



Índice de Diversidad Heterogénea: una medida de variabilidad y disparidad

Heterogeneous Diversity Index: A measure of variability and disparity

Luz Judith Rodríguez Esparza*, Julio César Macías Ponce** y
Roberto Alejandro Kú Carrillo***

Información del artículo	Resumen
Recibido: 16 octubre 2023	En este artículo se propone un nuevo índice de heterogeneidad que compara dos vectores, uno de observación y otro el esperado. Un caso particular resulta cuando el vector esperado tiene las mismas entradas en todos sus elementos, entonces, el índice propuesto resulta ser de desigualdad. La construcción geométrica y analítica del índice propuesto le confiere robustez matemática, además de su amplia gama de aplicaciones en distintas disciplinas como economía, estadística y sociología. En este trabajo hemos aplicado el índice en tópicos como el reparto de bienes, el desarrollo humano y el salario mínimo, obteniéndose resultados bastantes satisfactorios.
Aceptado: 01 noviembre 2024	
Clasificación JEL: C433, D60, I14, I24.	
Palabras clave: heterogeneidad, desigualdad, Gini, distribución de bienes, índice de desarrollo humano.	
Article information	Abstract
Received: 16 October 2023	This paper proposes a new heterogeneity index that compares two vectors, one for observations and the other for expected values. A particular case arises when the expected vector has the same entries in all of its elements, thus, the proposed index serves as a measure of inequality. The geometric and analytical construction of the proposed index gives it mathematical robustness, in
Accepted: 01 November 2024	

* Universidad Autónoma de Aguascalientes, luz.rodriguez@edu.uaa.mx, <https://orcid.org/0000-0003-2241-1102>.

** Universidad Autónoma de Aguascalientes, cesar.macias@edu.uaa.mx, <https://orcid.org/0000-0001-5141-7074>.

*** Universidad Autónoma de Aguascalientes, alejandro.ku@edu.uaa.mx, <https://orcid.org/0000-0003-0425-0122>.



JEL Classification:
C433, D60, I14, I24.

Keywords:
heterogeneity,
inequality, Gini,
allocation of goods,
human development
index.

addition to its wide range of applications in different disciplines such as economics, statistics and sociology. In this work, we have applied the index to topics such as the allocation of goods, and human development, and minimum wage analysis, obtaining quite satisfactory results.

Introducción

De acuerdo a la Real Academia Española (RAE), la heterogeneidad se define como la mezcla de partes de distinta naturaleza. Este concepto es crucial en la investigación de diversas áreas como la economía, la salud, la estadística y la ecología, por mencionar algunas. La heterogeneidad no solo es fundamental para entender la complejidad de los sistemas naturales y sociales, sino que también tiene implicaciones prácticas en la toma de decisiones y en la formulación de políticas. Su impacto se extiende a paradigmas de casos y controles, estudios epidemiológicos, análisis de desigualdad económica y la gestión de recursos naturales, entre otros. Generalmente, la heterogeneidad se mide a través de la varianza, una herramienta estadística que captura la diversidad dentro de un conjunto de datos. Es por esto que medir y analizar la diversidad o variabilidad en un conjunto de datos heterogéneo es un problema importante, que para resolverlo, podemos hacer uso de la relación estrecha entre la heterogeneidad y la desigualdad. Nunes, Trappenberg y Alda (2020) presentan a la heterogeneidad como el grado en que un sistema se desvía de la conformidad perfecta.

Por otro lado, los índices son medidas estadísticas que representan la magnitud de atributos de un ente o fenómeno y los podemos encontrar en distintos contextos. Existen diversos tipos de índices, tales como los índices sociales, económicos, de salud, ambientales, entre otros. Además, estos ofrecen la versatilidad de compactar varias categorías en un sólo valor. Un ejemplo destacado es el Índice de Desarrollo Humano (IDH) (Salazar y García, 2014) calculado por el Programa de las Naciones Unidas para el Desarrollo (PNUD)¹, el cual mide los logros de un país en su desarrollo humano, fundamentalmente a través de tres aspectos: salud, educación e ingresos; pudiéndose así clasificar personas, bienes, naciones, entre otros, con respecto a las variables analizadas. Particularmente, si nos enfocamos en índices que miden la heterogeneidad y desigualdad,

¹ <https://report.hdr.undp.org>

podemos encontrar que existen varios utilizados en diferentes contextos y disciplinas. Algunos de los índices más comunes son: el índice de Gini (Gini y Salvemini, 1912), el cual se utiliza para medir la desigualdad económica y social en una población; el índice de Theil, el cual se utiliza para medir la desigualdad en la distribución de ingresos o de cualquier otra variable, éste proporciona información sobre la contribución de cada estrato a la desigualdad total (Cotler, 2007); el Coeficiente de Variación, el cual mide la variabilidad relativa de una variable en relación con su media, además se utiliza para evaluar la heterogeneidad de una muestra o una población (Eliazar, 2017); y el índice de Simpson (Simpson, 1949), que es una medida de la diversidad que tiene en cuenta tanto el número de especies presentes como la abundancia relativa de cada una. Éste se define como la probabilidad de que dos individuos seleccionados al azar de una muestra pertenezcan a la misma especie. Este índice varía entre 0 y 1, donde valores cercanos a 0 indican alta diversidad y valores cercanos a 1 sugieren baja diversidad. El índice de Simpson se utiliza ampliamente en ecología y biología de la conservación para evaluar la biodiversidad de un hábitat, pero su aplicación se extiende a otras disciplinas, como la economía y la sociología, para analizar la distribución y la desigualdad en diferentes contextos.

En particular, los índices de desigualdad son puntajes cuantitativos que toman valores en el intervalo unitario, con el puntaje de 0 que caracteriza la igualdad perfecta, y el valor de 1 que significa la desigualdad total. El índice de Gini es el índice más popular que se emplea para medir la concentración de los ingresos entre los habitantes de una región en un tiempo determinado. Plata, Sánchez y Sánchez (2015) dan una caracterización del índice de Gini, como la única función que satisface las propiedades de invariancia de escala, simetría, proporcionalidad y separabilidad comonótona en clasificaciones similares. La identificación de éstas es importante, más allá de la caracterización, pues dan sentido a su aplicación en un contexto particular.

Ahora discutamos la aplicación de un índice de desigualdad al siguiente problema. Si nos preguntamos qué tan desiguales son los vectores $(0,1,1,1)$ y $(1,0,0,0)$ pensaríamos que tienen la misma desigualdad —quizá en un contexto cualitativo—, pues ambos tienen tres entradas iguales (independientemente del valor que contiene cada entrada) y una diferente. Sin embargo, el índice de Gini les otorga los valores de $\frac{1}{4}$ y $\frac{3}{4}$, respectivamente, por lo tanto, se puede argumentar que el índice de Gini no captura la parte cualitativa de los datos. Esto ocurre, pues el índice de

Gini otorga un mayor peso a los valores más pequeños de la muestra, de este modo, destaca más la desigualdad si existe una porción mayor de la muestra con valores menores que cuando la mayoría tiene valores altos. Dado que el índice de Gini es 0 cuando todos los valores de la muestra son iguales, podemos decir que, su referencia es la igualdad total pero sesgada hacia la "pobreza".

Otro aspecto interesante de los índices de desigualdad es que se calculan teniendo en cuenta únicamente los datos observados. Por lo que, una cuestión relevante sería proponer un índice que compare el vector de datos observado con un vector de datos esperado, de tal forma que, entre mayor sea el valor obtenido del índice, mayor es la diferencia/distancia entre el vector observado y el esperado. Este es el concepto que conocemos de métrica entre dos vectores de datos, pero en nuestro caso, se busca además cumplir con propiedades y características propias de un índice. Note que tendríamos un índice de desigualdad, pero respecto de una referencia o vector esperado (que puede ser distinto del vector con todos sus elementos iguales), es decir, un índice de heterogeneidad. Un caso particular sería cuando el vector esperado esté formado por 1's, esto se podría interpretar como que todos los datos del vector observado tienen el mismo "peso", y entonces se tendría homogeneidad completa o nula desigualdad.

Así pues, nuestra principal motivación en este trabajo es proponer un índice de heterogeneidad que le otorgue a los vectores observados $(0,1,1,1)$ y $(1,0,0,0)$ el mismo valor, considerando el vector esperado $(1,1,1,1)$, de tal manera que se obtengan analogías y diferencias con el índice de Gini. Pero además, que el índice propuesto se pueda calcular para cualquier vector esperado, y así mida la variabilidad de los datos y comparación entre grupos, de tal forma que permita la identificación de patrones.

Este artículo está organizado de la siguiente manera. En la Sección 1 se presentan los antecedentes. En la Sección 2 se presenta la propuesta de la definición del nuevo índice de heterogeneidad. Aplicaciones de los resultados a problemas reales se presentan en la Sección 3, mientras que los comentarios finales se presentan en la Sección 4.

1. Marco Teórico

Los índices permiten obtener resultados cuantitativos a las distintas problemáticas que se presentan en la sociedad. Estos índices se crean precisamente para dar un diagnóstico de las situaciones y de esta manera

ser capaces de darles seguimiento, hasta lograr anticipar ciertas tendencias de los fenómenos reales.

