Versione 1.0.1 | Studio

Autore: Vojtech Royce

Traduzione dell'originale inglese. English original

Come la ponderazione cambia la graduatoria

Un test riproducibile dei punteggi compositi di uguaglianza

Documento di lavoro 1.0.1. In attesa di revisione da parte dell'autore. Tutti i profili numerici di questo studio sono esempi didattici artificiali.

Riassunto

Un indice composito può invertire l'ordine di due profili senza che cambi una sola osservazione sottostante. Questo studio dimostra il meccanismo utilizzando due profili esplicitamente artificiali misurati su due dimensioni. Il Profilo A presenta valori di 90 e 40; il Profilo B presenta valori di 60 e 80. Con una media aritmetica ponderata, il loro ordine si inverte quando il peso della prima dimensione attraversa la soglia di quattro settimi. Il calcolo è esatto e riproducibile. Non costituisce un'evidenza relativa ad alcun paese o gruppo demografico. Utilizziamo l'esempio per definire uno standard pratico di presentazione dei risultati per World Equality Index: pubblicare i valori delle componenti, definire i pesi ammissibili, mostrare le soglie di inversione, distinguere la sensibilità del modello dall'incertezza campionaria e spiegare se i punti di forza possano compensare le debolezze. Un quadro che presenta separatamente le dimensioni rimane il riferimento empirico principale. Un punteggio aggregato, ove giustificabile, è un'interpretazione dichiarata di tali dimensioni.

Perché è importante per la ricerca sull'uguaglianza

Diritti, opportunità, esiti, contributo, oneri e resilienza rispondono a domande diverse. Una giurisdizione può ottenere buoni risultati su una dimensione e risultati scarsi su un'altra. Sommare queste dimensioni introduce un'ulteriore affermazione: che le compensazioni tra esse abbiano senso su una scala comune e che i rapporti di scambio scelti siano adeguati allo scopo dichiarato. Una graduatoria può apparire precisa anche quando questa interpretazione è la parte più controversa del calcolo.

Il manuale dell'OCSE e del Centro comune di ricerca della Commissione europea tratta la selezione degli indicatori, la normalizzazione, la ponderazione e l'aggregazione come scelte che richiedono una giustificazione trasparente. La sua trattazione della robustezza raccomanda di esaminare come l'incertezza e le ipotesi alternative influenzino un indice. Questo studio adotta tale orientamento metodologico generale e presenta un proprio semplice esempio analitico e una proposta di regole per la presentazione dei risultati di WEI. Manuale OCSE, UE ed EC-JRC, 2008, in particolare le fasi 6 e 7

In questa dimostrazione non vi è alcun dataset di osservazioni sui paesi. I nomi Profilo A e Profilo B sono deliberatamente artificiali. Le etichette delle dimensioni, Diritti ed Esiti, sono segnaposto illustrativi, non misurazioni di ordinamenti giuridici o di esiti demografici. I loro valori numerici sono stati scelti per creare una compensazione trasparente, analizzabile con un calcolo a mano. Non devono mai essere visualizzati come punteggi nazionali, inseriti in uno strumento di esplorazione empirica come osservazioni reali o descritti come risultati relativi a gruppi sociali.

Il modello

Si assumono due punteggi, uno per dimensione, su una scala fissa da 0 a 100, con valori maggiori considerati migliori secondo la convenzione di attribuzione dei punteggi adottata a scopo illustrativo. Il Profilo A ha Diritti = 90 ed Esiti = 40. Il Profilo B ha Diritti = 60 ed Esiti = 80. Sia w il peso dei Diritti e sia 1 meno w il peso degli Esiti. L'intervallo ammissibile va da 0 a 1.

Il punteggio aritmetico è:

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

Sostituendo i valori di ciascun profilo si ottiene:

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

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

Sottraendo i due punteggi si ottiene:

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

I due profili sono a pari merito quando 70w è uguale a 40, ossia w = 4/7, circa 0,571429. Al di sotto di tale soglia B si colloca più in alto. Al di sopra si colloca più in alto A. Alla soglia sono a pari merito. Le osservazioni e il calcolo sono fissi; il cambiamento di ordine dipende interamente dalla scelta del peso tra quelli ammessi.

