



# CREACIÓN DEL ÍNDICE DE VULNERABILIDAD SOCIOECONÓMICA DEL PROBEMS

Resumen: De acuerdo a la literatura económica, las medidas de la riqueza de los hogares pueden reflejarse en la información sobre el ingreso, el consumo o el gasto. Sin embargo, la recopilación de datos precisos sobre ingresos y consumo requiere de amplios recursos para la realización de encuestas y la validación de datos de los encuestados. Dado que el análisis de componentes principales (ACP) ha sido validado como un método para describir índices socioeconómicos dentro de una población -ayudando a tener un mayor acercamiento sobre el nivel de riqueza de los hogares- en el presente documento, mediante un ACP, se construyó un índice socioeconómico con los datos del equipamiento y servicios básicos del hogar de los alumnos solicitantes de las becas PROBEMS. Asimismo, se abordan cuestiones relacionadas con la elección de las variables adecuadas y posibles problemas metodológicos, como el agrupamiento y truncamiento de datos. Finalmente, se discute la interpretación de los resultados y la clasificación de los hogares en grupos socioeconómicos para su mejor interpretación.

**PROBEMS**  
**Programa de Becas de Educación Media Superior**  
**Octubre 2018**

## **Introducción**

El Programa de Becas de Educación Media Superior (PROBEMS) opera becas por registro que requieren el llenado de la Encuesta Socioeconómica de Estudiantes de Educación Media Superior (ENCSEEMS) por parte de los solicitantes. En dicha encuesta se solicita, entre otras variables, el ingreso del hogar y el número de integrantes del mismo, con las cuales, se obtiene el ingreso per cápita (IPC).

Por Reglas de Operación, el IPC es una variable que prioriza a los solicitantes (es decir, los estudiantes que reportan menor ingreso tienen mayor probabilidad de obtener un apoyo) por lo que estos tienden a subreportar el ingreso de su hogar, esperando tener mayores oportunidades de recibir una beca.

Los incentivos perversos a mentir para obtener un apoyo, generan un problema de endogeneidad en la selección de beneficiarios, lo que impide identificar a los alumnos cuya necesidad económica es real dentro del grupo de solicitantes. Asimismo, la información sobre el ingreso no capta el hecho de que las personas pueden tener ingresos en especie, lo que representa un problema adicional para la cuantificación y reporte de los ingresos.

Por lo anterior, el PROBEMS concluyó que es necesario establecer mecanismos que permitan una mejor aproximación al ingreso/gasto de los hogares para robustecer la priorización de los alumnos solicitantes de un apoyo económico.

En este sentido, la presente investigación consistió en construir un índice socioeconómico que permitiera “rankear” a los hogares de los alumnos solicitantes de acuerdo a su capacidad de consumo o gasto. Dicho índice corresponde a una aproximación al valor de los activos y servicios básicos del hogar, como determinante del estándar de vida del hogar, y, en última instancia, del gasto/ingreso. Para la creación del índice se utiliza un Análisis de Componentes Principales (ACP) mediante la recopilación de variables que capturan los niveles de vida, como: la propiedad de bienes duraderos (por ejemplo, televisión, automóvil), la accesibilidad a servicios básicos (agua, luz, drenaje) y, la calidad y espacios de la vivienda (tipo de techo, tipo de piso, hacinamiento, etc), los resultados derivaron en la metodología para definir un índice socioeconómico “base” que hiciera comparable las poblaciones solicitantes de las diversas modalidades.

A partir del ciclo escolar 2017-2018, el Programa comenzó a definir perfiles de vulnerabilidad por modalidad de beca, tomando en cuenta las características específicas de cada población. Hasta el momento, se ha creado un índice específico para las modalidades de beca para Prepa en Línea y para la modalidad de Transporte.

El presente documento está distribuido de la siguiente manera: en el primero apartado, se explica de manera general qué es un Análisis de Componente Principales (ACP); en la segunda sección se analiza la viabilidad del uso de las variables de equipamiento y servicios básicos del hogar para la creación de un índice socioeconómico; en la tercera sección se aplica el ACP para la creación del índice socioeconómico y se realizan pruebas para la correcta interpretación del mismo; en la cuarta sección se muestran la creación de los índices específicos para la modalidad de Prepa en Línea y Transporte. Finalmente, se presentan las conclusiones.

## I. ¿Qué es un Análisis de Componentes Principales (ACP)?

El ACP es una técnica estadística multivariable<sup>1</sup> utilizada para reducir el número de variables en un conjunto de datos en un menor número de "dimensiones". Es una forma de identificar patrones en los datos y expresarlos de tal manera que resalten sus semejanzas y diferencias.

En términos matemáticos, a partir de un conjunto inicial de  $n$  variables correlacionadas, el ACP crea índices o componentes no correlacionados, donde cada componente es una combinación lineal ponderada de las variables iniciales.

$$PC_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1n}X_n$$

⋮

$$PC_m = a_{m1}X_1 + a_{m2}X_2 + \dots + a_{mn}X_n$$

Donde  $a_{mn}$  representa el peso para el componente principal de  $m$  y la  $n$ -ésima variable. Los pesos para cada componente principal están dados por los vectores propios de la matriz de correlación. La varianza ( $\lambda$ ) para cada componente principal viene dada por el autovalor del autovector correspondiente.

Los componentes están ordenados de manera que el primer componente (PC1) explique la mayor cantidad posible de variación en los datos originales, sujeto a la restricción de que la suma de los pesos cuadrados ( $a_{11}^2 + a_{12}^2 + \dots + a_{1n}^2$ ) es igual a uno.

