

Dagnino Pastore, José María ; Libonatti, Luis María

*Lidiando con las estadísticas internacionales en
las ciencias sociales*

Ensayos de Política Económica Año VIII, Vol. II, N° 2, 2014

Este documento está disponible en la Biblioteca Digital de la Universidad Católica Argentina, repositorio institucional desarrollado por la Biblioteca Central "San Benito Abad". Su objetivo es difundir y preservar la producción intelectual de la Institución.

La Biblioteca posee la autorización del autor para su divulgación en línea.

Cómo citar el documento:

Dagnino Pastore, J. M., Libonatti, L. M. (2014). Lidiando con las estadísticas internacionales en las ciencias sociales [en línea], *Ensayos de Política Económica*, 2(2). Disponible en: <http://bibliotecadigital.uca.edu.ar/repositorio/revistas/lidiando-estadisticas-internacionales-ciencias.pdf> [Fecha de consulta:.....]

LIDIANDO CON LAS ESTADÍSTICAS INTERNACIONALES EN LAS CIENCIAS SOCIALES

José María Dagnino Pastore y Luis María Libonatti²⁶

Resumen

Este artículo busca señalar algunos inconvenientes que surgen a partir de un extenso trabajo de comparación de la medición de variables en los distintos países, en el que se resaltan diferencias visibles. Frente a esto, se busca aclararlas con el fin de poder realizar comparaciones más acertadas. Para eso, el trabajo se divide en cuatro partes: la definición de los conceptos en las ciencias sociales, la descripción de las variables, el tratamiento de ellas y de datos cuantitativos, y por último el repaso por las diferencias en las estadísticas internacionales

Abstract

This article tries to convey some inconveniences that result after an extensive work which compares the ways in which the variables of different countries are measured, due to the fact that there are visible differences amongst them. Thus, the aim is pointed at the ability to make more precise comparisons. To do so, this work is divided into four parts: la definition of concepts of the social sciences, the description of the variables, the treatment of these and of quantitative data, and finally the scan on the main differences found in international statistics.

Keywords: concepto, variable, método, estadística

JEL codes: B410

Introducción

Durante la preparación de un trabajo basado en comparaciones internacionales de variables económicas, políticas y sociales²⁷ se tropieza en cada paso con problemas inesperados. Este artículo ordena las enseñanzas recogidas al abocarse a resolverlos, que quizás sean útiles a quienes tengan que recorrer caminos similares.

²⁶ Miembro de la Academia Nacional de Ciencias Económicas, Profesor, Pontificia Universidad Católica Argentina "Santa María de los Buenos Aires" (UCA) y Asistente de Investigación, respectivamente. dagnino.pastore@fibertel.com.ar

²⁷ Dagnino Pastore (2014).

Consta de cuatro partes. La primera se refiere a la definición de los conceptos en las ciencias sociales. La segunda, a su representación en conceptos mensurables - variables -. La tercera afronta las cuestiones más prácticas de los datos cuantitativos disponibles y de su tratamiento. Finalmente, la cuarta repasa algunas anomalías encontradas en las estadísticas internacionales.

I. De los conceptos en las ciencias sociales

Una ley de la lógica es la relación inversa entre la extensión y la comprensión - contenido, poder explicativo - de los conceptos; en las ciencias sociales suele transformarse en un problema, ya que emplean 'grandes conjuntos' [que] con el ánimo de abarcar todo terminan al cabo explicando muy poco (Botana, 2012)²⁸.

En particular, a veces hay elementos integrantes de un gran concepto que forman parte de otro - p.e.: la "libertad económica" es parte de la "libertad" (en general) y también de la "competitividad económica" -: el conjunto libertad (en general) se intersecta con el conjunto competitividad económica.

Esto no le quita utilidad al uso de ambos conceptos - todo nuestro lenguaje es prueba de ello -, pero sí tiene por lo menos 2 consecuencias:

- a) si se emplean juntos en un grupo de conceptos, la libertad económica está sobre-representada; y
- b) la libertad (en general) y la competitividad económica tendrían de salida una correlación, aunque ella represente la "realidad".

I.1 De conceptos y palabras

Admitamos *prima facie* que los "conceptos"²⁹ existen. ¿Por qué existen? Porque nos permiten funcionar mejor en nuestro entorno natural y cultural, dentro de nuestras posibilidades y limitaciones físicas y mentales.

¿Qué son los conceptos? Para describirlos con cierta pretensión lógica, un concepto define uno o más atributos que un objeto debe - o no - cumplir para ser un elemento de aquél³⁰.

La selección de atributos se adecua al concepto buscado, pero ¿Para qué y cómo se elige el concepto buscado entre infinitos posibles?

El "para qué" está claro: para mejorar nuestro funcionamiento (p. e.: los conceptos de "león" y "bisonte" permitían un mejor ajuste del hombre primitivo a su entorno).

²⁸ Botana (2012).

²⁹ Enseguida los describimos.

³⁰ Los conceptos pueden ser más o menos nítidos o "borrosos". Ver Zadeh (1965); pp.339-53 y Klaua (1965); pp.859-76.

