Nota Metodológica

Índice de Percepção da Corrupção (IPC)

Nota metodológica

Histórico

O Índice de Percepção da Corrupção (IPC) foi criado em 1995 como um indicador composto utilizado para medir as percepções de corrupção no setor público em diferentes países ao redor do globo. Nos últimos 20 anos, tanto as fontes usadas para compor o índice quanto a metodologia dele têm sido reajustadas e refinadas. Em 2012, importantes mudanças foram feitas para que fosse possível comparar as notas ao longo do tempo[1] – algo que passou a ser feito a partir daquele ano.

Metodologia

A metodologia do IPC segue quatro etapas básicas: seleção dos dados, redimensionamento dos dados, agregação dos dados redimensionados e por fim uma mensuração de incerteza. O cálculo também incorpora um rígido controle de qualidade, que consiste em cálculos paralelos independentes feitos por dois pesquisadores internos e dois consultores acadêmicos externos, sem relação com a Transparência Internacional.

1. Seleção de Fontes de Dados:

O IPC utiliza 13 fontes de dados que avaliam a percepção de especialistas e executivos sobre determinados comportamentos corruptos no serviço público, incluindo:

- Suborno;
- Desvio de dinheiro público;
- Uso do cargo público para ganho privado;
- Nepotismo no serviço civil;
- Captura do Estado

[1] A metodologia usada para calcular o IPC se baseia no trabalho que avalia abordagens alternativas de construção do IPC, conduzido por Andrew Gelman, professor do Departamento de Estatística e do Departamento de Ciência Política da Universidade de Columbia; e de Piero Stanig, pesquisador do Instituto de Metodologia da London School of Economics and Political Science. Este trabalho foi apresentado à Transparência Internacional em um documento disponível mediante solicitação. Por favor, mande um e-mail para cpi@transparency.org.

Algumas dessas fontes também olham para os mecanismos disponíveis de prevenção da corrupção em determinado país, como:

- A capacidade do governo de aplicar seus mecanismos de integridade;
- A responsabilização penal efetiva de agentes públicos corrutos;
- Entraves burocráticos e carga burocrática excessiva;
- A existência de leis adequadas sobre divulgação de informações financeiras, prevenção de conflitos de interesse e acesso à informação;
- Proteção legal para denunciantes, jornalistas e investigadores.

Cada uma das fontes de dados usadas para calcular o IPC é avaliada de acordo com os seguintes critérios:

A) **Confiabilidade metodológica e reputação institucional:** Para uma fonte ser incluída no IPC, é necessário que ela garanta a qualidade e adequação de sua abordagem metodológica. Por essa razão, cada fonte deve ser proveniente de uma instituição profissional que registre claramente a forma como coleta seus dados e sua abordagem de mensuração. A Transparência Internacional, então, avalia a solidez da metodologia.

B) **Alinhamento conceitual dos dados:** Como se trata de uma medida da corrupção no setor público, todas as fontes de dados utilizadas na construção do IPC devem estar explicitamente ligadas aos níveis de corrupção ou os riscos de corrupção no setor público. As perguntas podem estar relacionadas a um determinado tipo de corrupção (por exemplo, a "corrupção de pequena escala") ou à efetividade dos mecanismos de prevenção da corrupção, que também podem ser usados como um *proxy* para o nível de percepção de corrupção em um país.

C) **Granularidade quantitativa:** As escalas usadas pelas fontes de dados devem permitir diferenciação suficiente dos dados (ou seja, numa escala de pelo menos quatro pontos) nos níveis percebidos de corrupção entre diferentes países, permitindo que possam ser reescalonadas para uma escala de 0 a 100 do IPC.

D) **Comparabilidade entre países:** Como o IPC classifica os países em relação uns aos outros, os dados das fontes também devem ser também legitimamente comparáveis entre países e não especificos a um único contexto nacional. Além disso, a fonte deve medir o mesmo conceito nos países e com a mesma escala. Embora atualmente não exista um critério

mínimo exigido para que uma fonte seja incorporada ao IPC, a fonte de dados com menor cobertura fornece pontuações para um total de 16 países.

E) **Disponibilidade de dados ao longo de múltiplos anos:** Como o IPC mede a corrupção entre países e ao longo do tempo, são excluídas as fontes que captam percepções de corrupção em um único ponto no tempo e que não foram concebidas para serem repetidas ao longo dos anos.

Com o objetivo de realizar esse processo de garantia de qualidade, a Transparência Internacional entra em contato com cada uma das instituições fornecedoras de dados a fim de verificar a metodologia utilizada para a geração de suas pontuações. Como algumas das fontes de dados não estão disponíveis publicamente, a Transparência Internacional também solicita autorização para publicar as pontuações reescalonadas de cada fonte, juntamente com aa pontuação composta do IPC. A Transparência Internacional, porém, não tem permissão para divulgar ao público geral as pontuações originais fornecidas por fontes privadas.