La transición entre estos índices tradicionales y los compuestos o sintéticos (Commision *et al*, 2008; Rodríguez, Reyes y Favela, 2016) radica principalmente en que los analistas deciden los parámetros, variables, además de la metodología a utilizar, esto con el objetivo de facilitar la comprensión de información. Cabe señalar que estos indicadores no sustituyen a los tradicionales ya existentes, sino más bien los complementan. Un índice compuesto muy importante hoy en día es el Índice de Desarrollo Humano (IDH) presentado desde 1990 por PNUD (PNUD, 2015). El IDH se calcula tomando en cuenta tres dimensiones principales del desarrollo humano: la salud, la educación y el nivel de vida. Estas dimensiones se reflejan en tres indicadores específicos: la esperanza de vida al nacer, la educación (medida por la tasa de alfabetización de adultos y la tasa bruta de matriculación en educación primaria, secundaria y superior) y el ingreso nacional bruto per cápita. El estudio del IDH y su cálculo está estrechamente relacionado con la economía del desarrollo, que se enfoca en comprender y analizar los procesos económicos y sociales que influyen en el bienestar humano y el desarrollo de los países. Adicionalmente, el IDH ofrece una medida compuesta del desarrollo humano que facilita la comparación y clasificación de los países en función de su nivel de desarrollo.

Además de la economía, el cálculo de los índices también se cruza con otras disciplinas, como la sociología, la demografía, la estadística y la política, ya que aborda cuestiones relacionadas con la calidad de vida, la equidad social y el desarrollo sostenible. Dependiendo el área de estudio, las unidades de medida cambian, por ejemplo, los indicadores económicos utilizan unidades monetarias y productos, mientras que los índices sociales utilizan personas; y los ambientales, recursos naturales (Mondragón, 2002).

Los índices en general se construyen sólo teniendo como insumo los valores de las variables involucradas, y así obtienen valores entre 0 y 1, pudiéndose comparar entre países, entidades, etcétera. Sin embargo, una alternativa interesante sería cuando se obtiene un índice construido a través de la comparación entre lo observado y lo esperado o referencia. En estadística, por ejemplo, está la prueba X^2 (o Chi-cuadrada) que se utiliza para determinar si existe una diferencia estadísticamente significativa entre la frecuencia esperada y las frecuencias observadas en una o más categorías de una tabla de contingencia (Corder y Foreman,

2011). En Deeks, Higgins, Altman y C.S.M. Group (2019) se presentan diferentes pruebas estadísticas para medir la heterogeneidad estadística.

Por otro lado, un índice de desigualdad es una medida cuantitativa que calibra la divergencia económica de la distribución de la riqueza de una sociedad de la perfecta equidad. En la literatura se han propuesto muchos índices (Eliazar, 2018), por ejemplo, el índice Pietra, el cual es utilizado para medir las disparidades de ingresos, la desigualdad en las cotizaciones, la volatilidad de los activos financieros o el diagnóstico médico (Sarabia y Jordá, 2014). Otro índice muy importante es el índice de Gini, el cual mide cómo se distribuye el ingreso en una sociedad (Medina, 2001). Este índice no combina diferentes dimensiones, es decir, no considera otros aspectos del desarrollo humano, como la educación, la salud o el nivel de vida, como sí lo hace el IDH. Este último, tiene un enfoque más amplio y abarca múltiples dimensiones del desarrollo humano, mientras que el índice de Gini se utiliza principalmente en el ámbito económico y social para analizar la desigualdad de ingresos.

Además de los índices mencionados anteriormente, existen muchos otros índices que son utilizados en distintos campos y situaciones. Por ejemplo, la referencia de Esparza, Lazcano, Ponce y Cortés (2020) presenta una amplia variedad de índices adicionales, como el índice de Atkinson y el índice de Robin Hood, proporcionando una descripción detallada, así como las ventajas y desventajas de cada uno. Esta referencia es una excelente fuente para explorar más a fondo la variedad de índices disponibles y comprender mejor sus aplicaciones en el análisis económico.

La propuesta de nuevos índices de desigualdad y heterogeneidad es importante para identificar brechas y disparidades en la sociedad, ya que evalúan el impacto de políticas y programas sociales, además de promover la equidad y la justicia social.

Existen principalmente tres tipos de heterogeneidad (Deeks *et al.*, 2019): clínica (diferencia en los participantes, intervenciones o los resultados), metodológica (diferencias en los diseños de estudio, sesgo) y estadística (variación en los efectos o resultados de la intervención). Estudiar estas diferencias es importante hoy en día, pues se llega a una mayor comprensión de los factores que influyen en la intervención.

En particular, los índices de heterogeneidad nos brindan información de cuánto un sistema se aleja de una idoneidad perfecta (Eliazar, 2018). Evaluar el rendimiento de dicho sistema se beneficiaría al contrastar los

niveles de heterogeneidad en momentos específicos. En Olatunji, Love, Shim, Ferrua y East (2017) se presenta un índice de heterogeneidad, que incluye métodos para modelar, visualizar y cuantificar la heterogeneidad, describiendo, por ejemplo, las distribuciones de temperatura y humedad, innovando en el área de la ingeniería. Nunes *et al.* (2020), por su parte, analizaron enfoques psiquiátricos para cuantificar la heterogeneidad, incluidos, entre otros, estudios de heterogeneidad de síntomas, biodiversidad de microbiomas, conteo de grupos y análisis de series temporales.

En la siguiente sección se propondrá un índice de heterogeneidad cuyo objetivo principal será el de medir y analizar la desigualdad en diferentes aspectos de una sociedad. Este índice permitirá cuantificar la distribución de recursos, ingresos o bienes entre los individuos de una población y además nos proporcionará información valiosa para comprender la disparidad existente y sus implicaciones.

2. Metodología

En esta sección, definimos un índice de heterogeneidad para vectores en $\mathbb{R}_{\geq 0}^2$ con una interpretación geométrica y extendemos su definición a $\mathbb{R}_{\geq 0}^n$. Para estos índices se muestran relaciones con la norma euclidiana y el índice de Gini, además, se prueban propiedades deseables, que finalmente ilustramos con algunos ejemplos numéricos.

Partimos de la idea que los índices de desigualdad son aplicables en el contexto de distribuciones generales de tamaños: cantidades no negativas como conteo, longitud, área, volumen, masa, energía y duración (Eliazar, 2018). En la vida real, las variables consideradas para construir índices tienen diferente magnitud. Dado que requerimos construir índices entre 0 y 1, lo primero que haremos será considerar vectores estandarizados con entradas no negativas.

Definición 2.1 Sea \mathbf{u} un vector en $\mathbb{R}_{\geq 0}^n$, decimos que \mathbf{u} está estandarizado si $\sum_{i=1}^n u_i = 1$.

Consideremos que la dimensión del vector dado representa el número de variables que lo componen, por ejemplo, para $n = 2$, se tendrían dos variables.

Definición 2.2 Sean $\mathbf{u}, \mathbf{v} \in \mathbb{R}_{\geq 0}^2$, vectores estandarizados, definimos el índice de heterogeneidad A de $\mathbf{u} = (u_1, u_2)$ y con respecto $\mathbf{v} = (v_1, v_2)$ como

$$A(\mathbf{u}, \mathbf{v}) := \begin{cases} \min\{u_1, v_1\}|u_2 - v_2| + |u_1 - v_1||u_2 - v_2| \\ \quad + \min\{u_2, v_2\}|u_1 - v_1| & \text{si } \mathbf{u}, \mathbf{v} > \mathbf{0} \\ 0 & \text{en otro caso} \end{cases} \quad (1)$$

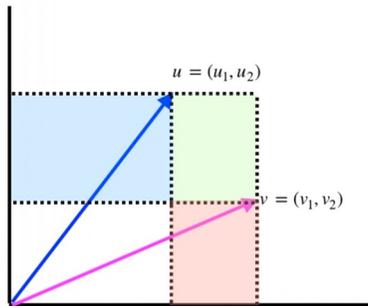
donde $\mathbf{u}, \mathbf{v} > \mathbf{0}$ significa que las entradas de los vectores son estrictamente positivas y $\mathbf{0}$ denota el vector con todas las entradas ceros.

De acuerdo a los valores de las entradas de $\mathbf{u} = (u_1, u_2)$ y $\mathbf{v} = (v_1, v_2)$ pueden existir ocho casos:

Caso 1	$u_1 < u_2,$	$u_1 < v_1,$	$u_2 < v_2,$
Caso 2	$u_1 < u_2,$	$u_1 < v_1,$	$u_2 > v_2,$
Caso 3	$u_1 < u_2,$	$u_1 > v_1,$	$u_2 < v_2,$
Caso 4	$u_1 < u_2,$	$u_1 > v_1,$	$u_2 > v_2,$
Caso 5	$u_1 > u_2,$	$u_1 < v_1,$	$u_2 < v_2,$
Caso 6	$u_1 > u_2,$	$u_1 < v_1,$	$u_2 > v_2,$
Caso 7	$u_1 > u_2,$	$u_1 > v_1,$	$u_2 < v_2,$
Caso 8	$u_1 > u_2,$	$u_1 > v_1,$	$u_2 > v_2.$

Gráficamente, el índice $A(\mathbf{u}, \mathbf{v})$ considerando el caso 2, sería la suma de las tres áreas coloreadas de la figura 1. De manera análoga se verían gráficamente los otros casos.

Figura 1
Esquema de las áreas para el cálculo del índice A



La definición se inspira en el índice de Gini, el cual se interpreta como un cociente de áreas, pero la motivación geométrica de esta definición surge al observar que si $(u_1, u_2) \rightarrow (v_1, v_2)$, entonces $A(\mathbf{u}, \mathbf{v}) \rightarrow 0$. Note, además, que al ser \mathbf{u} y \mathbf{v} vectores estandarizados no pueden ser el vector cero, y que el índice A toma el valor de 1 cuando $\mathbf{u} = (0,1)$ y $\mathbf{v} = (1,0)$, o viceversa. Luego, A toma valores entre 0 y 1.