Naturalmente, la creación de conceptos va teniendo en cuenta los ya existentes y se van conformando agrupaciones y redes.

“Para la formación de categorías [*vgr.*: conceptos y sus agrupamientos] se proponen 2 principios generales y básicos. El primero... afirma que la función de los sistemas de categorías es proveer la máxima información con el menor esfuerzo cognitivo; el segundo sostiene que el mundo percibido viene como información estructurada más que como atributos arbitrarios o impredecibles. Así la máxima información con el menor esfuerzo cognitivo se obtiene si las categorías copian la estructura percibida del mundo lo más precisamente posible... ya sea ligando las categorías a ciertos atributos de las estructuras, o definiendo y redefiniendo los atributos para lograr un conjunto dado de categorías apropiadamente estructurado”³¹.

Esto en cuanto a la finalidad de la creación de conceptos.

En cuanto al “cómo” de la formación de conceptos, un pionero de la psicología cognitiva la ha descrito como “la búsqueda y el listado de atributos que pueden ser usados para distinguir entre ejemplares y no ejemplares de una categoría”³².

Una definición más actual: “En psicología, es el desarrollo de ideas basadas en las propiedades comunes de objetos, eventos a calidades usando los procesos de abstracción y generalización”³³.

Pero el gran progreso de la psicología cognitiva en cuanto al cómo, no es sustituto de la comprensión de la finalidad que guía la creación de conceptos.

Esta “economía conceptual” culmina con la asignación al concepto de un signo que lo identifique: seña, guiño, gruñido, palabra, dibujo rupestre, carácter, letra, etc.

En este sentido, p. e., “la premisa inicial de la teoría es que la base del lenguaje natural es conceptual. Esta base es inter-lingüística; sus elementos son conceptos, no palabras. El contenido conceptual subyacente a, p.e., un gruñido o palabra es su significado.... Las relaciones de dependencia entre conceptos forman una red que constituye la base conceptual”³⁴.

Pero el lenguaje es mucho más que una ayuda mnemónica interna de las estructuras conceptuales en cada persona; su enorme contribución es como vehículo de comunicación interpersonal, lo que requiere que un signo tenga “significado” comprensible para los interlocutores. Aunque de hecho, en los lenguajes usuales, la relación entre palabras y conceptos no son biunívocas: aquéllas son en general “polisémicas” - tienen varios significados -.

Un pionero en estos temas³⁵ ve el lenguaje como una convención - regularidad de conductas y/o creencias que resuelve problemas de coordinación, *vgr.*: soluciones a

³¹ Roach (1978). El texto entre corchetes, y el subrayado, son nuestros.

³² Bruner y otros (1967).

³³ *The American Heritage* (2007).

³⁴ Schank (1972); pp. 552-631.

³⁵ Lewis (1969).

“juegos de coordinación” -, y se pregunta: ¿Cómo emergen las convenciones? ¿Por qué la gente se comporta conforme las convenciones?

En la jerga de la teoría de los juegos, puede haber múltiples soluciones a un juego de coordinación, pero las matrices de *pay-off* premian la coordinación en sí por sobre cuál de las múltiples soluciones se elige. *Vgr.*: lo importante es hablar el mismo idioma, más que cuál se elige - si se eligen distintos no hay comunicación -.

La elección de la solución puede ser aleatoria, o condicionada históricamente, o impuesta, o por otros motivos que van definiendo un “punto focal” al que se converge³⁶.

I.2 Del rigor conceptual

El problema mencionado por Botana se inscribe, en el enfoque de Roach, como un desajuste entre las categorías, expresadas en lenguajes, y la estructura percibida del mundo.

Sin pretender una taxonomía de tales desajustes, es claro que pueden ser de origen (estáticos) o ir surgiendo con el paso del tiempo (dinámicos). También es claro que son más probables cuando se trata de conceptos de grandes conjuntos, de difícil falseabilidad.

Tanto los cambios naturales y sociales por el paso del tiempo como el uso de términos con intencionalidad política los pueden alejar de la estructura percibida del mundo, violando principios de la formación de categorías conceptuales.

El mayor interés actual en la superación de tales desajustes proviene no sólo de inquietudes metodológicas, sino de la realidad, internacional y local.

En particular, la percepción pública de los conceptos está afectada por el contenido emocional y valorativo que a los diferentes relatos políticos les pueda convenir.

La construcción social de la realidad³⁷ es un proceso continuo, que se origina en cambios naturales y sociales y resulta en modificaciones también continuas en esa realidad.

Son palpables la salida y entrada de uso de las palabras, los cambios en los significados de la misma palabra, etc.

De allí la importancia de precisar el contenido y la extensión de los grandes conceptos seleccionados para el análisis, así de los subconjuntos que los integran, y de señalar las intersecciones entre ellos, todo lo cual sufre una necesaria adaptación a la disponibilidad de información cuantitativa.

³⁶ Esto explica la lucha por liderar mercados como, p.e.: los vehículos de redes sociales (“*winner takes all*”).

