



Universidad de Valladolid

**Facultad de Ciencias
Económicas y Empresariales
Trabajo de Fin de Grado**

Grado en Economía

*Medidas de dependencia alternativas al coeficiente de correlación
lineal de Pearson. Aplicación a la medición del bienestar*

Presentado por:

Luis Rodríguez Miguel

Tutelado por:

Ana Pérez Espartero

Valladolid, 21 de julio 2022

RESUMEN

Este trabajo estudia posibles alternativas al coeficiente de correlación lineal de Pearson para cuantificar el grado de dependencia entre las dimensiones utilizadas para calcular el Índice de Desarrollo Humano (IDH). Lo primero que haremos será definir el IDH, qué mide y de qué dimensiones se compone. Después, definimos el coeficiente de correlación lineal de Pearson y los coeficientes alternativos presentados, basados en rangos o en modificaciones robustas del coeficiente lineal de Pearson. En la aplicación empírica, aplicamos todos los coeficientes para medir las interrelaciones entre los pares de las dimensiones del IDH utilizando datos de 142 países para los años 1990, 1995, 2000, 2005, 2010, 2015 y 2019. Todos los coeficientes que hemos estudiado indican la fuerte relación que existe entre las tres dimensiones que conforman el IDH, y entre éstas, la mayor relación de dependencia en los primeros años se da claramente entre la dimensión de la salud y de economía, pero en los últimos años, la dependencia es más fuerte entre la dimensión de la educación y la economía.

Palabras clave: IDH, correlación, desarrollo humano, Spearman, Kendall

ABSTRACT

This paper studies possible alternatives to Pearson's linear correlation coefficient to quantify the degree of dependence between the dimensions used to calculate the Human Development Index (HDI). The first thing we will do is to define the HDI, what it measures and what dimensions it is composed of. Then, we define Pearson's linear correlation coefficient and the alternative coefficients presented, based on ranges or robust modifications of Pearson's linear coefficient. In the empirical application, we applied all coefficients to measure the interrelationships the pairs of HDI dimensions using data from 142 countries for the years 1990, 1995, 2000, 2005, 2010, 2015 and 2019. All the coefficients we have studied indicate the strong relationship that exists between the three dimensions that make up the HDI, and among these, the strongest dependency relationship in the early years is clearly between the health and economy dimensions, but in the later years, the dependency is stronger between the education and economy dimensions.

Key words: HDI, correlation, human development, Spearman, Kendall

ÍNDICE

1. INTRODUCCIÓN	4
2. EL ÍNDICE DEL DESARROLLO HUMANO	5
2.1. Cálculo del IDH	6
2.2. Interpretación del IDH	10
2.3. Críticas al IDH	11
3. MEDIDAS DE DEPENDENCIA	12
3.1. Coeficiente de correlación lineal de Pearson	12
3.2. Coeficiente de correlación por rangos de Spearman	15
3.3. Coeficiente Tau de Kendall	17
3.4. Coeficiente de correlación basado en mediana	19
4. APLICACIÓN EMPÍRICA	20
4.1. Descripción de los datos	20
4.2. Evolución del IDH	21
4.3. Distribución marginal y conjunta de las dimensiones del IDH	22
4.4. Medidas de dependencia	25
5. CONCLUSIONES	30
6. BIBLIOGRAFÍA	31

1. INTRODUCCIÓN

Hay un consenso general de que el bienestar y el desarrollo humano, son conceptos multidimensionales que no solo dependen de los ingresos sino también de dimensiones no monetarias como la educación o la salud. Por lo tanto, la mejoría en el bienestar dependerá del nivel de avance de cada dimensión y del grado de asociación entre ellas. Como señalan Atkinson y Bourguignon (1982), para dos distribuciones conjuntas con los mismos valores, pero con distinto grado de asociación, una menor asociación entre las dimensiones del bienestar es socialmente preferible. Por ejemplo, si tuviéramos un ranking de países del mundo en el que un país ocupa el primer lugar en todas las dimensiones del bienestar, otro país fuera segundo en todas las dimensiones del bienestar y así sucesivamente, este mundo gozará de menor bienestar que aquel en el cual todas las posiciones de los países en todas las dimensiones estén menos correlacionadas.

Por ello, sería interesante medir el nivel de dependencia entre las dimensiones que componen uno de los índices de bienestar más utilizados, el índice del desarrollo humano, el índice del desarrollo humano (IDH), y este es el objetivo del trabajo. Estas dimensiones, son la salud, la educación y la economía. Para el análisis bivariado, existen varias medidas de dependencia, entre las cuales, la más conocida es el coeficiente de correlación lineal de Pearson, adecuada para distribuciones normales bivariantes. Sin embargo, en el caso de variables cuya distribución no es Normal, sería más adecuado utilizar medidas de dependencia basadas en rangos, como el coeficiente por rangos de Spearman o el coeficiente Tau de Kendall, los cuales se centran más en las posiciones de cada individuo en las dimensiones que en los valores específicos. En este trabajo analizaremos estos coeficientes y también calcularemos un coeficiente de correlación “robusto” basado en la mediana que trata de evitar la influencia de datos atípicos.

Nuestros resultados muestran que todos los coeficientes que hemos estudiado indican la fuerte relación que existe entre las tres dimensiones que conforman el IDH, y entre éstas, la mayor relación de dependencia en los primeros años se da claramente entre la dimensión de la salud y de economía, pero en los últimos años, la dependencia es más fuerte entre la dimensión de la educación y la economía. Además, el coeficiente de

correlación de Pearson ofrece valores que no distan mucho de los valores del coeficiente por rangos de Spearman mientras que el coeficiente Tau de Kendall siempre presenta valores más pequeños y todos ellos difieren bastante de índice basado en la mediana.

El trabajo se estructura de la siguiente manera. En el capítulo 2, introduciremos el IDH, las dimensiones que lo forman con sus respectivos índices y la forma de calcularlo. En el capítulo 3, comenzamos por describir el coeficiente de correlación de Pearson y después, presentaremos las alternativas a este, en el orden siguiente: el coeficiente de correlación por rangos de Spearman, el coeficiente Tau de Kendall y la correlación basada en la mediana. En el capítulo 4 aplicaremos todos estos coeficientes para medir la dependencia entre los posibles pares de los índices que conforman el IDH (índice de salud, índice de ingresos e índice de educación) para los años 1990, 1995, 2000, 2010, 2015 y 2019. En el capítulo 5 resumimos las principales conclusiones.

2. EL ÍNDICE DEL DESARROLLO HUMANO

La Organización de Naciones Unidas (ONU) es una organización internacional formada por 192 países, que puede tomar medidas sobre los problemas que enfrenta la humanidad y tiene cuatro propósitos: Mantener la paz y la seguridad internacionales; Fomentar las relaciones de amistad entre naciones; Favorecer la cooperación internacional para solucionar problemas entre naciones y estimular el respeto a los derechos humanos; armonizar los esfuerzos de las naciones por alcanzar estos propósitos. En concreto, uno de los programas de la ONU, el Programa de las Naciones Unidas para el Desarrollo (PNUD), se encarga de estudiar las cuestiones esenciales para mejorar la calidad de vida de las personas entre las que destacamos, la erradicación de la pobreza, la desigualdad o los impactos de cambio climático. El IDH fue propuesto por el PNUD en el año 1990. Tras estudiar el trabajo del investigador MahBub al Haq, el cual se basaba en las ideas desarrolladas por el Nobel Amartya Sen, se llegó a la conclusión de que el desarrollo de un país no se podía medir únicamente por el nivel de ingresos de este y para que el estudio del desarrollo de cualquier país se ajustara más a la realidad, se debían incluir también variables que midieran la salud y la educación.

Como acabamos de apuntar, el IDH mide el nivel de desarrollo de un país, en función de tres componentes (salud, educación y economía) que determinan el nivel de vida medio en un territorio. A continuación, repasaremos cómo se mide cada componente de dicho índice (PNUD, 1990):

- La componente de la salud se mide en términos de esperanza de vida al nacer.
- La componente de la educación tiene dos subcomponentes, la tasa de alfabetización de adultos y la tasa bruta combinada de matriculación en los tres niveles de enseñanza: primaria, secundaria y terciaria.
- La componente de la economía, que recoge la riqueza de un país, se mide por el ingreso nacional bruto (INB) per cápita.

El PNUD, a la hora de definir el IDH, considera que su *“objetivo no es crear un indicador incuestionable del bienestar, sino reorientar la atención hacia un desarrollo enfocado en el ser humano y alimentar el debate sobre cómo propiciar el progreso de las sociedades”* (PNUD, 2010)

Cabe mencionar que el IDH no es la única métrica que el PNUD considera para medir el progreso de un país. Junto a este indicador, existen otros importantes como el IDH ajustado por desigualdad o el Índice de Desigualdad de Género (Rosales, 2017; Ventura, 2015).

2.1. Cálculo del IDH

Como hemos citado anteriormente, el IDH es un índice que refleja tres dimensiones básicas del desarrollo: salud, educación y riqueza.

