Version 1.0.1 | Étude

Auteur: Vojtech Royce

Traduction de l'original anglais. English original

Comment la pondération modifie le classement

Un test reproductible des scores composites d'égalité

Document de travail 1.0.1. En attente de relecture par l'auteur. Tous les profils numériques de cet article sont des exemples pédagogiques fictifs.

Résumé

Un indice composite peut inverser l'ordre de deux profils sans qu'aucune observation sous-jacente ne change. Cet article en démontre le mécanisme à l'aide de deux profils explicitement fictifs, décrits sur deux dimensions. Le profil A prend les valeurs 90 et 40 ; le profil B prend les valeurs 60 et 80. Avec une moyenne arithmétique pondérée, leur classement s'inverse lorsque le poids de la première dimension franchit quatre septièmes. Le calcul est exact et reproductible. Il n'apporte aucun élément probant concernant un pays ou un groupe démographique. Nous utilisons cet exemple pour définir une norme pratique de présentation pour le World Equality Index : publier les valeurs des composantes, définir les pondérations admissibles, montrer les seuils d'inversion, distinguer la sensibilité du modèle de l'incertitude d'échantillonnage et expliquer si des points forts peuvent compenser des points faibles. Un tableau de bord présentant les dimensions séparément demeure le principal support des résultats empiriques. Un score agrégé, lorsqu'il est défendable, constitue une interprétation explicite de ces dimensions.

Pourquoi cette question compte pour la recherche sur l'égalité

Les droits, les chances, les résultats, la contribution, les charges et la résilience répondent à des questions différentes. Un territoire peut obtenir de bons résultats sur une dimension et de mauvais résultats sur une autre. Additionner ces dimensions revient à formuler une affirmation supplémentaire : les arbitrages entre elles ont un sens sur une échelle commune et les rapports de compensation retenus conviennent à l'objectif annoncé. Un classement peut paraître précis même lorsque cette interprétation constitue la partie la plus contestée du calcul.

Le manuel de l'OCDE et du Centre commun de recherche de la Commission européenne considère la sélection des indicateurs, la normalisation, la pondération et l'agrégation comme des choix nécessitant une justification transparente. Son analyse de la robustesse recommande d'examiner comment l'incertitude et d'autres hypothèses influencent un indice. Cet article adopte cette orientation méthodologique générale et fournit son propre exemple analytique simple ainsi que des règles proposées pour la présentation des résultats du WEI.

Manuel de l'OCDE, de l'UE et du JRC de la Commission européenne, 2008, notamment les étapes 6 et 7

Cette démonstration ne comporte aucun jeu de données observées sur des pays. Les noms profil A et profil B désignent délibérément des profils fictifs. Les libellés des dimensions, Droits et Résultats, servent à l'illustration ; ils ne correspondent pas à des mesures de dispositions juridiques ou de résultats démographiques. Leurs valeurs numériques ont été choisies pour créer un arbitrage transparent dont le seuil peut être calculé à la main. Elles ne doivent jamais être affichées comme des scores nationaux, intégrées à un outil d'exploration empirique comme des observations réelles ou décrites comme des résultats concernant des groupes sociaux.

Le modèle

Supposons que les scores de deux dimensions se situent sur une échelle fixe de 0 à 100, les valeurs plus élevées étant considérées comme meilleures selon la convention de notation de cet exemple. Le profil A a pour valeurs Droits = 90 et Résultats = 40. Le profil B a pour valeurs Droits = 60 et Résultats = 80. Soit w le poids des droits, et 1 moins w le poids des résultats. L'intervalle admissible va de 0 à 1. Dans les formules, Rights désigne les droits et Outcomes les résultats.

Le score arithmétique est :

$S(w) = w \times \text{Rights} + (1 - w) \times \text{Outcomes}$

En substituant les valeurs de chaque profil, on obtient :

$S_A(w) = 90w + 40(1 - w) = 40 + 50w$

$S_B(w) = 60w + 80(1 - w) = 80 - 20w$

La soustraction des deux scores donne :

$S_A(w) - S_B(w) = 70w - 40$

Les deux profils sont à égalité lorsque 70w vaut 40, soit w = 4/7, environ 0,571429. En dessous de ce seuil, B se classe devant A. Au-dessus, A se classe devant B. Au seuil, ils sont à égalité. Les observations et le calcul sont fixes ; le changement d'ordre résulte entièrement du choix de pondération autorisé.