³⁷ Berger y Luckmann (1966).

II. De la representación de conceptos por variables cuantitativas

Localmente, la persistente preocupación por la calidad de las series estadísticas se acrecentó a partir de principios de 2007 con relación a los índices de precios y también de sus efectos en otras variables, pero también avivó el interés en aclarar qué se quiere medir.

En el orden internacional, la proliferación de mediciones de supuestos conceptos, en particular los sociales y políticos, si bien supuestamente orientadoras, también levantan dudas sobre su significado, así como la asignación de variables para medir conceptos en principio bien definidos³⁸.

No es nuevo: implícita en la famosa *Methodenstreit* - la "disputa del método" - en la Alemania de fines del Siglo XIX entre Karl Menger³⁹ y Gustav von Schmöller⁴⁰, estaba la diferencia, traducida al lenguaje actual, entre la "nitidez" de los conceptos deductivos y la "borrosidad" de los conceptos inductivos. Aunque cuando se baja a la arena empírica, a la medición de variables, muchos conceptos deductivos requieren ser precisados o redefinidos.

La disponibilidad de series estadísticas de variables sociales comparables internacionalmente era muy escasa antes de la creación, en 1944-5, de los grandes organismos públicos internacionales - Naciones Unidas [**NU**], Fondo Monetario Internacional [**FMI**], Banco Mundial [**BM**], etc. - los que empezaron a establecer metodologías (p.e.: la del sistema de cuentas nacionales en 1952) y a propiciar su producción en los países.

Desde entonces fue aumentando gradualmente, incluyendo la provisión por fuentes internacionales privadas (p.e.: *World Values Surveys* desde 1981), hasta que con la irrupción de la computación se aceleró, con un notable florecimiento hacia mediados de los 90s y aún mayor en este siglo.

Las estadísticas disponibles pueden clasificarse en "duras" y "blandas". Las duras se obtienen a partir de registros que aspiran a cubrir la totalidad del universo representado. P.e.: la tasa de homicidios intencionales, en base a datos policiales.

Las blandas se elaboran a partir de encuestas, que recogen muestras del universo representado, o de opiniones sobre él. P.e.: la percepción de corrupción.

Debido a sus recursos y a su poder de inducir la producción de datos homogéneos en los países, las principales fuentes de series estadísticas duras comparables internacionalmente son los grandes organismos públicos internacionales. También lo son de las blandas, pero en éstas la participación de organismos internacionales privados es mayor.

³⁸ Esto llevó a la realización de un Seminario sobre "La medición en las ciencias sociales", en 2010 en la UCA, coordinado por Carlos Hoevel, algunas de cuyas contribuciones fueron publicadas en la Revista Cultura y Economía.

³⁹ Menger (1883).

⁴⁰ Schmöller, Gustav von (1883).

Más allá de representar aceptablemente los conceptos de las ciencias sociales a analizar, las variables cuantitativas seleccionadas a los efectos de comparaciones internacionales (de alcance mundial) deben cumplir con, entre otros, estos requisitos: ser comparables en el tiempo (para el período analizado, en nuestro caso desde 1980) y en el espacio (datos para un alto número de países, p.e.: unos 100, ½ del total) y con la frecuencia requerida.

III. Del tratamiento de los datos cuantitativos

En general las comparaciones internacionales de series estadísticas tratan de aclarar 3 cuestiones clásicas: ¿Cómo está un país dado en el mundo? ¿Cómo ha evolucionado? ¿Cómo ha avanzado o retrocedido en el mundo?

Los datos disponibles de las variables seleccionadas presentan dificultades para su interpretación y para su comparación inter-temporal e internacional y entre variables. A continuación exponemos los criterios usados para resolverlas o atenuarlas.

a) Se emplean cifras decenales de años que terminan en 0, porque para facilitar su comparación internacional, las NU recomiendan que los países hagan los censos, decenales, en los mismos años, y eligió los que terminan en 0. En algunos casos se dispone de datos posteriores a 2010; los hemos empleado sistemáticamente si corresponden a 2013 o aún a 2012. Si las series comienzan más tarde, p.e.: en los años 90; hemos usado datos quinquenales. Si las observaciones de países vienen dadas en "ondas" o "ciclos" (p.e.: 2009-13), los datos pueden requerir ciertos ajustes para centrarlos (p.e.: 2011) y compatibilizarlos con los años de las series quinquenales (p.e.: 2010). **(Ver # 1. b).**

b) Casi nunca la "muestra" de países - el conjunto de los que se tienen cifras - de un año es exactamente igual, en cantidad y en composición, a la de otro año. Esto afecta los promedios y los *rankings* y plantea problemas para su interpretación y comparación inter-temporal e internacional y entre variables. **(Ver # 2).**

c) Cuando los datos son de organismos internacionales públicos, se deben basar en los provistos por cada país; no siempre totalmente fidedignos o de similar calidad. En algunos casos lo hacemos notar y aportamos visiones a partir de datos de otras fuentes.

d) Los indicadores provistos por las fuentes usan escalas distintas. Los hemos uniformado en la escala de 0 (o 1) a 100 - aunque alguna serie original venía normalizada, lo que impide comparaciones absolutas -. Algunos indicadores conocidos, al transformarlos a la escala uniforme pierden la facilidad de interpretación que da el uso frecuente. En las Tablas de puntajes la escala va de 0 (peor) a 100 (mejor), salvo que se aclare lo contrario. En las Tablas de *rankings* (percentiles) **(Ver # 2)** la escala va de 1 (mejor) a 100 (peor).

