Version 1.0.2 | Studie

Autor: Vojtech Royce

Übersetzung des englischen Originals. English original

Wie Gewichtungen die Rangfolge verändern

Ein reproduzierbarer Test zusammengesetzter Gleichheitswerte

Studie 1.0.2. Alle numerischen Profile in dieser Studie sind künstliche Lehrbeispiele.

Zusammenfassung

Ein zusammengesetzter Index kann die Reihenfolge zweier Profile umkehren, ohne dass sich eine einzige zugrunde liegende Beobachtung ändert. Diese Studie veranschaulicht den Mechanismus anhand zweier ausdrücklich künstlicher Profile mit Werten in zwei Dimensionen. Profil A hat die Werte 90 und 40, Profil B die Werte 60 und 80. Bei einem gewichteten arithmetischen Mittel kehrt sich ihre Reihenfolge um, wenn das Gewicht der ersten Dimension die Schwelle von vier Siebteln überschreitet. Die Berechnung ist exakt und reproduzierbar. Sie liefert keine Belege über ein Land oder eine demografische Gruppe. Wir nutzen das Beispiel, um einen praktischen Berichtsstandard für den World Equality Index zu formulieren: Komponentenwerte veröffentlichen, zulässige Gewichte definieren, Schwellen für einen Rangwechsel zeigen, Modellsensitivität von Stichprobenunsicherheit unterscheiden und erläutern, ob Stärken Schwächen ausgleichen dürfen. Der zentrale empirische Ausgangspunkt bleibt eine Übersicht der getrennten Dimensionen. Ein Gesamtwert ist, sofern vertretbar, eine offengelegte Interpretation dieser Dimensionen.

Warum dies für die Gleichheitsforschung wichtig ist

Rechte, Chancen, Ergebnisse, Beitrag, Belastung und Resilienz beantworten unterschiedliche Fragen. Ein Land oder Gebiet kann in einer Dimension gut und in einer anderen schlecht abschneiden. Diese Dimensionen zu addieren, begründet eine zusätzliche Aussage: dass ihre Abwägungen auf einer gemeinsamen Skala sinnvoll sind und dass die gewählten Austauschverhältnisse dem angegebenen Zweck entsprechen. Eine Rangfolge kann präzise erscheinen, selbst wenn diese Interpretation der umstrittenste Teil der Berechnung ist.

Das Handbuch der OECD und der Gemeinsamen Forschungsstelle der Europäischen Kommission behandelt Indikatorauswahl, Normalisierung, Gewichtung und Aggregation als Entscheidungen, die eine transparente Begründung erfordern. Seine Erörterung der Robustheit empfiehlt zu untersuchen, wie Unsicherheit und alternative Annahmen einen Index beeinflussen. Diese Studie übernimmt diese allgemeine methodische Ausrichtung und liefert ein eigenes kleines analytisches Beispiel sowie vorgeschlagene WEI-Berichtsregeln. Handbuch von OECD, EU und EC-JRC, 2008, insbesondere Schritte 6 und 7

Diese Demonstration enthält keinen beobachteten Länderdatensatz. Die Bezeichnungen Profil A und Profil B sind bewusst künstlich. Die Dimensionsbezeichnungen Rechte und

Ergebnisse sind illustrative Platzhalter, keine gemessenen Rechtsordnungen oder demografischen Ergebnisse. Ihre Zahlenwerte wurden so gewählt, dass die Abwägung zwischen den Dimensionen nachvollziehbar ist und sich von Hand berechnen lässt. Sie dürfen niemals als nationale Indexwerte dargestellt, als reale Beobachtungen in ein Werkzeug zur Untersuchung empirischer Daten eingespeist oder als Befunde über gesellschaftliche Gruppen beschrieben werden.

Das Modell

Wir nehmen an, dass zwei Dimensionswerte auf einer festen Skala von 0 bis 100 liegen, wobei höhere Werte nach der illustrativen Bewertungskonvention als besser gelten. Profil A hat Rechte = 90 und Ergebnisse = 40. Profil B hat Rechte = 60 und Ergebnisse = 80. Dabei sei w das Gewicht für Rechte und 1 minus w das Gewicht für Ergebnisse. Das zulässige Intervall reicht von 0 bis 1.

