar patrones de elección (García, 2016).

Si bien estos enfoques han sido valiosos para demostrar la asociación entre el origen social y los resultados educativos, la presente investigación se justifica por la necesidad de explorar y validar una metodología alternativa, robusta y con información reciente, específicamente diseñada para la naturaleza de los datos disponibles en encuestas como la de COMIPEMS. Muchas de las variables relevantes para caracterizar el contexto socioeconómico (servicios, bienes en el hogar, ingreso reportado en rangos, escolaridad de padres, condiciones de vivienda) son de naturaleza categórica. El Análisis de Correspondencias Múltiples (MCM) es una técnica estadística particularmente adecuada para analizar este tipo de datos (Vélez, 2015, Peña, 2002), ya que permite:

- Sintetizar información compleja: El MCM puede reducir un gran número de variables categóricas interrelacionadas a un conjunto más pequeño de dimensiones (ejes factoriales) que capturan la estructura subyacente de los datos, como el nivel socioeconómico.
- Manejar datos categóricos de forma nativa: A diferencia del PCA, que asume variables continuas o requiere transformaciones (como la creación de variables dummy), el MCM trabaja directamente con las categorías, respetando la naturaleza original de la información.
- Identificar patrones y oposiciones: Permite visualizar cómo se asocian las diferentes categorías de las variables, revelando patrones (por ejemplo, qué combinación de características tiende a definir a los hogares de ISE bajo versus alto).
- Construir un índice fundamentado: Al utilizar el primer eje factorial, que típicamente captura la mayor varianza relacionada con la dimensión socioeconómica principal, se puede construir un índice continuo de ISE basado empíricamente en las interrelaciones observadas en los datos específicos de la población COMIPEMS.

Por lo tanto, este estudio no solo busca construir un índice socioeconómico, sino que aporta valor al campo al aplicar y evaluar la idoneidad del MCM para esta tarea específica con datos de COMIPEMS. Validar esta metodología puede ofrecer una herramienta refinada y estadísticamente sólida para futuras investigaciones sobre equidad educativa en este contexto, complementando los enfoques utilizados previamente y ofreciendo una

perspectiva más matizada sobre la dimensión socioeconómica basada en la estructura intrínseca de los datos categóricos disponibles.

El método de correspondencias múltiples

De acuerdo con (Vélez, 2015), el MCM construye “dimensiones”, es decir, combinaciones lineales de las categorías de las variables, y cada una explica una fracción de la variabilidad total —denominada *inercia*, aunque no existe una regla exacta de cuánto se considera una buena cantidad explicada de inercia.

En cada categoría se examinan dos métricas clave: (1) la *calidad de ajuste*, que varía entre 0 (sin ajuste) y 1 (ajuste perfecto), y (2) la *contribución*, que refleja el porcentaje de inercia que esa categoría aporta a la dimensión. Además, las *coordenadas* —positivas o negativas— indican la manera en que cada categoría incide en el índice. Por ejemplo, si la categoría “dispone de internet” presenta una coordenada positiva y “no dispone de internet” una negativa, los hogares con internet mostrarán un valor más alto en el índice que los que no tienen internet.

Es fundamental que, dentro de una misma dimensión, las variables conserven una orientación coherente. Cuando dos variables se comportan en sentidos opuestos —por ejemplo, la falta de drenaje se asocia positivamente con el índice, mientras que tener horno de microondas lo hace de forma negativa—, resulta recomendable excluir la variable cuya contribución a la inercia sea menor, con el fin de preservar la consistencia interpretativa del modelo (Velez, 2015).

Metodología

El presente trabajo propone construir un índice socioeconómico estandarizado (ISE) para los estudiantes que presentan el examen COMIPEMS, a partir de diversas variables categóricas disponibles que describen su contexto (servicios, bienes en el hogar, ingreso reportado en rangos, escolaridad de padres, condiciones de vivienda). Utilizaremos el método de correspondencias múltiples (MCM) para sintetizar estas variables en una dimensión continua principal que represente el nivel socioeconómico.

En particular, nos enfocaremos en el primer eje factorial del MCM, que suele capturar la mayor parte de la variabilidad asociada al nivel socioeconómico (por ejemplo, un eje que opone hogares con mayor acumulación de bienes/servicios/capital cultural vs. hogares con mayores carencias). A partir de las coordenadas de cada individuo en este primer eje, definiremos un índice continuo de nivel socioeconómico. Posteriormente, clasificaremos este índice en 5 niveles ordinales (por ejemplo: muy bajo, bajo, medio, alto y muy alto) para facilitar su interpretación y uso en análisis posteriores.

Una consideración metodológica importante será la selección de las variables activas a incluir en el MCM. Se evaluará la pertinencia de cada variable categórica, excluyendo aquellas que sean constantes o casi constantes (con muy baja variabilidad), ya que no

aportarían información significativa (inercia) al análisis y podrían introducir ruido. Asimismo, tras ejecutar el MCM, se analizará la contribución de cada variable y sus categorías a la formación del primer eje factorial para asegurar que el índice resultante capture efectivamente la dimensión socioeconómica de interés y evaluar la relevancia de las variables incluidas.

Fuente de datos

Para la construcción del ISE propuesto en esta investigación utilizaremos los datos de la encuesta COMIPEMS 2021. Esta encuesta recopila información demográfica, académica y socioeconómica de los aspirantes que presentan el examen de ingreso al nivel medio superior en la Zona Metropolitana del Valle de México (ZMVM). La encuesta cuenta con respuestas de aproximadamente 275,347 aspirantes, de los cuales 246,176 tuvieron una asignación a alguna preparatoria, entre ellos 121,272 hombres y 124,904 mujeres, proporcionando una amplia base estadística para el análisis y comprensión de las desigualdades educativas en la ZMVM.

