



FACULTAD
DE CIENCIAS
ECONÓMICAS



Universidad
Nacional
de Córdoba

REPOSITORIO DIGITAL UNIVERSITARIO (RDU-UNC)

Construcción de un índice de desigualdad de género, aplicación del análisis de componentes principales

María Carolina Togliero, Martín Saino

Ponencia presentada en XLVI Coloquio Argentino de Estadística-XLVI CAE y 4ta Jornada de Educación Estadística realizado en 2018 en la Facultad de Ciencias Económicas - Universidad Nacional de Río Cuarto. Río Cuarto. Córdoba, Argentina



Esta obra está bajo una [Licencia Creative Commons Atribución 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

que trabaja con el total de las componentes. Otra ventaja del método VB es que permite obtener las mismas estimaciones máximo-verosímiles que con EM, si se utiliza como a posteriori (denominadas estimaciones MAP).

Referencias bibliográficas

- Bishop, C. M., & Mitchell, T. M. (2014). *Pattern Recognition and Machine Learning*. New York: Springer.
- Kingma, D. P., & Welling, M. (2014). Auto-encoding variational bayes. *International Conference on Learning Representations Proceedings*. Banff, Alberta, Canada.
- McLachlan, G., & Krishnan, T. (2007). *The EM algorithm and extensions* (Vol. 382). John Wiley & Sons.,
- Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal processing magazine*, 13(6), 47-60.
- Orbanz, P. (2009). Construction of nonparametric Bayesian models from parametric Bayes equations. In *Advances in neural information processing systems* (pp. 1392-1400).
- Tzikas, D. G., Likas, A. C., & Galatsanos, N. P. (2008). The variational approximation for Bayesian inference. *IEEE Signal Processing Magazine*, 25(6), 131-146.

Palabras Clave: Aprendizaje Automático - EM- Inferencia Bayesiana Variacional

Área de Aplicación: Otras Aplicaciones (Aprendizaje Automático).

CONSTRUCCIÓN DE UN ÍNDICE DE DESIGUALDAD DE GÉNERO: APLICACIÓN DEL ANÁLISIS DE COMPONENTES PRINCIPALES.

María Carolina Trogliero, Martín Saino

Facultad de Ciencias Económicas, Universidad Nacional de Córdoba

martinsaino@gmail.com

Resumen

El análisis de las desigualdades, y particularmente, la medición de las inequidades de la sociedad, se ha centrado, en general, en las carencias y exclusiones que afectan a la mayoría en términos de ingresos, de educación y de participación. Con esta investigación se propone modificar el foco de atención y reconocer la existencia de otro tipo de inequidad imperante y transversal a todas las dimensiones de la sociedad: la desigualdad de género. El principal problema con este tipo de desigualdades, es que no solo implican intrínsecamente una injusticia, sino que también limitan el logro de una sociedad más igualitaria, cohesionada e incluyente (CEPAL, 2001).

De aquí surge la motivación de la construcción, mediante el uso de técnicas de análisis multivariante, de un índice de desigualdad de género para los aglomerados que conforman las provincias argentinas. La revisión de la literatura evidencia la novedad en la temática y la necesidad de explotar y aprovechar la información vigente en cada país. Hoy en día, existe una fuerte inclinación hacia la construcción de índices con propósitos de comparación internacional lo que obliga a que el mismo sea primitivo y a veces poco representativo de las sociedades más desarrolladas, dada la limitada información en algunos países subdesarrollados o en vías de desarrollo.

Se considera de gran importancia aportar herramientas que permitan medir en forma más objetiva la realidad Argentina, explorando los diferentes usos de los métodos multivariados tanto para realizar una fotografía descriptiva de la realidad como para construir índices sintéticos.

Objetivo: Construcción y medición de un índice de desigualdad de género de los aglomerados de las provincias argentinas y el total nacional para aportar herramientas de sondeo de la inequidad existente.