Der arithmetische Indexwert lautet:

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

Durch Einsetzen der jeweiligen Profile ergibt sich:

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

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

Die Subtraktion der beiden Indexwerte ergibt:

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

Die beiden Profile sind gleichauf, wenn 70w gleich 40 ist, also bei w = 4/7, ungefähr 0,571429. Unterhalb dieser Schwelle liegt B höher. Oberhalb liegt A höher. An der Schwelle besteht Gleichstand. Die Ausgangswerte und die Rechenregeln bleiben unverändert. Die veränderte Reihenfolge ergibt sich ausschließlich aus der zulässigen Wahl des Gewichts.

Ergebnisse des Beispiels

Gewicht für Rechte	Gewicht für Ergebnisse	Indexwert Profil A	Indexwert Profil B	Reihenfolge
0,00	1,00	40,00	80,00	B vor A
0,25	0,75	52,50	75,00	B vor A
0,50	0,50	65,00	70,00	B vor A
4/7	3/7	68,571429	68,571429	Gleichstand
0,75	0,25	77,50	65,00	A vor B
0,80	0,20	80,00	64,00	A vor B
1,00	0,00	90,00	60,00	A vor B

Jeder Wert in dieser Tabelle wird aus den beiden künstlichen Profilen berechnet. Es wurde keine Schätzung aus einem gesellschaftsbezogenen Datensatz abgerufen. Bei gleichen Gewichten liegt B um fünf Punkte vorn. Bei einem Gewicht für Rechte von 0,80 liegt A um sechzehn Punkte vorn. Eines der beiden Ergebnisse als die objektiv richtige Gesamtrangfolge zu bezeichnen, würde die inhaltliche Entscheidung verschleiern, die im Gewicht zum Ausdruck kommt.

Der relevante Befund ist aussagekräftiger als die Feststellung, dass Gewichte eine Rolle spielen. Er bestimmt genau, wo sich die Antwort ändert. Hätte ein Forschungsauftrag ein Gewicht für Rechte zwischen 0,20 und 0,50 begründet, würde B im gesamten angegebenen Intervall vorn liegen. Hätte er ein Intervall von 0,65 bis 0,90 begründet, würde A durchgehend vorn liegen. Umfasst das zulässige Intervall Werte beiderseits von vier Siebteln, hängt die Reihenfolge innerhalb der Modellfamilie von der Spezifikation ab. Die Vertretbarkeit eines Intervalls muss beurteilt werden, bevor die bevorzugte Rangfolge ausgewählt wird.

Robustheit hat einen definierten Geltungsbereich

Eine Reihenfolge ist nur relativ zu einer angegebenen Menge zulässiger Annahmen robust. Eine Rangfolge als robust zu bezeichnen, nachdem ein Gewicht zwischen 0,49 und 0,51 variiert wurde, sagt wenig über einen plausiblen Bereich von 0,10 bis 0,90 aus. Das Intervall, die Normalisierungsmethode, die Aggregationsregel und der Umgang mit fehlenden Daten gehören unmittelbar zum Robustheitsergebnis. Das Ergebnis ist nur zusammen mit dem Experiment aussagekräftig, aus dem es hervorgegangen ist.

Ein nützlicher Sonderfall ist die komponentenweise Dominanz. Wenn ein Profil in jeder Dimension mindestens so hoch und in einer Dimension mit positivem Gewicht strikt höher liegt, weist ihm ein arithmetischer Index mit nichtnegativen Gewichten einen höheren Rang zu. Unter diesen Annahmen kehrt keine Abwägung diese Reihenfolge um. Unser Beispiel weist absichtlich keine Dominanz auf: A führt bei einer Komponente und B bei der anderen. Der Rangwechsel ist eine Eigenschaft dieses Gegensatzes und kein Fehler der Arithmetik.