Construcción del índice socioeconómico

A continuación, se describe el proceso de construcción de dicho índice. El proceso comenzó con un conjunto inicial de 23 variables extraídas de la encuesta:

- **Escolaridad familiar:**
 - Escolaridad del jefe del hogar del aspirante (1-14)
 - Escolaridad del padre del aspirante (1-14)
 - Escolaridad de la madre del aspirante (1-14)
 - Régimen de la secundaria de origen (0,1)
- **Patrimonio material del hogar**
 - Línea telefónica fija en el hogar del aspirante (0,1)
 - Lavadora de ropa en el hogar del aspirante (0,1)
 - Refrigerador en el hogar del aspirante (0,1)
 - Horno de microondas en el hogar del aspirante (0,1)
 - Internet en el hogar del aspirante (0,1)
 - Televisión de paga en el hogar del aspirante (0,1)
 - Tableta en el hogar del aspirante (0,1)
 - Drenaje o alcantarillado en el hogar del aspirante (0,1)
 - Agua potable entubada en el hogar del aspirante (0,1)
 - Luz eléctrica de servicio público en el hogar del aspirante (0,1)
 - Gas natural, de tanque o de tubería para cocinar (0,1)
 - Número de computadora en el hogar del aspirante (1-5)
 - Número de automóvil en el hogar del aspirante (1-5)
- **Condiciones habitacionales** (cuartos, baños, hacinamiento).
 - Número cuartos en el hogar del aspirante (sala, comedor, recámaras, estudio, etc., sin contar cocina, baños, y pasillos) (1-10)
 - Número de baños completos (con regadera) en el hogar del aspirante (1-5)
 - Número de personas que viven en el hogar del aspirante (1-9)
- **Capital cultural**

- Número de libros que hay en el hogar del aspirante (1-8)
- **Experiencias de ocio**
 - Número de veces que el aspirante salido de vacaciones dentro de la República Mexicana en los últimos 2 años (1-5)
- **Ingreso familiar mensual**
 - Ingreso mensual de toda la familia en el hogar del aspirante (1-21)

En el análisis inicial se identificaron 107 dimensiones para las variables estudiadas. Para alcanzar un 30% de varianza explicada acumulada, fue necesario considerar hasta la dimensión 17, fenómeno atribuible a la abundancia de variables y categorías diferentes. Esta situación provoca que cada variable o categoría aporte una fracción pequeña de variabilidad, distribuyendo la inercia entre numerosas dimensiones. Como consecuencia, la primera dimensión solo representa un 4.96% de la varianza total, requiriendo avanzar hasta la dimensión 17 para superar el umbral del 30%.

Si trabajamos con variables dicotómicas (de solo dos categorías) y en menor cantidad (por ejemplo, 10 variables), el número total de categorías se reduciría significativamente en comparación con variables que presentan cinco o más categorías. La regla es clara: a mayor número de categorías por variable y más variables incluidas, mayor será la dispersión de la variabilidad entre las distintas dimensiones.

Para optimizar el análisis, implementamos una estrategia de simplificación mediante:

- La eliminación de variables con escasa capacidad de diferenciación, como aquellas con categorías muy sesgadas o que concentran la mayoría de casos en una sola categoría.
- La recodificación de variables, fusionando variables o categorías pertinentes.
- La eliminación de variables sin relevancia teórica o con contribución mínima a las primeras dimensiones del análisis.

Eliminación de variables poco diferenciadas

Tras revisar las distribuciones de todas las variables, se identificaron aquellas con limitada capacidad discriminadora debido a la alta concentración de respuestas en una sola categoría (ver Figura 1). Estos casos presentan poco valor diferenciador para el análisis. Por ejemplo:

- **Lavadora:** El 85% de los aspirantes reporta tener lavadora en el hogar, lo que resulta en una variable con distribución muy sesgada.
- **Refrigerador:** Aún más pronunciado es el caso del refrigerador, donde el 95.7% de los aspirantes indica contar con este electrodoméstico.

Las variables marcadas en rojo en el análisis original tienen una contribución marginal al índice debido a su escasa variabilidad entre los aspirantes o están sesgadas a una categoría. Su inclusión podría introducir ruido sin añadir poder explicativo, por lo que se recomienda su exclusión.

- **Capital educativo familiar:**
 - Escolaridad del jefe del hogar del aspirante (1-14)
 - Escolaridad del padre del aspirante (1-14)
 - Escolaridad de la madre del aspirante (1-14)
 - Régimen de la secundaria de origen (0,1)
- **Patrimonio material del hogar**
 - Línea telefónica fija en el hogar del aspirante (0,1)
 - Lavadora de ropa en el hogar del aspirante (0,1)
 - Refrigerador en el hogar del aspirante (0,1)
 - Horno de microondas en el hogar del aspirante (0,1)
 - Internet en el hogar del aspirante (0,1)
 - Televisión de paga en el hogar del aspirante (0,1)
 - Tableta en el hogar del aspirante (0,1)
 - Drenaje o alcantarillado en el hogar del aspirante (0,1)
 - Agua potable entubada en el hogar del aspirante (0,1)
 - Luz eléctrica de servicio público en el hogar del aspirante (0,1)
 - Gas natural, de tanque o de tubería para cocinar (0,1)
 - Número de computadora en el hogar del aspirante (1-5)
 - Número de automóvil en el hogar del aspirante (1-5)
- **Condiciones habitacionales** (cuartos, baños, hacinamiento).
 - Número cuartos en el hogar del aspirante (sala, comedor, recámaras, estudio, etc., sin contar cocina, baños, y pasillos) (1-10)
 - Número de baños completos (con regadera) en el hogar del aspirante (1-5)
 - Número de personas que viven en el hogar del aspirante (1-9)
- **Capital cultural**
 - Número de libros que hay en en el hogar del aspirante (1-8)
- **Experiencias de ocio**
 - Número de veces que el aspirante salido de vacaciones dentro de la República Mexicana en los últimos 2 años (1-5)
- **Ingreso familiar mensual**
 - Ingreso mensual de toda la familia en el hogar del aspirante (1-21)