Como la suma de los valores propios es igual al número de variables, la proporción de la variación total en el conjunto de datos original representada por cada componente principal viene dada por  $(\lambda/n)$ . El segundo componente (PC2) está completamente no correlacionado con el primer componente, y explica la variación adicional pero menor que el primer componente, sujeto a la misma restricción.

Los componentes subsiguientes no están correlacionados con los componentes anteriores, por lo tanto, cada componente captura una dimensión adicional en los datos, mientras que explica proporciones cada vez más pequeñas de la varianza de las variables originales. Cuanto mayor es el grado de correlación entre las variables originales en los datos, menos componentes se requieren para capturar información común.

## II. Seleccionar las variables para la creación del índice socioeconómico PROBEMS.

El ACP funciona mejor cuando las variables están correlacionadas, pero también cuando la distribución de las variables varía entre los individuos que conforman la población. Para el caso específico de un índice socioeconómico, los activos que se distribuyen más desigualmente entre los individuos reciben más peso en el ACP (McKenzie 2003), esto significa que las variables con baja desviación estándar tendrán un bajo peso en el análisis.

---

<sup>1</sup> Las técnicas estadísticas multivariadas permiten establecer, a partir de numerosos datos y variables, ciertas relaciones y ensayar diversas maneras de organizar dichos datos, transformándolos y presentándolos bajo una forma nueva más asequible, reduciéndolos, sin perder demasiada información inicial, hasta componer un resumen lo más completo posible del conjunto de datos original, habitualmente bastante complejo.

Por ejemplo, un activo que todos los individuos poseen o que ningún individuo posee (es decir, desviación estándar cero) no exhibiría ninguna variación entre los individuos y tendría una ponderación cero y, por tanto, sería de poca utilidad para diferenciar el índice.

En nuestro caso, para la creación del índice socioeconómico PROBEMS, como primer paso, se realizaron análisis descriptivos de todas las variables, considerando medias, frecuencias y desviaciones estándar (ver Tabla 1). El análisis descriptivo puede informar las decisiones sobre qué variables incluir en el análisis y resaltar los problemas de gestión de datos, como la codificación de variables y los “missing values”. Como podemos observar, todas las variables tienen un comportamiento similar y desviaciones estándar entre 0.3 y 0.5, por lo que pueden ser consideradas de utilidad para la elaboración del índice, en un primer acercamiento.

Tabla 1: Estadística descriptiva de las variables de Equipamiento en el Hogar

Variable	Obs	Media	Desviación Estándar	Min	Max
<b>Equipamiento</b>					
Refrigerador	799315	0.8240	0.3808	0	1
Lavadora automática	799315	0.3492	0.4767	0	1
Reproductor de video	799315	0.1544	0.3613	0	1
Vehículo	799315	0.2408	0.4276	0	1
Teléfono fijo	799315	0.4022	0.4903	0	1
Horno	799315	0.2377	0.4257	0	1
Computadora	799315	0.2523	0.4343	0	1
Estufa parrilla	799315	0.8607	0.3463	0	1
Calentador de agua	799315	0.3329	0.4713	0	1
Internet	799315	0.2705	0.4442	0	1
Teléfono móvil	799315	0.7208	0.4486	0	1
Televisión	799315	0.7349	0.4414	0	1
Televisión digital	799315	0.2361	0.4247	0	1
Televisión de paga	799315	0.2037	0.4027	0	1
<b>Calidad y espacios en la vivienda</b>					
Carencia piso	799315	0.0646	0.2458	0	1
Carencia techo	799315	0.0856	0.2798	0	1
Carencia muro	799315	0.0306	0.1723	0	1
Carencia hacinamiento	799315	0.0874	0.2824	0	1
<b>Servicios básicos de la vivienda</b>					
Carencia agua	799313	0.2669	0.4423	0	1
Carencia drenaje	799315	0.1370	0.3438	0	1
Carencia electricidad	799315	0.0078	0.0878	0	1
Carencia combustible	799315	0.1598	0.3664	0	1

Fuente: elaboración propia con datos de ENCSEEMS

Cabe destacar que tener un amplio número de activos para la elaboración del índice permite estimaciones más robustas. McKenzie (2003) destaca que un reto importante para los índices basados en ACP es asegurar que la gama de variables de activos incluidas sea suficientemente amplia para evitar problemas de "agrupamiento" y "truncamiento". El agrupamiento se describe cuando los individuos se agrupan en pequeños y distintos “clusters”. El

truncamiento se observa en la distribución del índice, cuando se hace difícil la diferenciación entre los grupos socioeconómicos (por ejemplo, cuando las distribuciones se sesgan hacia un lado u otro y no se puede distinguir entre los pobres y los muy pobres). Sin embargo, la clave es incluir variables adicionales que capturen la desigualdad entre los hogares.

Siguiendo con la selección de variables, se realizó la matriz de correlaciones para asegurar que todos los datos tengan el mismo peso. Por ejemplo, si se utilizara el número de habitaciones para dormir que es una variable cuantitativa y tiene una mayor varianza que las otras variables binarias, dominaría ésta en el primer componente principal. En nuestro caso, la matriz de correlaciones presenta altas correlaciones entre los componentes (ver anexo 1).

Después de la selección de variables (activos en el hogar), se procedió a la estimación del ACP. El resultado de este análisis es una tabla de puntajes de factores o ponderaciones para cada variable.