Risultati della dimostrazione

Peso dei Diritti	Peso degli Esiti	Punteggio del Profilo A	Punteggio del Profilo B	Ordine
0,00	1,00	40,00	80,00	B precede A
0,25	0,75	52,50	75,00	B precede A
0,50	0,50	65,00	70,00	B precede A
4/7	3/7	68,571429	68,571429	Parità
0,75	0,25	77,50	65,00	A precede B
0,80	0,20	80,00	64,00	A precede B
1,00	0,00	90,00	60,00	A precede B

Ogni valore di questa tabella è calcolato a partire dai due profili artificiali. Nessuna stima è stata ricavata da un dataset di dati sociali. Con pesi uguali B è in vantaggio di cinque punti. Con un peso dei Diritti pari a 0,80, A è in vantaggio di sedici punti. Definire uno dei due risultati come la graduatoria complessiva oggettivamente corretta nasconderebbe la scelta sostanziale incorporata nel peso.

Il risultato rilevante è più informativo della semplice affermazione che i pesi contano. Identifica esattamente dove cambia la risposta. Se un documento di impostazione della ricerca avesse giustificato un peso dei Diritti compreso tra 0,20 e 0,50, B sarebbe in vantaggio in tutto quell'intervallo specificato. Se avesse giustificato un intervallo da 0,65 a 0,90, A sarebbe in vantaggio in tutto l'intervallo. Se l'intervallo ammissibile si estende su entrambi i lati di quattro settimi, l'ordine dipende dalla scelta effettuata all'interno della famiglia di modelli. La giustificazione di un intervallo deve essere valutata prima di selezionare la graduatoria preferita.

La robustezza vale entro un ambito definito

Un ordine è robusto soltanto rispetto a un insieme dichiarato di ipotesi ammesse. Definire robusta una graduatoria dopo aver variato un peso da 0,49 a 0,51 dice poco su un intervallo plausibile da 0,10 a 0,90. L'intervallo, il metodo di normalizzazione, la regola di aggregazione e il trattamento dei dati mancanti devono accompagnare il risultato sulla robustezza. Il risultato non può essere separato dall'esperimento che lo ha prodotto.

Un caso particolare utile è la dominanza componente per componente. Se un profilo ha valori almeno altrettanto alti su ogni dimensione e strettamente più alti su una dimensione con peso positivo, un punteggio aritmetico con pesi non negativi lo colloca più in alto. Nessuna compensazione inverte tale ordine entro queste ipotesi. Il nostro esempio è intenzionalmente privo di dominanza: A è in vantaggio su una componente e B sull'altra. L'inversione dell'ordine dipende da questo contrasto, non da un difetto dell'aritmetica.

Con più di due dimensioni, un singolo cursore non copre più tutte le combinazioni di pesi. Una futura implementazione empirica dovrebbe definire una regione ammissibile dei pesi ed esaminare le graduatorie al suo interno. Una griglia a maglie larghe può mostrare esempi, ma non può stabilire automaticamente che una posizione estrema sia impossibile tra i punti campionati. Le affermazioni sui limiti delle posizioni raggiungibili dovrebbero essere sostenute da un'ottimizzazione esatta o da una procedura di approssimazione documentata.

Che cosa significa e che cosa non significa un cursore

Un controllo interattivo dei pesi può aiutare il lettore a esplorare le ipotesi. Il controllo dovrebbe mostrare l'intero vettore dei pesi e normalizzarlo in modo coerente, anziché modificare altri valori nascosti. Dovrebbe mantenere sullo schermo i valori non aggregati delle componenti e consentire l'esportazione di uno scenario. Uno scenario salvato documenta il calcolo scelto da un lettore, non sostituisce il metodo di riferimento pubblicato.

Se uno strumento campiona i pesi uniformemente, la proporzione di campioni in cui A è in vantaggio è una proprietà di quello schema di campionamento. Non è la probabilità che A presenti realmente un maggiore livello di uguaglianza, la probabilità di consenso pubblico o un livello di confidenza statistica. In questo esempio bidimensionale, con una distribuzione

uniforme di w nell'intervallo da 0 a 1, A prevale nei tre settimi superiori dell'intervallo. Ciò descrive una proporzione geometrica secondo una distribuzione arbitraria dei pesi. Un'altra distribuzione modifica la proporzione lasciando invariati tutti i valori dei profili.

Gli intervalli di sensibilità non sono neppure intervalli di incertezza campionaria. Un intervallo di confidenza riguarda l'incertezza di una quantità stimata secondo una procedura statistica. Un intervallo di punteggi al variare dei pesi riguarda le conseguenze della modifica della regola di attribuzione del punteggio. Entrambi possono essere rilevanti nel lavoro empirico, ma combinarli in una fascia di incertezza priva di etichetta renderebbe il grafico più difficile da interpretare.