Para el cálculo de los indicadores de cada dimensión, en primer lugar, se tendrán que determinar valores mínimos y máximos, lo que se denomina “valores límite”, para transformar cada uno de los indicadores (expresados en diferentes unidades) en números índices con valores entre 0 y 1. La siguiente tabla muestra los valores límite de cada indicador de cada una de las dimensiones:

Dimensión	Indicador	Mínimo	Máximo
Salud	Esperanza de vida (años)	20	85
Educación	Esperanza de años de escolarización	0	18
	Media de años de escolarización	0	15
Ingreso	Producto interior bruto per cápita (2011 PPA \$)	100	75000

Figura 1: Tabla de indicadores (PNUD, 2010).

Así, para un país dado, la forma de calcular cada indicador de dimensión es la siguiente (González García, 2016):

$$\text{Índice de dimensión} = \frac{\text{valor actual} - \text{valor mínimo}}{\text{valor máximo} - \text{valor mínimo}}$$

En el caso de la educación, esta fórmula se aplicará a cada uno de sus dos subcomponentes y luego se calculará el promedio de ambos. En los ingresos, aplicaremos logaritmos naturales de los valores mínimos y máximos reales ya que la contribución de ingresos al nivel de desarrollo del país presenta rendimientos marginales decrecientes.

Dimensión de la salud:

Como ya habíamos indicado antes, esta dimensión se mide con la esperanza de vida al nacer. En concreto para el país i , el indicador de salud, denominado IEV_i , se calcula como:

$$IEV_i = \frac{EV_i - EV_{min}}{EV_{max} - EV_{min}};$$

donde EV_i representa la esperanza de vida del país i y EV_{min} y EV_{max} son los valores mínimo y máximo, respectivamente, de este indicador (véase Figura 1)

Dimensión de la educación:

Como ya habíamos mencionado antes, esta componente se va a medir a través de dos variables.

- *Media de años de escolaridad (MEE)*: Es un cálculo del promedio de años de educación que reciben las personas de 25 o más años a lo largo de su vida, basados en los niveles de educación alcanzados por la población. El indicador de este subcomponente, para el país i se denomina $IMEE_i$, se calcula como:

$$IMEE_i = \frac{MEE_i - MEE_{min}}{MEE_{max} - MEE_{min}};$$

donde MEE_i es la media de años de escolaridad del país i y MEE_{min} y MEE_{max} son los valores mínimo y máximo, respectivamente, de esta subdimensión (véase Figura 1).

- *Años esperados de escolaridad (AEE)*: Es el número de años esperados en los que un niño de edad de matricularse en el primer nivel de educación, pase en los diferentes niveles de educación. Es la suma de las tasas de matriculación primaria, secundaria, post- secundaria no terciaria y terciaria. Para el país i , este indicador, que denominamos como $IAEE_i$, se calcula como:

$$IAEE_i = \frac{AEE_i - AEE_{min}}{AEE_{max} - AEE_{min}};$$

donde AEE_i es la media de años de escolaridad del país i y AEE_{min} y AEE_{max} son los valores mínimo y máximo, respectivamente, de esta subdimensión (véase Figura 1).

Una vez que tenemos los dos subindicadores anteriores, procedemos al cálculo del indicador de educación, que para un país i vamos a denominar IE_i , de la siguiente forma:

$$IE_i = \frac{IAEE_i + IMEE_i}{2}$$

Dimensión de la economía:

Como hemos indicado anteriormente, para calcular el indicador de esta dimensión utilizaremos el PIB per cápita ajustado PPA expresado en logaritmos. Así, para el país i , este indicador, que denominaremos $IPIB_i$, se calcula como:

$$IPIB_i = \frac{\log(PIB_i) - \log(100)}{\log(75000) - \log(100)}$$

donde PIB_i representa el PIB per cápita del país i , y 100 y 75000 son los valores máximos y mínimos, expresados en dólares, de este indicador (véase Figura 1).

Cuando ya hemos obtenidos los tres índices anteriores, uno para cada dimensión, se calcula el IDH del país i como la media geométrica de dichos índices con la siguiente fórmula:

$$IDH_i = \sqrt[3]{IIEV_i * IE_i * IPIB_i}$$

La siguiente figura sintetiza el procedimiento del cálculo del IDH:

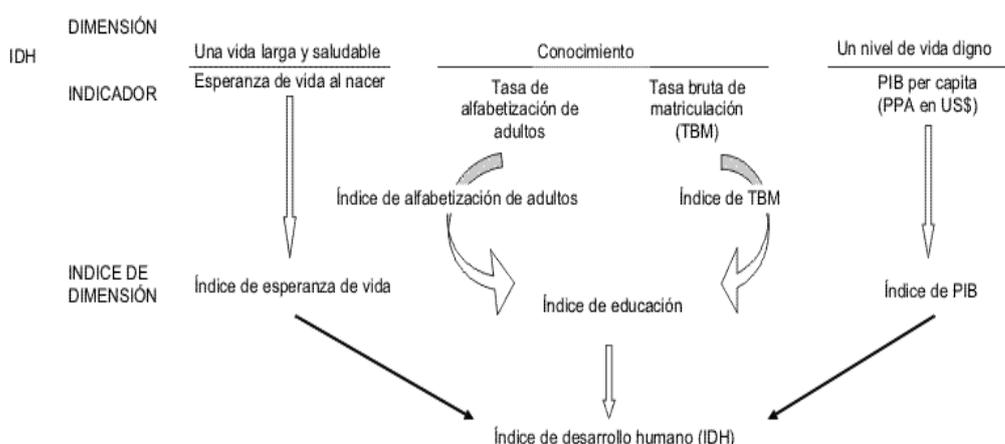


Figura 2: Componentes IDH (Eustat, 2021).

2.2. Interpretación IDH

El IDH asigna a cada país un valor del 0 a 1, desde la calificación más baja hasta la más alta, de modo que cada país será clasificado en el grupo correspondiente según su nivel de desarrollo (Sánchez, 2022), distinguiéndose los siguientes grupos (PNUD, 1990):

- Desarrollo Humano muy alto: países con IDH superior a 0,80.
- Desarrollo humano alto: países con IDH entre 0,70 y 0,80.
- Desarrollo humano medio: países con IDH entre 0,50 y 0,70
- Desarrollo humano bajo: países con IDH por debajo de 0,55.

Un ejemplo del resultado podemos verlo en la Figura 3 con la categorización de 6 países (Iberdrola, 2022):

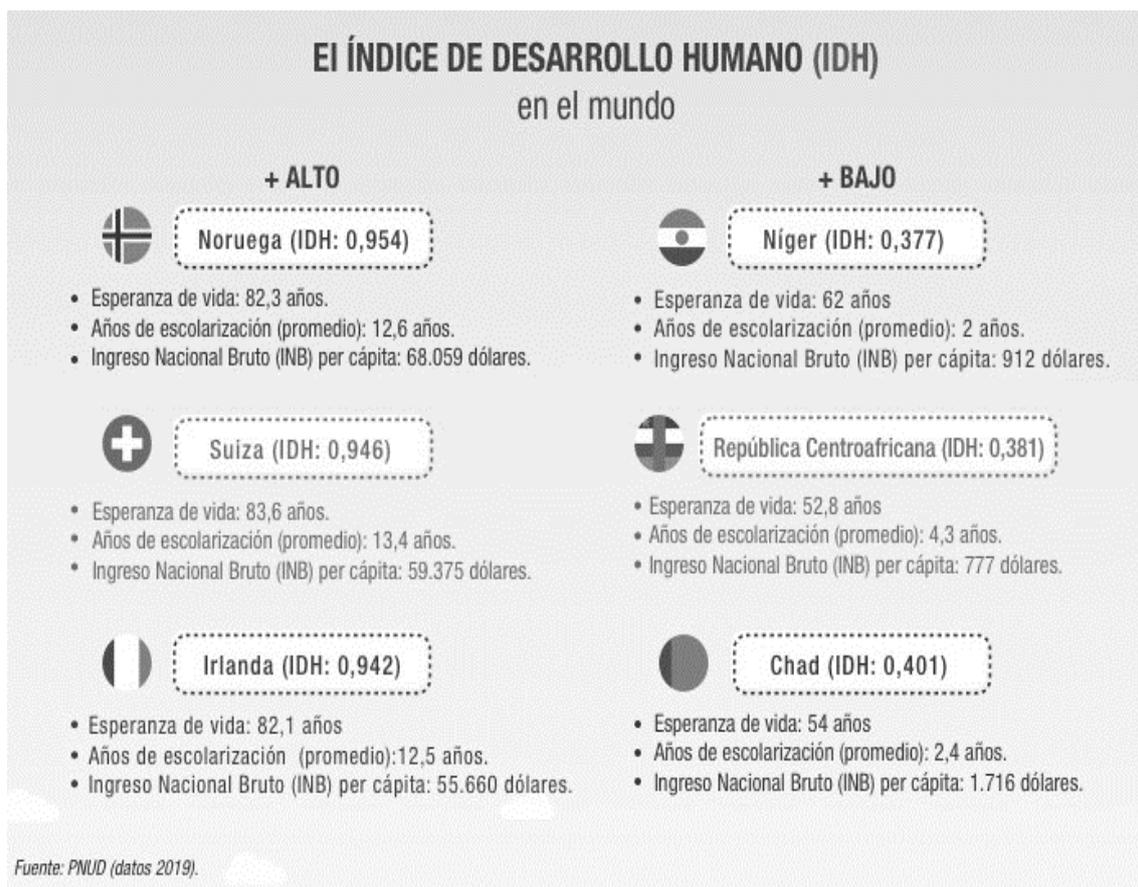


Figura 3: Ejemplo de países según su IDH (Iberdrola, 2022).