El *eigenvalor* o “valor propio” para cada componente principal indica el porcentaje de variación en el total de datos explicados. Porcentajes que no son altos, podría reflejar la complejidad de las correlaciones entre variables, ya que cada variable incluida puede tener un determinante propio distinto en el estatus socioeconómico de cada estudiante.

Asimismo, los datos de varianza explicada son muy importantes para saber cuántos componentes principales vamos a utilizar en nuestro análisis. No hay una regla definida sobre el número que se debe utilizar, con lo cual deberemos decidir en función del número de variables y de la proporción de varianza explicada acumulada. También podemos representar un gráfico de sedimentación (scree plot) de los valores propios como el de la figura 1. Típicamente el gráfico muestra la clara ruptura entre la pronunciada pendiente de los factores más importantes y el descenso gradual de los restantes (los sedimentos).

En este caso, los 3 primeros componentes tienen varianzas (eigenvalores) mayores que 1<sup>2</sup>, con ellos se explica el 34% de la varianza y, a partir del tercer componente, el descenso se estabiliza. Lo que indica que sólo debemos preocuparnos de los tres primeros componentes ya que los siguientes tienen poca capacidad explicativa.

Al final de la tabla de resultados, nos aparecen las correlaciones de cada componente principal con cada variable. Para interpretar los componentes consideramos sus coordenadas en las variables. En este caso, se observa que el primer componente es un factor de tamaño, ya que es una media ponderada de todas las variables, con mayor peso en internet, computadora, estufa y lavadora; por lo tanto, es claro que la posesión de servicios digitales y ciertos electrodomésticos tiene mayor peso para identificar un índice socioeconómico más alto; contrariamente las carencias por servicios básicos y calidad y espacios en la vivienda tienen signo negativo, es decir, tener carencias en la vivienda disminuye el valor del índice socioeconómico. En cuanto al PC2, es un factor de forma y contrapone los indicadores de internet y computadora frente a activos como televisión y estufa.<sup>3</sup>

---

<sup>2</sup> (Criterio de Kaiser, 1960): cada una de las variables observadas contribuyen al menos con una unidad de variación a la variación total; así, si es mayor a 1 el componente principal explica al menos lo mismo que una variable observada.

<sup>3</sup> Cuando existe una alta correlación positiva entre todas las variables, el primer componente principal puede interpretarse como un promedio ponderado de todas las variables, o un factor global de “tamaño”. Los restantes

Tabla 2: Salidas del Análisis de Componentes Principales

Observaciones= 799313  
Componentes= 22

Rotation: (unrotated = principal)			Rho= 1	
Componente	Eigenvalor	Diferencia	Proporción	Acumulada
Comp1	4.405	2.490	0.200	0.200
Comp2	1.915	0.728	0.087	0.287
Comp3	1.187	0.102	0.054	0.341
Comp4	1.085	0.023	0.049	0.391
Comp5	1.063	0.097	0.048	0.439
Comp6	0.966	0.045	0.044	0.483
Comp7	0.921	0.029	0.042	0.525
Comp8	0.892	0.008	0.041	0.565
Comp9	0.884	0.046	0.040	0.605
Comp10	0.838	0.020	0.038	0.643
Comp11	0.818	0.015	0.037	0.681
Comp12	0.803	0.003	0.037	0.717
Comp13	0.800	0.046	0.036	0.753
Comp14	0.755	0.043	0.034	0.788
Comp15	0.712	0.025	0.032	0.820
Comp16	0.687	0.022	0.031	0.851
Comp17	0.664	0.009	0.030	0.882
Comp18	0.655	0.033	0.030	0.911
Comp19	0.623	0.024	0.028	0.940
Comp20	0.599	0.213	0.027	0.967
Comp21	0.385	0.041	0.018	0.984
Comp22	0.344	.	0.016	1.000

Principal components (eigenvectors)

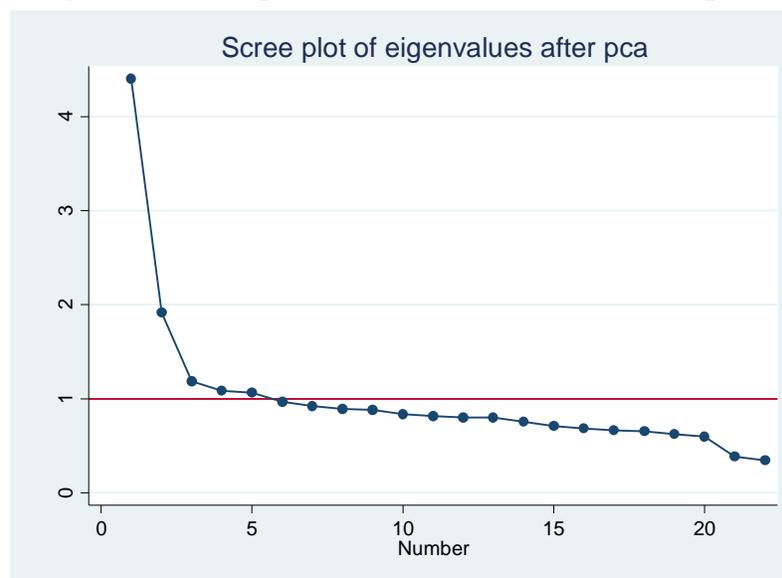
Variable	Comp1	Comp2	Comp3	Comp4	Comp5
Refrigerador	0.2321	-0.1701	0.0776	0.1862	-0.006
Lavadora automática	0.2805	0.1699	0.1424	-0.0206	0.0333
Reproductor video	0.1952	0.2042	0.288	-0.2568	0.0546
Vehículo	0.2456	0.204	0.0434	0.0455	-0.0791
Teléfono fijo	0.2017	0.1021	-0.5118	0.1945	-0.0995
Horno	0.2532	0.2051	0.1509	-0.0635	0.0382
Computadora	0.2685	0.2403	-0.2715	0.0489	-0.124
Estufa parrilla	0.2738	-0.3127	0.067	0.1312	0.3101
Calentador agua	0.2394	0.0841	-0.1156	0.03	0.1055
Internet	0.2958	0.2585	-0.3343	0.1142	-0.1197