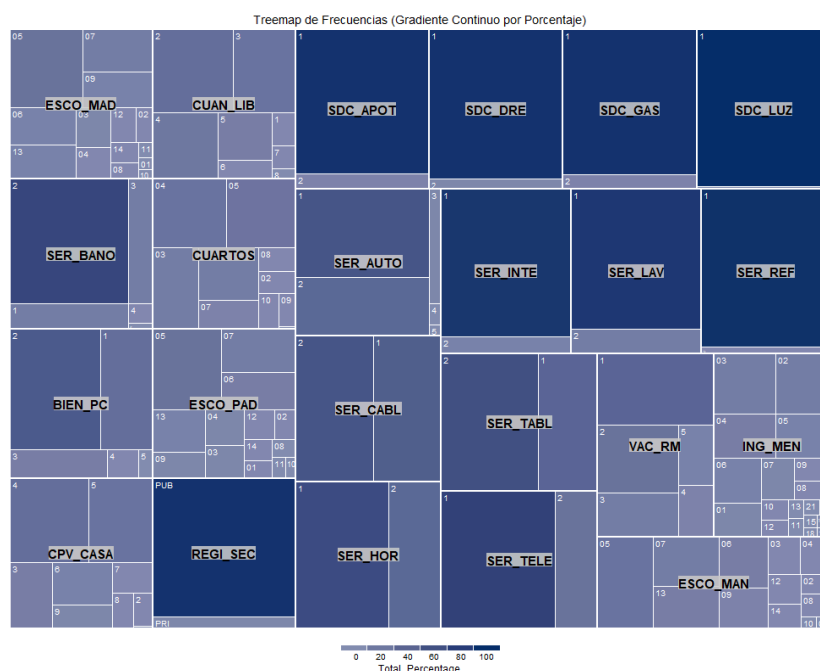


Figura 1. Distribuciones de todas las variables. El área de cada rectángulo es proporcional a la frecuencia (o porcentaje) de respuestas en esa categoría específica. El gradiente de color (de claro a oscuro) representa visualmente dicho porcentaje, donde los tonos más oscuros indican una mayor concentración porcentual. Esto facilita la identificación de variables sesgadas, como aquellas donde una sola categoría agrupa un alto porcentaje (p. ej., >80%) de las respuestas.

Recodificación de variables

Se realizó una recodificación estratégica de ciertas variables para mejorar la eficiencia del análisis. Por ejemplo:

- **Cuartos:** En cuanto al número de cuartos la categoría “Ninguno” tiene muy pocos casos, sugiriendo la conveniencia de fusionar la categoría con la categoría que sigue.
- **Escolaridad:** En cuanto al nivel educativo máximo (jefe del hogar, madre o padre), la categoría “No estudió” tiene muy pocos casos, sugiriendo la conveniencia de fusionar la categoría con "Hasta Primaria" para obtener una distribución más equilibrada.
- **Ingresos:** El ingreso mensual de toda la familia en el hogar del aspirante se simplificó, pasando de 21 a 6 categorías.
- **Computadora, automóvil y vacaciones:** El número de computadoras en el hogar del aspirante, número de automóvil en el hogar del aspirante y número de veces que el aspirante ha salido de vacaciones dentro de la República Mexicana en los últimos 2 años, se pasó de 5 a 3 categorías.

A continuación se presenta el listado completo de variables, destacando en azul aquellas que fueron sometidas a procesos de recodificación:

- **Capital educativo familiar:**

- Escolaridad del jefe del hogar del aspirante (1-7)
- Escolaridad del padre del aspirante (1-7)
- Escolaridad de la madre del aspirante (1-7)
- Régimen de la secundaria de origen (0,1)
- **Patrimonio material del hogar**
 - Línea telefónica fija en el hogar del aspirante (0,1)
 - Horno de microondas en el hogar del aspirante (0,1)
 - Televisión de paga en el hogar del aspirante (0,1)
 - Tableta en el hogar del aspirante (0,1)
 - Número de computadora en el hogar del aspirante (1-3)
 - Número de automóvil en el hogar del aspirante (1-3)
- **Condiciones habitacionales** (cuartos, baños, hacinamiento).
 - Número cuartos en el hogar del aspirante (sala, comedor, recámaras, estudio, etc., sin contar cocina, baños, y pasillos) (1-4)
 - Número de baños completos (con regadera) en el hogar del aspirante (1-3)
 - Número de personas que viven en el hogar del aspirante (1-3)
- **Capital cultural**
 - Número de libros que hay en en el hogar del aspirante (1-4)
- **Experiencias de ocio**
 - Número de veces que el aspirante salido de vacaciones dentro de la República Mexicana en los últimos 2 años (1-3)
- **Ingreso familiar mensual**
 - Ingreso mensual de toda la familia en el hogar del aspirante (1-6)

Eliminación de variables sin relevancia teórica o con contribución mínima a las primeras dimensiones del análisis.