III.1 Temas de datos

Si se estudian tendencias de largo plazo, es probable que se emplee una periodicidad quinquenal (o decenal).

a. Datos de países

En las estadísticas de un país se pueden dar varios casos:

- a) Hay datos para todos los años requeridos para la serie, p.e.: quinquenal: no hay problema.
- b) No hay datos para algún año de la serie quinquenal, y hay datos para el año precedente y/o el año siguiente al de la serie quinquenal: se toma el valor si sólo hay datos de uno de ellos (precedente o siguiente), o se toma el valor promedio de ellos si hay datos de los dos (precedente y siguiente):

$$Y_{t(1)} = (\delta_i Y_{t+1} + \delta_i Y_{t-1}) / \sum \delta_i \quad , \text{ donde:}$$

$Y_{t(1)}$ = Valor estimado de la variable en el año t de la serie quinquenal, en base a datos del(os) año(s) contiguos(s)

δ_i = Delta de Kronecker: $\delta_i = 1$ si hay datos de Y_{t+1} , $\delta_i = 1$ si hay datos de Y_{t-1} ; $\delta_i = 0$ si no hay datos de Y_{t+1} , $\delta_i = 0$ si no hay datos de Y_{t-1}

Y_{t+1}, Y_{t-1} = Valores observados de la variable en los años t+1 y t-1

- c) No hay datos para el año precedente al de la serie quinquenal, ni para el año siguiente al de la serie quinquenal, pero sí hay datos para el año anterior al precedente y/o para el año posterior al siguiente: se toma el valor si sólo hay datos de uno de ellos (anterior al precedente o posterior al siguiente), o se toma el valor promedio de ellos si hay datos de los 2 (anterior al precedente y posterior al siguiente):

$$Y_{t(2)} = (\delta_i Y_{t+2} + \delta_i Y_{t-2}) / \sum \delta_i \quad , \text{ donde:}$$

$Y_{t(2)}$ = Valor estimado de la variable en el año t de la serie quinquenal, en base a datos del(os) año(s) contiguos(s)

δ_i = Delta de Kronecker: $\delta_i = 1$ si hay datos de Y_{t+2} , $\delta_i = 1$ si hay datos de Y_{t-2} ; $\delta_i = 0$ si no hay datos de Y_{t+2} , $\delta_i = 0$ si no hay datos de Y_{t-2}

Y_{t+2}, Y_{t-2} = Valores observados de la variable en los años t+2 y t-2

- d) No hay datos ni para los 2 años precedentes al de la serie quinquenal, ni para los 2 años siguientes al de la serie quinquenal: se admite que no hay datos del país disponibles para el año de la serie quinquenal.
- e) No hay datos para algún año de la serie quinquenal, pero hay una serie similar que cubre los años faltantes, y los datos faltantes corresponden a años no extremos de la serie quinquenal: se siguen los procedimientos expuestos en b). En nuestra experiencia no hemos necesitado de otros métodos, aunque podría usarse la correlación entre ambas series:

$Y_{t(c)} = a + b X_t + e$, donde:

$Y_{t(c)}$ = Valor estimado de la variable en el año t de la serie quinquenal, en base a la correlación entre las variables Y y X, y el valor de la variable X en el año t.

a, b = Coeficientes de la regresión de Y en X

X_t = Valor de la variable X en el año t

- f) No hay datos para algún año de la serie quinquenal, pero hay una serie similar que cubre los años faltantes, y los datos faltantes corresponden a el(los) año(s) extremo(s) (p.e.: el último) de la serie quinquenal: se "engancha" la serie nueva a la vieja:

$Y_{t(e)} = Y_{t-n} + [(X_t - X_{t-n}) \sigma_Y / \sigma_X]$, donde:

$Y_{t(e)}$ = Valor estimado de la variable en el año t de la serie quinquenal, en base a su "enganche" con la variable X

Y_{t-n} = Valor observado de la variable Y en el año t-n

X_t, X_{t-n} = Valores observados de la variable X en los años t y t-n, respectivamente

σ_Y = desviación estándar de la variable Y en el período de Y_{t-n} a Y_{t-n-m}

σ_X = desviación estándar de la variable X en el período de X_t a X_{t-n}

En lo posible, se trata que $m = n$.

b. Datos internacionales

Las estadísticas internacionales suelen proveer datos con periodicidad variable y agruparlas en "ondas" o "ciclos" quinquenales. Tal agrupación tiene la ventaja de aumentar el tamaño de la muestra y la desventaja de perder la sincronía perfecta entre los datos de los distintos países (**ver # 1.a**).