componentes se interpretan como factores “de forma” y típicamente tienen coordenadas positivas y negativas, que implica que contraponen unos grupos de variables frente a otros. Estos factores de forma pueden frecuentemente escribirse como medias ponderadas de dos grupos de variables con distinto signo y contraponen las variables de un signo a las del otro

Teléfono móvil	0.1657	-0.0233	0.364	-0.0929	-0.1628
Televisión	0.144	-0.1524	0.3411	0.4818	-0.4147
Televisión digital	0.2005	0.2129	0.0894	-0.4733	0.2763
Televisión paga	0.2217	0.2127	0.2741	0.0822	-0.2101
Carencia piso	-0.225	0.3276	0.0045	0.0569	-0.0759
Carencia techo	-0.0897	0.2468	0.1328	0.3954	0.3351
Carencia muro	-0.1069	0.2365	0.0866	0.3944	0.3483
Carencia hacinamiento	-0.1328	0.1162	0.1452	0.1009	0.1785
Carencia agua	-0.1829	0.1815	0.0597	-0.0012	-0.2218
Carencia drenaje	-0.1241	0.2112	0.1376	0.0227	-0.0775
Carencia electricidad	-0.1046	0.2381	-0.0034	0.1181	0.2818
Carencia estufa	-0.2954	0.3025	-0.0079	-0.1009	-0.3413

Fuente: Elaboración propia con datos de ENCSEEMS

Figura 1: Gráfico para seleccionar el número de componentes

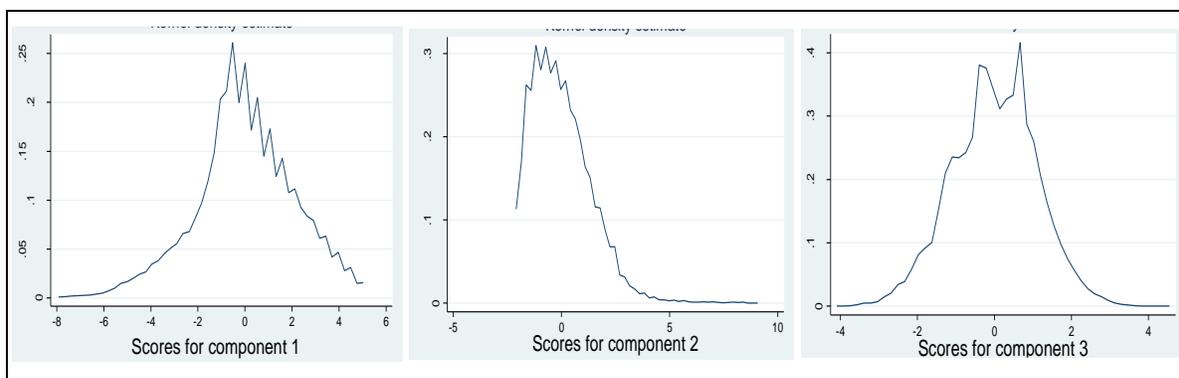


Fuente: Elaboración propia con datos de ENCSEEMS

### III. Índice Socioeconómico PROBEMS

Como se menciona en apartados anteriores, el índice elaborado a través de componentes principales, debe presentar una distribución normal para tener una mejor diferenciación entre los grupos socioeconómicos. En este caso, la distribución de las puntuaciones de los componentes tiende a seguir una curva normal para los componentes 1 y 3; para el caso del componente 2, la población está fuertemente sesgada a la derecha, poniendo de relieve el grado de aglomeración que pudiera dificultar la diferenciación entre los grupos socioeconómicos.

Figura 2: Distribución de los 3 primeros componentes principales



Fuente: Elaboración propia con datos de ENCSEEMS

De esta manera, utilizando las puntuaciones del primer componente principal, dado que es el que capta mayor varianza y tiene una distribución normal, se puede construir una variable dependiente para cada alumno ( $Y_1$ ). Esta variable dependiente puede considerarse como el puntaje socioeconómico de los hogares, y cuanto mayor sea el puntaje socioeconómico del hogar, mayor será el índice socioeconómico implícito de ese individuo.

#### *Clasificación de los alumnos solicitantes en grupos socioeconómicos.*

Para una mejor interpretación del índice, otros estudios han utilizado puntos de corte para diferenciar los hogares en categorías socioeconómicas, por ejemplo, la división de los hogares en quintiles (Gwatkin et al., 2000). Para el presente análisis, se clasificó a los hogares en quintiles y se calculó el puntaje socioeconómico medio para cada grupo (Tabla 3). Se puede observar que la diferencia es pequeña entre los tres quintiles más pobres, ya que cada grupo tiene una puntuación media similar. En el caso de los quintiles más altos, también se presenta una puntuación media similar y, en este caso, mayores diferencias.

El análisis de agrupamiento en quintiles ajustó generalmente los patrones encontrados a partir de la distribución de las puntuaciones socioeconómicas de los individuos mostradas en los histogramas. Asimismo, es importante notar que el ingreso per-cápita sigue la misma distribución por quintiles del puntaje socioeconómico. Es decir, la tendencia del índice socioeconómico obedece la misma tendencia del ingreso per-cápita de los alumnos.

