

Indice de perception de la corruption 2018 : Note technique de méthodologie

Contexte

L'indice de perception de la corruption (IPC) a été établi en 1995 en tant qu'indicateur composite utilisé pour mesurer les perceptions de la corruption dans le secteur public dans différents pays du monde. Au cours des 20 dernières années, les sources retenues pour constituer l'indice et la méthodologie ont été ajustées et affinées. En 2012, la méthodologie a fait l'objet de profonds changements pour permettre d'effectuer des comparaisons de notes dans le temps¹, ce qui n'avait pas été possible avant 2012.

Méthodologie

La méthodologie suit quatre étapes fondamentales : la sélection des données sources, le rééchantillonnage des données sources, l'agrégation des données rééchantillonnées, puis la déclaration d'une mesure d'incertitude. Le processus de calcul intègre également un mécanisme strict de contrôle de la qualité qui consiste en une collecte parallèle de données indépendantes et en des calculs effectués par deux chercheurs internes et deux conseillers universitaires sans affiliation à Transparency International.

1. Sélection des sources de données

L'IPC s'appuie sur [13 sources de données](#) qui saisissent l'évaluation par des experts et des chefs d'entreprise d'un certain nombre de comportements corrompus dans le secteur public, y compris :

- Le versement de pots-de-vin
- Le détournement de fonds publics
- L'utilisation de la fonction publique à des fins privées
- Le népotisme dans la fonction publique
- La captation de l'État

Certaines sources examinent également les mécanismes disponibles pour prévenir la corruption dans un pays, tels que :

- La capacité du gouvernement à faire appliquer des mécanismes d'intégrité
- La poursuite en justice efficace des fonctionnaires corrompus
- La paperasserie et le fardeau bureaucratique excessif

¹La méthodologie utilisée pour calculer l'IPC 2018 s'appuie sur le travail d'examen d'autres approches de constitution de l'IPC réalisé par le Prof. Andrew Gelman : professeur au Département de statistique et Département de science politique, Université de Columbia et le Dr Piero Stanig : Membre de l'Institut de méthodologie, London School of Economics and Political Science. Ce travail a été présenté à Transparency International dans un rapport disponible sur demande. Veuillez adresser un e-mail à cpi@transparency.org.

- L'existence de lois pertinentes sur la divulgation financière, la prévention des conflits d'intérêts et l'accès aux informations
- La protection juridique des lanceurs d'alerte, des journalistes et des enquêteurs

Chacune des sources de données utilisées pour calculer l'IPC est évaluée en fonction des critères suivants :

- A) Fiabilité méthodologique et réputation institutionnelle** : pour qu'une source soit incluse dans l'IPC, il est nécessaire d'assurer la qualité et la fiabilité de son approche méthodologique. Pour cette raison, chaque source doit provenir d'une institution professionnelle qui explique clairement, documents à l'appui, ses méthodes de collecte de données et son approche de mesure. Transparency International évalue ensuite la fiabilité de la méthodologie. Dans le cas, par exemple, d'un « avis d'expert », il est nécessaire de s'assurer de l'authenticité des qualifications de cet expert. En ce qui concerne les enquêtes auprès des entreprises, Transparency International examine si l'échantillon de l'enquête est suffisamment important pour être représentatif.
- B) Alignement conceptuel des données** : comme il s'agit d'une mesure de la corruption dans le secteur public, toutes les sources de données retenues pour constituer l'IPC doivent être explicitement liées aux niveaux de corruption présente dans le secteur public. Les questions peuvent se rapporter à un « type » défini de corruption (par exemple, la petite corruption) ou à l'efficacité des mécanismes de prévention de la corruption, qui peuvent également être utilisés comme indicateur du niveau de corruption perçue dans un pays.
- C) Niveau de granularité** : les échelles sur lesquelles s'appuient les sources de données doivent permettre une différenciation suffisante des données (c'est-à-dire au moins une échelle de quatre points) sur les niveaux perçus de corruption entre les divers pays afin de pouvoir les rééchelonner sur une échelle de 0 à 100 de l'IPC.
- D) Comparabilité entre les pays** : du fait que l'IPC classe les pays les uns par rapport aux autres, les données des sources doivent également être légitimement comparables entre les pays et ne pas être spécifiques à un pays en particulier. De plus, la source doit mesurer le même concept dans l'ensemble des pays et selon la même échelle. Bien qu'il n'existe actuellement aucun critère concernant la couverture minimale qu'une source doit avoir pour faire partie de l'IPC, la source de données ayant la couverture la plus faible attribue des notes à un total de 16 pays.
- E) Disponibilité des données sur plusieurs années** : étant donné que l'IPC vise à mesurer la corruption dans l'ensemble des pays et dans la durée, les sources qui saisissent les perceptions de la corruption à un moment donné, mais qui ne sont pas conçues pour être répétées au fil du temps, sont exclues.