Para adaptar los datos de los países al año central del "ciclo" vale lo expuesto (**Ver # 1. a**), pero puede ocurrir que dicho año no coincida con el año "correspondiente" de la serie quinquenal.

En tal caso, las soluciones alternativas serían:

- Tomar las cifras del año central del ciclo (*vgr.*: 2011) como si correspondieran al año de la serie quinquenal - implica suponer que no hubo cambios entre 2010 y 2011 -, o
- Correr el ciclo (*vgr.*: de 2009-23 a 2008-12), de modo que su año central coincida con el correspondiente de la serie quinquenal - lo que exige un recalcular las cifras de todos los países -.

III.2 Temas de métodos

La cuantificación de las variables "representativas" de aspectos seleccionados de la situación y e(in)volución de la sociedad argentina y de su entorno zonal e internacional se manifiesta en puntajes y sus consecuentes *rankings*, cuya interpretación y

comparación, temporal, espacial y conceptual, plantea cuestiones de métodos, a las que me aboco.

III.2.a Temas de puntajes

Clasificamos las cuestiones de puntajes en 2 grupos: las de escalas y las de muestras.

a. 1. Diferentes escalas

Con referencia a los puntajes provistos por las fuentes, hay 2 casos: o sus valores tienen un "techo" - límite superior - pasible de alcanzar o no lo tienen. Nótese que hay variables cuyos puntajes tienen un techo pero inalcanzable (p.e.: tasa de homicidio del 100%) que anula su utilidad como límite de la escala, pues las observaciones se aglomerarían en el 5% inferior). Además hay diferencias entre las escalas de los puntajes provistas en las series originales y la escala uniforme adoptada para este trabajo: de 0 (peor) a 100 (mejor).

Escala original con límite superior

En este caso, la única tarea es transformar la escala usada en los datos originales a la escala uniforme adoptada para este trabajo: de 0 (peor) a 100 (mejor). Las diferencias entre ambas escalas pueden provenir de si son as(o des) cendientes; de si sus rangos son más (o menos) extensos y de si sus medias son mayores (o menores). La fórmula para transformar la escala original en la de este trabajo, si aquélla es ascendente - ésta siempre lo es - es:

[1] $y_i = [(y_z - y_A) / (x_z - x_A)] [x_i + (\bar{y} - \bar{x})]$, donde:

y_i = Puntaje del país i, en la escala de este trabajo

y_z = Puntaje máximo 100 (mejor) de la escala

y_A = Puntaje mínimo 0 (peor) de la escala

x_i = Puntaje del país i, en la escala original

y_z = Puntaje máximo de la escala original

y_A = Puntaje mínimo de la escala original

$\bar{y} = y_z - y_A$, Puntaje medio (5) de la escala de este trabajo

$\bar{x} = x_z - x_A$, Puntaje medio de la escala original

La fórmula para transformar la escala original en la de este trabajo, si aquélla es descendente (p.e.: la escala original del coeficiente de "inequidad" Gini se transforma en una escala de "equidad") es:

[2] $y_i = 100 - x_i$

Escala original sin límite superior. En este caso, la 1ª tarea sería fijar un techo razonable a la escala. Como no hemos encontrado una solución satisfactoria, usamos la escala superior y sacrificamos la comparabilidad (los casos son pocos).

a. 2. Diferentes muestras

El hecho de que la muestra de países del año 1 incluya(ex)cluya países (no)estaban en la muestra de países del año 0 crea problemas para interpretar los cambios ocurrido en puntajes promedio entre el año 0 y el año 1. ¿Cuánto de estos cambios se debe a cambios en los puntajes de los países y cuánto a la in(ex)clusión de países en la muestra?

La descomposición de los cambios en los niveles de los puntajes promedios

Para elucidarlo definimos:

Pt = Puntaje promedio los países en el año t,

p_{i,t} = Puntaje del país i en el año t,

c_i = País i (i = 1...N), donde:

a) Si hay cifras del país c_i, c_i = 1, y

b) Si no hay cifras del país c_i, c_i = 0; y

t = Periodos (t = 0...T).

[4] $P1 = (\sum_{i=1}^N c_{i1} p_{i1}) / (\sum_{i=1}^N c_{i1})$

[5] $P0 = (\sum_{i=1}^N c_{i0} p_{i0}) / (\sum_{i=1}^N c_{i0})$

[6] $P1 / P0 = (P1 / PA) / (PA / P0)$; donde:

$PA = (\sum_{i=1}^N c_{i0} p_{i1}) / (\sum_{i=1}^N c_{i0})$

[7] $P1 / P0 = \frac{(\sum_{i=1}^N c_{i1} p_{i1}) / (\sum_{i=1}^N c_{i1})}{(\sum_{i=1}^N c_{i0} p_{i1}) / (\sum_{i=1}^N c_{i0})} \cdot \frac{(\sum_{i=1}^N c_{i0} p_{i1}) / (\sum_{i=1}^N c_{i0})}{(\sum_{i=1}^N c_{i0} p_{i0}) / (\sum_{i=1}^N c_{i0})}$