Podemos extender la definición 2.2 al caso donde los vectores \mathbf{u} ó \mathbf{v} no estén estandarizados, definiendo al índice A de la siguiente manera:

$$A(\mathbf{u}, \mathbf{v}) := A\left(\frac{\mathbf{u}}{\sum_{i=1}^n u_i}, \frac{\mathbf{v}}{\sum_{i=1}^n v_i}\right)$$

Por ejemplo, si $\mathbf{u} = (a, a)$ para $a \in \mathbb{R}_{>0}$ y $\mathbf{v} = (1,1)$, entonces

$$A(\mathbf{u}, \mathbf{v}) = A\left(\frac{(a,a)}{2a}, \frac{(1,1)}{2}\right) = 0$$

Ahora bien, puesto que el índice A se asemeja a una métrica, nos preguntamos si existe una relación con la distancia euclidiana, que es la métrica más usada al comparar vectores. Recordemos que esta distancia está definida por:

$$E(\mathbf{u}, \mathbf{v}) = (\sum_{i=1}^n (u_i - v_i)^2)^{1/2}, \text{ para } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n.$$

La respuesta a esta pregunta es afirmativa y está dada en el siguiente teorema.

Teorema 2.1 *Para vectores estandarizados $\mathbf{u}, \mathbf{v} \in \mathbb{R}_{\geq 0}^2$ cualesquiera, se cumple que*

$$A(\mathbf{u}, \mathbf{v}) = \frac{E(\mathbf{u}, \mathbf{v})}{\sqrt{2}} = |u_1 v_2 - u_2 v_1|.$$

Más aún, como $u_1 + u_2 = 1$ y $v_1 + v_2 = 1$, entonces tenemos que

$$A(\mathbf{u}, \mathbf{v}) = \frac{E(\mathbf{u}, \mathbf{v})}{\sqrt{2}} = |u_1 - v_1|.$$

Demostración: supongamos que estamos en el caso 2, i.e., $u_1 < u_2$, $u_1 < v_1$ y $u_2 > v_2$ (los otros casos se analizan de manera análoga), entonces

$$\begin{aligned} A(\mathbf{u}, \mathbf{v}) &= u_1 |u_2 - v_2| + |u_1 - v_1| |u_2 - v_2| + v_2 |u_1 - v_1| \\ &= u_1 (u_2 - v_2) + (v_1 - u_1) (u_2 - v_2) + v_2 (v_1 - u_1) \\ &= |u_1 v_2 - u_2 v_1|. \end{aligned}$$

Y por otro lado,

$$\frac{E(\mathbf{u}, \mathbf{v})}{\sqrt{2}} = \left(\frac{1}{2} \sum_{i=1}^n (u_i - v_i)^2 \right)^{1/2} =^{p.d.} |u_1 v_2 - u_2 v_1|,$$

luego

$$\left[\frac{1}{2} ((u_1 - v_1)^2 + (u_2 - v_2)^2) \right] =^{p.d.} (u_1 v_2 - u_2 v_1)^2$$

$$(u_1 - v_1)^2 + (u_2 - v_2)^2 =^{p.d.} 2(u_1 v_2 - u_2 v_1)^2.$$

Como $u_1 + u_2 = 1$ y $v_1 + v_2 = 1$, entonces

$$(u_1 - v_1)^2 + ((1 - u_1) - (1 - v_1))^2 = 2(u_1(1 - v_1) - (1 - u_1)v_1)^2$$

$$(u_1 - v_1)^2 + (v_1 - u_1)^2 = 2(u_1 - u_1 v_1 - v_1 + u_1 v_1)^2$$

$$2(u_1 - v_1)^2 = 2(u_1 - v_1)^2.$$

Finalmente, como $u_1 + u_2 = 1$ y $v_1 + v_2 = 1$, entonces

$$A(\mathbf{u}, \mathbf{v}) = \frac{E(\mathbf{u}, \mathbf{v})}{\sqrt{2}} = |u_1(1 - v_1) - (1 - u_1)v_1| = |u_1 - v_1|.$$

■

Ahora veremos la relación que tiene el índice propuesto A con el índice de Gini, definido para un vector $\mathbf{x} \in \mathbb{R}_{\geq 0}^n$ por:

$$\text{Gini}(\mathbf{x}) = \frac{1}{n} \left[n + 1 - 2 \frac{\sum_{i=1}^n (n+1-i)x_i^*}{\sum_{i=1}^n x_i^*} \right],$$

donde \mathbf{x}^* es el vector ordenado (de menor a mayor) de \mathbf{x} (Plata-Pérez *et al.*, 2015).

Teorema 2.2 Sea $\mathbf{u} \in \mathbb{R}_{\geq 0}^2$ un vector estandarizado cualquiera y $\mathbf{v} = \left(\frac{1}{2}, \frac{1}{2} \right)$, entonces:

$$A(\mathbf{u}, \mathbf{v}) = \frac{E(\mathbf{u}, \mathbf{v})}{\sqrt{2}} = \text{Gini}(\mathbf{u}) = \frac{1}{2} |u_1 - u_2|.$$

Como $u_1 + u_2 = 1$, entonces

$$A(\mathbf{u}, \mathbf{v}) = \frac{E(\mathbf{u}, \mathbf{v})}{\sqrt{2}} = \text{Gini}(\mathbf{u}) = \left| u_1 - \frac{1}{2} \right|.$$

Demostración: sólo resta ver que $\text{Gini}(\mathbf{u}) = \frac{1}{2}|u_1 - u_2|$, pues las otras igualdades se obtienen directamente del teorema 2.1.

Supongamos, sin pérdida de generalidad, que $u_1 < u_2$, i.e., $u_2 > \frac{1}{2}$, luego

$$\begin{aligned} \text{Gini}(\mathbf{u}) &= \frac{1}{2} [2 + 1((2 + 2 - 1)u_1 + (2 + 1 - 2)u_2)] \\ &= \frac{1}{2} [3 - 2(2u_1 + u_2)] \\ &= \frac{1}{2} [2 + (u_1 + u_2) - 2(u_1 + (u_1 + u_2))] \\ &= \frac{1}{2} [2 + u_1 + u_2 - 2u_1 - 2] \\ &= \frac{1}{2} |u_1 - u_2|. \end{aligned}$$

Como $u_1 + u_2 = 1$, entonces

$$A(\mathbf{u}, \mathbf{v}) = \frac{E(\mathbf{u}, \mathbf{v})}{\sqrt{2}} = \text{Gini}(\mathbf{u}) = \frac{1}{2} |u_1 - (1 - u_1)| = \left| u_1 - \frac{1}{2} \right|.$$

■

Así pues, en $\mathbb{R}_{\geq 0}^2$ hemos visto que el índice A coincide con el índice de Gini, cuando $\mathbf{v} = \left(\frac{1}{2}, \frac{1}{2}\right)$ y es proporcional a la distancia euclidiana. Ahora extenderemos la definición de este índice A en $\mathbb{R}_{\geq 0}^n$, para $n > 2$.

Definición 2.3 Sean $\mathbf{u}, \mathbf{v} \in \mathbb{R}_{\geq 0}^n$ para $n > 2$, y $c = \binom{n}{2}$ el número de combinaciones de n en 2. Denotemos a \mathbf{C} , una matriz cuyos renglones están dados por dichas combinaciones, i.e., $\mathbf{C} \in \mathcal{M}_{n \times 2}$. El índice de heterogeneidad A_n de \mathbf{u} con respecto a \mathbf{v} en $\mathbb{R}_{\geq 0}^n$ lo definimos de la siguiente manera:

$$A_n(\mathbf{u}, \mathbf{v}) := \frac{1}{c} \sum_{i=1}^c A(\mathbf{u}_i, \mathbf{v}_i), \tag{2}$$

donde $\mathbf{u}_i := \frac{(u_{c_{i1}}, u_{c_{i2}})}{u_{c_{i1}} + u_{c_{i2}}}$ y $\mathbf{v}_i := \frac{(v_{c_{i1}}, v_{c_{i2}})}{v_{c_{i1}} + v_{c_{i2}}}$, para $i = 1, \dots, c$, ambos vectores estandarizados y $A(\cdot, \cdot)$ está definido en la ecuación (1).

Observe que este nuevo índice es el promedio de los índices de los vectores en $\mathbb{R}_{\geq 0}^2$ formados por las combinaciones de las entradas de \mathbf{u} y \mathbf{v} .

Ejemplo 2.1 Supongamos $n = 4$, es decir, se tienen los vectores $\mathbf{u} =$

$$(u_1, u_2, u_3, u_4) \text{ y } \mathbf{v} = (v_1, v_2, v_3, v_4). \text{ Entonces } c = \binom{4}{2} = 6, \text{ y } \mathbf{C} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 2 & 3 \\ 2 & 4 \\ 3 & 4 \end{pmatrix},$$

luego $A_n = \frac{1}{6} [A(\mathbf{u}_1, \mathbf{v}_1) + \dots + A(\mathbf{u}_6, \mathbf{v}_6)]$, donde $\mathbf{u}_1 = \frac{(u_1, u_2)}{u_1 + u_2}$; $\mathbf{u}_2 = \frac{(u_1, u_3)}{u_1 + u_3}$; $\mathbf{u}_3 = \frac{(u_1, u_4)}{u_1 + u_4}$; $\mathbf{u}_4 = \frac{(u_2, u_3)}{u_2 + u_3}$; $\mathbf{u}_5 = \frac{(u_2, u_4)}{u_2 + u_4}$; $\mathbf{u}_6 = \frac{(u_3, u_4)}{u_3 + u_4}$; análogamente para $\mathbf{v}_i, i = 1, \dots, 6$.