2. Padronização das fontes de dados

Cada uma das fontes incluídas no IPC é padronizada para permitir sua agregação na pontuação do IPC. A padronização converte todos os pontos de dados para uma escala de 0 a 100, em que 0 representa o maior nível percebido de corrupção e 100 o menor nível percebido de corrupção.

Revisão da direção das fontes e conversão daquelas que são opostas

Embora a maioria das fontes que compõem o IPC seja codificada na mesma direção (com menores pontuações indicando níveis mais elevados de percepção de corrupção), três fontes de dados são escalonadas na direção oposta – ou seja, com pontuações menores indicando menores níveis de corrupção. Para fins de comparabilidade, essas fontes de dados são convertidas por meio da multiplicação de cada pontuação por –1.

As fontes que precisam ser convertidas são:

- O indicador de corrupção da *Economist Intelligent Unit;*
- A pontuação de percepções de corrupção da *Political and Economic Risk Consultanty Asian Intelligence;*
- O índice de corrupção política do *Varieties of Democracy Project* (V-Dem).

Obtenção da média e do desvio padrão do ano–base t

Como muitas das fontes utilizadas no IPC não possuem cobertura global, os valores ausentes são imputados para o ano-base t. O ano-base t corresponde ao ano mais antigo possível em que a cobertura e a metodologia das fontes não sofreram alterações. Antes do ano corrente, o ano-base t era 2017. Em 2024, o ano-base t foi alterado em razão de um novo método de agregação metodológica adotado pelo Fórum Econômico Mundial em seu *Forum Executive Survey*. Os novos parâmetros são estatisticamente idênticos aos originais de 2012, o que assegura a comparabilidade ao longo do tempo, mas afeta a comparabilidade das fontes subjacentes.

O processo de imputação é realizado por meio do software estatístico STATA , mais especificamente utilizando o seu comando "impute". Esse comando emprega regressões lineares para estimar um valor para cada ponto do dado ausente, utilizando apenas aquelas fontes de dados que cobrem ao menos 50% do número total de países incluídos no IPC no ano-base t[2]. A única exceção a esse procedimento são os dados do índice de transformação da Fundação Bertelsmann, que não s]ao utilizados para a imputação dos Indicadores de Governança Sustentável da Fundação Bertelsmann.

[2] Conforme a regra de cobertura de 50% dos países mencionada acima, as seguintes sete fontes de dados foram utilizadas para o processo de imputação do IPC 2024:

Projeto Varieties of Democracy 2024 (n= 178)
Índice de Transformação 2014 da Fundação Bertelsmann 2014 (n = 137)
Notas 2014 da Economist Intelligence Unit Country Risk (n = 131)
Notas da S&P / Global Insights Country Risk (n = 207)
Guia 2024 do Political Risk Services International Country Risk (n = 141)
Pesquisa de Opinião entre Executivos 2024 do Fórum Econômico Mundial (n = 117)
Índice de Estado de Direito 2017-2018 do World Justice Project (n=142)

Após a conclusão do processo de imputação, são calculados a média e o desvio padrão de cada fonte de dados no ano-base t, que passam a ser utilizados como parâmetros para a padronização dos dados originais em pontos z.

Padronização dos dados em pontos z com média 0 e desvio padrão de 1

Esse processo é realizado subtraindo-se a média do ano-base de cada fonte da pontuação original de cada país e dividindo-se o resultado pelo desvio padrão do ano-base da respectiva fonte. Como resultado, obtém-se um conjunto de dados centrados em 0 e com desvio padrão igual a 1.

É importante observar que o conjunto completo de dados com valores imputados é utilizado apenas para a geração dos parâmetros do ano-base 7, não sendo os valores imputados empregados na agregação final que resulta nas pontuações do IPC.

Reescalonamento dos dados para a escala 0 a 100

Os pontos z são então transformados para se ajustarem à escala de 0 a 100 do IPC. Esse procedimento é realizado por meio de uma fórmula de reescalonamento que fixa a média global do conjunto de dados padronizados em 45 e o desvio padrão global em 20, parâmetros definidos para o ano de referência de 2012[3]. Isso garante que as pontuações do IPC sejam comparáveis ano a ano desde o ano de referência. A seguinte fórmula é utilizada para o cálculo dos indicadores reescalonados:

$$\frac{Original\ Indicator X_{Country} - mean\ imputed\ Indicator X^{t=baseline}}{standard\ \ deviation\ imputed\ Indicator X^{t=baseline}} * 20 + 45$$

[3] Até 2016, os parâmetros globais de 2012 foram usados como linha de base. Com a introdução dos dados da VDEM no IPC em 2016, os dados faltantes para 2016 foram imputados novamente. A comparação entre os parâmetros globais de 2012 e 2016, no entanto, se provou estatisticamente insignificante, o que sginifica que a mudança na linha de base não afetou a comparabilidade entre os diferentes anos. Em 2017, a VDEM expandiu a cobertura de países para mais de 50% dos países avaliados no IPC. Por essa razão, os parâmetros globais tiveram que ser recalculados de novo. Em 2024, por conta de mudanças na metodologia de agregação dos dados da WEF para incluir controles de qualidade adicionais, os parâmetros tiveram que ser calculados mais uma vez, mas as mudanças na média global e no desvio padrão, mais uma vez, foram estatisticamente insignificantes. Como resultado, efetivamente o ano referência continua sendo 2012.