[8] $P1 / P0 = \frac{(\sum_{i=1}^N c_{i1} p_{i1})}{(\sum_{i=1}^N c_{i0} p_{i1})} \cdot \frac{(\sum_{i=1}^N c_{i0})}{(\sum_{i=1}^N c_{i1})} \cdot \frac{(\sum_{i=1}^N c_{i0} p_{i1})}{(\sum_{i=1}^N c_{i0} p_{i0})}$

Definimos:

$$ILP1 = \frac{\sum_{i=1}^N C_{i0} p_{i1}}{\sum_{i=1}^N C_{i0} p_{i0}}, \text{ Índice Laspeyres de puntaje promedio}$$

$$IEC1 = \frac{(\sum_{i=1}^N C_{i1} p_{i1}) \cdot (\sum_{i=1}^N C_{i0})}{(\sum_{i=1}^N C_{i0} p_{i1}) \cdot (\sum_{i=1}^N C_{i1})}, \text{ Índice de efectos países, donde:}$$

$$IPC1 = \frac{\sum_{i=1}^N C_{i1} p_{i1}}{\sum_{i=1}^N C_{i0} p_{i1}}, \text{ Índice Paasche de países}$$

$$M1 = \sum_{i=1}^N C_{i1}, \text{ N}^{\circ} \text{ de países con cifras en el periodo 1}$$

$$M0 = \sum_{i=1}^N C_{i0}, \text{ N}^{\circ} \text{ de países con cifras en el periodo 0}$$

Por lo tanto: el cambio relativo en el nivel del puntaje promedio es el producto de:

- a) un índice Laspeyres de puntaje promedio (ILP), y
- b) un índice Paasche de países (IPC) ajustado por el N° de países con cifras (M0 / M1):

$$[9] P1 / P0 = ILP1 \times IPC1 \times (M0 / M1)$$

El cambio % en el nivel del puntaje promedio es:

$$[10] dP1\% = (dP1 / P0) 100 = ([ILP1 \times IPC1 \times (M0 / M1)] - 1) 100, \text{ donde:}$$

dP1% = Cambio % en el nivel del puntaje promedio.

El índice "encadenado" de niveles de puntajes promedio

Pero la utilización reiterada del Índice de Laspeyres de puntaje promedio tiene el inconveniente de mantener fija la composición de la muestra de países, alejándolo de la aproximación a la realidad que permite su adecuación - según la disponibilidad de datos -.

Por ello se recurre al encadenado del índice:

$$[11] ILP_{1,0} = \frac{\sum_{i=1}^N C_{i0} p_{i1}}{\sum_{i=1}^N C_{i0} p_{i0}}, \text{ Índice del año 1, base año 0}$$

$$[12] ILP_{2,1} = \frac{\sum_{i=1}^N C_{i1} p_{i2}}{\sum_{i=1}^N C_{i1} p_{i1}}, \text{ Índice del año 2, base año 1}$$

$$[13] ILPE_{2,0} = \frac{\sum_{i=1}^N C_{i0} P_{i1}}{\sum_{i=1}^N C_{i0} P_{i0}} \cdot \frac{\sum_{i=1}^N C_{i1} P_{i2}}{\sum_{i=1}^N C_{i1} P_{i1}}, \text{ Índice encadenado del año 2,}$$

Nótese que el valor del Índice encadenado difiere del valor del Índice "directo", que hace uso reiterado de la composición de la muestra inicial de países.

$$[14] ILP_{2,0} = \frac{\sum_{i=1}^N C_{i1} P_{i2}}{\sum_{i=1}^N C_{i1} P_{i1}}, \text{ Índice directo del año 2.}$$

En particular, la condición para que el valor del Índice encadenado y del índice directo del año 2 deban coincidir:

$$[15] C_{i1} = C_{i0}, \text{ para todo } c_i (1...N)$$

es trivial - que la composición de la muestra de países no varíe -.

III.2.b Temas de *rankings*

También clasificamos las cuestiones de *rankings* en las de escalas, y las de muestras. Pero en este caso, tratamos primero b) y luego a).

b. 1. Diferentes muestras

La in(ex)clusión de países también afecta los *rankings*, por 2 vías: por extensión o contracción de la muestra, y según la posición de entrada en o salida, en la muestra, de los países.

La 1ª vía devela que la escala (Nº de países) tiene un techo "virtual" (la cantidad total de países) pero que en cada observación la escala de los *rankings* "absolutos" es el tamaño de la muestra.

Pero pasamos a tratar la 2ª vía. Como en el caso de los puntajes, el hecho de que la muestra de países del año 1 in(ex)cluya países que (no)estaban en la muestra de países del año 0 crea problemas para interpretar los cambios ocurrido en los *rankings* entre el año 0 y el año 1. ¿Cuánto de estos cambios se debe a cambios en los puntajes de los países y cuánto a la in(ex)clusión de países en la muestra?