2.3. Críticas al IDH

El IDH ha sido duramente criticado desde su creación. A continuación, se expondrán algunos de los argumentos principales entorno a esto.

Como indicador, no puede captar toda la realidad y no incluye variables que puedan reflejar la difícil situación del país. Por ejemplo, en un país con una fuerte caída del PIB per cápita, cuya población comienza a pasar hambre, sus fluctuaciones no se verán reflejadas automáticamente en el IDH. La esperanza de vida, en términos de la primera variable, descenderá, pero de forma paulatina.

Además, no hay una razón lógica para decir que cualquiera de tres componentes del IDH es más importante, más débil o igualmente importante que las otras dos. Hay muchas más variables que se pueden incluir en cualquier medida de desarrollo, ya que pueden afectar a la calidad de vida directa o indirectamente a través de terceras componentes. Por ejemplo, de nada sirve un buen sistema educativo para las mujeres (probablemente la mitad de la población) si existe una extrema desigualdad de género (Montes de Oca Noriega, 2018).

Otro punto importante es el ajuste de la desigualdad. Considere un país con una población de 100. Por ejemplo, una persona que tenía una esperanza de vida de 40 años puede fallecer con 50 y una que tenía una esperanza de vida de 90 fallece con 50. La expectativa de vida promedio es de 65 años, pero la brecha es amplia. Lo mismo puede ocurrir con la educación o los ingresos. El ajuste por desigualdad es un gran paso en el desarrollo de este indicador (Ibáñez, Mujica y Castillo, 2017).

Otra cuestión sujeta a críticas es la composición del índice. Se podrían agregar más indicadores, indicadores que reflejen los niveles de educación, salud e ingresos de una manera más relevante y confiable. De nada sirve tener un promedio de 10 años de escolaridad si la educación es de mala calidad. Del mismo modo, una alta esperanza de vida no garantiza que esos años de vida sean de alta calidad (López, 2022).

En último lugar, cabe señalar, como ya mencionamos en la introducción, que el bienestar no solo depende de los avances en las distintas dimensiones que lo conforman, sino de la dependencia entre dichas dimensiones; véase Atkinson y Bourguignon (1982). Sin embargo, el IDH, como indicador de bienestar, sólo tiene en cuenta las distribuciones marginales de sus tres dimensiones (salud, educación y economía), pero ignora la dependencia entre ellas, siendo este uno de sus inconvenientes (Pérez y Prieto-Alaiz, 2016). Por ello, en la siguiente sección vamos a calcular, utilizando distintos coeficientes, la dependencia bivalente entre dichas dimensiones.

3. MEDIDAS DE DEPENDENCIA

En este apartado definiremos varios coeficientes que nos permiten medir la dependencia entre pares de variables. Decimos que dos variables son independientes cuando el hecho de que una ocurra no altera las probabilidades de ocurrir de la otra, por tanto, en este caso, la distribución conjunta de ambas variables es igual al producto de las distribuciones marginales de cada una de ellas. En caso contrario decimos que las variables son dependientes. Para poder cuantificar el grado de dependencia entre los pares de variables, se pueden utilizar muy diversos coeficientes. El más conocido y utilizado es el coeficiente de correlación lineal de Pearson, pero existen otros más apropiados para cuando las variables no siguen distribuciones normales.

Antes de definir los siguientes coeficientes, introducimos la siguiente notación. Suponemos que tenemos dos variables aleatorias X e Y, y disponemos de n pares de observaciones de dichas variables, $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

3.1. Coeficiente de correlación lineal de Pearson

Entre los numerosos autores que han tratado de cuantificar este grado de dependencia mediante un coeficiente matemático, destacamos a Karl Pearson, quién propuso un coeficiente, muy sencillo de calcular, que permite cuantificar el grado de correlación lineal entre dos variables. Este coeficiente se define como:

$$r_{xy} = \frac{s_{xy}}{s_x * s_y}, \quad (1)$$

teniendo en cuenta que:

- s_{xy} es la covarianza entre X e Y definida como:

$$s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n} \quad (2)$$

- s_x es la desviación típica de X definida como:

$$s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (3)$$

- s_y es la desviación típica de Y definida como:

$$s_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}} \quad (4)$$

- \bar{x} es la media de X definida como:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (5)$$

- \bar{y} es la media de Y definida como:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (6)$$

El coeficiente de correlación lineal r_{xy} toma valores entre -1 y 1, y su interpretación es la siguiente (Etxebarria, 1999):

- Si $r_{xy} < 0$, entre las dos variables se da una correlación negativa: si una de las variables crece, la otra tiende a decrecer y, si una variable decrece, la otra tiende a crecer. A medida que r_{xy} se aproxime a -1, mayor será la intensidad de esa relación negativa.
- Si $r_{xy} = -1$, entre las dos variables se da una correlación perfecta negativa: si una de las variables crece, la otra sufrirá un descenso proporcional conforme a una relación lineal exacta negativa.
- Si $r_{xy} > 0$, entre las dos variables se da una correlación positiva: si una de las variables crece la otra tiende a crecer y, si una variable decrece la otra tiende

decrecer también. A medida que r_{xy} se aproxime a 1, mayor será la intensidad de esa relación positiva.

- Si obtenemos $r_{xy} = 1$, entre las dos variables se da una relación lineal exacta positiva: cuando una de las variables crece, la otra lo hace proporcionalmente y conforme a una relación lineal exacta creciente.
- Si $r_{xy} = 0$, no hay relación lineal entre las variables por lo que podemos decir que serían dos variables incorrelacionadas, lo que no implica que sean independientes, como ilustramos a continuación

El siguiente gráfico muestra una serie de diagramas de dispersión de diferentes pares de variables dependiendo de su coeficiente de correlación de Pearson:

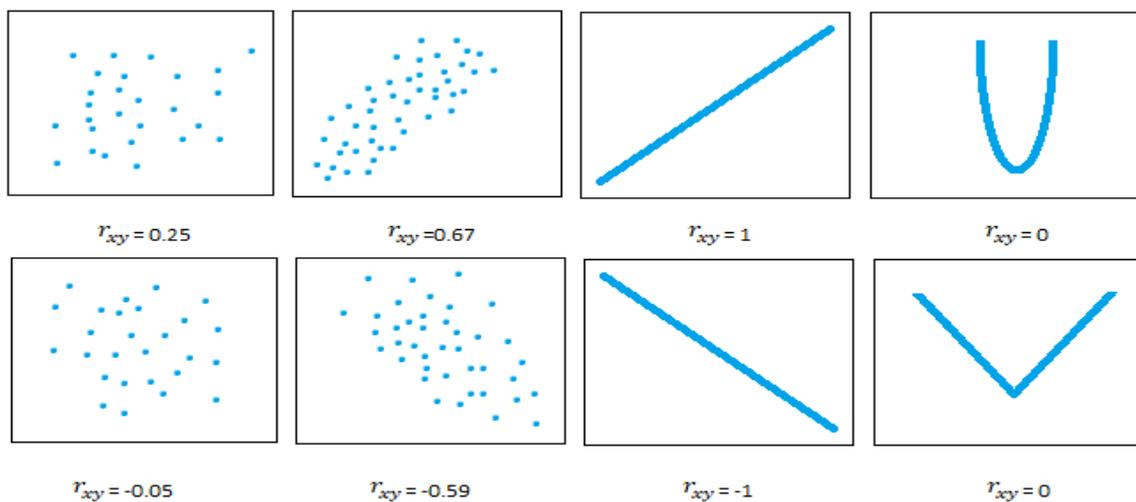


Gráfico 1: Diagramas de dispersión y coeficiente de correlación lineal asociado.

En este gráfico tenemos los diferentes escenarios que encontraríamos según el nivel de dependencia lineal entre las variables. Observamos que el coeficiente r_{xy} va de 0 a 1 cuando la relación es creciente (la pendiente de la nube de puntos es positiva) y, de 0 a -1 cuando la relación es decreciente (la pendiente de la nube de puntos es negativa).

En el gráfico 1 se representan dos casos que ilustran la propiedad antes mencionada, en la que las dos variables están funcionalmente relacionadas ($Y=|X|$) o ($Y=X^2$), y por tanto son totalmente dependientes, aunque el coeficiente lineal de Pearson sea nulo porque

la relación entre ellas no es lineal. Por lo tanto, son dos casos en los que observamos la incapacidad del coeficiente r_{xy} para recoger cualquier dependencia que no sea lineal.