Metodología: Se hace uso de herramientas de análisis exploratorio de datos y de técnicas de análisis multivariante con el propósito de medir la desigualdad de género dentro de cada unidad de análisis y, luego, comparar la situación de cada unidad respecto al total del país. Dado que el problema de la desigualdad de género es transversal a todos los campos de una sociedad, se definen cuatro dimensiones teóricas de análisis: dimensión económica, dimensión educativa, dimensión política y dimensión de salud.

La fuente de datos es diversa. Se utilizará principalmente la base de datos del tercer trimestre del año 2013 de la Encuesta Permanente de Hogares (EPH) y la Encuesta sobre Trabajo No Remunerado y Uso del Tiempo realizada por el INDEC el tercer trimestre del mismo año. Además, se recolectan datos de las Estadísticas Vitales, los mapas de género de la Oficina de la Mujer de la Corte Suprema de Justicia y de las Cámaras de Diputados y Senadores de las provincias argentinas. Para aquellas variables/indicadores que no se encuentran contempladas en las EPH se utilizará información provincial como proxy.

La investigación fue dividida en dos etapas. La primera etapa está asociada a la construcción de mapas factoriales mediante el uso de Análisis de Componentes

Principales (ACP) con el propósito de situar y describir el estado de cada dimensión teórica previamente definida.

El ACP es un método multivariante de reducción de la dimensión, es decir, permite disminuir la dimensionalidad de un problema. Tiene como objetivo explicar la mayor parte de la variabilidad total con el menor número posible de componentes, de forma que se produzca la menor pérdida de información (Peña, 2002). El mismo transforma un conjunto de variables originales (p) correlacionadas entre sí, en un nuevo conjunto de variables (k), denominadas Componentes principales, los cuales se caracterizan por estar incorrelacionadas entre sí; donde $k < p$.

Se aplica el ACP a cada dimensión teórica considerando un total de 22 variables ($p=8$ en la dimensión económica, $p=6$ en la dimensión educativa, $p=4$ en la dimensión política y $p=4$ en la dimensión de salud). Al emplear el ACP se disminuye cada campo teórico a dos dimensiones ($R=2$), permitiendo graficar en los ejes cartesianos los mapas factoriales y el círculo unitario. Cada mapa factorial evidencia disparidades existentes entre hombres y mujeres en cada jurisdicción en los diversos campos teóricos definidos.

En una segunda etapa, se aplica el ACP con el propósito de obtener los ponderadores para la construcción del índice sintético. La fortaleza del método surge de brindar la posibilidad de solucionar dos típicos problemas asociados a la elaboración de índices. Por un lado, evita la doble contabilización de la información que puede existir en la agregación del sistema de indicadores. Por otro lado, asigna valores concretos a los ponderadores asociados a cada variable (Saino, 2016).

Existen diversos procedimientos que utilizan el ACP para obtener los valores de los índices sintéticos. Algunos autores proponen como índice sintético al primer componente principal, el cual es una combinación lineal de las variables originales donde las mismas están ponderadas por los vectores característicos. De esta forma, se consigue establecer un criterio objetivo de ponderación. Otros autores, sugieren utilizar como ponderadores de las variables originales estandarizadas el producto del porcentaje de la varianza explicada por cada componente y los coeficientes de cada variable original asociados a la definición del componente. Otra posibilidad es establecer como ponderador el producto entre los autovalores – varianza del componente principal- y los vectores propios. Se opta por el tercer procedimiento para definir los ponderadores.

