

# Índice de Percepción de la Corrupción 2018: Nota técnica sobre la metodología

## Antecedentes

El Índice de Percepción de la Corrupción (IPC) se creó en 1995 como un indicador compuesto para medir las percepciones sobre corrupción en el sector público en distintos países del mundo. Durante los últimos 20 años, tanto las fuentes empleadas para compilar el índice como la metodología se han ido modificando y perfeccionando. En 2012 se incorporaron cambios importantes a la metodología para que se pudieran comparar las puntuaciones en el tiempo<sup>1</sup>, algo que no era posible antes.

## Metodología

La metodología comprende cuatro pasos básicos: seleccionar las fuentes de datos, ajustar la escala de las fuentes de datos, combinar dichas fuentes una vez ajustadas y luego determinar una medida de incertidumbre. El proceso de cálculo incluye asimismo un estricto mecanismo de control de calidad que consiste en recopilar datos independientes paralelos y cálculos llevados a cabo por dos investigadores internos y dos asesores académicos que no pertenecen a Transparency International.

### **1. Seleccionar las fuentes de datos**

El IPC utiliza [13 fuentes de datos](#) que recogen la evaluación, por parte de expertos y ejecutivos de empresas, sobre una variedad de comportamientos corruptos en el sector público, como los siguientes:

- Soborno
- Desvío de fondos públicos
- Uso de la función pública para beneficio personal
- Nepotismo en la administración pública
- Captura del Estado

Algunas de las fuentes consideran también los mecanismos que existen para prevenir la corrupción en un país, como por ejemplo:

---

<sup>1</sup> La metodología usada para calcular el IPC 2018 toma como base el estudio de enfoques alternativos realizado para generar el IPC por el profesor Andrew Gelman (profesor, Departamento de Estadísticas y Departamento de Ciencias Políticas, Universidad de Columbia), y el Dr. Piero Stanig (profesor visitante, Instituto de Metodología, London School of Economics and Political Science). Este trabajo se presentó a Transparency International en un informe que puede consultarse a pedido. Enviar solicitud por correo electrónico a [cpi@transparency.org](mailto:cpi@transparency.org).

- La capacidad del gobierno de hacer cumplir los mecanismos de integridad
- El juzgamiento efectivo de funcionarios corruptos
- Trabas administrativas y requisitos burocráticos excesivos
- La existencia de leyes adecuadas sobre divulgación financiera, prevención de conflictos de intereses y acceso a la información
- Protección legal para denunciantes, periodistas e investigadores

Cada una de las fuentes de datos utilizadas para calcular el IPC se evalúa en función de los siguientes criterios:

- A) Fiabilidad metodológica y prestigio institucional:** Para que se incluya una fuente en el IPC, es necesario comprobar la calidad y la pertinencia de su enfoque metodológico. Por ese motivo, cada fuente debe originarse en una institución profesional que documente de manera clara cuáles son sus métodos de recopilación de datos y su criterio de medición. A continuación, Transparency International evalúa la seriedad de la metodología. Por ejemplo, si se proporciona la “opinión de un experto”, es necesario constatar la idoneidad del experto. En cuanto a las encuestas empresariales, Transparency International analiza si la muestra tomada para la encuesta es lo suficientemente grande como para ser representativa.
- B) Alineación conceptual de los datos:** Ya que se trata de una medición de la corrupción en el sector público, todas las fuentes de datos utilizadas para preparar el IPC deben estar directamente asociadas con los niveles de corrupción en ese sector. Las preguntas pueden estar vinculadas con un “tipo” de corrupción determinado (p. ej., concretamente, los actos de corrupción menores) o con la eficacia de los mecanismos de prevención de la corrupción, y esto a su vez puede usarse como indicador del nivel de corrupción que se percibe en un país.
- C) Precisión cuantitativa:** Las escalas empleadas por las fuentes deben permitir una diferenciación suficiente de los datos (es decir, 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. Asimismo, la fuente debe medir el mismo concepto entre los distintos países y aplicando la misma escala. Si bien no hay actualmente criterios relativos a cuál debe ser la cobertura mínima de una fuente para formar parte del IPC, la fuente de datos con la menor cobertura asigna puntuaciones a un total de 16 países.
- E) Disponibilidad a lo largo de varios años:** Dado que se supone que el IPC mide la corrupción en distintos países y momentos, se excluyeron las fuentes que captan las percepciones de corrupción en un único momento histórico y no fueron diseñadas para repetirse en el tiempo.