Una vez recodificada y eliminada ciertas variables corremos nuevamente el análisis de MCM donde se identifica que las dos primeras dimensiones explican ya, de manera conjunta, un 18.4% de la inercia. La primera dimensión explica el 12.3% de la inercia y la segunda explica el 6.1%. Aquí podemos eliminar algunas variables debido a su contribución mínima a la primera dimensión del análisis, lo que indica que su relevancia teórica y empírica es reducida en el contexto del análisis realizado. Variables como número de personas en casa (contribución de 0.2%) o si tienen línea telefónica fija , televisión de paga (contribución de 3.52%), número de vacaciones (contribución de 3.6%), tableta como ipad, samsung (contribución de 3.9%) y horno de microondas (contribución de 4.1%) presentan una baja inercia explicada (ver Figura 2), que en total aportan 18.4% a la inercia, lo que sugiere que aportan significativamente poco al modelo y pueden ser consideradas ruido. Al eliminar estas variables, se optimiza el modelo, mejorando la claridad interpretativa y reduciendo la complejidad sin perder relevancia en la explicación de las dimensiones principales. Observar que no eliminamos la variable que tiene que ver con el régimen de la secundaria de origen, ya que a nivel teórico juega un papel importante de si el estudiante viene de escuela pública o privada: se ha encontrado que si vienen de escuelas privadas tienen mejores puntuaciones en el examen (Blanco, Solís, y Robles, 2014). Las variables eliminadas se encuentran en rojo:

- **Capital educativo familiar:**
 - Escolaridad del jefe del hogar del aspirante (1-7)
 - Escolaridad del padre del aspirante (1-7)
 - Escolaridad de la madre del aspirante (1-7)
 - Régimen de la secundaria de origen (0,1)
- **Patrimonio material del hogar**
 - Línea telefónica fija en el hogar del aspirante (0,1)
 - Horno de microondas en el hogar del aspirante (0,1)
 - Televisión de paga en el hogar del aspirante (0,1)
 - Tableta en el hogar del aspirante (0,1)
 - Número de computadora en el hogar del aspirante (1-3)
 - Número de automóvil en el hogar del aspirante (1-3)
- **Condiciones habitacionales** (cuartos, baños, hacinamiento).
 - Número cuartos en el hogar del aspirante (sala, comedor, recámaras, estudio, etc., sin contar cocina, baños, y pasillos) (1-4)
 - Número de baños completos (con regadera) en el hogar del aspirante (1-3)
 - Número de personas que viven en el hogar del aspirante (1-3)
- **Capital cultural**
 - Número de libros que hay en en el hogar del aspirante (1-4)
- **Experiencias de ocio**
 - Número de veces que el aspirante salido de vacaciones dentro de la República Mexicana en los últimos 2 años (1-3)
- **Ingreso familiar mensual**
 - Ingreso mensual de toda la familia en el hogar del aspirante (1-6)

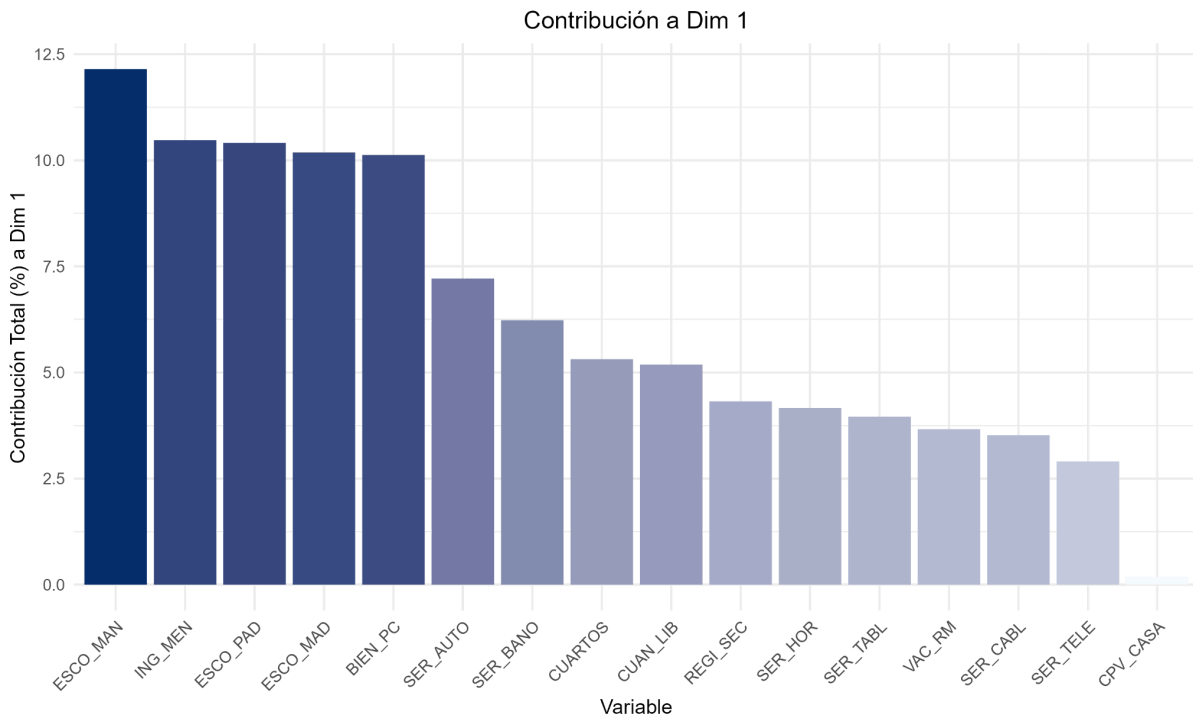


Figura 2. Porcentaje que contribuye cada variable a la inercia total explicada en el análisis. La escolaridad del jefe del hogar contribuye al 12% de la inercia total mientras que el número de personas que viven en casa al 0.2%.