Tabla 3: Puntajes promedio del índice socioeconómico por quintiles.

Variable	Quintiles				
	1	2	3	4	5
Refrigerador	-2.5833	-0.9243	-0.0283	1.0655	2.9216
Lavadora automática	-2.2694	-0.9227	0.0067	1.1003	3.0156
Reproductor video	-2.4177	-0.9160	0.0183	1.1292	3.2566
Vehículo	-2.3130	-0.9284	0.0364	1.1339	3.1121
Teléfono fijo	-2.6627	-0.8831	0.0268	1.1157	3.0277
Horno	-2.3192	-0.9234	0.0206	1.1260	3.1287
Computadora	-2.3274	-0.9338	0.0440	1.1562	3.0685
Estufa parrilla	-2.2557	-0.9350	-0.0309	1.0629	2.9181
Calentador agua	-2.3296	-0.9113	-0.0028	1.0918	3.0729
Internet	-2.2714	-0.8913	0.1006	1.1793	3.0246
Teléfono móvil	-2.7306	-0.9057	-0.0129	1.0774	2.9641

Televisión	-2.6556	-0.9167	-0.0212	1.0675	2.9600
Televisión digital	-2.5453	-0.9201	0.0088	1.1256	3.1504
Televisión paga	-2.3638	-0.9343	0.0179	1.1291	3.1580
Carencia piso	-4.0991	-1.0935	-0.1091	0.9682	2.5300
Carencia techo	-3.8022	-1.0084	-0.0300	1.0637	2.8183
Carencia muro	-4.0676	-1.0236	-0.0417	1.0636	2.7184
Carencia hacinamiento	-3.5421	-1.0754	-0.0889	1.0056	2.5350
Carencia agua	-3.3246	-1.0251	-0.0832	1.0137	2.6475
Carencia drenaje	-3.4950	-1.0288	-0.0737	1.0282	2.7422
Carencia electricidad	-5.5847	-1.0656	-0.0764	1.0299	2.5123
Carencia estufa	-3.4231	-1.1154	-0.1085	0.9702	2.4409
<b>Ingreso Per Cápita</b>	<b>\$489.62</b>	<b>\$639.66</b>	<b>\$747.40</b>	<b>\$891.14</b>	<b>\$1286.73</b>

Fuente: Elaboración propia con datos de ENCSEEMS

Finalmente, para demostrar el grado de correlación que los pesos de los componentes principales (y por consiguiente, el índice) mantienen con el ingreso per-cápita de los alumnos, se corrió una regresión de Mínimos Cuadrados Ordinarios. Los resultados muestran que los 3 componentes analizados son significativos al 1%. Asimismo, se observa que el componente 1 mantiene el mayor peso para ser indicativo del ingreso.

Tabla 4: Estimación del Ingreso per-cápita mediante componentes principales.

VARIABLES	INGRESO_PERCAPITA
Scores for component 1	133.63*** (.3847)
Scores for component 2	35.94*** (.5835)
Scores for component 3	-35.03*** (.7413)
Constant	819.3*** (0.770)
Observations	799313
R-squared	0.1368

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Por otro lado, a partir del ciclo 2017-2018, se utiliza como variable de vulnerabilidad, el Índice de Rezago Social (IRS) que realiza el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) a nivel localidad. Esta variable, al ser continua, presenta una gran varianza comparada con las variables dicotómicas de carencias en el hogar, esto significa que tiene un gran peso en el ACP, lo cual significa que índice socioeconómico incluye factores de vulnerabilidad externa propias de la localidad y factores de capacidad económica medida a través del equipamiento del hogar.

#### **IV. Especificación de índices socioeconómicos por modalidad de beca**

Derivado de la metodología anterior, el índice socioeconómico general describe de manera adecuada el grado de equipamiento en el hogar, sin embargo, dado la diversidad de poblaciones objetivo que atiende el Programa por modalidad, es necesario especificar cada perfil de vulnerabilidad, tomando en cuenta información específica de los solicitantes de acuerdo con las características de la beca solicitada.

A partir de la segunda mitad del ciclo escolar 2017-2018, el PROBEMS ha comenzado a crear perfiles de vulnerabilidad específica, las primeras modalidades para las cuales se han hecho modificaciones son la beca para el Sistema Prepa en Línea y la modalidad de Beca de Transporte.