Como $0 \leq A(\cdot, \cdot) \leq 1$, entonces $0 \leq A_n(\cdot, \cdot) \leq 1$, es decir, hemos definido un índice de heterogeneidad en $\mathbb{R}_{\geq 0}^n$ entre 0 y 1.

Observación 2.1 La norma euclidiana puede interpretarse como un índice de heterogeneidad al medir la distancia del vector \mathbf{u} con respecto a vector esperado \mathbf{v} .

A continuación, se presentan algunos resultados importantes de este índice con relación al índice de Gini.

2.1 El índice de heterogeneidad y el índice de Gini

El resultado que a continuación se presenta, se relaciona con la propiedad de estandarización en el sentido del índice de Gini.

Teorema 2.3 Supongamos que tenemos un vector $\mathbf{y} \in \mathbb{R}_{\geq 0}^n$, estandarizado, cuyas entradas son sólo de dos tipos: m componentes con entradas $a \in \mathbb{R}_{\geq 0}$ y $m - n$ componentes con entradas $b \in \mathbb{R}_{\geq 0}$ (sin importar el orden). Sin pérdida de generalidad, podemos reescribir este vector \mathbf{y} como sigue

$$\mathbf{y} = \underbrace{(a, \dots, a)}_{a \in \mathbb{R}_{\geq 0}^m}, \underbrace{(b, \dots, b)}_{a \in \mathbb{R}_{\geq 0}^{m-n}}.$$

Tomando al vector esperado como $\mathbf{e} = (1, \dots, 1) \in \mathbb{R}^n$, entonces

$$A_n(\mathbf{y}, \mathbf{e}) = \frac{m(n-m)}{n} |b - a| \frac{1}{n-1} \frac{1}{a+b} \quad (3)$$

$$\text{Gini}(\mathbf{y}) = \frac{m(n-m)}{n} |b - a|. \quad (4)$$

En dado caso de que $\frac{1}{n-1} \frac{1}{a+b} = 1$ se tendría la igualdad entre los índices A_n y Gini .

Dem: Sin pérdida de generalidad, supongamos que $a < b$, como $m \cdot a + (n - m) \cdot b = 1$, entonces

$$a < b = \frac{1-m \cdot a}{n-m}$$

$$(n - m) \cdot a < 1 - m \cdot a$$

$$a < 1/n.$$

Sea $c = \binom{n}{2} = \frac{n!}{(n-2)!2!}$, como $A((a, a), (1, 1)) = 0$, entonces

$$\begin{aligned} A_n(\mathbf{y}, \mathbf{e}) &= \frac{m(n-m)}{c} A((a, b), (1, 1)) \\ &= \frac{m(n-m)}{c} \frac{1}{2} \frac{|b-a|}{a+b} \\ &= \frac{m(n-m)}{n} |b - a| \frac{1}{n-1} \frac{1}{a+b}. \end{aligned}$$

Por otro lado,

$$\begin{aligned} \text{Gini}(\mathbf{y}) &= \frac{1}{n} [n + 1 - 2(\sum_{i=1}^m (n + 1 - i)a + \sum_{j=m+1}^n (n + 1 - j)b)] \\ &= \frac{1}{n} [n + 1 - 2((n + 1)am - \sum_{i=1}^m ia + (n + 1)b(n - m) - \\ &\quad b \sum_{i=1}^{n-m} (i + m))] \\ &= \frac{1}{n} [-(n + 1) + am(m + 1) + b(n - m)((n - m) + 1) + \\ &\quad 2bm(n - m)] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} [-n + am^2 + b(n-m)^2 + 2bm(n-m)] \\
&= \frac{1}{n} [-n + am^2 + b(n-m)(2m + (n-m))] \\
&= \frac{1}{n} [-n + am \cdot m + b(n-m)(n+m)] \\
&= \frac{1}{n} [-n + am \cdot m + bm(n-m) + nb(n-m)] \\
&= \frac{1}{n} [-n + m(am + b(n-m)) + nb(n-m)] \\
&= \frac{1}{n} [-n + m + nb(n-m)] \\
&= \frac{1}{n} [(n-m)(nb-1)] \\
&= \frac{1}{n} [(n-m)(m(b-a))] = \frac{m}{n} (n-m)(b-a).
\end{aligned}$$

■

De la ecuación (3), si $a = 0$ entonces

$$\mathbf{y} = (\underbrace{0, \dots, 0}_{m \text{ veces}}, \underbrace{b, \dots, b}_{(n-m) \text{ veces}}), \quad (5)$$

tal que, para tener un vector estandarizado se requiere que $(n-m)b = 1$, luego

$$A_n(\mathbf{y}, \mathbf{e}) = \frac{m(n-m)}{n(n-1)}$$

Si por ejemplo

$$\mathbf{y}' = (\underbrace{0, \dots, 0}_{(n-m) \text{ veces}}, \underbrace{b', \dots, b'}_{m \text{ veces}}), \quad (6)$$

tal que $mb' = 1$, luego

$$A_n(\mathbf{y}', \mathbf{e}) = \frac{(n-m)(n-(n-m))}{n(n-1)} = \frac{m(n-m)}{n(n-1)} = A_n(\mathbf{y}, \mathbf{e}).$$

Mientras que $\text{Gini}(\mathbf{y}) = \frac{m}{n}$ y $\text{Gini}(\mathbf{y}') = \frac{n-m}{n}$.

Ejemplo 2.2 Supongamos que $n = 50$. Para construir los vectores \mathbf{y} y \mathbf{y}' de las ecuaciones (5) y (6), respectivamente, definimos $b = \frac{1}{n-m}$ y $b' = \frac{1}{m}$, con m tomando distintos valores. Se calcularon los índices de Gini y A_n ; los resultados se presentan en la tabla 1.

Tabla 1
Comparación entre el índice de Gini y el índice A_n para vectores de dos grupos de datos, fijando el tamaño $n = 50$

m	Gini(\mathbf{y})	Gini(\mathbf{y}')	$A_n(\mathbf{y}', \mathbf{e}) = A_n(\mathbf{y}, \mathbf{e})$
5	0.1	0.9	0.09183673
10	0.2	0.8	0.16326531
15	0.3	0.7	0.21428571
20	0.4	0.6	0.24489796
25	0.5	0.5	0.25510204

Note que el índice de Gini siempre tiene en cuenta el número de ceros en el vector, a diferencia del índice A_n , que no lo toma en consideración.

Ejemplo 2.3 Supongamos $n = 4$, en la tabla 2 se presentan diferentes vectores no estandarizados \mathbf{y} y el cálculo de su índice de Gini y el $A_n(\mathbf{y}, \mathbf{e})$.

Tabla 2
Comparación del índice de Gini y el índice A_n considerando diferentes tipos de vectores

\mathbf{y}	Gini(\mathbf{y})	$A_n(\mathbf{y}, \mathbf{e})$
(0,1,1,1)	0.2500000	0.2500000
(1,0,0,0)	0.7500000	0.2500000
(1,1,1,100)	0.7208738	0.2450495
(1,1,100,100)	0.4900990	0.3267327
(1,100,100,100)	0.2466777	0.2450495
(1,1,1,1000)	0.7470090	0.2495005
(1,1,1000,1000)	0.4990010	0.3326673
(1,1000,1000,1000)	0.2496668	0.2495005

Notemos en la tabla 2 que no importando el valor 100 ó 1000 en el vector \mathbf{y} el índice A_n les da aproximadamente el mismo valor a los vectores presentados. Sin embargo, el índice de Gini toma en cuenta cuántos tuvieron el valor de 1 y cuántos el valor 100 ó 1000. Si suponemos, por ejemplo, que estos vectores representan el ingreso que tienen 4 personas, el índice de desigualdad de Gini toma en cuenta el número de personas que tienen poco ingreso, pero no así el índice A_n . Así pues, hemos propuesto un índice de heterogeneidad, que a su vez está midiendo desigualdad cualitativa.

2.2 Propiedades

A continuación, se presentan algunas propiedades que satisface el índice de heterogeneidad propuesto. La mayoría de estas propiedades ya han sido probadas, y merece especial atención la propiedad de separabilidad comonótona, que enunciaremos como un teorema.

Sean $\mathbf{u}, \mathbf{v} \in \mathbb{R}_{\geq 0}^n$.

- **Independencia de escala**

Para $\lambda > 0$ y $\beta > 0$, se cumple que

$$A_n(\lambda \mathbf{u}, \beta \mathbf{v}) = A_n(\mathbf{u}, \mathbf{v})$$

pues por definición se tiene que

$$A_n(\lambda \mathbf{u}, \beta \mathbf{v}) = \frac{1}{c} \sum_{i=1}^c A(\hat{\mathbf{u}}_i, \hat{\mathbf{v}}_i)$$

donde

$$\hat{\mathbf{u}}_i = \frac{(\lambda u_{c_{i1}}, \lambda u_{c_{i2}})}{\lambda u_{c_{i1}} + \lambda u_{c_{i2}}} = \frac{(u_{c_{i1}}, u_{c_{i2}})}{u_{c_{i1}} + u_{c_{i2}}} = \mathbf{u}_i$$

y

$$\hat{\mathbf{v}}_i = \frac{(\beta v_{c_{i1}}, \beta v_{c_{i2}})}{\beta v_{c_{i1}} + \beta v_{c_{i2}}} = \frac{(v_{c_{i1}}, v_{c_{i2}})}{v_{c_{i1}} + v_{c_{i2}}} = \mathbf{v}_i$$

Si en particular $\mathbf{u} = (a, a, \dots, a) \in \mathbb{R}_{> 0}^n$ y $\mathbf{v} = \mathbf{e}$, entonces

$$A_n(\mathbf{u}, \mathbf{v}) = 0$$

pues

$$A_n(\mathbf{u}, \mathbf{v}) = A_n(a\mathbf{e}, \mathbf{e}) = A_n(\mathbf{e}, \mathbf{e}) = 0.$$

- **Simetría**

Sea S_n el conjunto de permutaciones en n . Para cada $\theta \in S_n$ se cumple que

$$A_n(\mathbf{u}, \mathbf{v}) = A_n(\theta(\mathbf{u}), \theta(\mathbf{v})).$$

La demostración de este resultado se sigue directamente de la definición del índice A que se presenta en la ecuación (1), pues si \mathbf{u} y \mathbf{v} son vectores estandarizados, se cumple que $A(\mathbf{u}, \mathbf{v}) = A(\mathbf{v}, \mathbf{u})$.