Selección de variables que se diferencian de las otras que contribuyen mejor a las dimensiones

Nos quedamos con las siguientes variables para construir el índice:

- **Capital educativo familiar:**
 - Escolaridad del jefe del hogar del aspirante (1-7)
 - Escolaridad del padre del aspirante (1-7)
 - Escolaridad de la madre del aspirante (1-7)
 - Régimen de la secundaria de origen (0,1)
- **Patrimonio material del hogar**
 - Número de computadora en el hogar del aspirante (1-3)
 - Número de automóvil en el hogar del aspirante (1-3)
- **Condiciones habitacionales** (cuartos, baños, hacinamiento).
 - Número cuartos en el hogar del aspirante (sala, comedor, recámaras, estudio, etc., sin contar cocina, baños, y pasillos) (1-4)
 - Número de baños completos (con regadera) en el hogar del aspirante (1-3)
- **Capital cultural**
 - Número de libros que hay en en el hogar del aspirante (1-4)
- **Ingreso familiar mensual**
 - Ingreso mensual de toda la familia en el hogar del aspirante (1-6)

Con los datos listos, podemos ejecutar nuevamente el MCM, lo cual nos arroja el porcentaje de variabilidad de cada dimensión (ver Figura 3). Nos interesa principalmente la Dimensión 1, que en este caso explica la mayor proporción de la variabilidad: el 13.5% de la inercia total. Esto sugiere que existe un eje dominante (Dim 1) que captura una proporción considerada de la variabilidad conjunta de las variables socioeconómicas, lo cual típicamente corresponde al gradiente de nivel socioeconómico. La Dimensión 2 explica un 7.6% adicional (podría representar otro factor).

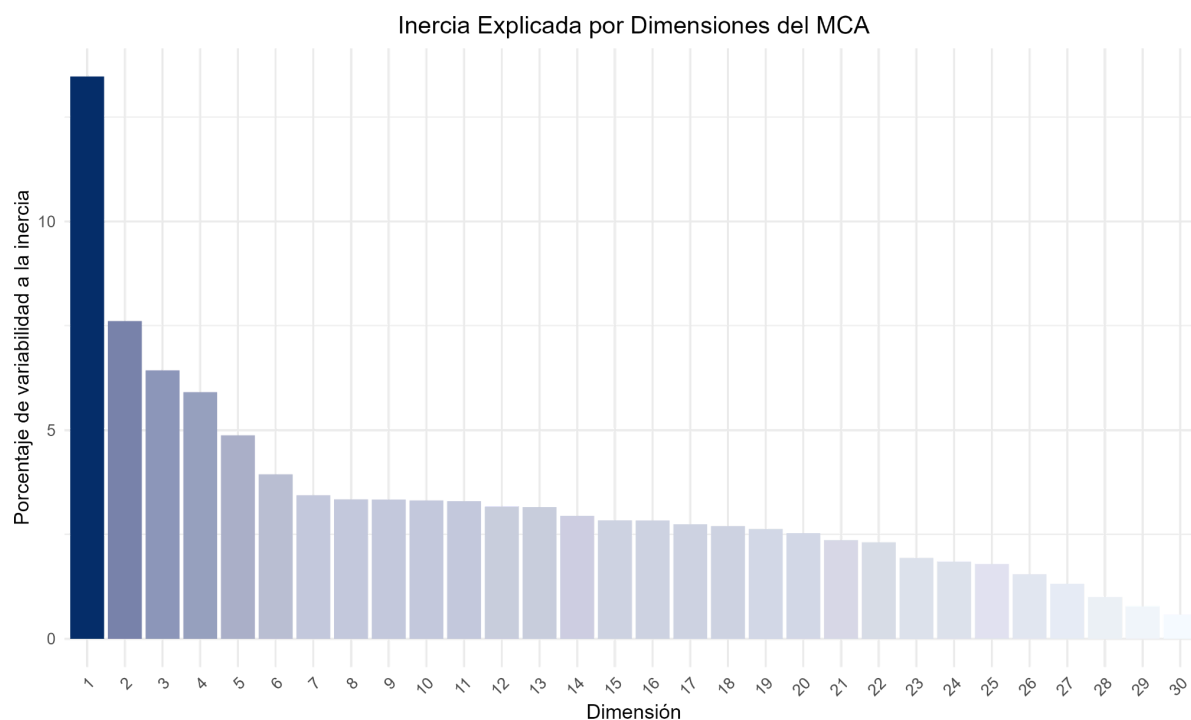


Figura 3. Porcentaje de variabilidad en cada dimensión.

Aunque podríamos usar múltiples dimensiones para diversos análisis, para construir un índice socioeconómico único nos enfocaremos en la Dimensión 1. Podemos explorar las coordenadas de las categorías en las primeras dimensiones y sus contribuciones (ver Tabla 1). Las coordenadas indican la posición de cada categoría. Las contribuciones (% de inercia) indican cuánto aporta cada categoría a la formación de ese eje (ver Tabla 2). Una categoría con contribución alta en Dim 1 significa que dicha categoría es importante para definir el eje socioeconómico.

	Dim 1	Dim 2
ESCO_MAN_No estudió o hasta Primaria	-0.943311825	1.27718216
ESCO_MAN_Hasta Secundaria	-0.697037565	0.38342149
ESCO_MAN_Hasta nivel medio superior	-0.038739801	-0.77146522
ESCO_MAN_Hasta nivel superior	1.153985097	-0.19567531
ESCO_MAN_Posgrado	2.148066924	2.74534570
ESCO_MAD_No estudió o hasta Primaria	-0.878135474	1.06911713
ESCO_MAD_Hasta Secundaria	-0.598970208	0.26782801
ESCO_MAD_Hasta nivel medio superior	0.063757743	-0.63495220
ESCO_MAD_Hasta nivel superior	1.193260114	-0.06515898
ESCO_MAD_Posgrado	2.168661110	2.64807626
ESCO_PAD_No estudió o hasta Primaria	-0.872046837	1.03755813
ESCO_PAD_Hasta Secundaria	-0.566925904	0.24141005
ESCO_PAD_Hasta nivel medio superior	0.052490746	-0.69430685
ESCO_PAD_Hasta nivel superior	1.188611345	-0.10476255
ESCO_PAD_Posgrado	2.166658122	2.75223935
PRI	1.707999451	0.60249721
PUB	-0.138241588	-0.04876475
BIEN_PC_0	-0.784552927	0.36842215
BIEN_PC_1	-0.003335584	-0.32482508
BIEN_PC_2 o más	1.200269720	0.30472002
SER_AUTO_0	-0.470801181	0.03800272
SER_AUTO_1	0.453653604	-0.19087860
SER_AUTO_2 o más	1.137778895	0.57742716
0-2	-0.739500320	0.33034427
3-4	-0.060645325	-0.18605643
5-6	0.493485722	-0.08835017
7 o más	0.760364599	0.33650513
SER_BANO_0	-0.969517411	0.62870901
SER_BANO_1	-0.021047904	-0.20099069
SER_BANO_2 o más	0.888130342	0.30378986
0 a 10	-0.598018419	0.20784139
11 a 100 (de medio a 3 estantes)	0.171292638	-0.22050150
101 a 500 (de 5 a 10 estantes)	1.277037103	0.50077476
más de 500	1.607455560	1.36683750
de \$1748 a \$3496	-0.739113191	0.42004985
de \$3496 a \$6992	-0.373913554	-0.12037282
de \$6992 a \$10488	0.104358637	-0.40160108
de \$10488 a \$15732	0.628036236	-0.30269091
de \$15732 a \$20976	1.149930057	0.09651871
más \$20976	1.726977955	0.98786828