Compensazione e scala

La formula aritmetica consente la compensazione. Un guadagno su una dimensione può compensare una perdita sull'altra. Alcuni scopi di ricerca possono ammettere questa compensazione; altri possono richiedere una soglia minima per un diritto o un servizio essenziale. Una regola con soglie obbligatorie risponde a una domanda diversa rispetto a una media senza vincoli. WEI dovrebbe dichiarare la compensazione consentita anziché trattare una formula familiare come un'opzione predefinita neutrale.

Anche la scala scelta è importante. Passare da una percentuale a un conteggio per una dimensione, senza adeguarne la scala, può modificarne drasticamente l'influenza numerica. Anche pesi nominalmente uguali non garantiscono un'influenza pratica uguale quando le distribuzioni delle componenti differiscono. Prima di assegnare i pesi, uno studio deve definire il costrutto, la direzione, le unità, la trasformazione e la popolazione di riferimento per ogni componente. Un punteggio elevato deve avere un'interpretazione precisa in quel contesto.

Per le misure di uguaglianza, persino la direzione può essere controversa. Una minore differenza tra gruppi in un esito osservato non dimostra parità di trattamento, e risultati disuguali possono derivare da più processi istituzionali. La scelta di premiare la parità degli esiti dovrebbe essere documentata come il costrutto a cui viene attribuito il punteggio. Una misura dei diritti giuridici deve derivare da evidenze relative ai diritti, anziché essere inferita tacitamente da una distribuzione occupazionale.

Una proposta di standard di pubblicazione WEI

Ogni indicatore composito empirico dovrebbe avere uno scopo esplicito e un inventario delle componenti. L'inventario dovrebbe includere le versioni delle fonti, la copertura geografica, i periodi, i denominatori, i valori mancanti, le trasformazioni e il significato sostanziale dei valori più elevati. I lettori dovrebbero poter esaminare una componente senza accettare l'interpretazione aggregata.

Il metodo dovrebbe poi identificare un vettore di pesi di riferimento e giustificarne l'uso. I pesi ammissibili alternativi e le regole di aggregazione dovrebbero essere specificati prima di esaminare quali entità ne traggano vantaggio. I risultati dovrebbero riportare confronti componente per componente, punteggi di riferimento, soglie di inversione ove calcolabili e intervalli di posizione o di punteggio secondo le alternative dichiarate. I casi in cui l'ordine

cambia meritano una segnalazione visibile, non una nota a piè di pagina che descriva l'indice come generalmente robusto.

Le componenti mancanti richiedono una regola che preservi il confronto dichiarato. Rinormalizzare i pesi per un'entità mantenendo quelli originali per un'altra può portare a confrontare costrutti diversi. Un calcolo dovrebbe dichiarare quando le entità vengono escluse, quando i dati vengono imputati e se un risultato dipende da tale scelta. Un valore mancante non deve diventare silenziosamente uno zero o la media campionaria.

Infine, la pubblicazione dovrebbe includere calcoli eseguibili e un registro delle decisioni editoriali. Le correzioni dovrebbero identificare se siano cambiate le evidenze sottostanti, se sia stato corretto un errore di programmazione o se sia cambiata una decisione normativa di attribuzione del punteggio. Una nuova regola di ponderazione costituisce una revisione metodologica anche se il dataset rimane identico. File scaricabili con versioni identificate offrono al lettore un riferimento stabile da citare e rendono separabili i disaccordi interpretativi da quelli sui calcoli.

Limiti e utilizzo

Questa dimostrazione prova un'inversione dell'ordine all'interno di un modello aritmetico deliberatamente semplice. Non valuta gli indici di uguaglianza esistenti, non stabilisce pesi preferibili, non misura alcuna giurisdizione reale e non fornisce evidenze a sostegno di una conclusione politica. La sua riproducibilità è esatta perché gli input sono esempi definiti, non perché la misurazione sociale sia priva di incertezza.

Lo script di accompagnamento ricrea tutti i punteggi visualizzati e verifica la parità a quattro settimi. L'implicazione pratica è un obbligo di pubblicazione: quando l'ordine cambia al variare delle ipotesi ammissibili, mostrare il cambiamento e spiegare le ipotesi. Quando le componenti rispondono a domande diverse, mantenerne visibili i significati. Un indicatore composito può sintetizzare una prospettiva dichiarata, mentre una piattaforma trasparente consente al lettore di vedere come tale prospettiva abbia plasmato il risultato.