- **Estandarización** (en el sentido de Gini (Plata-Pérez *et. al*, 2015)).

Se tiene que

$$A_n \left(\left(0, \frac{1}{n-m} \right), \mathbf{e} \right) = \frac{m}{n}.$$

Notemos que el índice Gini de dos grupos dada en la ecuación (4) y el índice A_n presentado en la ecuación (3) son iguales si y sólo si

$$\frac{1}{n-1} \frac{1}{a+b} = 1,$$

tal que $ma + (n - m)b = 1$. Así pues, la única solución para $a = 0$ es

$$m = 1, b = \frac{1}{n-1}.$$

Luego, si $\mathbf{y} = \left(0, \frac{1}{n-1}, \dots, \frac{1}{n-1} \right)$, de la ecuación (3) tenemos que:

$$A_n(\mathbf{y}, \mathbf{e}) = \frac{1}{n} = Gini(\mathbf{y}).$$

- **Separabilidad comonótona** (en el sentido de Gini (Plata-Pérez *et. al*, 2015)). Esta propiedad la enunciamos como un teorema.

Teorema 2.4 Sea $\beta \in (0,1)$, y $\mathbf{x}, \mathbf{y} \in \mathbb{R}_{\geq 0}^n$, dos vectores estandarizados, entonces, el índice A_n cumple con la propiedad de separabilidad comonótona si

$$A_n(\beta\mathbf{x} + (1 - \beta)\mathbf{z}, \mathbf{e}) = \beta A_n(\mathbf{x}, \mathbf{e}) + (1 - \beta)A_n(\mathbf{z}, \mathbf{e}) \tag{7}$$

tales que se cumplen las siguientes condiciones: $\sum_{i=1}^n x_i = \sum_{i=1}^n z_i$, en particular, $x_i + x_j = z_i + z_j$ para $i \neq j$ y además $(x_i - x_j)(z_i - z_j) \geq 0$, $i \neq j$.

Demostración: tenemos que, la parte izquierda de la ecuación (7) está dada por

$$\begin{aligned}
\beta A_n(\mathbf{x}, \mathbf{e}) + (1 - \beta)A_n(\mathbf{z}, \mathbf{e}) &= \frac{\beta}{c} \sum_{i=1}^c A(\mathbf{x}_i, \mathbf{e}_i) + \frac{1-\beta}{c} \sum_{i=1}^c A(\mathbf{z}_i, \mathbf{e}_i), \\
&= \frac{1}{c} \sum_{i=1}^c \beta A(\mathbf{x}_i, \mathbf{e}_i) + \frac{1}{c} \sum_{i=1}^c (1 - \beta)A(\mathbf{z}_i, \mathbf{e}_i), \\
&= \frac{1}{c} \sum_{i=1}^c [\beta A(\mathbf{x}_i, \mathbf{e}_i) + (1 - \beta)A(\mathbf{z}_i, \mathbf{e}_i)], \\
&= \frac{1}{2c} \sum_{i=1}^c \left[\beta \frac{|x_{c_{i1}} - x_{c_{i2}}|}{x_{c_{i1}} + x_{c_{i2}}} + (1 - \beta) \frac{|z_{c_{i1}} - z_{c_{i2}}|}{z_{c_{i1}} + z_{c_{i2}}} \right],
\end{aligned}$$

donde $\mathbf{x}_i = \frac{(x_{c_{i1}}, x_{c_{i2}})}{x_{c_{i1}} + x_{c_{i2}}}$, $\mathbf{z}_i = \frac{(z_{c_{i1}}, z_{c_{i2}})}{z_{c_{i1}} + z_{c_{i2}}}$ y $\mathbf{e}_i = \left(\frac{1}{2}, \frac{1}{2}\right)$, para $i = 1, \dots, c$. En la última igualdad se utilizó el teorema 2.2.

Por otro lado, la parte derecha de (7) está dada por

$$A_n(\beta \mathbf{x} + (1 - \beta)\mathbf{z}, \mathbf{e}) = \frac{1}{c} \sum_{i=1}^c A(\beta \mathbf{x}_i + (1 - \beta)\mathbf{z}_i, \mathbf{e}_i)$$

donde

$$\begin{aligned}
\beta \mathbf{x}_i + (1 - \beta)\mathbf{z}_i &= \frac{(\beta x_{c_{i1}} + (1-\beta)z_{c_{i1}}, \beta x_{c_{i2}} + (1-\beta)z_{c_{i2}})}{[\beta x_{c_{i1}} + (1-\beta)z_{c_{i1}} + \beta x_{c_{i2}} + (1-\beta)z_{c_{i2}}]} \\
&= \frac{(\beta x_{c_{i1}} + (1-\beta)z_{c_{i1}}, \beta x_{c_{i2}} + (1-\beta)z_{c_{i2}})}{[\beta(x_{c_{i1}} + x_{c_{i2}}) + (1-\beta)(z_{c_{i1}} + z_{c_{i2}})]}.
\end{aligned}$$

Luego,

$$\begin{aligned}
A_n(\beta \mathbf{x} + (1 - \beta)\mathbf{z}, \mathbf{e}) &= \frac{1}{2c} \sum_{i=1}^c \frac{[\beta x_{c_{i1}} + (1-\beta)z_{c_{i1}}] - [\beta x_{c_{i2}} + (1-\beta)z_{c_{i2}}]}{[\beta(x_{c_{i1}} + x_{c_{i2}}) + (1-\beta)(z_{c_{i1}} + z_{c_{i2}})]} \\
&= \frac{1}{2c} \sum_{i=1}^c \frac{|\beta(x_{c_{i1}} - x_{c_{i2}}) + (1-\beta)(z_{c_{i1}} - z_{c_{i2}})|}{[\beta(x_{c_{i1}} + x_{c_{i2}}) + (1-\beta)(z_{c_{i1}} + z_{c_{i2}})]} \\
&= \frac{1}{2c} \sum_{i=1}^c \frac{\beta|x_{c_{i1}} - x_{c_{i2}}| + (1-\beta)|z_{c_{i1}} - z_{c_{i2}}|}{[\beta(x_{c_{i1}} + x_{c_{i2}}) + (1-\beta)(z_{c_{i1}} + z_{c_{i2}})]}
\end{aligned}$$

esto último pues si $x_{c_{i1}} \geq x_{c_{i2}}$ entonces $z_{c_{i1}} \geq z_{c_{i2}}$. Como además $x_{c_{i1}} + x_{c_{i2}} = x_{c_{i1}} + x_{c_{i2}}$ entonces

$$A_n(\beta \mathbf{x} + (1 - \beta)\mathbf{z}, \mathbf{e}) = \frac{1}{2c} \sum_{i=1}^c \left(\beta \frac{|x_{ci1} - x_{ci2}|}{x_{ci1} + x_{ci2}} + (1 - \beta) \frac{|z_{ci1} - z_{ci2}|}{z_{ci1} + z_{ci2}} \right)$$

$$= \beta A_n(\mathbf{x}, \mathbf{e}) + (1 - \beta) A_n(\mathbf{z}, \mathbf{e}).$$

■

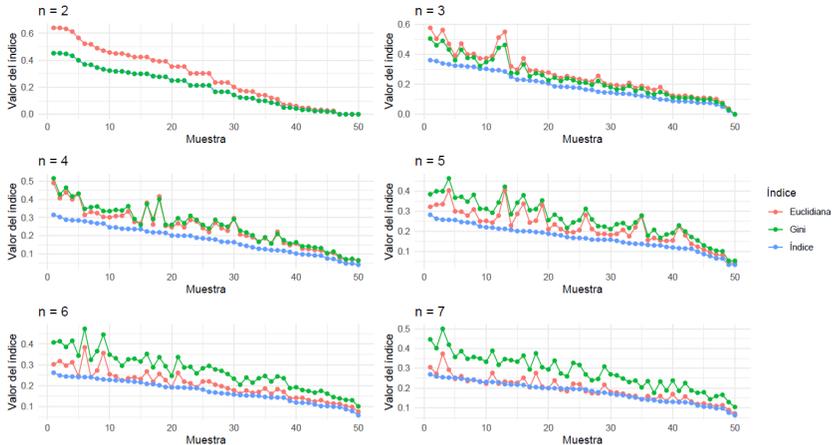
En el siguiente ejemplo mostramos cómo se comporta el índice de heterogeneidad propuesto, visto como un índice de desigualdad, es decir, considerando el vector esperado como \mathbf{e} y vectores observados aleatorios.