Résultats de la démonstration

Poids des droits	Poids des résultats	Score du profil A	Score du profil B	Classement
0,00	1,00	40,00	80,00	B devant A
0,25	0,75	52,50	75,00	B devant A
0,50	0,50	65,00	70,00	B devant A
4/7	3/7	68,571429	68,571429	Égalité
0,75	0,25	77,50	65,00	A devant B
0,80	0,20	80,00	64,00	A devant B
1,00	0,00	90,00	60,00	A devant B

Chaque valeur de ce tableau est calculée à partir des deux profils fictifs. Aucune estimation n'a été extraite d'un jeu de données sociales. À pondérations égales, B devance A de cinq points. Avec un poids des droits de 0,80, A devance B de seize points. Qualifier l'un ou l'autre résultat de classement global objectivement correct masquerait le choix de fond inscrit dans la pondération.

Le résultat apporte davantage que la simple affirmation que les pondérations comptent. Il indique exactement où la réponse change. Si une note de recherche avait justifié un poids des droits compris entre 0,20 et 0,50, B serait en tête sur tout cet intervalle spécifié. Si elle avait justifié un intervalle de 0,65 à 0,90, A serait en tête sur tout l'intervalle. Si l'intervalle admissible s'étend de part et d'autre de quatre septièmes, le classement dépend du choix effectué au sein de la famille de modèles. La justification d'un intervalle doit être évaluée avant de choisir le classement préféré.

Le domaine de validité de la robustesse

Un classement n'est robuste que par rapport à un ensemble déclaré d'hypothèses admissibles. Présenter un classement comme robuste après avoir fait varier un poids de 0,49 à 0,51 renseigne peu sur une plage plausible allant de 0,10 à 0,90. L'intervalle, la méthode de normalisation, la règle d'agrégation et le traitement des données manquantes doivent figurer à côté du résultat de robustesse. Ce résultat ne peut être dissocié de l'expérience qui l'a produit.

Un cas particulier utile est la dominance composante par composante. Si un profil atteint une valeur au moins aussi élevée sur chaque dimension et strictement plus élevée sur une dimension recevant un poids positif, un score arithmétique à pondérations non négatives le classe devant l'autre. Aucun arbitrage n'inverse cet ordre dans le cadre de ces hypothèses. Notre exemple ne présente volontairement aucune dominance : A est en tête sur une composante et B sur l'autre. Le croisement résulte de cette opposition, et non d'un défaut du calcul arithmétique.

Au-delà de deux dimensions, un seul curseur ne couvre plus toutes les combinaisons de pondérations. Une future mise en œuvre empirique devrait définir une région de pondérations admissibles et examiner les classements dans l'ensemble de cette région. Une grille peu fine peut illustrer des exemples, mais elle ne permet pas automatiquement d'établir qu'un rang extrême est impossible entre les points échantillonnés. Les affirmations sur les bornes des rangs atteignables devraient s'appuyer sur une optimisation exacte ou sur une procédure d'approximation documentée.

Ce qu'un curseur permet ou non d'interpréter

Une commande interactive de pondération peut aider le lecteur à explorer les hypothèses. Elle devrait afficher le vecteur complet des poids et le normaliser de manière cohérente, plutôt que de modifier des valeurs associées cachées. Elle devrait conserver les composantes brutes à l'écran et permettre l'exportation d'un scénario. Un scénario enregistré est l'expression du calcul choisi par le lecteur, et non un substitut à la méthode de référence publiée.

Si un outil échantillonne les pondérations uniformément, la proportion des tirages dans lesquels A arrive en tête est une propriété de cette méthode d'échantillonnage. Ce n'est ni la probabilité que A soit réellement plus égalitaire, ni la probabilité d'un accord du public, ni un niveau de confiance statistique. Dans cet exemple à deux dimensions, si w suit une distribution uniforme sur l'intervalle de 0 à 1, A arrive en tête dans les trois septièmes supérieurs de l'intervalle. Cela décrit une proportion géométrique sous une distribution arbitraire des pondérations. Une autre distribution modifie cette proportion tout en laissant intactes toutes les valeurs des profils.

Les plages de sensibilité ne sont pas non plus des intervalles liés à l'échantillonnage. Un intervalle de confiance porte sur l'incertitude d'une quantité estimée dans le cadre d'une procédure statistique. Une plage de scores obtenue en faisant varier les pondérations décrit les conséquences d'un changement de règle de notation. Les deux peuvent compter dans un travail empirique, mais les réunir dans une bande d'incertitude sans en préciser la nature rendrait le graphique plus difficile à interpréter.