A fin de cumplir este proceso de control de calidad, Transparency International se contacta con cada una de las instituciones que proporcionan datos, con el fin de verificar la metodología que emplearon para generar sus puntuaciones. Como algunas de las fuentes no están disponibles en forma pública, Transparency

International solicita autorización para publicar las puntuaciones ajustadas a la nueva escala junto con la puntuación compuesta del IPC. No obstante, no se permite que Transparency International difunda al público general las puntuaciones originales asignadas por fuentes privadas.

## 2. Estandarizar las fuentes de datos

Cada una de las fuentes incluidas en el IPC se estandarizó para permitir que sea incorporada a la puntuación del IPC. La estandarización convierte todos los puntos 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. Aunque la mayoría de las fuentes subyacentes del IPC también están codificadas de la misma forma (las puntuaciones más bajas implican mayores niveles de corrupción percibida), hay cuatro fuentes que tienen una escala en sentido contrario, es decir, las puntuaciones más bajas denotan niveles más bajos de corrupción. Para permitir la comparación, estas cuatro fuentes se invierten multiplicando cada puntuación por -1.

Las fuentes que se invierten son las siguientes:

- Indicador de corrupción de Economist Intelligence Unit
- Puntaje de corrupción de Naciones en Transición de Freedom House
- Puntuación sobre percepciones de corrupción de Asian Intelligence de Political and Economic Risk Consultancy
- Índice de corrupción política del Proyecto Variedades de Democracia

Dado que muchas de las fuentes utilizadas para el IPC no tienen alcance global, los valores faltantes para estas fuentes se imputan al año de referencia<sup>2</sup>. Este proceso se lleva a cabo usando el paquete de software estadístico STATA y, más específicamente, el comando “imputar” del programa. El comando “imputar” de STATA calcula un valor para cada punto de datos faltante, usando únicamente las fuentes de datos donde estén al menos el 50 % del total de los países alcanzados por el IPC en un año determinado<sup>3</sup>. Esto no se aplica para los datos del Índice de

---

<sup>2</sup> Hasta 2016, se usaban los parámetros globales de 2012 como valor de referencia. Con la incorporación de los datos de VDEM al IPC en 2016, los valores faltantes para ese año se imputaron nuevamente. Sin embargo, una comparación de los parámetros globales de 2012 y 2016 resultó ser insignificante en términos estadísticos, lo que significa que el cambio en el valor de referencia no afectó la comparabilidad en el tiempo. En 2017, VDEM amplió la cantidad de países cubiertos a más del 50 % de los países del IPC. Por este motivo, los parámetros globales debieron ser recalculados una vez más. Los cambios en la media global y la desviación estándar tampoco ahora fueron estadísticamente significativos. Como resultado, el año de referencia efectivo sigue siendo el 2012.

<sup>3</sup> Conforme a la regla de un 50 % de países alcanzados mencionada antes, las siete fuentes de datos siguientes se usaron para el proceso de imputación del IPC 2017:

- Índice de Transformación 2018 de Bertelsmann Foundation (62,32 % de cobertura de la muestra)
- Calificaciones de riesgo país de la Economist Intelligence Unit de 2017 (63,29 % de cobertura de la muestra)
- Calificaciones de Riesgo País de Global Insight (98,55 % de cobertura de la muestra)
- Guía Internacional sobre Riesgo País de Political Risk Services (67,63 % de cobertura de la muestra)
- Encuesta de Opinión Ejecutiva del Foro Económico Mundial (64,42 % de cobertura de la muestra)

Transformación de Bertelsmann Foundation, que no se utilizan para imputar los Indicadores sobre Gobernabilidad Sostenible de esa fundación.

Tras el proceso de imputación, la media y la desviación estándar para cada fuente de datos se calculan y utilizan como parámetros para estandarizar los datos originales en puntuaciones z. Las puntuaciones z estandarizadas se calculan restando la media de cada fuente a la puntuación de cada país y dividiendo el valor por la desviación estándar de la respectiva fuente. Esto tiene como resultado una serie de datos centrados en el valor 0 y con una desviación estándar de 1. Un dato fundamental es que las puntuaciones z se calculan usando los parámetros sobre media y desviación estándar de las fuentes imputadas del año de referencia. Esto asegura que las puntuaciones del IPC sean comparables de un año a otro, a partir del año de referencia.

Es importante destacar que la serie de datos completa con valores imputados se usa únicamente para generar los parámetros globales de línea de base y los valores imputados propiamente dichos no se usan en el cálculo final que genera las puntuaciones del IPC.