Cuando la distribución conjunta de las dos variables X e Y es Normal, la independencia y la incorrelación lineal son propiedades equivalentes puesto que entre variables normales bidimensionales solo puede existir una relación lineal (Embretchs, McNeil y Straumann, 2002). Sin embargo, en otros casos, como ilustran los dos ejemplos del Gráfico 1 antes mencionados, dependencia e incorrelación no son equivalentes. De hecho, el coeficiente de Pearson puede no ser el más apropiado en distribuciones que se desvían de la normalidad, por ejemplo, por ser asimétricas o por presentar observaciones atípicas. En estos casos, pueden ser más apropiados los coeficientes basados en rangos o cuantiles, algunos de los que veremos a continuación.

3.2. Coeficiente de correlación por rangos de Spearman

El coeficiente de correlación de Spearman es una medida no paramétrica de correlación por rangos, que mide el grado de asociación entre dos variables calculando el coeficiente de correlación de Pearson entre los rangos correspondientes. Para obtener los rangos, se ordenarán las variables de menor a mayor y se asignarán números enteros positivos (rangos) a los valores ordenados de las variables, empezando por el 1 para el valor más pequeño, el 2 para el siguiente, y así hasta n para el mayor valor observado.

Cuando tenemos calculados los rangos, los compararemos de tal manera que, si los rangos de ambas variables son iguales, estaremos hablando de una relación perfecta positiva entre ellos y, si los rangos son contrarios, habrá una relación negativa perfecta entre ellos. Entre estas dos situaciones extremas, encontraremos situaciones intermedias de mayor o menor relación positiva o negativa.

Cuando ya tenemos los rangos asignados, calculamos la diferencia entre ellos para la misma observación de cada variable. Por ejemplo, si tenemos las observaciones (X_i, Y_i) y le asignamos sus rangos $R(x_i)$ y $R(y_i)$, respectivamente, la diferencia entre los rangos, será, $d_i = R(x_i) - R(y_i)$. Podemos observar dicho procedimiento en la Tabla 1:

X	Y	R (X)	R (Y)	d
50	51	1	1	0
53	52	2	2	0
72	61	5	3	2
64	63	3	4	-1
75	69	6	5	1
65	74	4	6	-2
83	77	8	7	1
95	79	10	8	2
82	90	7	9	-2
87	95	9	10	-1

Tabla 1: Variables X e Y, con sus valores, rangos y diferencia entre rangos.

Una vez tenemos los rangos asignados y la diferencia entre estos, el coeficiente de correlación de Spearman (r_s), se calcula de la siguiente forma:

$$r_s = \frac{6 \sum_{i=1}^n d_i^2}{n(n^2-1)} \quad (7)$$

Al igual que el coeficiente de Pearson, el coeficiente de rangos de Spearman toma valores entre -1 y 1, y su interpretación es la siguiente (Fernández Cuesta y Fuentes García, 1995):

- $r_s = 0$: Asociación nula entre las variables.
- $r_s > 0$: Asociación positiva entre las variables, los individuos que están mejor posicionados en una variable, también lo están en la otra.
- $r_s = 1$: los rangos de las variables son iguales, hay asociación perfecta positiva, los individuos ocupan las mismas posiciones en ambas variables.
- $r_s < 0$: Asociación negativa entre las variables, los individuos mejor posicionados en una variable, son los peores posicionados en la otra.
- $r_s = -1$: los rangos son completamente contrarios, el primer individuo en el ranking en una variable es el último en la otra y viceversa, hay asociación perfecta negativa.

En el ejemplo de la Tabla 1 hemos obtenido $r_s = 0,878$, por lo tanto tenemos una asociación positiva entre los rangos de las variables. Para los mismos datos de la Tabla

1, calculamos el coeficiente de correlación de Pearson y obtenemos: $r_{xy} = 0,837$, una correlación también positiva y fuerte, similar al coeficiente de Spearman.

Entre las ventajas del coeficiente de Spearman frente al de Pearson, podemos destacar, que permite medir la correlación entre dos variables cuando las mediciones se realizan en una escala ordinal y que es robusto a “outliers”, es decir, a observaciones que se desvían del patrón del resto de los datos.

Ahora procedemos a calcular de nuevo los coeficientes, pero cambiando el valor de una de las variables por un valor atípico. En concreto, utilizamos $Y_{10} = 950$ en lugar de $Y_{10} = 95$ y obtenemos como resultado, $r_{xy} = 0,378$ y $r_s = 0,878$. Como podemos observar, debido al valor atípico que hemos introducido, el coeficiente de correlación lineal de Pearson sufre un gran cambio y se reduce notablemente, pero el coeficiente de Spearman no varía ya que Y_{10} antes de que cambiáramos el valor ya se encontraba en la posición más alta.

3.3. Coeficiente Tau de Kendall

Para definir el coeficiente Tau de Kendall, el concepto más relevante es el par concordante. Si tenemos dos pares de observaciones (x_i, y_i) y (x_p, y_p) decimos que los pares son concordantes cuando $(x_i - x_p)(y_i - y_p) > 0$ es decir, cuando ocurre simultáneamente o bien $\{(x_i > x_p) \text{ y } (y_i > y_p)\}$ o bien $\{(x_i < x_p) \text{ y } (y_i < y_p)\}$. Veremos a continuación un ejemplo con los datos de la siguiente tabla:

X	Y
1	1
2	3
3	2

Tabla 2: Datos del ejemplo.

Veamos cuáles de los pares son concordantes y cuáles no:

- Los pares de observaciones (x_1, y_1) y (x_2, y_2) son concordantes porque $(x_1 < x_2)$ e $(y_1 < y_2)$
- Los pares de observaciones (x_1, y_1) y (x_3, y_3) son concordantes porque $(x_1 < x_3)$ e $(y_1 < y_3)$.

- Los pares de observaciones (x_2, y_2) y (x_3, y_3) son discordantes porque $(x_2 < x_3)$ e $(y_2 > y_3)$.

Una vez contados los pares concordantes, el coeficiente de Tau de Kendall se calculará restando la proporción de pares concordantes menos la proporción de pares discordantes y dividiendo entre el número de posibles pares de la siguiente forma:

$$\tau = \frac{(\text{número de pares concordantes}) - (\text{número de pares no concordantes})}{\binom{n}{2}} \quad (8)$$

donde $\binom{n}{2} = \frac{n(n-1)}{2}$ es el número de todos los posibles pares que se pueden formar con n observaciones.

Como los anteriores coeficientes estudiados, el coeficiente Tau de Kendall toma valores entre -1 y 1 y su interpretación es la siguiente (Liebetreau, 1983):

- $\tau = 0$: Ausencia total de relación de las variables.
- $\tau > 0$: La mayoría de los pares son concordantes, por lo tanto, hablamos de una relación positiva entre las variables, los individuos bien posicionados en unas variables tienden a estar bien posicionados en la otra.
- $\tau = 1$: Todos los cruces son concordantes, por lo tanto, hablamos de relación positiva perfecta entre las variables, la ordenación de los individuos en ambas variables es idéntica.
- $\tau < 0$: La mayoría de los pares son disconcordantes, por lo tanto, hablamos de una relación negativa entre las variables, los individuos bien posicionados en una variable tienden a estar mal posicionados en la otra.
- $\tau = -1$: No hay ningún par concordante, por lo tanto, hablamos de relación negativa perfecta entre las variables.

En nuestro ejemplo de la Tabla 1, hemos obtenido $\tau = 0,733$, por lo cual podemos afirmar que hay una relación positiva entre las variables. Recordamos que, para estos mismos datos de esta tabla, el coeficiente de correlación de Pearson era $r_{xy} = 0,837$ y el de Spearman $r_s = 0,878$. Por lo tanto, en este caso, el coeficiente de correlación lineal de Pearson y el de Spearman nos indican una correlación más fuerte que el coeficiente Tau de Kendall. Siempre que las variables de estudio no tengan una distribución conjunta

Normal, será más adecuado el coeficiente Tau de Kendall o el de Spearman que el coeficiente de correlación lineal de Pearson.

Calculamos el coeficiente Tau de Kendall con nuestro dato atípico, $Y_{10} = 950$, y obtenemos que $\tau = 0,733$. Al igual que el coeficiente de Spearman, tampoco varía, lo que indica que este coeficiente también es más robusto que el coeficiente de correlación lineal de Pearson ante la presencia de datos atípicos.

3.4. Coeficiente de correlación basado en mediana

Como acabamos de ilustrar, un problema del coeficiente lineal de correlación de Pearson es que, al estar basado en promedios, es muy sensible a datos atípicos, es decir, a datos que difieren mucho del resto y no parecen provenir de la misma distribución. Un dato atípico podría aparecer como consecuencia de, por ejemplo, una guerra, una crisis mundial como la derivada de la pandemia de la COVID-19, etc y por ello, para evitar la influencia de estos datos atípicos surgen medidas de correlación robustas basadas en cuartiles, además de las basadas en rangos calculadas previamente.