Compensation et échelle

La formule arithmétique autorise la compensation. Un gain sur une dimension peut compenser une perte sur l'autre. Certains objectifs de recherche peuvent accepter cet arbitrage ; d'autres peuvent exiger un seuil minimal pour un droit ou un service essentiel. Une règle imposant des seuils obligatoires répond à une question différente de celle d'une moyenne sans restriction. Le WEI devrait préciser l'arbitrage autorisé au lieu de considérer une formule familière comme un choix par défaut neutre.

L'échelle choisie compte également. Faire passer une dimension d'un pourcentage à un effectif sans remise à l'échelle peut modifier considérablement son influence numérique. Même des poids égaux en apparence ne garantissent pas une influence effective égale lorsque les distributions des composantes diffèrent. Avant d'attribuer des poids, un article doit définir, pour chaque composante, le concept mesuré, le sens de variation, les unités, la transformation et la population de référence. Un score élevé doit avoir une interprétation précise dans ce contexte.

Pour les mesures d'égalité, même le sens de variation peut être contesté. Un écart plus faible entre groupes pour un résultat observé ne démontre pas une égalité de traitement, et des résultats inégaux peuvent découler de plusieurs processus institutionnels. Si le score valorise la parité des résultats, ce choix doit être explicité dans la définition du concept mesuré. Une mesure des droits juridiques doit provenir d'éléments relatifs aux droits, plutôt que d'être tacitement déduite d'une répartition de l'emploi.

Une proposition de norme de publication pour le WEI

Chaque indice composite empirique devrait avoir un objectif explicite et un inventaire de ses composantes. Cet inventaire devrait inclure les versions des sources, la couverture géographique, les périodes, les dénominateurs, les valeurs manquantes, les transformations et la signification concrète des valeurs plus élevées. Les lecteurs devraient pouvoir examiner une composante sans accepter l'interprétation agrégée.

La méthode devrait ensuite identifier un vecteur de pondérations de référence et justifier son utilisation. Les autres pondérations et règles d'agrégation admissibles devraient être spécifiées avant d'examiner quelles entités en bénéficient. Les résultats devraient présenter les comparaisons composante par composante, les scores de référence, les seuils d'inversion lorsqu'ils sont calculables, ainsi que les plages de rangs ou de scores selon les autres choix déclarés. Les cas où l'ordre change méritent d'être signalés de manière bien visible, et non dans une note de bas de page décrivant l'indice comme globalement robuste.

Les composantes manquantes nécessitent une règle préservant la comparaison annoncée. Renormaliser les pondérations pour une entité tout en conservant les pondérations initiales pour une autre peut conduire à comparer des concepts différents. Un calcul devrait indiquer quand des entités sont exclues, quand des données sont imputées et si le résultat dépend de ce choix. Une valeur manquante ne doit pas devenir silencieusement un zéro ou la moyenne de l'échantillon.

Enfin, la publication devrait inclure des calculs exécutables et un relevé des décisions éditoriales. Les corrections devraient préciser si le changement concerne les éléments probants sous-jacents, une erreur de codage ou une décision normative de notation. Une nouvelle politique de pondération est une révision de méthode même si le jeu de données reste identique. Des fichiers téléchargeables dont les versions sont conservées donnent au lecteur un document stable à citer et permettent de distinguer les désaccords d'interprétation des désaccords sur le calcul.

Limites et utilisation

Cette démonstration prouve une inversion de classement au sein d'un modèle arithmétique délibérément simple. Elle n'évalue pas les indices d'égalité existants, ne détermine pas de pondérations préférables, ne mesure aucun territoire réel et ne fournit aucun élément probant à l'appui d'une conclusion politique. Sa reproductibilité est exacte parce que les données d'entrée sont des exemples définis, et non parce que la mesure des phénomènes sociaux serait exempte d'incertitude.

Le script d'accompagnement recrée tous les scores affichés et vérifie l'égalité à quatre septièmes. Il en découle une obligation de présentation : lorsque l'ordre change sous des

hypothèses admissibles, montrer ce changement et expliquer les hypothèses. Lorsque les composantes répondent à des questions différentes, garder leur signification explicite. Un indice composite peut résumer une perspective déclarée, tandis qu'une plateforme transparente permet au lecteur de voir comment cette perspective a façonné le résultat.