Ejemplo 2.4 Con diferentes dimensiones de \mathbb{R}^n , $n = 2, \dots, 7$, generamos 50 vectores $\mathbf{u} \in \mathbb{R}_{\geq 0}^n$ cuyas entradas fueron valores aleatorios entre 1 y 20. Calculamos su índice $A_n(\mathbf{u}, \mathbf{e})$ (ecuación (2)), su distancia euclidiana $E(\mathbf{u}, \mathbf{e})$ y su índice de Gini (\mathbf{u}) . Los resultados se muestran en la figura 2.

Observemos que para $n = 2$, el índice A_n y el índice de Gini coinciden (teorema 2.2). A medida que el valor de n se incrementa, es decir, la dimensión de los vectores, se observa una mayor disparidad entre la distancia euclidiana y el índice de Gini, mientras que el índice A_n se asemeja cada vez más a la distancia euclidiana.

Figura 2

Comparación del índice de Gini, el índice A_n y la distancia euclidiana para diferentes vectores



3. Aplicaciones

En esta sección, presentamos aplicaciones reales para obtener el índice de heterogeneidad de distintos datos. Una aplicación aborda la clasificación de profesores en una universidad según su desempeño académico. La siguiente aplicación se centra en el IDH con Servicios (IDHs). Por último, se analiza la distribución del salario en México de 2005 a 2024.

3.1 Clasificación de los profesores

El Programa de Estímulo al Desempeño del Personal Docente (PEDPD) es un programa federal que tiene como propósito “reafirmar el trascendente papel de los educadores en el proceso de enseñanza y aprendizaje” (SEP, 2018) en las universidades públicas de México. Cada universidad o Institución de Educación Superior (IES), emite una convocatoria anual con base en un reglamento propio. Dicho reglamento sigue las reglas de operación emitidas por la Secretaría de Educación Pública (SEP).

La IES define, mediante un sistema de puntaje, los criterios para repartir el presupuesto asignado por la Secretaría de Hacienda y Crédito Público. Dicho estímulo se reparte en múltiplos de una Unidad de Medición y Actualización (UMA), es decir, al nivel 1 se le otorga 1 UMA, al nivel 2, 2 UMA, y así sucesivamente hasta el nivel 7.

El Reglamento del PEDPD (ver, por ejemplo, UAA (2008)) señala que deberán asignarse dos puntajes a cada profesor: el Puntaje Total (T) y el Puntaje de Calidad (C), donde el rubro de calidad está contenido en el puntaje total T . Este último se compone de tres grandes categorías: Calidad (C), Dedicación (De), y Permanencia (P), es decir,

$$T = C + De + P. \quad (8)$$

Por ejemplo, los requisitos de los profesores de la Universidad Autónoma de Aguascalientes para participar en este programa se encuentran en la convocatoria 2022-2023 (UAA, 2022). Al cumplir los requisitos de participación, el personal docente se evalúa en distintas categorías:

- Permanencia, P (máximo 100 puntos).
- Dedicación a la docencia, De (máximo 300 puntos).
- Calidad, C (máximo 600 puntos). Esta categoría es la más importante y se refiere a las actividades esenciales para desempeñarse adecuadamente como docente universitario: *Docencia, Do* (máximo 240 puntos), *Investigación, I* (máximo 160

puntos), *Tutoría*, Tu (máximo 100 puntos) y *Cuerpos Colegiados*, CC (máximo 100 puntos). Así, el rubro de calidad está dado por $C = Do + I + Tu + CC$.

Con base en las variables T y C , se asigna el nivel de estímulo que recibirá el profesor, de acuerdo con la tabla 3.

Tabla 3
Nivel asignado a cada profesor de acuerdo a calidad y total de puntos

Total (puntos)	Calidad (puntos)	Nivel
301-400	181-240	1
401-500	241-300	2
501-600	301-360	3
601-700	361-420	4
701-800	421-480	5
801-900	481-540	6
901-1000	541-600	7

Note que el puntaje total, T , asigna un nivel, y el puntaje de calidad, C , podría asignar otro distinto. Sin embargo, el nivel asignado a un profesor es el mínimo asignado por estos puntajes. Luego, podemos definir una función nivel que depende del nivel obtenido con el puntaje de Calidad y el nivel obtenido por el puntaje Total dado por:

$$N = \min\{N_T, N_C\}, \tag{9}$$

donde N_T es el nivel obtenido por el puntaje Total y N_C el nivel obtenido por el puntaje de Calidad.

Ejemplo 3.1

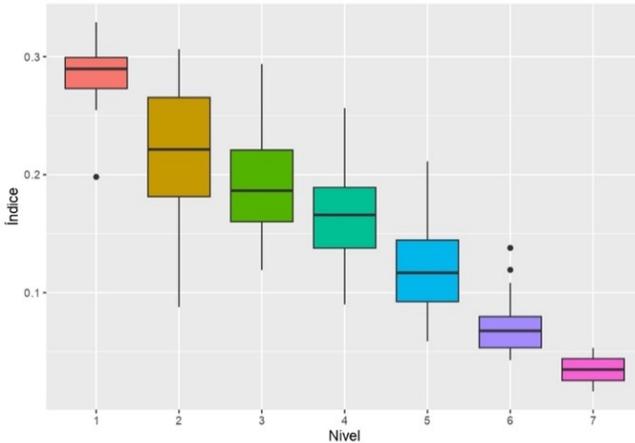
Sea $n = 6$, que hace referencia al número de variables a considerar: Do, I, Tu, CC, P y De ; y consideremos al vector esperado estandarizado dado por

$$v = (240/1000, 160/1000, 100/1000, 100/1000, 100/1000, 300/1000),$$

que se refiere a los pesos para las variables, respectivamente. En este vector se han puesto las proporciones de cada una de las variables: docencia, investigación, tutoría, cuerpos colegiados, permanencia y dedicación, con respecto al puntaje máximo que se puede obtener que son 1000 puntos, respectivamente.

Consideremos al vector \mathbf{u} (no estandarizado) como el puntaje de cada profesor en cada categoría. Supongamos que tenemos una muestra de 256 profesores. En la figura 3 se presenta un *boxplot* del índice A_n obtenido para los profesores utilizando la ecuación (2) así como el nivel real que la universidad le otorgó de acuerdo a la ecuación (9).

Figura 3
Boxplot del índice de heterogeneidad A_n y clasificación real de los profesores dado por su nivel



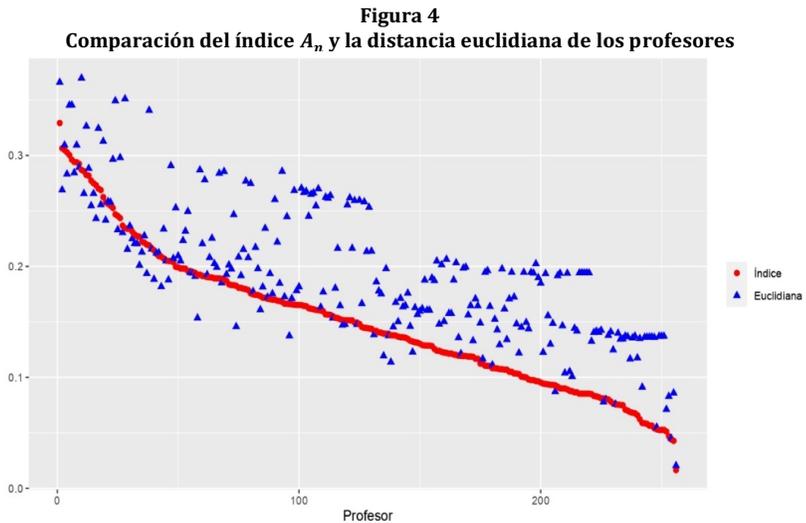
De manera general, en la figura 3 podemos observar que entre menor sea el valor del índice A_n su nivel es mayor, y si el valor de A_n es mayor, su nivel es menor. Se logra observar que algunos profesores obtienen valores parecidos del índice A_n pero su nivel varía significativamente. Por ejemplo, con un valor del índice A_n de 0.13 hay profesores en los niveles 2, 3, 4 y 5, sin embargo, son muy pocos casos donde sucede este fenómeno. De hecho, las únicas “cajas” que se traslapan son las de los niveles 2 y 3, 3 y 4; y 4 y 5. En la tabla 4 se presenta la media y desviación estándar del índice A_n por niveles.

Tabla 4
Media y desviación estándar del índice A_n de los profesores por niveles

Nivel	Media	Desviación estándar
1	0.28260	0.03076
2	0.21688	0.05765
3	0.19069	0.04569
4	0.16484	0.03774
5	0.11956	0.03410
6	0.07055	0.02317
7	0.03478	0.02599

Esto significa que, de acuerdo a la muestra considerada, un profesor en el nivel 7 obtuvo un índice A_n de aproximadamente 0.03478 ± 0.02599 . Este valor tan bajo indica que el vector observado del profesor es muy similar al vector esperado \mathbf{v} que requiere la universidad. Por lo tanto, a menor índice A_n , mayor nivel, ya que el vector observado se asemeja mucho al esperado.

En la figura 4 se presentan los resultados de esta aplicación considerando los índices A_n y la distancia euclidiana. En general, vemos que el índice A_n obtiene un valor menor que la distancia Euclidiana.



3.2 Índice de Desarrollo Humano con Servicios

Ahora vamos a calcular el índice de heterogeneidad utilizando como vector esperado \mathbf{e} (pesos iguales en todas las variables) y lo aplicaremos en desarrollo humano.