Tabla 1. Coordenadas de las categorías de las variables en las dimensiones 1 y 2.

	Dim 1
ESCO_MAN_No estudió o hasta Primaria	2.0131839418
ESCO_MAN_Hasta Secundaria	3.5340214818
ESCO_MAN_Hasta nivel medio superior	0.0141711713
ESCO_MAN_Hasta nivel superior	6.4320525783
ESCO_MAN_Posgrado	4.2880878639
ESCO_MAD_No estudió o hasta Primaria	1.8853213732
ESCO_MAD_Hasta Secundaria	2.8188062849
ESCO_MAD_Hasta nivel medio superior	0.0396325354
ESCO_MAD_Hasta nivel superior	5.7402361737
ESCO_MAD_Posgrado	3.0814222733
ESCO_PAD_No estudió o hasta Primaria	2.1356744700
ESCO_PAD_Hasta Secundaria	2.5731328366
ESCO_PAD_Hasta nivel medio superior	0.0247790413
ESCO_PAD_Hasta nivel superior	6.0192911622
ESCO_PAD_Posgrado	3.1450783733
PRI	5.4037060403
PUB	0.4373636662
BIEN_PC_0	4.4647139959
BIEN_PC_1	0.0001413906
BIEN_PC_2 o más	6.8813423409
SER_AUTO_0	3.0307584697
SER_AUTO_1	1.8505869888
SER_AUTO_2 o más	2.6830572739
0-2	2.9346377056
3-4	0.0414453145
5-6	1.3772595556
7 o más	1.4149860254
SER_BANO_0	3.2455224620
SER_BANO_1	0.0075802090
SER_BANO_2 o más	3.2929259390
0 a 10	3.0894856515
11 a 100 (de medio a 3 estantes)	0.4109291342
101 a 500 (de 5 a 10 estantes)	2.9554208624
más de 500	0.7280751332
de \$1748 a \$3496	3.1092648854
de \$3496 a \$6992	1.0333309282
de \$6992 a \$10488	0.0581455992
de \$10488 a \$15732	1.3056489438
de \$15732 a \$20976	1.9654437868
más \$20976	4.5333361369

Tabla 2. Contribución en porcentaje a la inercia por cada categoría de las variables.

A continuación vamos a verificar que la Dimensión 1 efectivamente representa un gradiente razonable de nivel socioeconómico. Por ejemplo, observamos que el número de automóvil en el hogar del aspirante = 2 o más, Número de computadora en el hogar del aspirante = 2 o más, Escolaridad máxima = Posgrado y Ingreso mensual de toda la familia en el hogar del aspirante = más \$20976 se ubican en un extremo de Dim1 (con coordenadas positivas altas) mientras Número de automóvil en el hogar del aspirante = no tiene, Número de computadora en el hogar del aspirante = no tiene, Escolaridad máxima = hasta primaria y Ingreso mensual de toda la familia en el hogar del aspirante = de \$1748 a \$3496 se encuentra en el extremo opuesto (coordenadas negativas altas). Vemos que hay coherencia, ya que el eje ordena las categorías de forma consistente (todas los indicadores de mayor bienestar van hacia un lado, las de menor hacia otro). Lo cual nos asegura que Dim 1 es una buena representación del nivel socioeconómico.

Cálculo del índice socioeconómico (primer eje factorial)

Una vez que todo tuvo coherencia, calculamos el índice socioeconómico para cada individuo. Este índice será simplemente la coordenada del individuo en la Dimensión 1 del MCM.

Por construcción del MCM, estas puntuaciones suelen tener media ~ 0 (porque el origen está centrado) y una varianza relacionada con el valor propio de Dim1. Los signos indican posición relativa en el eje: individuos con valores más altos en Dim1 (positivos) pertenecen al extremo definido por ciertas categorías, y valores negativos al extremo opuesto. Los valores positivos significan mayor nivel socioeconómico (por ejemplo, podrían ser positivos para quienes respondieron "el valor máximo de X bien").

Ahora cada estudiante tiene un valor de índice. Verificamos su distribución antes de categorizar (ver Figura 4).

