

# Metodologia e note tecniche

## WeWorld *Mai più invisibili* 2023

La nuova metodologia di calcolo dell'indice italiano 2023 riprende quanto sviluppato per l'indice mondiale 2022 con l'obiettivo d'indagare più accuratamente il livello d'inclusione di donne e bambini nelle diverse regioni italiane. Rispetto alla precedente edizione, possiamo ora tracciare in termini assoluti le prestazioni delle regioni e valutare punti di forza e debolezza in ciascuna delle dimensioni dell'indice. Con il nuovo metodo di calcolo le quindici dimensioni, i tre sottoindici e l'indice generale sono espressi come punteggio 0-100, fornendo per ciascun territorio un confronto assoluto rispetto a dei chiari scenari estremi di riferimento. Per realizzare l'indice è stato seguito un processo in cinque fasi che, a partire dai dati originali degli indicatori raccolti, permette di ottenere e confrontare i punteggi di ciascuna componente negli anni dal 2018 al 2022.

## 1 Raccolta dati

I dati dei trenta indicatori scelti per far parte dell'indice provengono quasi interamente dalla banca dati ISTAT. Nello specifico:

- venti sono contenuti nel rapporto ISTAT *Benessere equo e sostenibile*;
- nove provengono da altri rapporti e indagini ISTAT;
- uno è stato elaborato da WeWorld sulla base di dati grezzi ISPRA.

Per un elenco dettagliato degli indicatori, delle loro fonti e del loro aggiornamento<sup>1</sup> si rimanda alla tabella in appendice.

## 2 Imputazione valori mancanti

Per garantire un'adeguata integrità statistica dell'indice sono stati scelti indicatori con il minor numero possibile di osservazioni mancanti. Il campione di dati da utilizzare nel calcolo per ciascun territorio  $x$  (Regione/Provincia autonoma, Area o Italia) e anno  $i$  (2018–2022) in esame è stato determinato nel seguente modo:

- i. se presente si prende l'osservazione originale  $x_i$ ;

---

<sup>1</sup>Gli indicatori aggiornati al 2022 sono soltanto tre, in nove casi l'aggiornamento è al 2020 o 2019, i restanti diciotto sono aggiornati al 2021.

- ii. se  $x_i$  è mancante si esegue un'interpolazione lineare per colmare il valore assente a partire dalle osservazioni adiacenti o, se anch'esse mancanti, si propaga l'ultimo dato valido;<sup>2</sup>
- iii. se anche negli anni precedenti l'osservazione è mancante si utilizzano nell'ordine l'osservazione d'area presente nei dati originali o la media d'area pesata sulla popolazione.<sup>3</sup>

Soltanto per Molise e Basilicata nel caso dell'indicatore 9 si è reso necessario ricorrere ai dati d'area per assenza totale di osservazioni. Per le province autonome di Bolzano e Trento, dove mancante il dato a livello provinciale,<sup>4</sup> si è preferito quando possibile fare riferimento al dato regionale prima di qualsiasi altra imputazione.

## 3 Trasformazione

Prima di procedere alla normalizzazione degli indicatori sono stati necessari due tipi di trasformazioni: una trasformazione logaritmica in presenza di distribuzioni asimmetriche e un chiaro limite massimo nel caso della parità di genere.

### 3.1 Scala logaritmica

La trasformazione logaritmica si rende necessaria quando la distribuzione dell'indicatore si presenta fortemente asimmetrica e contiene valori estremi. Ad esempio un caso in cui è solitamente applicata è il tasso di omicidi che coinvolge gli indicatori 8 e 9. Questi indicatori sono quindi stati trasformati applicando la seguente funzione:

$$x' = \log(x + \alpha) \quad (1)$$

dove  $x$  è il dato grezzo,  $x'$  il dato trasformato e  $\alpha$  una costante positiva il cui valore è riportato nella tabella allegata. L'aggiunta di una costante positiva garantisce la possibilità di effettuare il logaritmo di tutti i valori nella distribuzione, inclusi eventuali valori nulli.

### 3.2 Limite massimo

Per gli indicatori 26, 29 e 30 relativi alle disuguaglianze di genere è stato imposto come valore massimo quello corrispondente alla parità.

---

<sup>2</sup>L'interpolazione viene eseguita solo in presenza di valori mancanti preceduti e seguiti da un dato valido. Esempio: se sono presenti il dato del 2018 e del 2019, ma mancano tutti i successivi, l'ultimo dato viene semplicemente propagato in avanti agli anni seguenti.

<sup>3</sup>La sola differenza fra i due casi è che il dato d'area o nazionale è fornito direttamente da ISTAT, mentre la media pesata viene calcolata in fase di creazione dell'indice.

<sup>4</sup>In alcuni casi il dato provinciale è semplicemente mancante in altri proprio non esistente, come per l'indicatore 30 che è definito a livello regionale.

## 4 Normalizzazione

Tutti gli indicatori sono stati normalizzati mediante una trasformazione min-max con valori limite fissati singolarmente. Questi valori, riportati nella tabella allegata, sono stati a seconda dei casi o scelti sulla base degli scenari estremi a livello teorico o delle osservazioni minime e massime registrate nella serie temporale dell'indicatore. Questo tipo di normalizzazione, al contrario della trasformazione z-score utilizzata in precedenza, permette di tenere traccia dell'andamento assoluto e di confrontare i territori non solo nel singolo anno, ma anche nel tempo. Ogni indicatore risulta dunque riportato su una scala 0-100 orientata positivamente mediante la seguente trasformazione:

$$x' = \begin{cases} 100 \cdot \frac{x - x_{\min}}{x_{\max} - x_{\min}} & , \text{ se } x \text{ è orientato positivamente} \\ 100 \cdot \left(1 - \frac{x - x_{\min}}{x_{\max} - x_{\min}}\right) & , \text{ se } x \text{ è orientato negativamente} \end{cases} \quad (2)$$

dove  $x$  è il dato originale dell'indicatore,  $x_{\min}$  e  $x_{\max}$  i suoi valori limite,  $x'$  il punteggio normalizzato utilizzato per l'aggregazione.