Las puntuaciones z luego se transforman para adecuarse a la escala de 0 a 100 del IPC. Para esto, se aplica una fórmula que permite ajustar la escala, y cuyo objetivo es fijar el valor medio del conjunto de datos estandarizados en 45 y la desviación estándar en 20. Para tal fin, se utiliza la siguiente fórmula:

$$\begin{aligned} & \text{Rescaled Indicator } X_{\text{Country}} \\ &= \frac{\text{Original Indicator } X_{\text{Country}} - \text{mean imputed Indicator } X^{t=\text{baseline}}}{\text{standard deviation imputed Indicator } X^{t=\text{baseline}}} * 20 + 45 \end{aligned}$$

Todas las puntuaciones redimensionadas que aceptan valores de menos de 0 se equiparan a 0, y a las puntuaciones redimensionadas que superan el valor de 100 se les impone un máximo de 100.

### 3. Combinar los datos ajustados a la nueva escala

La puntuación de cada país en el IPC se calcula como un promedio simple de todas las puntuaciones redimensionadas para ese país. (Cabe señalar que no usamos 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 mostrará junto con el error típico y el intervalo de confianza del 90 % que refleja la varianza en el valor de los datos que integran la puntuación del IPC. Asimismo, se identifican y notifican los cambios registrados en las puntuaciones del IPC que sean estadísticamente significativos.

Siguiendo la recomendación de la auditoría del [Centro Común de Investigación de la Comisión Europea](#) sobre la coherencia conceptual y estadística del IPC, se modificó la fórmula de cálculo de los errores típicos a partir del IPC 2018. La fórmula anterior calculaba el error típico como la desviación estándar de la fuente de datos ajustada a la nueva escala, dividida por la raíz cuadrada de la cantidad de fuentes. La nueva fórmula, expresada por la ecuación siguiente, tiene en cuenta una pequeña población de fuentes:

$$\Sigma = \sqrt{\frac{N-n}{N-1} \frac{\sigma}{\sqrt{n}}}$$

donde  $\Sigma$  es el error típico,  $N$  es el número total de fuentes utilizadas para calcular el IPC,  $n$  es el número de fuentes empleadas para calcular la puntuación del IPC de cualquier país dado y  $\sigma$  es la desviación típica de la puntuación del IPC para ese país.<sup>1</sup>

Aplicando este error típico, podemos calcular el intervalo de confianza del 90% y obtener los límites superior e inferior de la puntuación del IPC para cada país, presuponiendo una distribución normal.

Después de calcular los errores típicos, determinamos si la variación de la puntuación del IPC es estadísticamente significativa o no. Para ello, en primer lugar calculamos la dimensión del efecto de la diferencia entre la puntuación de un país en los dos años comparados utilizando la fórmula siguiente:

$$effect\ size = \frac{M_1 - M_2}{\sqrt{\frac{(N_1 - 1)SD_1^2 + (N_2 - 1)SD_2^2}{N_1 + N_2 - 2}}}$$

donde  $M_1$  es la puntuación del IPC para un país dado en un año determinado,  $M_2$  es la puntuación del IPC para ese mismo país en el año de comparación,  $N_1$  es el número de fuentes en el año dado,  $N_2$  es el número de fuentes en el año de comparación,  $SD_1^2$  es el cuadrado de la desviación estándar de la puntuación del IPC en el año dado y  $SD_2^2$  es el cuadrado de la desviación estándar de la puntuación del IPC en el año de comparación.

Una vez obtenida la dimensión del efecto, procedemos a calcular su desviación estándar ( $\sigma$ ) empleando la siguiente fórmula:

$$\sigma(d) = \sqrt{\frac{N_1 + N_2}{N_1 \times N_2} + \frac{d^2}{2 \times (N_1 + N_2)}}$$

donde  $d$  es la dimensión del efecto de las diferencias en las puntuaciones del IPC, y  $N_1$  y  $N_2$  son el número de fuentes disponibles para cada país. A continuación se calcula el intervalo de confianza del 90% mediante la fórmula siguiente:

$$CI_{0,90} = d \pm 1.96 \times \sigma(d)$$

Si este intervalo de confianza incluye el valor cero, no existen diferencias estadísticamente significativas entre los países. Si, por el contrario, el cero queda fuera del intervalo, la diferencia es estadísticamente significativa al nivel del 10%.

---

<sup>1</sup> En 2018, la población total de fuentes (N) fue de 13.