##### *Prepa en Línea*

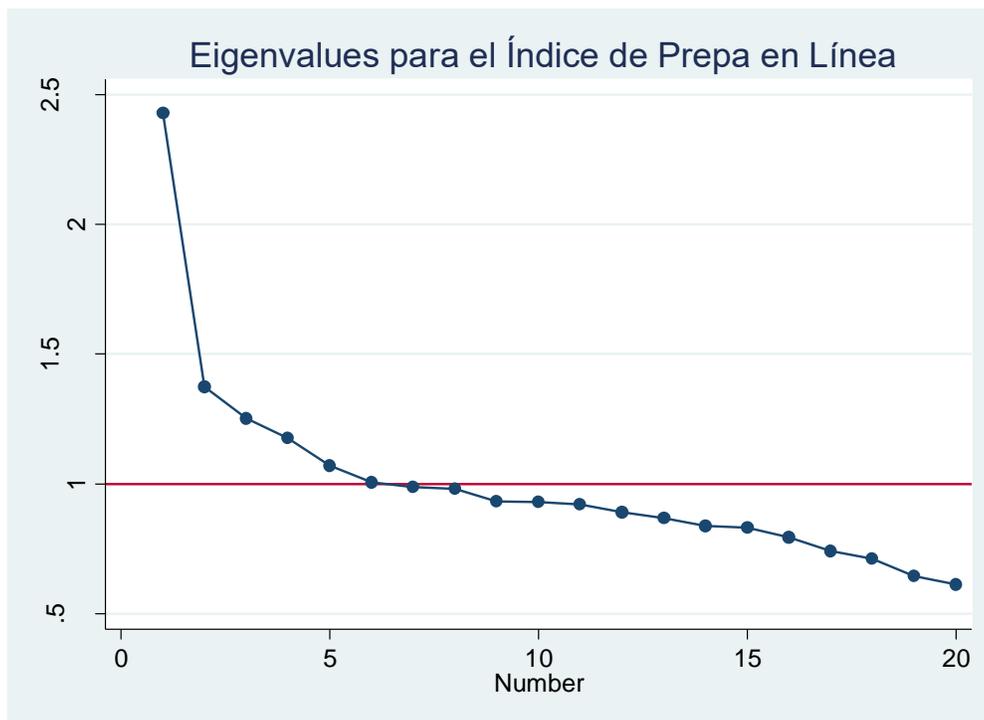
Para la creación del índice, el Programa trabajo en conjunto con miembros del Sistema Nacional de Bachillerato en Línea para delimitar el perfil específico de estudiantes que desertan, con ello se identificó variables que aumentaban la probabilidad de deserción y que fueron incluidas en el ACP adicional a las variables socioeconómicas del índice base. Dichas variables son:

- **Priorización de acuerdo a módulos estudiados:** Se identificó que los alumnos tienden a abandonar el Sistema en el primer módulo y que la probabilidad disminuye conforme se avanza en los módulos.
- **Grupos de edad:** De acuerdo a los datos, el grupo de edad con mayor deserción son los menores de 29 años, por lo tanto, se agregó una variable que toma el valor de 1 cuanto se cae en dicho grupo de edad.
- **Género:** Dada la particularidad del Sistema Prepa en Línea, el perfil que mayor deserción presenta son los hombres, es por ello que se tomó en cuenta esta variable para agregarla al índice de vulnerabilidad.

Como se mencionó en el apartado III, el índice elaborado a través de componentes principales para la modalidad de Prepa en Línea debe presentar una distribución normal para tener una mejor diferenciación entre los grupos socioeconómicos. Para este caso, la distribución de las puntuaciones del sigue una curva normal (Ver Gráfico 3), esto significa que este índice puede ser ocupado para diferenciar a los solicitantes de acuerdo a su perfil de vulnerabilidad específicos.



Por otro lado, el análisis de los valores de los vectores propios nos indica que es adecuado tomar el primer componte como índice que describe la mayor cantidad de varianza, esto significa que el primer componente describe de mejor manera las variables que se tomaron en cuenta para la construcción del índice.