Quaisquer pontuações reescalonadas que assumam valores inferiores a 0 são ajustadas para 0, e quaisquer pontuações reescalonadas que excedam 100 são limitadas a 100.

3. Agregação dos dados reescalonados

A pontuação do IPC de cada país é calculada como a média simples de todas as pontuações reescalonadas para aquele país. (Perceba que não utilizamos os valores imputados como pontuação na agregação final do IPC). Um país somente receberá uma pontuação caso existam pelo menos três fontes de dados disponíveis para o cálculo dessa média.

4. Apresentação de uma medida de incerteza

A pontuação do IPC é divulgada juntamente com o erro padrão e um intervalo de confiança de 90%, que refletem a variância dos valores das fontes de dados que compõem a pontuação do IPC. Além disso, são identificadas e reportadas as variações nas pontuações do IPC que sejam estatisticamente significantes.

Em conformidade com a recomendação da auditoria do *Joint Research Centre Centro de Pesquisas Comuns da Comissão Europeia* sobre a coerência conceitual e estatística do IPC, a fórmula para o cálculo dos erros padrão foi alterada a partir do IPC de 2018. A fórmula anterior calculava os termos de erros padrão como o desvio padrão dos dados reescalonados das fontes, dividido pela raiz quadrada do número de fontes. A nova fórmula, expressa na equação abaixo, leva em consideração o tamanho reduzido da população de de fontes:

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

Onde Σ representa o termo de erro padrão, N corresponde ao número total de fontes utilizadas no cálculo do IPC, n refere-se ao número de fontes

consideradas para a pontuação do IPC de um determinado país e σ indica o desvio padrão da pontuação do IPC desse país.[4]

Usando este erro padrão Σ, é possível calcular o intervalo de confiança de 90% utilizando o ponto z 1.645 e reportar os limites superior e inferior da pontuação do IPC para cada país, assumindo uma distribuição normal.

$$CI_{CPIx=country\ 0.90} = CPI_{x=country} \pm 1.645 \times \Sigma$$

As pontuações do IPC são estimativas pontuais e são apresentadas juntamente com seus intervalos de confiança de 90%. Ao interpretar a pontuação de um determinado país, devem ser considerados os limites inferior e superior desse intervalo. Com 90% de confiança, o valor populacional da percepção de corrupção de um país situa-se entre os limites inferior e superior apresentados.

Após o cálculo dos erros padrão, é determinado se a variação na pontuação do IPC de cada país é estatisticamente significante ou não. Para esse fim, calcula-se inicialmente o tamanho do efeito por meio da fórmula d de Cohen, aplicada à diferença entre a pontuação de um país nos dois anos comparados:

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

Onde $M1$ representa a pontuação do IPC de um determinado país em um ano específico , $M2$ corresponde à pontuação do IPC desse país no ano de comparação, $N1$ indica o número de fontes no ano considerado, $N2$ refere-se ao número de fontes no ano de comparação, SD^2_1 corresponde ao quadrado do desvio padrão da pontuação do IPC no ano considerado e SD^2_2 representa quadrado do desvio padrão da pontuação do IPC no ano de comparação.

Uma vez obtido o tamanho do efeito, calcula-se o desvio padrão ($\boldsymbol{\sigma}$) do tamanho do efeito utilizando o estimador não-viesado de Hedges para a variância do tamanho do efeito:

[4] Em 2024, o número total de fontes foi 13 (N)

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

Onde d corresponde ao tamanho do efeito das diferenças nas pontuações do IPC, e N_1 e N_2 representam o número de fontes disponíveis para cada país. Em seguida, o intervalo de confiança de 90% é calculado utilizando a fórmula abaixo:

$$CI_{0,90} = d \pm 1,645 \times \sigma(d)$$

Se esse intervalo de confiança incluir o valor zero, não há diferenças estatisticamente significativas entre os países. Por outro lado, se o zero estiver fora do intervalo, a diferença é considerada estatisticamente significanta ao nível de 10%. Esse critério aplica-se tanto às diferenças entre anos para um mesmo país quanto às diferenças entre pares de países. A frequência de diferenças significativas ao longo do tempo depende da janela temporal considerada. É comum observar poucas ou nenhuma diferença estatisticamente significativa entre dois anos consecutivos no IPC; no entanto, quando analisadas ao longo de períodos mais extensos, tornam-se perceptíveis mais informações e mudanças estatisticamente significativas.