La descomposición de los cambios en los niveles de los *rankings* absolutos

Para realizarla definimos: el cambio en la posición de un país en el *ranking* absoluto entre el año 0 y el año 1 atribuible exclusivamente a cambios de la posición del país es

la diferencia entre: su posición en el año 0 y su posición, usando la misma muestra de países, en el año 1.

[17] $dRRA_1 = RRA_{1/0} - RRA_{0/0}$, donde:

$dRRA_1$ = cambio en la posición de un país en el *ranking* absoluto entre el año 0 y el año 1 atribuible exclusivamente a cambios de la posición del país,

$RA_{1/0}$ = posición de un país en el *ranking* absoluto en el año 1, usando la misma muestra de países en el año 0,

$RA_{0/0}$ = posición de un país en el *ranking* absoluto en el año 0.

El cambio en la posición de un país en el *ranking* absoluto entre el año 0 y el año 1, atribuible puramente a cambios en la composición de la muestra de países es la diferencia (restante) entre el cambio (total) en la posición de un país en el *ranking* absoluto, y la parte de éste atribuible exclusivamente a cambios de la posición del país.

[18] $dCRA_1 = dRA_1 - dRRA_1 = RA_{1,1} - RA_{0,0} - dRRA_1$, donde:

$dCRA_1$ = cambio en la posición de un país en el *ranking* absoluto entre el año 0 y el año 1, atribuible puramente a cambios en la composición de la muestra de países,

$dRA_1 = RA_0 - RA_1$,

$RA_{1,1}$ = posición de un país en el *ranking* absoluto en el año 1, usando la muestra de países del año 1.

Esto permite computar qué porción del cambio en la posición del país en un periodo es atribuible exclusivamente a cambios de la posición del país y qué porción del cambio en la posición es atribuible puramente a la in(ex)clusión de países en la muestra.

[19] $SdRRA_1 = [(dRRA_1 / dRA_1) - 1]$, donde:

$SdRRA_1$ = porción del cambio en la posición del país atribuible exclusivamente a sus cambios de posición, y

[20] $SdCRA_1 = [(dCRA_1 / dRA_1) - 1]$, donde:

$SdCRA_1$ = porción del cambio en la posición del país atribuible a la in(ex)clusión de países en la muestra, respectivamente.

b. 2. Diferentes escalas

El uso de *rankings* absolutos tiene el inconveniente de su sensibilidad al tamaño de la muestra de países, que dificulta su interpretación y comparación (p.e.: un país que está en la mitad del *ranking* absoluto de una muestra de 40 países ocupa el puesto 20º; si está en la mitad del *ranking* absoluto de una muestra del 120 países ocupa el puesto 60º).

La transformación en *rankings* “percentiles”

Por ello, transformamos los datos de los *rankings* para que, en vez de representar posiciones absolutas (p.e.: 40º sobre 80 países), representen posiciones relativas - percentiles - (p.e.: $40/80 = 50\%$) porque así atenúamos, aunque no resolvemos, las distorsiones causadas por la in(ex)clusión de países. La escala de la posición relativa va del percentil 1% (mejor) al percentil 100% (peor).

Los cambios en los *rankings* percentiles pueden, a su vez, descomponerse en aquéllos atribuibles exclusivamente a cambios de la posición del país y aquéllos atribuibles puramente a la in(ex)clusión de países en la muestra empleando las ecuaciones [17] a [20].

[21] $dRRP_1 = SdRRA_1 \times dRRP_1$, donde:

$dRRP_1$ = Cambio en el RP atribuible exclusivamente a cambios de la posición del país, y

[22] $dCRP_1 = SdCRA_1 \times dCRP_1$, donde:

$dCRP_1$ = Cambio en el RP atribuible a la in(ex)clusión de países.

IV. Anomalías en estadísticas internacionales

Es justo comenzar este comentario sobre estadísticas internacionales con palabras de encomio, no sólo por el esfuerzo que conllevan y por la gran utilidad de sus resultados, sino también por su aporte de muchos años⁴¹ al método y la homogeneidad de su elaboración.

No obstante, al emplear series estadísticas internacionales para realizar comparaciones entre países hemos detectado anomalías sobre las cuales alertar. En su gran mayoría se refieren a datos vinculados con Argentina, el país que estudiamos, pero siembran dudas acerca de cifras relativas a otros países.

La mayoría de estos datos son provistos por organismos públicos internacionales y provienen de cifras oficiales de los países, en general miembros - aportantes, directivos - de esos organismos.

⁴¹ United Nations (1968).

Se presume que la calidad y la honestidad estadística del organismo nacional que las compila e informa es aceptable y uniforme en todos países, aunque de hecho los países con una infraestructura más(menos) desarrollada tienen mejores(peores) estimaciones y debido a que también sólo consideran ciertas partes del país.

A los organismos les resulta difícil - si no imposible -, cuando tienen dudas razonables sobre la calidad (en el sentido de reflejar la realidad) de los datos oficiales de un país, reemplazarlos por otros de distinto origen, pues eso les puede generar conflictos internos.

En la práctica, su opción es binaria: publicarlos o no hacerlo. O “desinforman” sobre un país y afectan la comparabilidad internacional de la muestra, o discontinúan la serie y alteran el tamaño y la composición de la muestra.