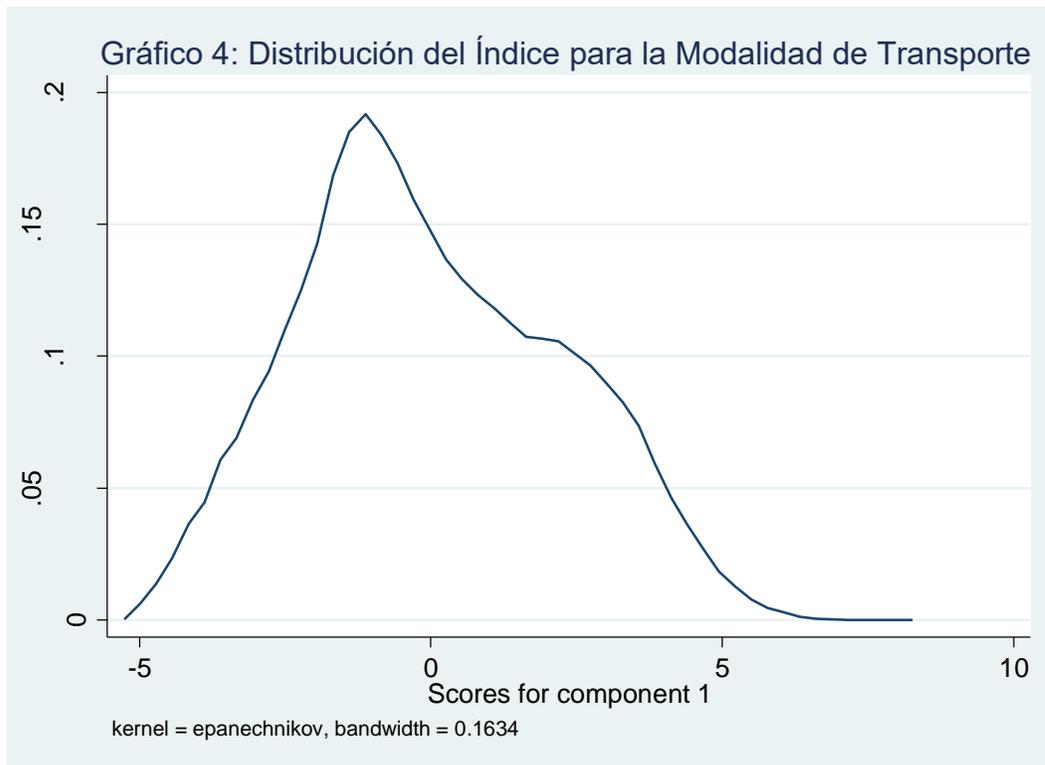


## Transporte

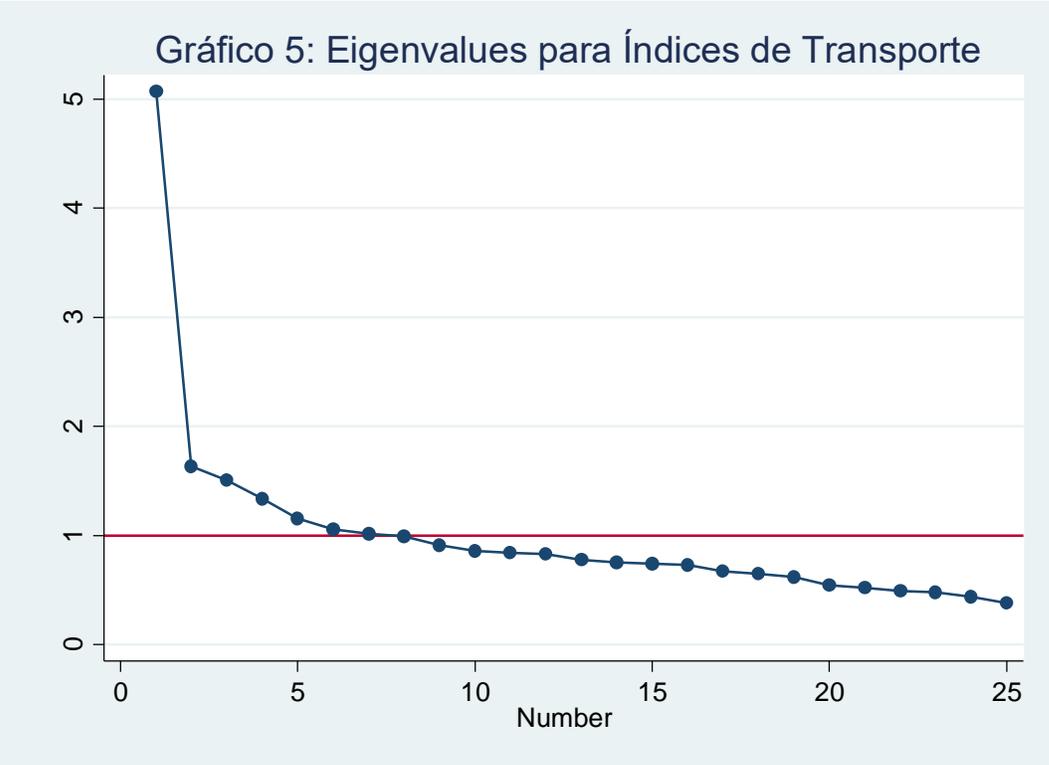
Las Becas de Transporte se enfocan en atender a la población que enfrentan una dificultad para trasladarse de su vivienda al plantel y cuyo gasto en transporte representa una proporción importante del gasto familiar. Para ello se toman en cuenta variables como la distancia y tiempo que hace el/la estudiante de su hogar al plantel, el porcentaje del gasto sumado a su grado de vulnerabilidad económica. Las variables que tomaron en cuenta fueron:

- Gasto en transporte: Se creó una variable que toma un valor igual a 1 si el gasto en transporte es mayor a la tercera parte del ingreso total familiar.
- Medio de transporte: Se agrega vulnerabilidad si el medio de transporte que utiliza el alumno es un animal doméstico como un burro, caballo, etc.
- Tiempo de traslado: Se toma en cuenta si el tiempo que tarde el/la estudiante en trasladarse de su hogar al plantel está por arriba de la media de solicitantes.
- Distancia de traslado: Se toma en cuenta si la distancia que tiene que recorrer el/la estudiante para trasladarse de su hogar al plantel está por arriba de la media de solicitantes.

Como se puede ver en el gráfico 4, la distribución del índice de Transporte se asemeja a una normal, con una ligera curvatura a la derecha, pero centrada en el cero, lo cual permite que los individuos puedan ser diferenciados entre diferentes quintiles de vulnerabilidad.



Como se puede ver en el gráfico 5, el primer componente tiene un alto valor como vector propio, por arriba de 1, que permite ser considerado como adecuado para ser tomado en cuenta.



## **V. CONCLUSIONES**

En el presente documento se obtiene un índice Socioeconómico para los alumnos solicitantes de una beca PROBEMS, llevando a cabo un ACP con variables del equipamiento en el hogar, accesibilidad a servicios básicos y calidad y espacios en la vivienda. La principal ventaja de este análisis radica en que evita muchos de los problemas de medición asociados con los métodos más tradicionales basados en el ingreso y el consumo, como el sesgo de recuerdo (los encuestados pueden olvidar cierta fuente de ingresos), el tiempo de recolección de datos (dado que pueden existir, a lo largo del tiempo, choques en el ingreso por factores exógenos que afectan el ingreso de manera temporal) y el sesgo por autoselección (una de las principales debilidades de la ENCSEEMS dado que el ingreso es autoreportado).

La base empírica de la técnica de componentes principales se basa en si el primer componente principal (PC<sub>1</sub>) puede predecir el estatus socioeconómico de los hogares. Esto dependerá de la naturaleza de los datos y de las correlaciones entre las variables que se están considerando, de la validez de las variables incluidas y de su fiabilidad (Vyas, S., & Kumaranayake, L. 2006).