Pour mener à bien ce processus d'assurance qualité, Transparency International se met en rapport avec chacune des institutions fournissant des données, afin de vérifier la méthodologie qu'elles utilisent pour générer leurs notes. Certaines des sources n'étant pas accessibles au public, Transparency International demande également la permission de publier les notes rééchelonnées de chaque source en même temps que la note composite de l'IPC. Transparency International n'est cependant pas autorisé à partager avec le grand public les notes d'origine fournies par des sources privées.

2. Standardiser les sources de données

Chacune des sources incluses dans l'IPC est standardisée pour permettre l'agrégation dans la note de l'IPC. La standardisation convertit tous les points de données à une échelle de 0 à 100, où 0 représente le niveau le plus élevé de corruption perçue et 100 le niveau le plus bas de corruption perçue. Même si la plupart des sources de l'IPC sous-jacentes sont également codées dans le même sens (les notes les plus basses indiquant des niveaux plus élevés de corruption perçue), les échelles utilisées pour quatre sources vont dans le sens opposé, c'est-à-dire que les notes les plus basses indiquent des niveaux inférieurs de corruption. À des fins de comparabilité, ces quatre sources sont inversées en multipliant chaque note par -1.

Les sources qui doivent être inversées incluent :

- l'indicateur de corruption de *The Economist Intelligence Unit*
- la note de corruption de *Nations in Transit* de *Freedom House*
- la note de perception de corruption de *Political and Economic Risk Consultancy Asian Intelligence*
- l'indice de corruption politique de *Varieties of Democracy Project*

Puisque de nombreuses sources utilisées pour l'IPC n'ont pas de couverture mondiale, les valeurs manquantes de ces sources sont imputées pour l'année de référence². Ce processus est réalisé à l'aide du progiciel statistique STATA et, plus particulièrement, de la commande « *impute* » du programme. Cette commande de STATA estime une valeur correspondant à chaque point de données manquant en utilisant uniquement les sources de données représentant au moins 50 % du nombre total de pays couverts par l'IPC au cours d'une année donnée³. À l'exception des données de l'indice de transformation de la Fondation Bertelsmann, qui ne sont pas utilisées pour l'imputation des indicateurs de gouvernance durable de cette même fondation.

Après le processus d'imputation, la moyenne et l'écart type pour chaque source de données sont calculés et utilisés comme paramètres pour standardiser les données

² Jusqu'en 2016, les paramètres mondiaux de 2012 ont été retenus comme référence. Avec l'introduction des données VDEM dans l'IPC en 2016, les valeurs manquantes pour 2016 ont été imputées à nouveau. Une comparaison des paramètres mondiaux de 2012 et 2016 s'est toutefois révélée non significative d'un point de vue statistique, ce qui signifie que le changement de référence n'a pas affecté la comparabilité dans le temps. En 2017, VDEM a étendu sa couverture nationale à plus de 50 % des pays couverts dans l'IPC. Pour cette raison, les paramètres mondiaux ont dû être recalculés une fois de plus. Les changements survenus dans la moyenne mondiale et l'écart-type n'étaient pas non plus significatifs d'un point de vue statistique. Par conséquent, l'année de référence effective demeure 2012.

³ Conformément à la règle de couverture de 50 % des pays mentionnée ci-dessus, les sept sources de données suivantes ont été utilisées pour le processus d'imputation pour l'IPC 2017:

- l'indice de transformation 2018 de la Fondation Bertelsmann (couverture de l'échantillon de 62,32 %)
- évaluations des risques des pays 2017 d'Economist Intelligence Unit (couverture de l'échantillon de 63,29 %)
- évaluations des risques des pays de Global Insights (couverture de l'échantillon de 98,55 %)
- guide des risques des pays de Political Risk Services International (couverture de l'échantillon de 67,63 %)
- sondage d'opinion des cadres 2017 du Forum économique mondial (couverture de l'échantillon de 64,42 %)
- indice de l'État de droit 2017-2018 du World Justice Project (couverture de l'échantillon de 54,59 %)
- Varieties of Democracy Project 2017 (couverture de l'échantillon de 83,58 %)

originales en notes z. Les notes z standardisées sont calculées en soustrayant la moyenne de chaque source de chaque note de pays et en divisant par l'écart type de la source respective. Ceci a pour résultat un ensemble de données centré autour de 0 et avec un écart-type de 1. Essentiellement, les notes z sont calculées en utilisant les paramètres de moyenne et d'écart-type des sources imputées de l'année de référence. Cela garantit que les notes de l'IPC sont comparables d'année en année à compter de l'année de référence.