Por ejemplo, Shevlyakov y Smirnov (2011) proponen una alternativa robusta al coeficiente de correlación lineal de Pearson que consiste en sustituir en la ecuación (1) todas las medidas basadas en promedios por las correspondientes medidas basadas en la mediana. Este coeficiente, que se denota r_{comed} , se define como:

$$r_{comed} = \frac{\text{med}((x - \text{med}(x))(y - \text{med}(y)))}{\text{MAD}(x)\text{MAD}(y)} \quad (9)$$

donde $\text{med}(x)$ y $\text{med}(y)$ son la mediana de x e y , respectivamente, y MAD es una medida robusta de dispersión definida como la mediana de las desviaciones absolutas a la mediana, es decir:

$$\text{MAD}(x) = \text{med}(|x - \text{med}(x)|) \quad (10)$$

$$\text{MAD}(y) = \text{med}(|y - \text{med}(y)|) \quad (11)$$

Podemos observar que, en el numerador de la ecuación (9), utilizamos las medianas en vez de las medias, como hacíamos en la ecuación (2), y hacemos lo mismo en el denominador, utilizamos el valor absoluto en vez de elevar al cuadrado, como se haría en la desviación típica en las ecuaciones (3) y (4). De este modo, amortiguamos el efecto adverso de las observaciones atípicas de gran magnitud.

Si calculamos este coeficiente con los datos de nuestro ejemplo de la Tabla 1, obtenemos $r_{\text{comed}} = 0,967$, mientras que después de introducir el dato atípico, obtenemos $r_{\text{comed}} = 0,967$. Como vemos, este coeficiente no varía al igual ocurría con el coeficiente por rangos de Spearman y el coeficiente Tau de Kendall coeficiente y a diferencia de lo que ocurría con el coeficiente de correlación lineal de Pearson.

4. APLICACIÓN EMPÍRICA

En esta sección vamos a aplicar los coeficientes que hemos desarrollado en la sección anterior para medir la dependencia por pares entre los índices de cada dimensión que componen el IDH y, una vez calculados, comentaremos los resultados obtenidos.

4.1. Descripción de los datos

Nuestro estudio lo realizaremos a partir de los datos procedentes de la aplicación de la base de datos del PNUD, “Human Development Report Office Statistical Data API” (http://ec2-54-174-131-205.compute-1.amazonaws.com/API/HDRO_API.php) en la que encontramos todos los datos que utilizan para sus estudios sobre el desarrollo humano.

Respecto al periodo de estudio, hemos tomado los años, 1990, 1995, 2000, 2005, 2010, 2015 y 2019 porque queríamos un periodo lo suficientemente amplio para poder estudiar una variación del IDH en un periodo que incluya eventos relevantes que pudieran afectar al desarrollo, como la “Gran Recesión de 2008” o la crisis de los años 90, por ejemplo.

Respecto a la selección de las variables, nos hemos centrado en las dimensiones que componen el IDH que, como hemos visto en la sección 2, son salud, ingresos y

educación. Tenemos descargados directamente de la web del PNUD (<https://hdr.undp.org/data-center/human-development-index#/indicies/HDI>) los tres indicadores, uno para cada dimensión, descritos en la sección 2, así como los propios valores del IDH para cada país y cada año analizado.

Una vez seleccionados los años y las variables, obtenemos una tabla para cada año para un máximo de 192 países y aquí, nos encontramos el primer problema. Hay países como, por ejemplo, Bután, que no ofrecen ningún dato de alguna de las variables seleccionadas para los años seleccionados por lo que en algunos años tenemos una lista con más países que en otros años. El segundo problema es que hay otros países, como Andorra, que no ofrecen datos solo en alguno de los años para las variables seleccionadas. Para solucionar estos problemas y hacer una tabla unificada con el mismo número de años analizados, países y variables, realizamos una depuración de datos y eliminamos todos los países que no ofrezcan todos los datos para todas las variables y todos los años analizados, obteniendo como resultado un total de 142 países que son sobre los que trabajaremos.

4.2. Evolución del IDH

En el *Gráfico 2* podemos observar la evolución de la distribución del Índice del desarrollo humano desde 1990 hasta 2019 representada a través de diagramas de caja, uno para cada año.

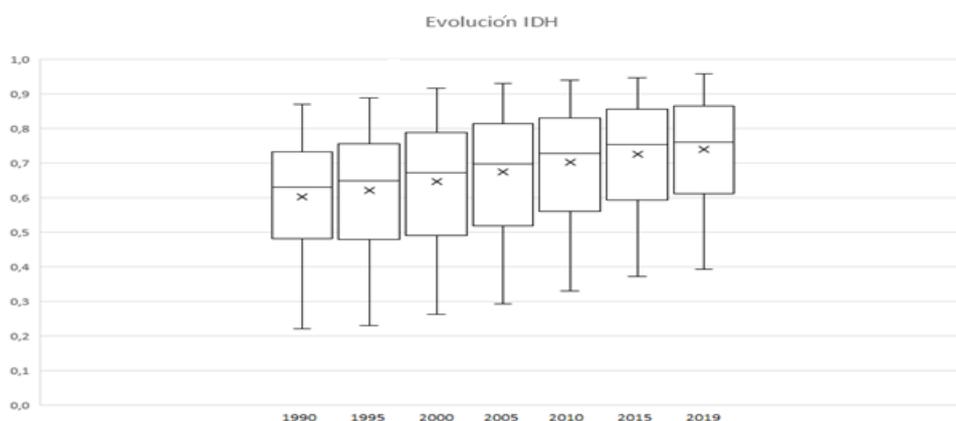


Gráfico 2: Evolución de la distribución del IDH

Hemos utilizado este tipo de gráfico para ver de manera más clara la forma de la distribución y sus principales características (posición, dispersión y forma) y su evolución a lo largo del tiempo. Cada diagrama de caja representa la distribución del IDH de los 142 países para un año concreto. En los extremos de la caja tenemos los cuartiles inferior (Q1) y superior (Q3), mientras que los extremos de los bigotes nos indican el máximo y mínimo sin valores atípicos. Dentro de la caja observamos un aspa que nos indica el valor de la media y la disección, que nos indica la mediana.

Como podemos observar, las distribuciones no son simétricas puesto que la media y la mediana no coinciden. Para todos los años analizados la mediana está por encima de la media. Además, la distancia entre el primer cuartil y la mediana es siempre mayor que la distancia entre que la mediana y el tercer cuartil y la distancia entre el mínimo y el primer cuartil es siempre mayor que la distancia entre el tercer cuartil y el máximo. Por lo tanto, la distribución es asimétrica a la izquierda. Nuestro gráfico no presenta valores atípicos.

A simple vista, observamos que el IDH crece en todo el periodo analizado. Observamos que los mínimos crecen a mayor velocidad que los máximos, por lo tanto, deducimos que el IDH de los países subdesarrollados aumenta a mayor ritmo que el de los países desarrollados. Observamos que, aunque habiéndose dado la crisis de la década de los años noventa y la crisis de 2008, no observamos ningún dato atípico en esos años.

En el periodo afectado por la crisis de los años noventa (1990 a 1995), el primer cuartil es el único que decrece por lo que podemos intuir que la crisis de 1990 afectó más a los países con un IDH bajo que a los de niveles mayores.

Respecto a las crisis de 2008, podemos observar que el IDH sigue aumentando, pero observamos que, los países con IDH bajo crecen más rápido que los países con un IDH alto y muy alto, por lo tanto, podemos deducir que esta crisis afectó más a los países desarrollados que a los no desarrollados, que no vieron un importante aumento de su desarrollo en ese periodo.

4.3. Distribución marginal y conjunta de las dimensiones del IDH

Antes de pasar al cálculo e interpretación de los coeficientes de dependencia presentados en la sección 3, se mostrarán una serie de gráficos de la distribución de los índices de las tres dimensiones que componen el IDH.