Bei mehr als zwei Dimensionen deckt ein einzelner Schieberegler nicht mehr alle Gewichtskombinationen ab. Eine künftige empirische Umsetzung sollte einen zulässigen Gewichtsraum definieren und Rangfolgen in diesem Raum untersuchen. Ein grobes Raster kann Beispiele veranschaulichen, aber nicht automatisch nachweisen, dass zwischen den untersuchten Punkten ein extremer Rang unmöglich ist. Aussagen über erreichbare Ranggrenzen sollten durch exakte Optimierung oder ein dokumentiertes Näherungsverfahren gestützt werden.

Was ein Schieberegler bedeutet und was nicht

Eine interaktive Gewichtssteuerung kann Leserinnen und Lesern helfen, Annahmen zu erkunden. Die Steuerung sollte den vollständigen Gewichtsvektor anzeigen und ihn konsistent normalisieren, statt verborgene zugehörige Werte zu verändern. Sie sollte die ursprünglichen Komponenten auf dem Bildschirm sichtbar halten und den Export eines Szenarios ermöglichen. Ein gespeichertes Szenario dokumentiert die von einer Leserin oder einem Leser gewählte Berechnung und ersetzt nicht die veröffentlichte Referenzmethode.

Wenn ein Werkzeug Gewichte gleichverteilt zieht, ist der Anteil der Ziehungen, bei denen A führt, eine Eigenschaft dieses Ziehungsverfahrens. Er ist weder die Wahrscheinlichkeit, dass A tatsächlich mehr Gleichheit aufweist, noch die Wahrscheinlichkeit öffentlicher Zustimmung oder ein statistisches Konfidenzniveau. In diesem zweidimensionalen Beispiel entfallen bei gleichverteiltem w im Intervall von 0 bis 1 die oberen drei Siebtel des Intervalls auf einen Vorsprung von A. Das beschreibt einen geometrischen Anteil unter einer willkürlichen Gewichtsverteilung. Eine andere Verteilung verändert den Anteil, während sämtliche Profilwerte unverändert bleiben.

Ebenso wenig sind Sensitivitätsbereiche Stichprobenintervalle. Ein Konfidenzintervall behandelt die Unsicherheit einer geschätzten Größe im Rahmen eines statistischen Verfahrens. Der Wertebereich bei unterschiedlichen Gewichten zeigt dagegen die Folgen einer veränderten Bewertungsregel. Beides kann in empirischer Arbeit relevant sein. Eine Zusammenführung zu einem nicht näher gekennzeichneten Unsicherheitsband würde die Interpretation der Grafik jedoch erschweren.

Kompensation und Skala

Die arithmetische Formel erlaubt Kompensation. Ein Zugewinn in einer Dimension kann einen Verlust in der anderen ausgleichen. Für manche Forschungszwecke mag diese Abwägung akzeptabel sein. Andere erfordern möglicherweise eine Mindestschwelle für ein Recht oder eine wesentliche Dienstleistung. Eine Regel mit verbindlichen Schwellen beantwortet eine andere Frage als ein uneingeschränkter Durchschnitt. WEI sollte die zulässige Abwägung offenlegen, statt eine vertraute Formel als neutralen Standard zu behandeln.

Auch die gewählte Skala ist relevant. Wird eine Dimension ohne Neuskalierung von einem Prozentanteil auf eine Anzahl umgestellt, kann sich ihr numerischer Einfluss stark verändern. Selbst ein nominell gleiches Gewicht gewährleistet keinen gleichen praktischen Einfluss, wenn sich die Verteilungen der Komponenten unterscheiden. Vor der Zuweisung von Gewichten muss eine Studie für jede Komponente das Konstrukt, die Richtung, die Einheiten, die Transformation und die Referenzpopulation definieren. Ein hoher Wert muss in diesem Kontext eine präzise Bedeutung haben.

