

Índice de Percepción de la Corrupción 2016

Nota técnica sobre la metodología utilizada

Antecedentes

El Índice de Percepción de la Corrupción (IPC) se creó en 1995 como un indicador compuesto utilizado para medir las percepciones sobre corrupción en el sector público, en distintos países de todo el mundo. Durante los últimos 18 años, tanto las fuentes empleadas para compilar el índice como la metodología se han ido modificando y perfeccionando. El último proceso de evaluación se efectuó en 2012¹ y se incorporaron algunos cambios importantes a la metodología. El método que se utilizó hasta 2012 para combinar las distintas fuentes de datos se ha simplificado, y ahora incluye los datos correspondientes a períodos de solamente un año de cada fuente de datos. Por sobre todo, este método nos permite ahora comparar las puntuaciones en el tiempo, una posibilidad que la metodología previa a 2012 no admitía.

Metodología

La metodología comprende 4 pasos básicos: seleccionar las fuentes de datos, ajustar la escala de las fuentes de datos, combinar las fuentes de datos y luego determinar una medida de incertidumbre.

1. Seleccionar las fuentes de datos

El IPC emplea una serie de fuentes disponibles que recogen las percepciones sobre corrupción. Cada fuente se evalúa teniendo en cuenta los criterios que se enumeran a continuación. Se ha contactado a cada una de las instituciones que aportan datos a fin de verificar la metodología que aplican al generar los puntajes y solicitar su autorización para publicar las puntuaciones de cada fuente que se ajustaron a la nueva escala, junto con la puntuación del índice compuesto.

A) Instituciones creíbles con metodologías y mecanismos de recolección de datos confiables: Es indispensable que confiemos en la validez de los datos que utilizamos. Por ende, cada fuente debe provenir de una institución profesional que documente claramente sus métodos de recolección de datos. Tales métodos deberán tener validez metodológica; por ejemplo, cuando se brinda una “consultoría especializada”, se deberá exigir que se demuestre la idoneidad del especialista, y cuando se realice una encuesta empresarial, la muestra de personas relevadas deberá ser representativa.

B) Datos vinculados con la corrupción en el sector público: La pregunta o el análisis debería estar relacionado con la percepción sobre el nivel de corrupción en el sector público de manera expresa. La pregunta puede estar vinculada con un “tipo” de corrupción determinado (p. ej., concretamente, los actos de corrupción menores) y, cuando corresponda, la efectividad de las acciones destinadas a prevenir la corrupción, ya que esto puede ser usado como indicador de la percepción sobre el nivel de corrupción en el país.

C) Precisión cuantitativa: Las escalas empleadas por las fuentes de datos deben permitir una diferenciación suficiente de los datos (p. ej., una escala de al menos cuatro puntos) sobre las percepciones de los niveles de corrupción correspondientes a distintos países, de modo que la información pueda adecuarse a la escala de 0 a 100 del IPC.

D) Comparabilidad entre países: Dado que el IPC clasifica a los países asignándoles una posición respecto de los demás, los datos de la fuente también deberán poder compararse válidamente entre distintos países y no ser exclusivos de un único país. En cada país al que se asigna una puntuación, la fuente debería medir un mismo aspecto, y en la misma escala.

E) Conjunto de datos correspondientes a varios años: Lo que procuramos es poder comparar la puntuación de un país, y obviamente el índice en general, de un año a otro. Por consiguiente, se

¹ La metodología empleada para calcular el IPC 2016 se basa en un trabajo de análisis de enfoques alternativos para la elaboración del IPC realizada por el Prof. Andrew Gelman del Departamento de Estadística y el Departamento de Ciencias Políticas de Columbia University, y por el Dr. Piero Stanig, becario del Instituto de Metodología de London School of Economics and Political Science. Este trabajo se presentó a Transparency International en un informe que se puede consultar a solicitud escribiendo a ssrinivasan@transparency.org.

excluyen aquellas fuentes que recogen las percepciones sobre corrupción en un momento histórico determinado y no han sido diseñadas para reiterarse en el tiempo.