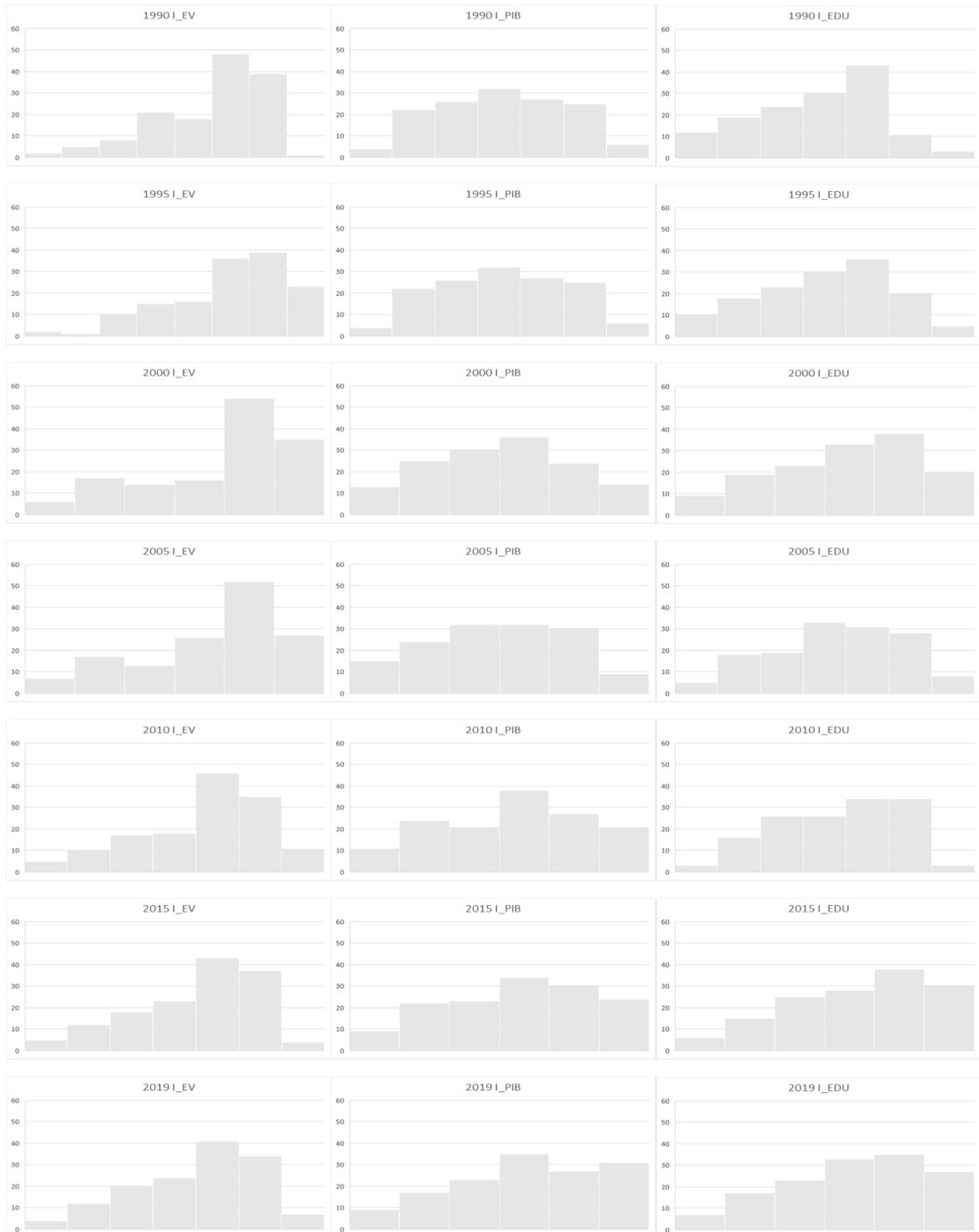


Gráfico 3: Histogramas de los índices que componen el IDH

El gráfico 3 muestra los diferentes histogramas de cada uno de los índices de las dimensiones que componen el IDH. Cada fila representa un año, desde 1990 (arriba) hasta 2019 (abajo). En cada columna representamos un índice de dimensión: en el panel izquierdo representamos los histogramas del índice de esperanza de vida, en el panel central del índice de ingresos y, en el panel derecho, el índice de educación.

Utilizando el gráfico 3, observamos, la falta de normalidad de los tres índices que componen el IDH. Por ejemplo, el indicador de salud presenta una clara asimetría a la izquierda, al igual que el de educación. Por lo tanto, parece justificado el cálculo de los coeficientes de dependencia presentados en la sección 3 para estudiar si las conclusiones que de ellos se obtienen difieren de las obtenidas utilizando únicamente el coeficiente de correlación lineal.

En el gráfico 4 de la siguiente página se muestran los diagramas de dispersión de las posibles combinaciones dos a dos de los tres índices que forman el IDH, uno por cada dimensión, desde 1990 (arriba) a 2019 (abajo). En la columna de la izquierda, tenemos el diagrama de dispersión entre el índice de salud y el índice de ingresos, en la columna central, el índice de salud frente al índice de educación y en la columna de la derecha, el índice de ingresos frente al índice de educación.

En primer lugar, observamos que todos los diagramas de dispersión presentan una tendencia decreciente: la mayoría de los puntos están en el cuadrante superior derecho y en el cuadrante inferior izquierdo y únicamente una minoría de puntos está lejos de la diagonal principal. Por lo tanto, podemos decir que, por lo general, todos los pares de índices representados presentan una relación positiva: a mayores niveles de educación, mayores niveles de ingresos y salud, y viceversa.

Además, observamos que, con el paso de los años, las nubes de puntos se han desplazado hacia la derecha y hacia arriba, especialmente en (salud, economía) y (salud, educación). Esto significa que los niveles de cada uno de estos indicadores han aumentado y por ello, el IDH ha aumentado.

Intuimos que la mayor correlación se dará, para la mayoría de años analizados, entre los índices de economía y educación ya que presentan datos más alineados y concentrados entorno a la diagonal principal que los otros dos pares de dispersiones.

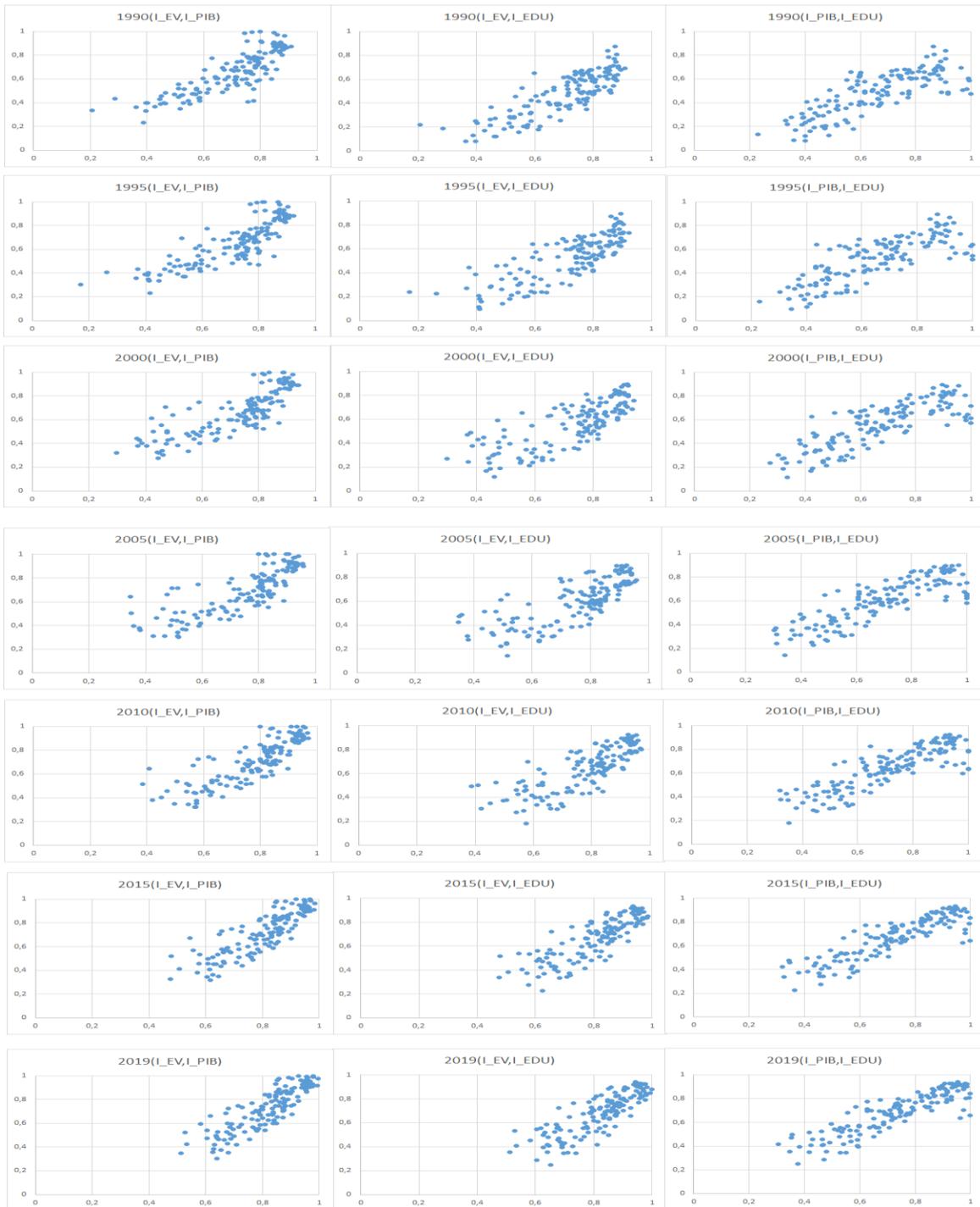


Gráfico 4: Diagramas de dispersión de las combinaciones dos a dos de los índices que conforman el IDH.

4.4. Medidas de dependencia

A continuación, calculamos para cada año los cuatro coeficientes de dependencia entre los pares de variables representados en el Gráfico 4. La Tabla 3 recoge los valores de dichos coeficientes.