Il est important de noter que l'ensemble de données complet assorti des valeurs imputées est utilisé uniquement pour générer les paramètres mondiaux de référence et que les valeurs imputées elles-mêmes ne sont pas utilisées dans l'agrégation finale qui produit les notes de l'IPC.

Les notes z sont ensuite transformées pour correspondre à l'échelle IPC de 0 à 100. Pour ce faire, on applique une formule de rééchelonnement, qui vise à définir la valeur moyenne de l'ensemble de données standardisé à 45 et l'écart-type à 20. La formule suivante est utilisée à cette fin :

$$\text{Indicateur rééchelonné}_{\text{pays}} = \frac{\text{Indicateur } X \text{ d'origine}_{\text{pays}} - \text{Indicateur } X \text{ imputée par la moyenne } t=\text{référence}}{\text{Indicateur } X^{t=\text{référence}} \text{ imputé par l'écart - type}} * 20 + 45$$

Toutes les notes rééchelonnées qui ont des valeurs inférieures à 0 sont ramenées à 0 et celles qui dépassent 100 sont plafonnées à 100.

3. Agréger les données rééchelonnées

La note IPC de chaque pays est calculée comme une moyenne simple de toutes les notes rééchelonnées disponibles pour ce pays. (Remarque : nous n'utilisons aucune des valeurs imputées en tant que note pour l'IPC agrégé). Un pays ne se verra attribuer une note que s'il y a au moins trois sources de données disponibles pour calculer cette moyenne.

4. Signaler une mesure d'incertitude

La note IPC est signalée de pair avec une erreur-type et un intervalle de confiance de 90 % qui reflètent la variance de la valeur des données des sources qui composent la note de l'IPC. En outre, les modifications des scores d'IPC statistiquement significatives sont identifiées et signalées.

Conformément à la recommandation du rapport d'audit effectué par le Centre commun de recherche de la Commission européenne examinant la cohérence conceptuelle et statistique de l'IPC, la formule de calcul des erreurs-types a été modifiée à partir de l'édition 2018 de l'IPC. La formule utilisée précédemment calculait le terme d'erreur-type comme l'écart-type des données sources rééchelonnées, divisé par la racine carrée du nombre de sources. La nouvelle formule, présentée dans l'équation ci-dessous, reflète un nombre restreint de sources :

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

où Σ est le terme d'erreur-type, N est le nombre total de sources utilisées pour le calcul de l'IPC, n est le nombre de sources sur lequel se fonde la note d'IPC un pays donné, et σ est l'écart-type du score IPC pour le pays en question⁴.

⁴ En 2018, le nombre total de sources (N) était de 13.

À partir de cette erreur-type, nous pouvons calculer l'intervalle de confiance de 90 %, et identifier les limites inférieures et supérieures du score d'IPC de chaque pays, en supposant une distribution normale.

Après avoir calculé les erreurs-types, nous examinons l'évolution du score de l'IPC pour déterminer s'il est significatif sur le plan statistique ou non. À cette fin, nous calculons d'abord la taille d'effet pour la différence de note d'un pays pour les deux années que l'on souhaite comparer, à l'aide de la formule ci-dessous :

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

où M_1 est le score IPC d'un pays donné l'année en question, M_2 est le score IPC de ce même pays pour l'année de référence, N_1 est le nombre de sources pour l'année en question, N_2 est le nombre de sources pour l'année de référence, SD_1^2 est le carré de l'écart-type du score IPC pour l'année en question et SD_2^2 est le carré de l'écart-type du score IPC pour l'année de référence.

Avec la taille d'effet ainsi obtenue, nous calculons alors l'écart-type (σ) de la taille d'effet avec la formule suivante :

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

où d est la taille d'effet des différences entre les notes d'IPC, et N_1 et N_2 représentent le nombre de sources disponibles pour chaque pays. L'intervalle de confiance de 90 % est ensuite calculé avec la formule :

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

Si cet intervalle de confiance comprend la valeur zéro, alors les différences entre les pays ne sont pas statistiquement significatives. Inversement, si la valeur zéro se trouve hors de l'intervalle, alors la différence est « statistiquement significative au niveau de 10 % ».