2. Estandarizar las fuentes de datos

A continuación, se estandariza cada una de las fuentes para lograr que sean compatibles con otras fuentes disponibles, a efectos de volcarlas a la escala del IPC. La estandarización convierte todas las fuentes de datos a una escala de 0 a 100, donde 0 equivale al nivel más alto de percepción de corrupción, y 100 implica el nivel más bajo de percepción de corrupción.

Cuando en una fuente determinada la escala prevea que las puntuaciones inferiores corresponden a los niveles más bajos de corrupción, se deberán invertir los datos. Para ello, se debe multiplicar cada puntuación en el conjunto de datos por -1.

Luego, cada una de las puntuaciones se estandariza (para conseguir una puntuación z) restando la media de los datos y dividiendo el valor por la desviación estándar. Esto permite obtener un conjunto de datos que parten de cero, con una desviación estándar de 1.

Para que estas puntuaciones z resulten comparables entre distintos conjuntos de datos, debemos definir antes los parámetros globales sobre la media y la desviación estándar. Por consiguiente, cuando un conjunto de datos abarca un espectro limitado de países, atribuimos puntuaciones a todos los demás países que no aparecen en ese grupo de datos. Para aplicar los valores faltantes a los países que no están incluidos en cada conjunto de datos, utilizamos el paquete de software estadístico STATA y, más concretamente, el comando *ice* del programa. Este comando emplea regresiones múltiples con todos los conjuntos de datos disponibles para estimar los valores correspondientes a cada uno de los países sobre los cuales falta información en cada conjunto de datos individual. El comando efectúa la imputación varias veces (10), y genera 10 valores estimados para cada puntuación "faltante". La media y la desviación estándar para el conjunto de datos se calcula como un promedio obtenido de los 10 conjuntos de datos completos y se usa como parámetro para estandarizar los datos no procesados. Cabe remarcar que la imputación de datos se efectúa únicamente para generar estos parámetros y no se utiliza como fuente de datos para determinar las puntuaciones de los países en el IPC.

Fundamentalmente, las puntuaciones z se calculan utilizando los parámetros sobre la media y la desviación estándar de las puntuaciones imputadas a 2012. Esto permite que 2012 se convierta en el año de referencia para los datos y que las puntuaciones ajustadas según la nueva escala puedan compararse de un año a otro. Cuando se incorporen nuevas fuentes al índice, y a fin de reflejar adecuadamente los cambios en el tiempo, la aplicación del cálculo para adecuarlas a la nueva escala permite que sean coherentes con los parámetros de referencia de 2012. Para ello, primero se estima si se produjo un cambio global en la media y la desviación estándar desde 2012, y luego se utilizan estos nuevos valores,² que posiblemente se hayan desviado respecto de 50 y 20, para ajustar a escala el nuevo conjunto de datos².

Las puntuaciones z luego se redimensionan para adecuarse a la escala del IPC, de 0 a 100. Para esto, se aplica una fórmula simple que permite ajustar la escala, y que fija el valor medio del conjunto de datos estandarizados en aproximadamente 45 y la desviación estándar en 20 aproximadamente. Se ajustará cualquier puntuación que exceda el límite de 0 a 100.

3. Combinar los datos ajustados a la nueva escala

La puntuación de cada país en el IPC se calcula estableciendo un promedio simple de todas las puntuaciones disponibles sobre ese país que se ajustaron a la escala (cabe destacar que no utilizamos ninguno de los valores imputados como puntuación para el IPC agregado). Solamente se asignará una puntuación a un país cuando existan al menos tres fuentes de datos disponibles a partir de las cuales se pueda calcular este promedio.

4. Establecer una medida de incertidumbre

La puntuación en el IPC se muestra junto con el error típico y el intervalo de confianza que refleja la varianza en el valor de los datos que integran la puntuación del IPC.

El error típico se calcula como la desviación estándar de la fuente de datos ajustada a la nueva escala, dividida por la raíz cuadrada de la cantidad de fuentes. Aplicando este error típico, podemos calcular el intervalo de confianza del 90%, presuponiendo una distribución normal.

² Dado que se incorporó una nueva fuente de datos al IPC, se usó el procedimiento anterior para controlar si hubo un cambio en la media y la desviación estándar desde 2012. Determinamos que la media y la desviación estándar no habían cambiado y, por lo tanto, se mantuvo la comparación de un año a otro de las puntuaciones del IPC.