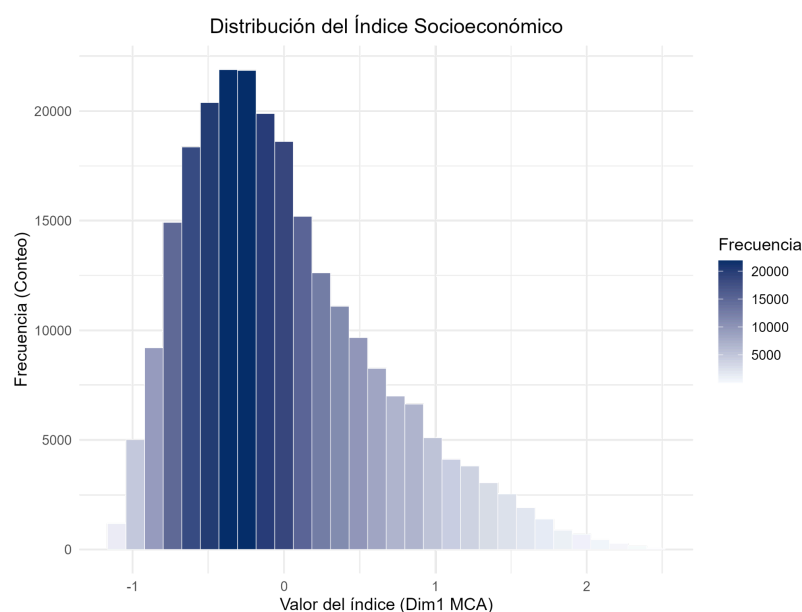


Figura 4. Distribución de los estudiantes de acuerdo a su puntuación en la dimensión 1.

Este histograma muestra la frecuencia de estudiantes en función de su puntuación en el índice socioeconómico, que viene directamente de la Dimensión 1 del MCM. Es decir, cada estudiante tiene un valor que resume su contexto socioeconómico con base en las variables que se incluyeron. La mayor parte de los estudiantes se agrupa hacia la izquierda del eje, en valores negativos o cercanos a cero. Esto sugiere que la mayoría de los aspirantes a COMIPEMS tiene un nivel socioeconómico muy bajo o bajo. A medida que el índice aumenta (hacia la derecha), la cantidad de estudiantes disminuye notablemente. Esto indica que sí existen aspirantes con alto nivel socioeconómico, pero son una minoría clara dentro del conjunto. Aunque no es perfectamente normal, esta distribución sí permite segmentar en 5 grupos (muy bajo, bajo, medio, alto, muy alto) porque se ve una dispersión razonable. No hay “bloques planos” o agrupamientos extremos que hagan imposible esa segmentación.

Clasificación del índice en 5 niveles socioeconómicos

El siguiente paso es convertir el índice continuo en una categoría ordinal de 5 niveles: Muy Bajo, Bajo, Medio, Alto, Muy Alto. Usamos el algoritmo de clustering (k-means) que permite identificar grupos *naturales* en la distribución del índice. Esto no garantiza porcentajes iguales, pero puede formar grupos más homogéneos internamente (ver Figura 5).

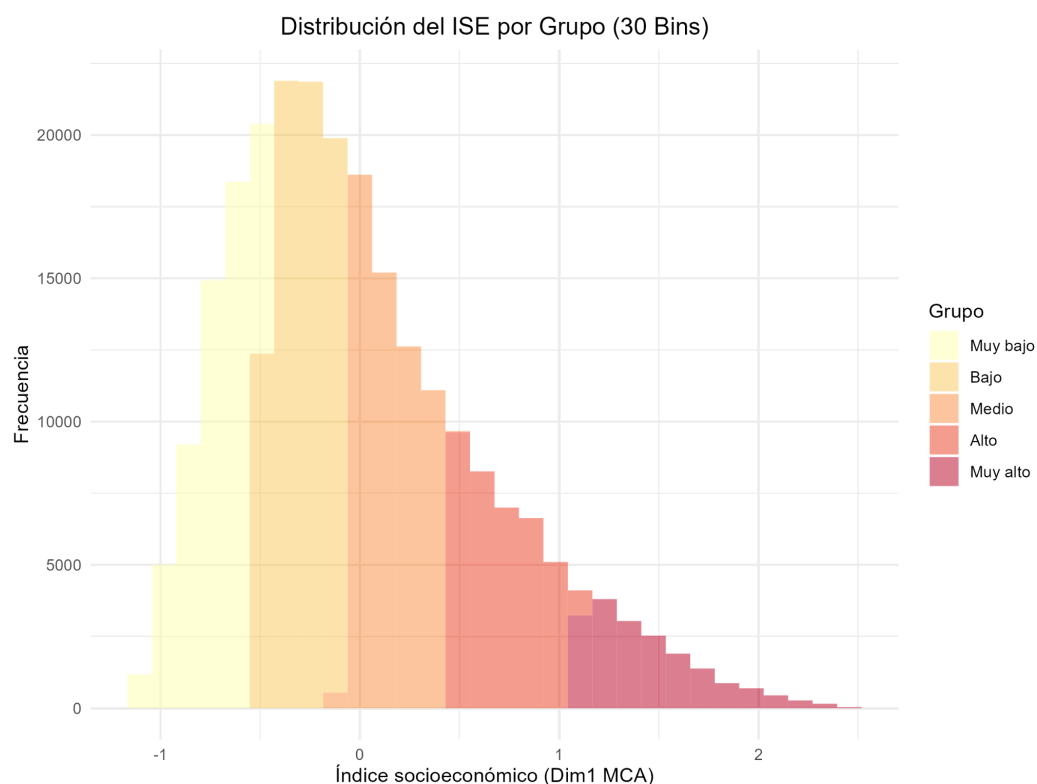


Figura 5. Distribución de los estudiantes de acuerdo a la segmentación por categoría ordinal: Muy Bajo, Bajo, Medio, Alto, Muy Alto.

Con este índice socioeconómico, los estudiantes se pueden comparar en términos de contexto socioeconómico de manera más integral que con cualquier variable individual. Además, la técnica es replicable: si se obtienen nuevos datos o se quiere aplicar en otros contextos (por ejemplo, otros años de COMIPEMS), se pueden seguir los mismos pasos para recalculer el índice y sus grupos. Al no depender de supuestos fuertes (más que la correlación existente entre las variables seleccionadas y el ISE), el MCM es una herramienta robusta para este propósito.