Año	Person: rPI (EVI_EDU)	Person: rPI (EVI_PIB)	Person: rPI (PIB_EDU)	Speeman: rSI (EVI_EDU)	Speeman: rSI (EVI_PIB)	Speeman: rSI (PIB_EDU)	Tau de Kendall: KI (EVI_EDU)	Tau de Kendall: KI (EVI_PIB)	Tau de Kendall: KI (PIB_EDU)	r _{COMED} (EVI_EDU)	r _{COMED} (EVI_PIB)	r _{COMED} (EDU_PIB)
1990	0,832	0,837	0,793	0,813	0,846	0,783	0,620	0,656	0,586	0,693	0,792	0,848
1995	0,796	0,838	0,792	0,797	0,858	0,783	0,598	0,668	0,592	0,648	0,709	0,895
2000	0,774	0,828	0,818	0,790	0,850	0,808	0,585	0,654	0,617	0,622	0,819	0,817
2005	0,767	0,812	0,840	0,795	0,836	0,833	0,590	0,642	0,640	0,701	0,887	0,912
2010	0,789	0,832	0,844	0,809	0,850	0,832	0,599	0,658	0,644	0,863	0,777	0,923
2015	0,831	0,856	0,877	0,843	0,870	0,871	0,640	0,679	0,694	0,868	0,882	0,887
2019	0,834	0,865	0,882	0,847	0,874	0,876	0,644	0,682	0,697	0,905	0,776	0,956

Tabla 3: Coeficientes de dependencia.

Para estudiar mejor la evolución de cada coeficiente para cada par de dimensiones relacionadas, presentaremos en el Gráfico 5 todos los coeficientes calculados para cada par de variables, de la siguiente manera.

En cada uno de los tres gráficos, presentamos todos los coeficientes para cada año que hemos estudiado. En el primer gráfico, tenemos todos los coeficientes que miden la relación entre salud (I_EV) y educación (I_EDU) a lo largo del periodo analizado. De la misma manera, los dos siguientes gráficos de columnas corresponden a la relación entre economía (I_PIB) y educación (I_EDU) y la relación entre salud (I_EV) y economía (I_PIB), respectivamente.

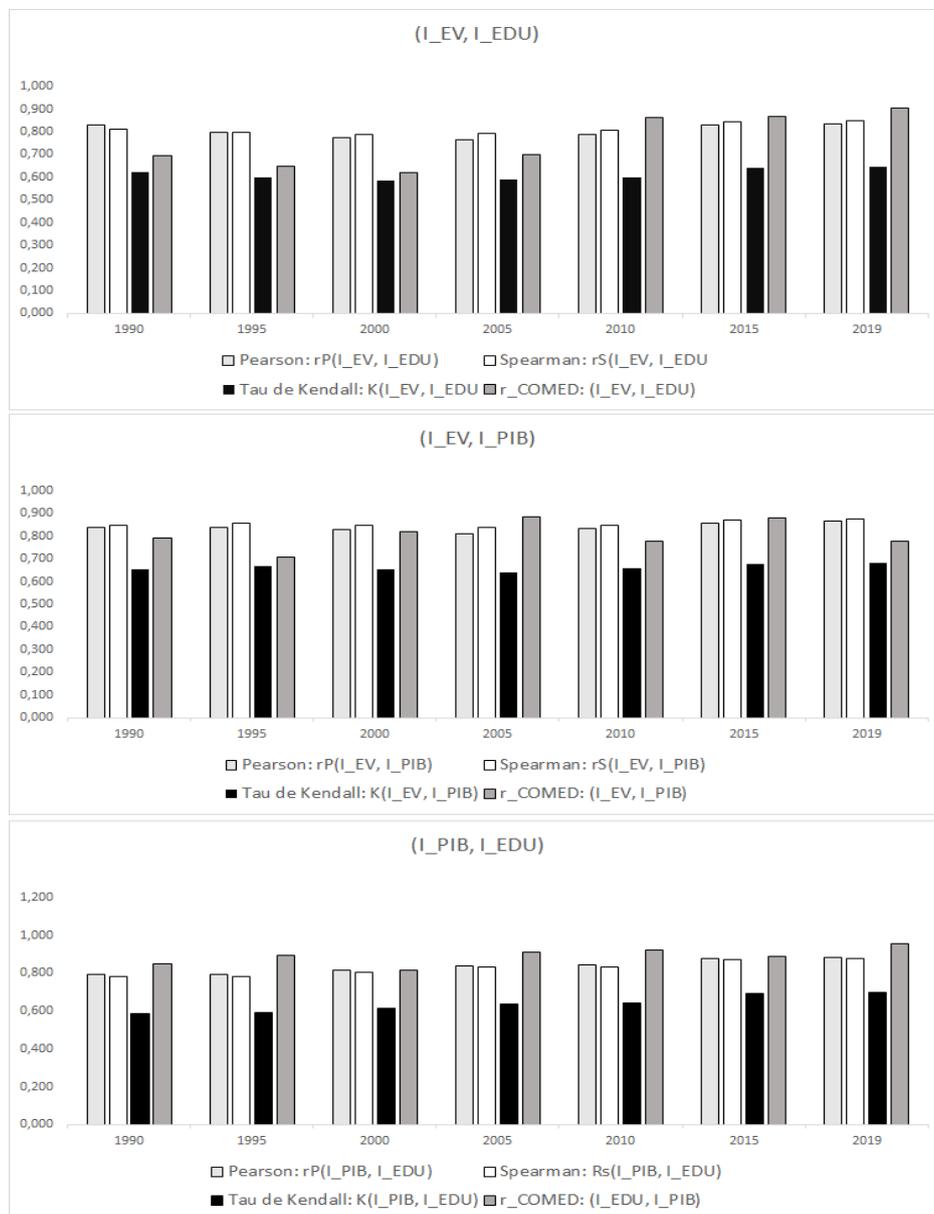


Gráfico 5: Evolución de los coeficientes de dependencia para cada par de dimensiones del IDH.

De los resultados presentados en la Tabla 3 y en el Gráfico 5 se deducen las siguientes conclusiones:

- Todos los coeficientes son positivos y elevados, es decir, tenemos una relación de dependencia fuerte y positiva en todos los pares de índices relacionados. Cuando un país, por ejemplo, tiene un índice de ingresos alto, también presentará índices de educación y salud altos mientras que, si su índice de PIB es bajo, los otros dos también lo serán.
- El coeficiente de correlación lineal de Pearson no toma valores muy dispares a los otros coeficientes estudiados y, en todas las series, toma valores muy parecidos al coeficiente de Spearman.
- Con los coeficientes de Pearson y Tau de Kendall, el grado de la dependencia entre economía y educación aumenta año tras año a lo largo del periodo analizado. Sin embargo, la dependencia entre salud y educación y entre salud y economía parece muy estable a lo largo de los años analizados.
- El coeficiente de Kendall es el menor en todos los casos. Recordemos que, para valores de Tau-Kendall mayor que cero, existe una relación con el coeficiente de Spearman que es la siguiente (Nelsen, 1991):

$$1,5\tau - 0,5 \leq r_s \leq 0,5 - \tau - 0,5 * \tau^2$$

- En general, la mayor relación de dependencia se observa siempre entre salud (I_EV) y economía (I_PIB), en todos los coeficientes utilizados, excepto en el caso de r_{comed} , que muestra mayor dependencia entre educación (I_EDU) y economía (I_PIB). No obstante, en los últimos años todos los coeficientes muestran su máximo valor en la asociación entre economía (I_PIB) y educación (I_EDU).

Para completar el análisis, en el Gráfico 6 representamos la evolución de cada uno de los coeficientes calculados para cada par de variables para los años analizados. Arriba a la izquierda presentamos la evolución del coeficiente de correlación lineal de Pearson, a su derecha el coeficiente de rangos de Spearman, abajo a la izquierda el coeficiente Tau de Kendall y a su derecha r_{comed} .