El Índice de Desarrollo Humano con Servicios (IDHs) fue por primera vez propuesto en Zamudio, Miranda y Tamez (2001) y Zamudio *et al.* (2017) presenta el cálculo de este índice para las Entidades Federativas de México, así como, para sus municipios. El IDHs se calcula como sigue:

$$\text{IDHs} = \frac{1}{3} \ln \text{Ser} + \frac{1}{3} \ln \text{Edu} + \frac{1}{3} \ln \text{Esp}, \quad (10)$$

donde *InSer* es a su vez el índice de servicios (que utiliza como datos el acceso al agua potable, al drenaje y a la electricidad), *InEdu* que es el índice de educación (que tiene como insumos a la matriculación infantil y a la alfabetización en adultos) y el *InEsp* que es el índice de esperanza de vida. Así, decimos pues que el IDHs tiene tres dimensiones: *InSer*, *InEdu* e *InEsp*, cada uno calculado de forma independiente y que a su vez toman valores entre 0 y 1 cada uno.

Tomando los datos de Zamudio *et al.* (2017) de las tres componentes del IDHs para el 2015 (vector observado), aplicamos el cálculo del A_n (ecuación (2) con $n = 3$) a las Entidades Federativas de México. El vector esperado fue $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ pues es el que se considera para el cálculo del IDHs (ver ecuación (10)).

En la figura 5 se presentan los valores que obtuvieron las Entidades Federativas en las tres dimensiones del IDHs y el índice A_n . De manera general, el índice de esperanza de vida es mucho menor al de los otros dos: servicios y educación. El *InSer* está mostrando valores muy altos (cercanos a 1), esto sugiere que, en general, las personas en México tuvieron un acceso adecuado a los servicios básicos hasta el año 2015.

Observamos que los Estados de Oaxaca, Guerrero, Chiapas, San Luis Potosí y Veracruz obtuvieron el menor valor del índice A_n , indicando que estos estados obtuvieron índices muy parecidos en *InEsp*, *InSer* y *InEdu*; y por otro lado tenemos que Aguascalientes, Colima, Distrito Federal (ahora Ciudad de México), Jalisco y Nuevo León obtuvieron un índice A_n mayor, indicando mayor heterogeneidad en las dimensiones del IDHs.

En la figura 6 se presentan el IDHs de las Entidades Federativas y su respectivo A_n , además se muestra la regresión lineal entre ambos. Se observa un comportamiento muy importante: entre mayor es el IDHs de los Estados, mayor es su índice A_n . Por otro lado, los Estados con mayor IDHs fueron: Aguascalientes, Nuevo León y Distrito Federal, mientras que los que obtuvieron menor IDHs fueron: Oaxaca, Chiapas y Guerrero.

3.3 El salario mínimo en México

El aumento en el salario mínimo (SM) ha cambiado profundamente la estructura de la distribución de las remuneraciones en México, duplicando en 20 años el porcentaje de personas que ganan hasta un salario mínimo y reduciendo a un mínimo histórico la proporción de ocupados que ganan más de tres salarios mínimos (INEGI, 2024).

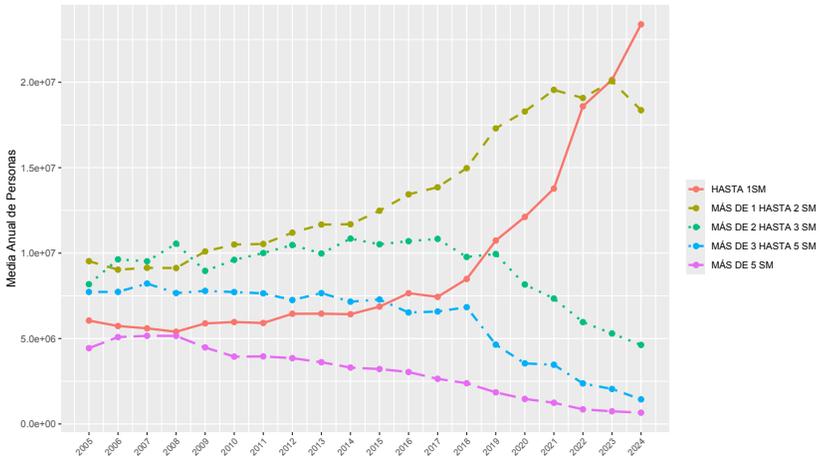
Esta composición tiene un impacto económico significativo por dos razones: primero, el aumento en el salario mínimo ha generado un crecimiento sin precedentes en los trabajos con menor remuneración, y segundo, esta información debe contrastarse con el aumento en el costo de vida. Los datos para este ejemplo –número de personas ocupadas– se obtuvieron de la Encuesta Nacional de Ocupación y Empleo (ENOE) del Instituto Nacional de Estadística y Geografía (INEGI), que se publica trimestralmente (excepto el segundo trimestre de 2020) y se clasifican de la siguiente manera:

- 1- Hasta un salario mínimo.
- 2- Más de 1 hasta 2 salarios mínimos.
- 3- Más de 2 hasta 3 salarios mínimos.
- 4- Más de 3 hasta 5 salarios mínimos.
- 5- Más de 5 salarios mínimos.

En la figura 7 se muestra el promedio anual (calculado a partir de los cuatro trimestres) del número de personas ocupadas en cada una de las categorías salariales previamente mencionadas. Se observa un incremento en el número de personas que ganan menos de un salario mínimo, así como un alto porcentaje de personas que ganan entre uno y dos salarios mínimos.

El número de personas pertenecientes a la clase media, representada por las categorías 3 y 4, se mantuvo constante entre 2005 y 2018; sin embargo, a partir de ese año ha disminuido. Este descenso se ha traducido en un aumento en el número de personas en la clase baja, correspondiente a las categorías 1 y 2, mientras que el número de personas en la clase alta, correspondiente a la categoría 5, también ha disminuido.

Figura 7
Media anual de personas ocupadas por categoría de salario mínimo en México

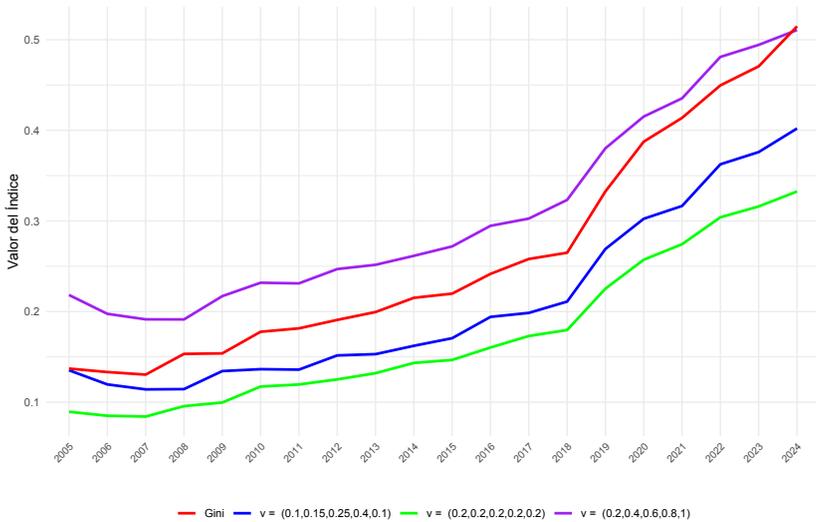


Ahora bien, utilizando los datos de las proporciones de las personas ocupadas en México en cada categoría de los salarios mínimos –vector observado–, se calculó el índice de Gini y el índice A_n con $n = 5$ y tres vectores esperados.

El vector $(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$ indica que se le dio el mismo peso a cada una de las categorías (es decir, se espera un porcentaje del 20% en todas las categorías); luego se consideró el vector $(0.10, 0.15, 0.25, 0.40, 0.10)$ que indica que el 10% de la población se espera gane hasta un salario mínimo al igual que más de cinco salarios mínimos, un 15% de uno a dos salarios mínimos, un 25% de dos a tres salarios mínimos y un 40% de tres a cinco salarios mínimos, así pues, se le está otorgando un mayor peso a la clase media.

También se consideró el vector esperado de $(\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{5}{5})$, el cual sugiere que se espera una mayor proporción de personas con un salario más alto. La figura 8 muestra los resultados.

Figura 8
Índice de heterogeneidad A_n e índice de Gini aplicado a la distribución de salario de las personas ocupadas en México



En la figura 8, se observa claramente un cambio abrupto en los índices desde 2005 hasta 2018, fecha de cambio de gobierno federal. Durante el período de 2005 a 2009, la distribución del número de personas en las cinco categorías del salario mínimo era relativamente homogénea (índice verde bajo). Si se esperaba un aumento en el porcentaje de personas con salarios mayores, esto resulta ser incorrecto (ver índice morado), ya que el índice de heterogeneidad muestra un incremento continuo, indicando un aumento en el porcentaje de personas en situación de pobreza (confirmado por la figura 7), a pesar del incremento del salario mínimo en los últimos años. Por otro lado, si se anticipaba una mayor proporción en la clase media (ver índice azul), el índice A_n fue relativamente bajo de 2005 a 2018, aunque en 2018 experimentó un aumento, reflejando una mayor heterogeneidad entre los datos reales y el vector esperado (0.10,0.15,0.25,0.40,0.10).

Conclusiones

La propuesta de nuevos índices de índole social, económico, ambiental, de salud, etcétera, ha sido una tarea continua y muy importante desde hace muchos años, ya que representan alternativas para la cuantificación de características de un problema y, por ende, de mejora en las condiciones de los seres vivos.

Las metodologías utilizadas en los índices de desigualdad proporcionan una amplia gama de herramientas para cuantificar la desigualdad intrínseca, es decir, la heterogeneidad estadística. La relevancia de la aplicación de los índices de desigualdad se extiende a todos los campos de la ciencia y la ingeniería, de ahí nuestro gran interés por estudiarla.