Los vectores propios (a_{ij}) del componente j , son los valores por los que esta ponderada cada variable en dicho componente principal. Es decir; $CP_j = a_{1j}Z_1 +$

$a_{2j}Z_2 + a_{3j}Z_3 + \dots + a_{pj}Z_p$. Mientras que los autovalores (λ_{ij}) de cada componente principal se obtienen sumando las cargas factoriales al cuadrado correspondientes a dicho componente, donde las cargas factoriales no son otra cosa que los coeficientes de correlación entre la variable y dicho componente principal (ρ_{ij}). La varianza del componente principal j es entonces: $\lambda_j = (p_{1j})^2 + (p_{2j})^2 + (p_{3j})^2 + \dots + (p_{pj})^2$. De esta forma, el ponderador de la variable original i queda definido:

$$w_i = \sum_{j=1}^k a_{ij} \cdot \lambda_j$$

A su vez, siguiendo la metodología del Programa de las Naciones Unidas para el Desarrollo (PNUD), se construyó un índice sintético para cada jurisdicción según sexo, para luego tomarle la diferencia y así medir la desigualdad de género existente. El índice para cada unidad de análisis queda definido como:

$$I_x^g = \sum_{i=1}^p w_i \cdot Z_{ix}^g$$

Donde Z_{ix} es el valor que asume la variable i en la jurisdicción x .

Entonces, para la construcción del índice de desigualdad de género se realiza la diferencia entre el valor que asume el índice para los valores asociados a los hombres y el valor que asume el mismo con la información femenina, se considera el valor absoluto de la diferencia:

$$IDG(X) = I_X^H - I_X^M = \sum_{i=1}^p w_i \cdot Z_{ix}^H - \sum_{i=1}^p w_i \cdot Z_{ix}^M = \sum_{i=1}^p w_i \cdot (Z_{ix}^H - Z_{ix}^M)$$

Resultados

Para obtener los valores del $IDG(X)$ se utiliza el Software R, se corre el ACP considerando las 22 variables definidas con anterioridad. Se realiza el Test de Esfericidad de Barlett y se evalúa la medida de adecuación muestral de Kaiser-Meyer-Olkin. En ambos casos, los resultados son más que aceptables para la aplicación del ACP.

Como criterio de retención del número de componentes, se utiliza *el criterio de la media aritmética con variables normalizadas*. Es decir, se retienen aquellos componentes con varianza mayor a 1. Por lo tanto, se mantienen los 7 primeros componentes principales. De esta forma, se reduce la dimensionalidad del problema de 22 a 7 a través del ACP, y luego, se construye un índice unidimensional que sintetiza el

80% de la información. Los valores que asume el índice de desigualdad de género para cada jurisdicción se observa, ordenado de mayor a menor, en la siguiente tabla:

Tabla 1. Índice de Desigualdad de Género

Jurisdicción	I_x^H	I_x^M	IDG (X)	IDG (X) normalizado
Catamarca	-14,77748	28,97081	43,74829	1,0000
Santiago del Estero	-20,41587	17,62054	38,03641	0,8694
Tucumán	-15,87357	17,35295	33,22652	0,7595
Chaco	-49,68736	-18,69194	30,99542	0,7085
Salta	-3,37682	26,65009	30,02691	0,6864
Santa Fe	-15,55514	11,81669	27,37182	0,6257
Total Nacional Urbano	-6,67785	19,26121	25,93906	0,5929
Córdoba	3,85203	28,56546	24,71343	0,5649
San Juan	-19,64678	2,33332	21,98010	0,5024
Entre Ríos	1,94341	23,63416	21,69075	0,4958
San Luis	-17,02892	3,13233	20,16125	0,4608
Santa Cruz	16,39091	-3,44762	19,83854	0,4535
Neuquen	-5,57805	14,10774	19,68579	0,4500
Río Negro	-13,65116	3,15603	16,80719	0,3842
Chubut	-8,15886	7,65133	15,81019	0,3614
Buenos Aires	-3,87253	10,04650	13,91904	0,3182
La Pampa	-13,99522	-0,77344	13,22178	0,3022
Corrientes	-11,89551	1,17931	13,07482	0,2989
Mendoza	0,60601	13,31705	12,71104	0,2905
Jujuy	-2,14281	9,72445	11,86726	0,2713
Formosa	-19,69636	-10,65519	9,04117	0,2067
CABA	29,51661	37,63955	8,12294	0,1857
Tierra del Fuego	4,54502	7,18350	2,63848	0,0603
La Rioja	-20,33702	-20,77913	0,44212	0,0101
Misiones	-11,87892	-11,60345	0,27547	0,0063

Es preciso explicitar que se realiza una aplicación del ACP considerando observaciones activas (hombres) y suplementarias (mujeres). La obtención de los ponderadores se realiza considerando solo aquellas observaciones activas. Con el propósito de que los índices para cada sexo sean comparables y consistentes entre sí, se utilizan los mismos ponderadores para las observaciones suplementarias.