En este estudio, después de realizar dichas pruebas de validación, se utiliza el primer componente principal para la creación del índice. Asimismo, se demuestra la coherencia que guardan las clasificaciones basadas en el ACP y los ingresos de los hogares de los alumnos solicitantes. De esta manera, se concluyó que el índice creado mediante ACP es un proxy robusto para “rankear” la riqueza o estatus socioeconómico de los hogares y obtener una mejor priorización de los grupos más vulnerables al momento de otorgar un apoyo económico. Finalmente, la adecuación del índice de acuerdo al tipo de modalidad, permite focalizar de mejor manera los apoyos de acuerdo a las características específicas de la población objetivo de cada beca.

## **BIBLIOGRAFÍA**

- Smith, L. I. (2002). A tutorial on principal components analysis. Cornell University, USA, 51, 52.
- Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: how to use principal components analysis. *Health policy and planning*, 21(6), 459-468.
- McKenzie, D. J. (2003). Measure inequality with asset indicators. Cambridge, MA: Bureau for Research and Economic Analysis of Development. Center for International Development, Harvard University.
- Gwatkin, D. R., Rustein, S., Johnson, K., Suliman, E., Wagstaff, A., & Amouzou, A. (2000). Socio-economic differences in Brazil. Washington, DC: HNP/Poverty Thematic Group of the World Bank.

## ANEXO 1

	Refrigerador	Lavadora automática	Reproductor video	Vehículo	Teléfono fijo	Horno	Computadora	Estufa parrilla	Calentador agua	Internet	Teléfono móvil	TV	TV digital	TV paga	Carencia piso
Refrigerador	1														
Lavadora automática	0.2274	1													
Reproductor video	0.0878	0.246	1												
Vehículo	0.183	0.3403	0.1869	1											
Teléfono fijo	0.152	0.1702	0.0979	0.179	1										
Horno	0.1829	0.3494	0.2696	0.281	0.155	1									
Computadora	0.17	0.2994	0.2181	0.296	0.265	0.2717	1								
Estufa parrilla	0.3165	0.2181	0.1248	0.164	0.154	0.1704	0.1617	1							
Calentador agua	0.1662	0.2642	0.1972	0.269	0.237	0.2757	0.2417	0.2187	1						
Internet	0.1882	0.337	0.208	0.314	0.423	0.3118	0.5686	0.1818	0.2569	1					
Teléfono móvil	0.1504	0.1745	0.1489	0.141	0.048	0.1514	0.1414	0.1723	0.0913	0.1358	1				
TV	0.2002	0.1319	0.1009	0.107	0.068	0.1104	0.0764	0.2002	0.0982	0.0809	0.146	1			
TV digital	0.1283	0.2699	0.2802	0.213	0.12	0.2504	0.2229	0.1371	0.1747	0.2452	0.1428	-0.089	1		
TV paga	0.1482	0.2916	0.2276	0.276	0.112	0.2644	0.2419	0.1334	0.1609	0.3238	0.1605	0.1925	0.2213	1	
Carencia piso	-0.2504	-0.1603	-0.089	-0.12	-0.12	-0.1232	-0.1295	-0.3637	-0.1541	-0.142	-0.138	-0.1552	-0.1031	-0.1031	1
Carencia techo	-0.0954	-0.0395	-0.018	-0.02	-0.05	-0.022	-0.0461	-0.1305	-0.0532	-0.0337	-0.066	-0.0597	-0.0239	-0.0055	0.1769
Carencia muro	-0.0978	-0.0569	-0.0364	-0.04	-0.06	-0.0423	-0.0549	-0.1452	-0.0677	-0.0552	-0.0738	-0.0753	-0.0424	-0.0327	0.1893
Carencia hacinamiento	-0.1383	-0.1028	-0.0502	-0.11	-0.11	-0.0922	-0.1036	-0.1488	-0.1306	-0.104	-0.0736	-0.0641	-0.0606	-0.0623	0.1522
Carencia agua	-0.1689	-0.1574	-0.0834	-0.1	-0.11	-0.1283	-0.1297	-0.2238	-0.1484	-0.1589	-0.1025	-0.1048	-0.1017	-0.0843	0.2164
Carencia drenaje	-0.1411	-0.0866	-0.0192	-0.06	-0.07	-0.0649	-0.0826	-0.1595	-0.0813	-0.091	-0.0634	-0.0706	-0.0422	-0.0378	0.1669
Carencia electricidad	-0.1428	-0.0541	-0.0283	-0.04	-0.05	-0.0411	-0.0444	-0.1421	-0.0511	-0.0481	-0.0714	-0.108	-0.0376	-0.0388	0.1836
Carencia estufa	-0.3064	-0.2496	-0.1401	-0.18	-0.18	-0.1992	-0.1987	-0.6453	-0.2478	-0.2305	-0.1621	-0.1742	-0.1536	-0.1452	0.4117

Fuente: Elaboración propia con datos de ENCSEEMS

## ANEXO 1

	Carencia techo	Carencia muro	Carencia hacinamiento	Carencia agua	Carencia drenaje	Carencia electricidad	Carencia estufa
Carencia techo	1						
Carencia muro	0.1824	1					
Carencia hacinamiento	0.0934	0.0977	1				
Carencia agua	0.0892	0.0986	0.0798	1			
Carencia drenaje	0.0991	0.0861	0.0734	0.1665	1		
Carencia electricidad	0.1026	0.1245	0.0842	0.1099	0.1126	1	
Carencia estufa	0.1331	0.1548	0.157	0.2888	0.2003	0.1419	1

Fuente: Elaboración propia con datos de ENCSEEM