## 5 Aggregazione

L'indice di ogni territorio è stato elaborato aggregando i punteggi dei suoi indicatori in tre diverse fasi. Innanzitutto sono calcolati i punteggi di ciascuna delle quindici dimensioni prendendo la media aritmetica dei punteggi dei due indicatori che la compongono. Successivamente, per evitare una piena compensabilità fra le dimensioni, il punteggio dei sottoindici è determinato dalla media geometrica delle dimensioni che ne fanno parte. La media geometrica è infine utilizzata anche per calcolare l'indice generale a partire dai tre sottoindici. Con un'aggregazione di questo tipo, una pessima prestazione in un aspetto giudicato fondamentale per l'inclusione non può venire del tutto o in parte compensato da un punteggio elevato in altre. Nello specifico i punteggi delle quindici dimensioni  $D_i$ , dei tre sottoindici  $S_j$  e dell'indice generale  $I$  sono calcolati come segue:

$$D_i = \frac{x_1 + x_2}{2} \quad (3a)$$

$$S_j = \sqrt[5]{D_1 \cdot D_2 \cdot D_3 \cdot D_4 \cdot D_5} \quad (3b)$$

$$I = \sqrt[3]{S_1 \cdot S_2 \cdot S_3} \quad (3c)$$

dove  $x_1$  e  $x_2$  sono i punteggi dei due indicatori in ciascuna dimensione,  $D_i$  è il punteggio di ognuna delle cinque dimensioni del sottoindice e  $S_j$  è il punteggio di ognuno dei tre sottoindici che compongono l'indice generale  $I$  di un territorio.

# Ulteriori precisazioni

## Punteggi d'area e punteggio nazionale

Le regioni e le due province autonome sono state raggruppate a fini statistici nelle cinque aree del NUTS 1:

- Nord-ovest: Piemonte, Valle d'Aosta, Liguria, Lombardia;
- Nord-est: Provincia Autonoma di Bolzano, Provincia Autonoma di Trento, Veneto, Friuli-Venezia Giulia, Emilia-Romagna;
- Centro: Toscana, Umbria, Marche, Lazio;
- Sud: Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria;
- Isole: Sicilia, Sardegna.

Ciascuna di queste aree e l'Italia intera sono state trattate come le singole regioni così da poterne calcolare l'indice con tutte le sue componenti. Qualora non già presenti nei dati originali, i valori d'area e il valore nazionale dei singoli indicatori sono stati ricavati da quelli regionali con una media pesata sulla popolazione residente nell'anno di riferimento.

## Gruppi d'inclusione

I territori sono stati suddivisi in sei gruppi d'inclusione in base al punteggio  $p$  ottenuto nell'indice generale e in ciascun sottoindice, seguendo i seguenti intervalli di valori:

Livello d'inclusione	Intervallo
Inclusione molto buona	$p \geq 85$
Inclusione buona	$75 \leq p < 85$
Inclusione sufficiente	$65 \leq p < 75$
Inclusione insufficiente	$55 \leq p < 65$
Esclusione grave	$45 \leq p < 55$
Esclusione molto grave	$p < 45$

Poiché la scala sottostante resta invariata, questa suddivisione permette di confrontare i gruppi tra i diversi anni.

# Tabelle

Tabella 1: Specifiche per la normalizzazione degli indicatori.

Indicatore	Orientamento negativo	Tipologia migliore	Tipologia peggiore	Valore migliore	Valore peggiore	$\alpha$	Limite massimo
1	sì	dato migliore	dato peggiore	8,93	27,2		no
2	sì	dato migliore	dato peggiore	327	697		no
3	sì	teorico	teorico	0	100		no
4	sì	teorico	teorico	0	100		no
5	no	teorico	teorico	100	0		no
6	no	dato migliore	teorico	29,7	0		no
7	sì	teorico	dato peggiore	0	1,99	1	no
8	sì	teorico	dato peggiore	0	22,4		no
9	sì	teorico	dato peggiore	0	0,78	1	no
10	sì	teorico	dato peggiore	0	66,8		no
11	sì	teorico	dato peggiore	0	41,7		no
12	no	dato migliore	dato peggiore	3,71	1,85		no
13	sì	teorico	teorico	0	100		no
14	sì	teorico	teorico	0	100		no
15	sì	teorico	dato peggiore	0	23,9		no
16	no	dato migliore	dato peggiore	63,3	2,66		no
17	no	dato migliore	teorico	74,9	0		no
18	no	dato migliore	teorico	51,9	0		no
19	sì	teorico	teorico	0	100		no
20	no	dato migliore	dato peggiore	5,1e+04	1,57e+04		no
21	no	teorico	teorico	100	0		no
22	no	dato migliore	dato peggiore	75,5	46,1		no
23	no	dato migliore	teorico	50,2	0		no
24	no	dato migliore	teorico	15,9	0		no
25	sì	teorico	dato peggiore	0	28,1		no
26	no	parità teorica	teorico	50	19,9		sì
27	no	dato migliore	teorico	49,5	0		no
28	no	teorico	teorico	100	0		no
29	no	parità teorica	teorico	50	13,6		sì
30	no	parità teorica	teorico	50	3,04		sì