Se realizó una normalización por el máximo. Puede observarse que la provincia con mayor desigualdad de género es Catamarca, le siguen en el ranking Santiago del Estero, Tucumán, Chaco, Salta y Santa Fe. En los seis casos nombrados, la desigualdad de género es mayor que la media a nivel nacional. Era de esperar que las primeras cinco jurisdicciones tuvieran los valores más altos, esto se debe a que están entre las provincias más conservadoras del país. Llama la atención la posición de Santa Fe. El resto de las jurisdicciones se encuentran con un nivel de desigualdad menor que la media nacional.

Es menester señalar que al construir el índice sintético de la forma en la que se lo hizo, y al obtener la desigualdad como una diferencia entre índices, no se puede inferir si un alto valor del $IDG(X)$ implica inequidad favorable para hombres o mujeres. Sin embargo, la pormenorización de cómo afecta esta desigualdad a cada uno de los géneros fue analizada en apartados anteriores.

Referencias bibliográficas

Bardhan, K., y Klasen, S. (1999). *UNDP's gender-related indices: A critical review*.

World Development, 27(6), pp. 985-1010.

- Bericat A., Eduardo (2012). *The European Gender Equality Index: Conceptual and Analytical Issues*. Social Indicators Research, Vol. 108, No. 1, pp. 1-28.
- Branisa, Boris; Klasen, Stephan; Ziegler, Maria (2009). *New Measures of Gender Inequality: The Social Institutions and Gender Index (SIGI) and its Subindices*. Courant Research Centre: Poverty, Equity and Growth. Discussion Papers, No. 10
- Castaño, Cecilia (1999). *Economía y género*. Universidad Complutense de Madrid. Política y Sociedad, 32, pp. 23-42.
- Dijkstra, A. G. (2002). *Revisiting UNDP's GDI and GEM: Towards an alternative*. Social Indicators Research, 57, pp. 301-338.
- Frias, S. (2008). *Measuring structural gender equality in Mexico: A state level analysis*. Social Indicators Research, 88, 215-2.
- Klasen, S. (2006). *UNDP's gender-related measures: Some conceptual problems and possible solutions*. Journal of Human Development, 7(2), pp. 243-274.
- Klasen, S. y D. Schüller (2011). *Reforming the Gender-related Development Index (GDI) and the Gender Empowerment Measure (GEM): Implementing some specific proposals*. Mimeo. University of Goettingen
- López-Claros, A. y S. Zahidi (2005). *Women's Empowerment: Measuring the Global Gender Gap*. Davos: World Economic Forum.
- Peña, Daniel (2002). *Análisis de Datos Multivariantes*,
- Permanyer, Iñaki (2010). *The Measurement of Multidimensional Gender Inequality: Continuing the Debate*. Social Indicators Research, Vol. 95, No. 2, pp. 181-198.
- Saino, O. Martín; Trogliero, M. Carolina y Badino, J.P. (2016). *Análisis de Componentes Principales para la construcción de índices sintéticos*. Universidad Nacional de Córdoba.
- Social Watch (2005). *Roars and Whispers Gender and Poverty: Promises versus Action*. Montevideo: Social Watch
- United Nations Economic Commission for Africa (2016). *The African Gender and Development Index: Phase 4*. Concept Notes.
- Wieringa, Saskia E. (1997). *Report of the Workshop on GDI/GEM Indicators*. Mimeo. The Hague: Institute of Social Studies.
- Palabras Clave:** Desigualdad de género, Análisis de componentes principales, Índice sintético