Bei Gleichheitsmaßen kann sogar die Richtung umstritten sein. Ein kleinerer Gruppenunterschied bei einem beobachteten Ergebnis belegt keine Gleichbehandlung, und ungleiche Ergebnisse können durch unterschiedliche institutionelle Prozesse entstehen. Die Entscheidung, Ergebnisparität positiv zu bewerten, sollte als Teil der Definition des bewerteten Konstrukts dokumentiert werden. Ein Maß rechtlicher Gleichberechtigung muss aus Belegen über Rechte hervorgehen, statt stillschweigend aus einer Erwerbstätigkeitsverteilung abgeleitet zu werden.

Ein vorgeschlagener WEI-Veröffentlichungsstandard

Jeder empirische zusammengesetzte Index sollte einen expliziten Zweck und ein Komponentenverzeichnis haben. Dieses Verzeichnis sollte Quellenversionen, geografische Abdeckung, Zeiträume, Nenner, fehlende Werte, Transformationen und die inhaltliche Bedeutung höherer Werte enthalten. Leserinnen und Leser sollten eine Komponente prüfen können, ohne die Interpretation des Gesamtwerts übernehmen zu müssen.

Die Methode sollte anschließend einen Referenzgewichtsvektor benennen und seine Verwendung begründen. Alternative zulässige Gewichte und Aggregationsregeln sollten festgelegt werden, bevor untersucht wird, welche Einheiten davon profitieren. Die Ergebnisse sollten komponentenweise Vergleiche, Referenzwerte, berechenbare Schwellen für einen Rangwechsel sowie Rang- oder Wertebereiche unter den offengelegten Alternativen berichten. Fälle mit wechselnder Reihenfolge verdienen eine deutlichen Hinweis und keine Fußnote, die den Index als allgemein robust bezeichnet.

Fehlende Komponenten benötigen eine Regel, die den angegebenen Vergleich wahrt. Werden Gewichte für eine Einheit neu normiert, während für eine andere die ursprünglichen Gewichte beibehalten werden, können unterschiedliche Konstrukte verglichen werden. Eine Berechnung sollte offenlegen, wann Einheiten ausgeschlossen werden, wann Daten imputiert werden und ob ein Ergebnis von dieser Entscheidung abhängt. Ein fehlender Wert darf nicht stillschweigend zu null oder zum Stichprobenmittelwert werden.

Schließlich sollte die Veröffentlichung ausführbare Berechnungen und eine Dokumentation redaktioneller Entscheidungen enthalten. Bei Korrekturen sollte deutlich werden, ob neue Belege, ein Programmierfehler oder eine geänderte normative Bewertungsentscheidung der Anlass sind. Eine neue Gewichtungsrichtlinie ist eine Methodenrevision, selbst wenn der Datensatz identisch bleibt. Versionierte Downloads geben den Leserinnen und Lesern eine eindeutig bestimmbare, zitierbare Fassung und ermöglichen es, Meinungsverschiedenheiten über die Interpretation von Meinungsverschiedenheiten über die Berechnung zu trennen.

Grenzen und Verwendung

Diese Demonstration beweist eine Rangumkehr innerhalb eines bewusst einfachen arithmetischen Modells. Sie bewertet keine bestehenden Gleichheitsindizes, legt keine bevorzugten Gewichte fest, misst kein reales Land oder Gebiet und liefert keine Belege für eine politische Schlussfolgerung. Die Ergebnisse sind exakt reproduzierbar, weil die Eingaben als Beispiele fest definiert sind, nicht weil die Messung gesellschaftlicher Phänomene frei von Unsicherheit wäre.

Das Begleitskript reproduziert alle dargestellten Indexwerte und überprüft den Gleichstand bei vier Siebteln. Die praktische Konsequenz ist eine Verpflichtung bei der Veröffentlichung: Wenn sich die Reihenfolge unter zulässigen Annahmen ändert, muss die Änderung gezeigt und müssen die Annahmen erläutert werden. Wenn Komponenten unterschiedliche Fragen beantworten, müssen ihre Bedeutungen sichtbar bleiben. Ein zusammengesetzter Index kann eine offengelegte Perspektive zusammenfassen, während eine transparente Plattform den Leserinnen und Lesern zeigt, wie diese Perspektive das Ergebnis geprägt hat.