1. En algunos casos de datos “semi-blandos”, basados en los censos y encuestas oficiales de cada país que, como las estimaciones del “costo de vida” son sensibles políticamente, existe la tentación de los gobiernos de influir sobre su cálculo, sin informar siquiera cómo lo hacen. El BM ha detenido la publicación de la vieja serie del Producto Bruto Interno [**PBI**] ajustado por la Paridad de Poder Adquisitivo [PPA] en el 2006 y no ha incorporado la nueva serie.

2. La muestra de países con series de pobreza excluyen a los países con estándares de vida elevados, quizás porque en ellos la pobreza no es un problema, lo cual sesga el análisis comparativo (p.e.: los *rankings*).

3. En casos de datos “duros”, basados en los registros oficiales (p.e.: policiales) de cada país, como los de diferentes actividades delictivas, aparecen resultados llamativos, como que en algunos países considerados más seguros (*vgr.*: Suecia) son mayores que en otros vistos como menos seguros (*vgr.*: Honduras), lo que hace sospechar que no se trata de menor criminalidad en ésta, sino de menor denuncia o inscripción de delitos.

4. Otra razón por la cual pueden aparecer esos resultados llamativos es que las definiciones de los diferentes delitos varíen en el tiempo y de país a país. P.e.: series de robos.

5. También se dan casos en que las metodologías recomendadas por los organismos públicos internacionales no contemplan rasgos diferenciales de los países. P.e.: el concepto de Gasto Público varía de país a país, siendo estos federales o unitarios, incluyendo o excluyendo la asistencia a empresas del Estado, lo que imposibilita una buena comparación; para Argentina, se han encontrado múltiples medidas que van desde un 25% hasta un 45% del PBI.

6. No todas las anomalías provienen de datos suministrados por los países. Ciertas cifras parecerían no estar del todo revisadas y habría inconsistencias entre las informadas por el organismo nacional encargado de su compilación y el organismo internacional que los utiliza como parte de su base de datos para comparaciones entre países. P.e.: la tasa de homicidios intencionales de Argentina en 2008-09: 5,45 por cada 100m habitantes vs. 3,4 por cada 100m habitantes.

7. Otra cuestión que provoca anomalías, particularmente en series de largo plazo, son los cambios de base los organismos realizan cada tanto. P.e.: el BM y las NU calculan los PBI en dólares de 2005 utilizando las tasas de crecimiento obtenidas de las series en moneda nacional constante. Cuando se realiza en éstas un cambio de base y se recalcula el crecimiento hacia atrás no siempre se incorporan las modificaciones dando como resultado diferentes tasas de crecimiento para los mismos años.

8. Finalmente, las meras magnitudes de algunas provocan incredulidad, o la búsqueda de una explicación difícil de hallar. P. e.: el BM estima los *Adjusted Net National Savings* como porcentaje *del Gross National Income* de Argentina para 2012 en - 99.829 millones de dólares corrientes (- 21,4% del PBI).

También los datos provistos por entidades internacionales privadas invitan a veces a una duda razonable. P.e.: las prestigiosas *Penn World Tables* registran para Argentina (versión 8) un crecimiento del PBI del 29,3% de 1995 a 1996. Esto invita a usar sistemáticamente la prueba de detección de *outliers*, para considerarlos antes de emplearlos en comparaciones o ejercicios econométricos.

V. Bibliografía

- Berger, Peter L. y Luckmann, Thomas (1966) *The Social Construction of Reality*. 1ª edición en español (1968) La construcción social de la realidad. Amorrurtu (2001).
- Botana, Natalio (2012) "El contrapoder sale a la calle", La Nación, 20 de Septiembre.
- Bruner, Jerome; Goodnow, J. J. and Austin, G. A. (1967) *A Study of Thinking*, Science, New York City.
- Klaauw, Dieter (1965) "Über einen Ansatz zur mehrwertigen Mengenlehre", Monatsb. Deutsch. Akad. Wiss., Berlin 7.
- Lewis, David (1969) *Conventions: a Philosophical Study*, Cambridge University Press.
- Menger, Karl (1883) Untersuchungen über die Methode der Sozialwissenschaften und der politischen Ökonomie insbesondere.
- Roach, Eleanor (1978) "Principles of Categorization", Capítulo 8 en Roach y Lloyd (1978).
- Roach, Eleanor y Lloyd, Barbara B. (1978) *Cognition and Categorization*, IFA y Erlbaum, Hillsdale, NJ.
- Schank, Roger C. (1972) "Conceptual Dependence: A Theory of Natural Language Understanding", Cognitive Psychology, 3(4), Octubre.
- Schmöller, Gustav von (1883) "Zur Methodologie der Staats und Sozialwissenschaften", Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im deutschen Reich.
- The American Heritage (2007) Medical Dictionary, Houghton Mifflin.
- United Nations (1968) *A System of National Accounts*, New York.
- Zadeh, Lofti A. (1965) "Fuzzy Sets", Information and Control 8(3).
-