Desigualdades de acceso según el índice socioeconómico (ISE)

El análisis del ISE revela una brecha contundente en el ingreso a las escuelas de mayor demanda, como las preparatorias de la UNAM. Del quintil con ISE muy alto logra ingresar el 57.2 % de quienes presentan el examen, mientras que en el quintil con ISE muy bajo la proporción desciende al 9.9 % (ver Figura 6). Dicho de otro modo, por cada estudiante de origen socioeconómico muy bajo que accede, cinco del estrato más favorecido obtienen un lugar. Esta relación 5:1 evidencia cómo la posición socioeconómica sigue siendo un factor decisivo para acceder a la oferta educativa de mayor prestigio.

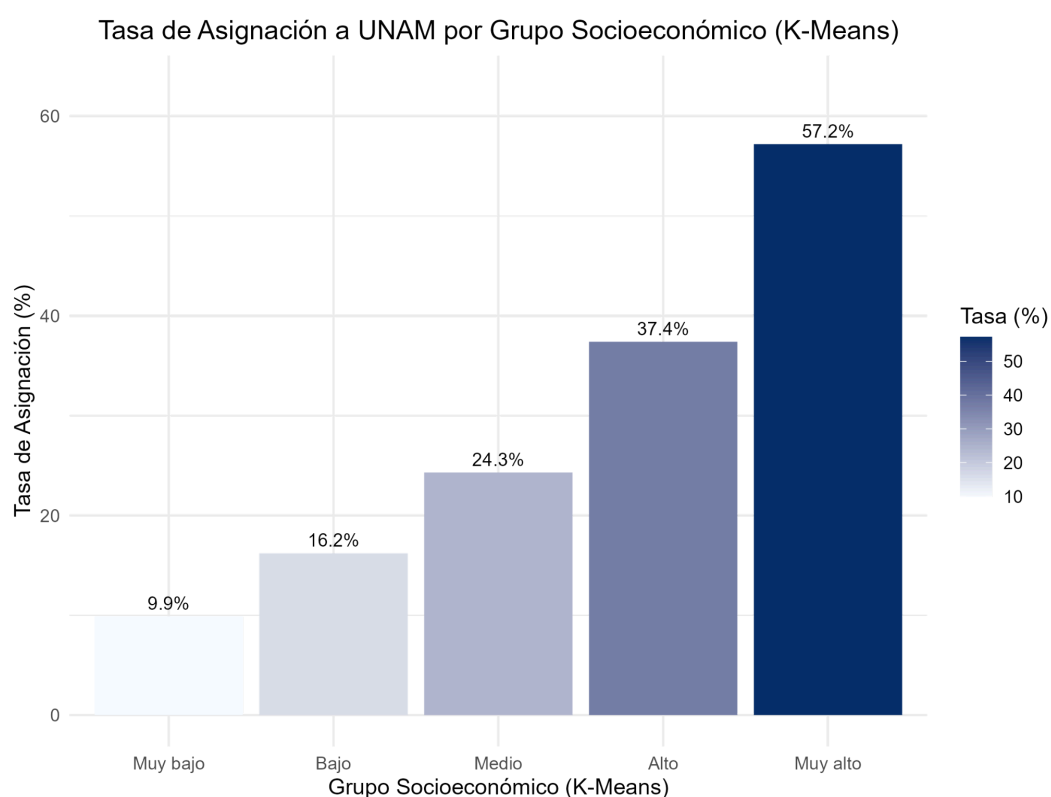


Figura 6. Tasa de asignación de los estudiantes de acuerdo al ISE para las escuelas de la UNAM.

Conclusiones

La construcción de un índice socioeconómico ISE mediante el Método de Correspondencias Múltiples (MCM) aquí propuesto no solo permite medir y clasificar a los estudiantes según su nivel socioeconómico, sino que también se convierte en una herramienta poderosa para analizar y comprender las desigualdades educativas en la ZMVM, y para informar políticas públicas que busquen mejorar la equidad en el sistema educativo.

Referencias

- Blanco, E. (2014). La desigualdad en el nivel medio superior de educación de la Ciudad de México. *Papeles de Población*, 20(80), 249-280.
- Solís, P., Rodríguez, E., & Brunet, N. (2013). Orígenes sociales, instituciones y decisiones educativas en la transición a la educación media superior. El caso del Distrito Federal. *Revista Mexicana en Investigación Educativa*, 18(50), 1103-1136.
- Rodríguez, E. (2015). El proceso de toma de decisiones en la transición de la secundaria a las instituciones públicas de educación media superior en la Ciudad de México: Un factor clave para entender el problema de la persistencia en la desigualdad de oportunidades educativas (Tesis de doctorado en Ciencias Sociales). *El Colegio de México*.
- García Pinzón, I. (2016). Patrones de elección de los participantes del concurso para el ingreso a la Educación Media Superior de la comipems. *Revista Mexicana de Investigación Educativa*, 21(68), 95-118.
- Estrada, R. (2017). *The effect of the increasing demand for elite schools on stratification* (CAF Working Paper No. 2017/09). Corporación Andina de Fomento (CAF).
- Mata Zúñiga, L. A., & Suárez Ibarra, J. J. (2016). Semejantes y diferentes. Selección social en el bachillerato de la Universidad Nacional Autónoma de México. *Universidades, UDUAL*, 68, abril-junio, 1-20.
- Vélez Grajales, R., Vélez Grajales, V., & Stabridis, O. (2015). *Construcción de un índice de riqueza intergeneracional a partir de la Encuesta ESRU de Movilidad Social en México (EMOVI)* (Documento de trabajo No. 02/2015). Centro de Estudios Espinosa Yglesias.
- Peña, D. (2002). *Análisis de datos multivariantes*. McGraw-Hill Interamericana de España. ISBN 978-8448136109.
- Blanco, E., Solís, P., & Robles, H. (Coords.). (2014). *Caminos desiguales: Trayectorias educativas y laborales de los jóvenes en la Ciudad de México*. Instituto Nacional para la Evaluación de la Educación (INEE) y El Colegio de México. ISBN 978-607-7675-61-7