Generalmente, los índices de desigualdad se definen al considerar un vector de datos observado que de manera implícita se contrasta con la conformidad o igualdad perfecta. Sin embargo, si permitimos que la conformidad perfecta no implica necesariamente igualdad perfecta, sino una distribución deseable o esperada, podemos construir índices de heterogeneidad. Fue así como en este artículo, hemos propuesto un nuevo índice de heterogeneidad, construido geoméricamente y cuyas propiedades se han probado analíticamente, que compara un vector observado con el esperado, cuyos tamaños están dados por el número de variables que componen al indicador. A diferencia de los índices tradicionales que asumen una conformidad perfecta basada en la igualdad absoluta, nuestro índice permite una definición más flexible de conformidad perfecta. Esto proporciona una medida más precisa de heterogeneidad en situaciones donde hay expectativas específicas sobre la distribución de las variables. Además, hemos demostrado analíticamente las propiedades del índice, garantizando su robustez y aplicabilidad en diferentes escenarios. En casos específicos donde el vector esperado está compuesto por los mismos elementos, nuestro índice se reduce a una medida de desigualdad tradicional, abarcando y extendiendo las aplicaciones de los índices tradicionales. La capacidad de definir vectores esperados específicos permite que el índice se adapte a diferentes disciplinas y contextos, ofreciendo una medida más relevante y contextualizada de la desigualdad o heterogeneidad.

Un caso particular se considera cuando el vector esperado está constituido por los mismos elementos en todas las entradas, por ejemplo, el vector e , que tiene 1's en todas sus entradas, en cuyo caso nuestro índice se convierte en uno de desigualdad. Sabemos que uno de los índices de desigualdad más relevantes es el índice de Gini, el cual considera, por ejemplo, los ingresos de los habitantes de una región determinada. Por lo tanto, podríamos afirmar que el índice de Gini es un indicador de desigualdad cuantitativa. En contraste, el índice propuesto en este estudio se definió como una medida de desigualdad más general.

Hemos aplicado este índice de heterogeneidad en distintas ramas académicas, obteniendo resultados consistentes con otros índices y con

clasificaciones de datos, de tal manera que nos brindan información relevante para la toma de decisiones.

Dado que el bienestar de las personas es de suma importancia, contar con este tipo de índices nos ayuda, por ejemplo, a realizar propuestas de nuevas políticas públicas de tal forma que se reduzca la brecha entre los grupos socioeconómicos, y a identificar patrones o tendencias de los datos; lo que puede ser útil para el análisis y la toma de decisiones.

Aunque el índice propuesto tiene una sólida base geométrica que facilita su interpretación, la construcción matemática subyacente puede ser compleja. Además, la elección del vector esperado puede influir significativamente en el valor del índice. Dependiendo de cómo se defina este vector, los resultados pueden variar, lo que podría introducir sesgos en la interpretación de los resultados.

Adicionalmente, la construcción del índice se basa en el número de combinaciones del tamaño del vector. Por ende, si este número es muy grande, el índice propuesto podría presentar dificultades operativas al momento de ser ejecutado. Potencialmente causaría que el cálculo falle debido a la complejidad computacional involucrada, por ejemplo, cuando $n = 15,000$.

Aunque la propuesta pudiera parecer abrumadora inicialmente debido a las fórmulas involucradas, se ha proporcionado en GitHub el código² necesario para reproducir este índice, demostrando que su implementación es realmente sencilla.

Agradecimientos

Agradecemos profundamente a los revisores por sus detalladas y constructivas observaciones. Sus comentarios han sido fundamentales para mejorar la calidad y claridad de nuestro artículo.

LJRE desea expresar su sincero agradecimiento a la Universidad Autónoma de Aguascalientes por el apoyo financiero brindado a través del proyecto PIM23-3.

² <https://github.com/judithr19/Indice-Hererogeneidad.git>

Referencias

- [1] Corder, G. W. y Foreman, D.I. (2011). "Nonparametric statistics for non-statisticians". Recuperado de [10.1002/9781118165881](https://doi.org/10.1002/9781118165881)
- [2] Cotler, P. (2007). *Políticas Públicas para un crecimiento incluyente*. Universidad Iberoamericana. Recuperado de <https://books.google.com.mx/books?id=y4EKHFigtncC>
- [3] Deeks, J. J., Higgins, J. P., Altman D. G., y C. S. M. Group. (2019). "Analysing data and undertaking meta-analyses". *Cochrane handbook for systematic reviews of interventions*, 241–284. Recuperado de <https://doi.org/10.1002/9781119536604.ch10>
- [4] Eliazar, I. (2017). "Investigating equality: the rényi spectrum". *Physica A: Statistical Mechanics and its Applications*, 481, 90–118. Recuperado de DOI: [10.1016/j.physa.2017.04.003](https://doi.org/10.1016/j.physa.2017.04.003)
- [5] Eliazar, I. (2018). "A tour of inequality". *Annals of Physics*, 389, 306–332. Recuperado de <https://doi.org/10.1016/j.aop.2017.12.010>
- [6] Esparza, L. J. R., Lazcano, D. A. O., Ponce, J. C. M., & Cortés, O. M. M. D. (2020). Bilateral Gini index: Application for regional studies and international comparisons. *RBEST Revista Brasileira de Economia Social e do Trabalho*, 2, e020010-e020010. <https://doi.org/10.20396/rbest.v2i.13481>
- [7] Gini, C. y Salvemini, T. (1912). "Variabilidad e mutabilidad reprinted in pizetti e". *Salvemini. Mem. Metodol. Stat.*
- [8] INEGI (2024). [Microdatos]. ENOE. Recuperado el 30 de julio de 2024 de <https://www.inegi.org.mx/programas/enoe/15ymas/#microdatos>
- [9] Mondragón, A. (2002). "¿Qué son los indicadores?". *Revista de información y análisis*, 19, 52–58. Recuperado de https://gc.scalahed.com/recursos/files/r161r/w23935w/PSM_U4_R2.pdf
- [10] Medina, F. (2001). *Consideraciones sobre el índice de Gini para medir la concentración del ingreso*. Disponible en: <https://hdl.handle.net/11362/4788>
- [11] Nunes, A., Trappenberg, T. y Alda, M. (2020). "The definition and measurement of heterogeneity". *Translational psychiatry*, 10(1), 299. Recuperado de <https://doi.org/10.1038/s41398-020-00986>
- [12] Olatunji, J., Love R., Shim, Y., Ferrua, M., y East, A. (2017). "Quantifying and visualising variation in batch operations: A new heterogeneity index". *Journal of Food Engineering*, 196, 81–93. Recuperado de <https://doi.org/10.1016/j.jfoodeng.2016.10.004>
- [13] Plata-Pérez, L., Sánchez-Pérez, J., y Sánchez-Sánchez, F. (2015). "An elementary characterization of the Gini index". *Mathematical Social Sciences*, 74, 79–83. Recuperado de <https://doi.org/10.1016/j.mathsocsci.2015.01.00>

- [14] PNUD. (2015). "Índice de desarrollo humano para las entidades federativas". Recuperado de: <http://www.pued.unam.mx/export/sites/default/archivos/SUCS/2015/180215RTG.pdf>
- [15] Rodríguez, R. A., Reyes, M., y Favela, M. F. (2016). "La importancia de los indicadores sintéticos en el desarrollo sustentable". Recuperado de <http://ru.iiec.unam.mx/id/eprint/3315>
- [16] Simpson, E. H. (1949). "Measurement of diversity". *Nature*, 163(4148), 688–688. Recuperado de <http://dx.doi.org/10.1038/163688a0>
- [17] Sarabia, J. M. y Jordá, V. (2014). "Explicit expressions of the Pietra index for the generalized function for the size distribution of income". *Physica A: Statistica Mechanica and its Applications*, 416, 582–595. Recuperado de <https://doi.org/10.1016/j.physa.2014.09.025>
- [18] Salazar, R. E. M., y García, J. M. J. P. (2014). "El índice de desarrollo humano como indicador social". *Nómadas. Critical Journal of Social and Juridical Sciences*, 44(4). Recuperado de <https://www.redalyc.org/pdf/181/18153>
- [19] SEP. (2018). "Programa de Estímulo al Desempeño Docente (2008-2009)". Recuperado de <https://dgesui.ses.sep.gob.mx/sites/default/files/2020-03/Manual%20ESDEPED%202008%20Rec.Ext.pdf>
- [20] UAA. 2008. "Programa de Estímulos al Desempeño Docente (2008-2009)". Recuperado de https://dgdg.uaa.mx/fep/pdfs/estim_docente/reglamento_est_desemp_personal_docente_2019.pdf
- [21] UAA. 2022. "Convocatoria Estímulos al Desempeño Personal Docente 2022-2023". Recuperado de https://dgdg.uaa.mx/fep/pdfs/estim_docente/Convocatoria-%20U006%202022%202023%20UAA.pdf
- [22] Zamudio Sánchez, F. J., Miranda, M. G. P., y Tamez, C. A.V. (2001). "Primer informe sobre desarrollo humano de México 1995". Guanajuato: Universidad Autónoma Chapingo y Centro de Desarrollo Humano de Guanajuato AC.
- [23] Zamudio Sánchez, F. J., Santibáñez Cortés, J., Viana Carrillo, S., Andrade Barrera, M. A., Jiménez Machorro, J., Rodríguez Esparza, L. J., y Ávalos Vargas, A. (2017). *Desarrollo Humano en México 2015: el impacto del fondo regional*. Universidad Autónoma Chapingo.