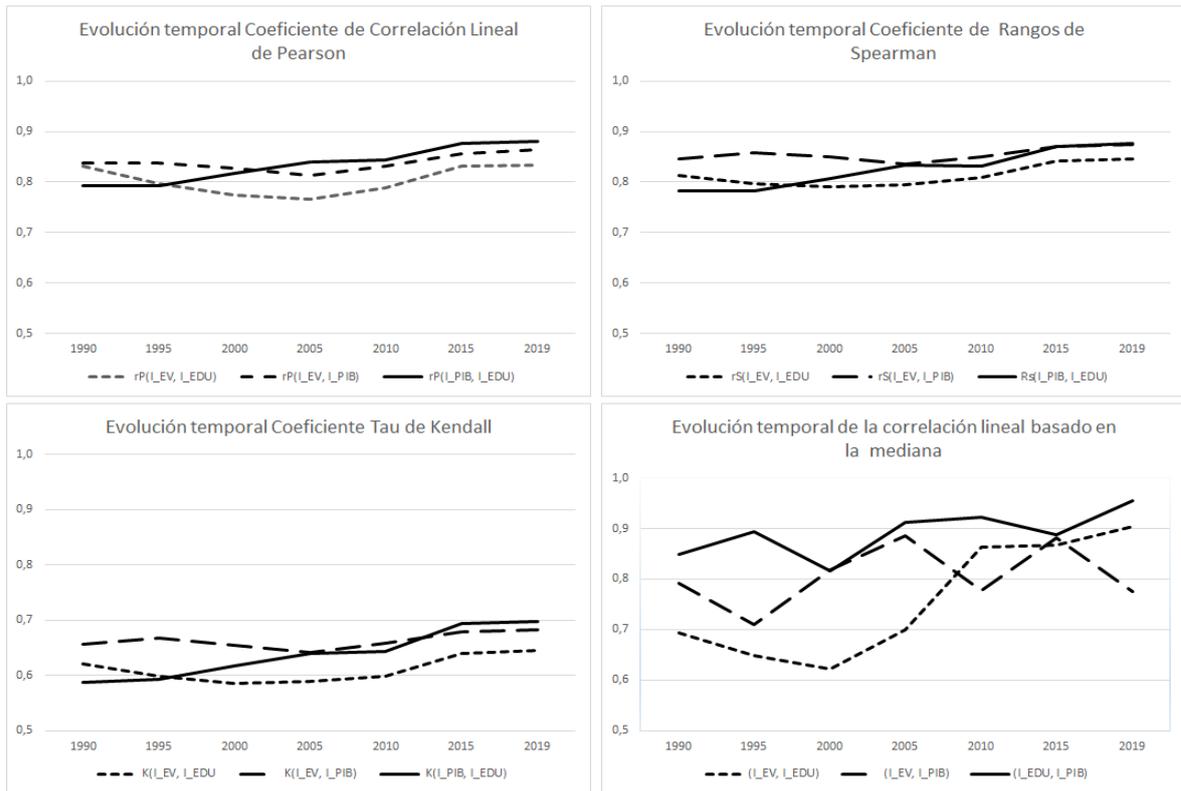


Gráfico 6: Evolución de los coeficientes de dependencia por tipo de coeficiente.

De este gráfico, destacaremos las siguientes características:

- Existe una gran similitud en la evolución y en los valores de los coeficientes de Pearson, de Spearman y de Kendall.
- Existe una gran diferencia, tanto en los valores como en la evolución temporal, de la correlación basada en la mediana respecto a las otras tres medidas.
- La relación de dependencia que más aumenta a lo largo del tiempo con todos los coeficientes, a excepción de r_{comed} , es siempre entre educación (I_{EDU}) y economía (I_{PIB}).
- Las mayores fluctuaciones en el tiempo las presenta r_{comed} .

5. CONCLUSIONES

En este trabajo hemos analizado el grado de dependencia entre los índices de cada una de las dimensiones que componen el IDH (el índice de la dimensión de la salud, el índice de la dimensión de la educación y el índice de la dimensión de la economía) para los años 1990, 1995, 2000, 2005, 2010, 2015 y 2019.

En primer lugar, hemos estudiado quién y para qué se creó IDH, qué variables lo componen y cómo se calcula. Después, hemos introducido el coeficiente de correlación lineal de Pearson y hemos discutido sus limitaciones para medir la dependencia entre variables cuando estas se alejan de la normalidad, lo cual hace necesario utilizar otros coeficientes. A continuación, explicamos las alternativas que proponemos para tratar de mejorar la cuantificación de la dependencia entre las dimensiones que componen el IDH. En concreto, hemos utilizado el coeficiente de asociación de Spearman, que está basado en rangos, el coeficiente Tau de Kendall, que se basa en el concepto de concordancia, y el coeficiente r_{comed} , que es una medida robusta alternativa al coeficiente de correlación de Pearson, basado en la mediana en lugar de en promedios.

En la aplicación empírica analizamos, en primer lugar, la falta de normalidad en los datos para justificar la posible falta de precisión del coeficiente de correlación lineal de Pearson y el cálculo de los otros coeficientes alternativos. Una vez calculados esos coeficientes, hacemos un análisis de los resultados. Nuestra primera conclusión es que el coeficiente de correlación de Pearson ofrece valores que no distan mucho de los valores del coeficiente por rangos de Spearman mientras que el coeficiente Tau de Kendall siempre presenta valores más pequeños. Para los pares de variables estudiados, claramente el que más fluctúa y se diferencia del resto de los coeficientes es " r_{comed} ", que está basado en la mediana.

Otra conclusión que se extrae de este trabajo es que todos los coeficientes que hemos estudiado indican la fuerte relación que existe entre las tres dimensiones que conforman el IDH, y entre éstas, la mayor relación de dependencia en los primeros años se da claramente entre la dimensión de la salud y de economía, pero en los últimos años, la dependencia es más fuerte entre la dimensión de la educación y la economía.

Por lo tanto, a pesar del incremento paulatino del IDH a lo largo de los años analizados, lo que significa que ha habido un aumento del bienestar en el mundo en los años analizados, podemos afirmar que los países más ricos tienden a ser los más sanos y los mejor educados mientras que los países más pobres tienden a tener niveles más bajos tanto en salud como en educación, y esta como la dependencia entre los pares de índices que componen el IDH ha ido en aumento.

6. BIBLIOGRAFÍA

Atkinson A.B y Bourguignon F. (1982), "The comparison of multidimensional distributions of economic status", *Review of Economic Studies* 12 183-201.

Embrechts, P., McNeil, A. J., y Straumann, D. (2002). Correlation and Dependence in Risk Management: Properties and Pitfalls. In M. A. H. Dempster (Ed.), *Risk Management: Value at Risk and Beyond* (pp. 176-223). Cambridge: Cambridge University Press.

<http://dx.doi.org/10.1017/CBO9780511615337.008>.

Etxeberria, J. (1999). *Regresión Múltiple*. Cuadernos de Estadística. Ed. La Muralla S.A. Espérides, Salamanca.

EUSTAT. (2021). Ficha metodológica: Índice de Desarrollo Humano. Disponible en https://www.eustat.eus/documentos/datos/PI_metod/IDH_IDH_c.asp.

Fernández Cuesta, C., Fuentes García, F. (1995). *Curso de Estadística Descriptiva: Teoría y práctica*. Editorial Ariel, Barcelona.

González García, J. (2016). "IDH de México y China: evolución y perspectivas 2030". *Revista Mexicana de Estudios Sobre la Cuenca del Pacífico*, 10(20), 203–227. Disponible en <https://udlap.idm.oclc.org/login?url=https://search-ebsohost-com.udlap.idm.oclc.org/login.aspx?direct=true&db=a2h&AN=120376095&lang=es&site=eds-live>.

Ibañez, N., Mujica, M., y Castillo, R. (2017). Componentes del desarrollo humano sustentable. *Negotium: Revista de Ciencias Gerenciales*, 12(36), 63-77.

Iberdrola. (2022) ¿Qué es el Índice de Desarrollo Humano? | Medición y Objetivos - Iberdrola. Retrieved April 22, 2022. Disponible en <https://www.iberdrola.com/compromiso-social/indice-desarrollo-humano>.

Liebetreau A. (1983): Measures of association, Sage Publications, Beverly Hills.

López. F. L. (2022).” Guía para calcular e interpretar el IDH” - Economipedia. Retrieved April 22, 2022. Disponible en <https://economipedia.com/guia/guia-para-calcular-e-interpretar-el-idh.html>.

Montes de Oca Noriega, D. F. (2018). La estimación de la calidad de vida: una alternativa al índice de desarrollo humano. Centro de Investigación y Docencia Económicas, A.C.

Nelsen, R.B. (1991). Copulas and Association. In: Dall’Aglio, G., Kotz, S., Salinetti, G. (eds) Advances in Probability Distributions with Given Marginals. Mathematics and Its Applications, vol 67. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-3466-8_3.

Pérez, A. y Prieto-Alaiz, M. (2016) “Measuring the Dependence Among Dimensions of Welfare: A SStudy -based on Spearman’s Footrule and Gini’s Gamma” International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 24, 1, 87-105.

PNUD (1990). Desarrollo humano. Informe. Conceptos y Medidas del Desarrollo Humano. Disponible en: http://hdr.undp.org/sites/default/files/hdr_1990_es_completo_nostats.pdf.32

PNUD (2010). Informe sobre Desarrollo Humano 2010. La verdadera riqueza de las naciones: Caminos al desarrollo humano. Disponible en https://hdr.dev.undp.org/sites/default/files/hdr_2010_es_complete_reprint.pdf.

Rosales, M. (2017). “ El desarrollo humano: una propuesta para su medición”. Aldea Mundo, vol.22, 65-75.

Sánchez. G.J. (2022) economipedia *Índice de desarrollo humano (IDH) | 2022 | Economipedia*. Retrieved April 22, 2022, Disponible en <https://economipedia.com/definiciones/indice-desarrollo-humano.html>.

Shevlyakov, G. y Smirnov, P. (2011): “Robust estimation of the correlation coefficient: an attempt of survey”, *Austrian Journal of Statistics*, 40, pp 147-150.

Ventura, J. P. (2015). Un nuevo indicador para medir el desarrollo: el Índice de Desarrollo Socioeconómico (IDSE). *Universidad de Zaragoza-AGE*, 2103-219.