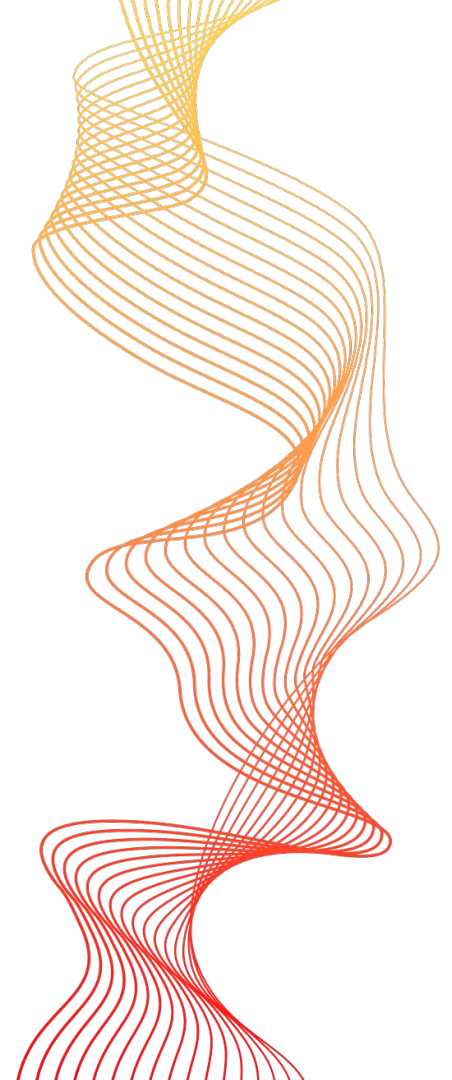




# Indicatori composti e normalizzazione

Questa presentazione spiega l'importanza della normalizzazione nella loro costruzione. Copre i diversi approcci all'aggregazione e i metodi di normalizzazione, con particolare attenzione all'impatto della scelta del metodo sui risultati.





# Introduzione agli indicatori composti

Gli indicatori composti sollevano la questione della compensabilità o sostituibilità dei singoli indicatori.

Gli indicatori sono sostituibili o non sostituibili in base al loro livello di compensabilità.

L'approccio di aggregazione può essere compensatorio, non compensativo o parzialmente compensativo.

Una questione fondamentale nella costruzione di indicatori composti è il grado di compensabilità o sostituibilità degli indicatori individuali.

Gli indicatori sono 'sostituibili' se un deficit in un indicatore può essere compensato da un eccesso in un altro( In altre parole, se una comunità ha un basso tasso di partecipazione all'attività fisica, questo può essere compensato da un alto tasso di adesione a una dieta sana, in modo da mantenere un livello generale di salute e benessere nella comunità), mentre sono 'non sostituibili' se un deficit in un indicatore non può essere compensato da un eccesso in un altro



## Approcci di aggregazione

- Gli approcci di aggregazione sono metodi utilizzati in diversi contesti per combinare le informazioni da più fonti in una singola misura sintetica
- La compensabilità è associata alla disuguaglianza tra gli indicatori.
- Gli approcci non compensativi o parzialmente compensativi bilanciano la dimensione del fenomeno e considerano la possibile disuguaglianza tra gli indicatori nelle funzioni di aggregazione.

gli approcci di aggregazione sono delle tecniche utilizzate per combinare diversi indicatori in un unico indice complessivo che rifletta le caratteristiche di un fenomeno. Esistono tre approcci di aggregazione: compensatorio, non compensatorio e parzialmente compensatorio (basato su compromesso tra i due approcci precedenti). La compensabilità è strettamente associata al concetto di 'disuguaglianza' (ovvero il fatto che alcuni indicatori possano essere più importanti di altri per la valutazione complessiva del fenomeno). In un approccio non compensatorio o parzialmente compensatorio, è necessaria un'equilibrio tra tutte le dimensioni del fenomeno considerate, e spesso si utilizza una funzione di aggregazione che tiene conto della disuguaglianza, in termini di penalizzazione. L'approccio compensatorio invece non considera la disuguaglianza tra gli indicatori individuali.

L'approccio da usare dipende dalla complessità del fenomeno e la necessità di tenere conto della disuguaglianza tra gli indicatori individuali



# Metodi di normalizzazione

- La normalizzazione rende gli indicatori comparabili e li porta sulla stessa scala di misurazione.
- La standardizzazione centra e allunga o accorcia la distribuzione degli indicatori per avere la stessa variabilità.
- Il riscalaggio porta tutti i valori in un intervallo comune ma perde un riferimento comune come la media.
- L'indicizzazione centra tutti gli indicatori intorno alla media ma non normalizza la loro variabilità.

La normalizzazione dei singoli indicatori è uno dei passaggi più importanti per la costruzione di un indicatore composito.

La normalizzazione mira a rendere gli indicatori comparabili e viene effettuata per portare gli indicatori a standard comuni.

Esistono diversi metodi di normalizzazione, tra cui Standardizzazione, Rescaling e Indicizzazione.





# Impatto della normalizzazione sulle corrispondenze

- La scelta del metodo di normalizzazione ha un forte impatto sulle corrispondenze tra indicatori e sulla definizione di equilibrio.
- La scelta errata del metodo di normalizzazione può portare a una distorsione inaccettabile dei risultati.

Il metodo di normalizzazione ha un forte impatto sui risultati perché crea un "sistema di corrispondenza" tra diversi indicatori.

Il sistema di corrispondenza definisce il livello di un indicatore che di solito corrisponde a livelli di altri indicatori. (ex'aspettativa di vita e il reddito pro capite, un aumento del reddito corrisponderà generalmente a un aumento dell'aspettativa di vita. )

Queste corrispondenze sono particolarmente importanti quando si segue un approccio non compensativo.

In tal caso, è necessario definire cosa si intende per "equilibrio". La definizione di equilibrio dipende dal metodo di normalizzazione adottato.

Ad esempio, se gli indicatori vengono convertiti in una scala comune con un intervallo compreso tra 0 e 1 (rescaling), allora il set di valori massimi e il set di valori minimi saranno considerati "bilanciati", mentre il set di valori medi potrebbe essere considerato "squilibrato". Al contrario, se gli indicatori vengono convertiti in una scala comune in cui il valore medio viene impostato a 100 (indicizzazione), il set di valori medi sarà considerato "bilanciato", mentre il set di valori massimi e il set di valori minimi potrebbero essere considerati "squilibrati".

Una scelta errata del metodo di normalizzazione può portare a una distorsione inaccettabile dei risultati.



# Griglia di Corrispondenza

E' un metodo utile per visualizzare i valori originali di diversi indicatori su una scala comune.

- Nel caso di normalizzazione, ogni metodo ha il suo concetto di bilanciamento, come :
- la distanza dalla media in termini di deviazioni standard nella standardizzazione.
- Nell' indicizzazione la distanza in percentuale dalla media.
- La normalizzazione mediante rescaling è l'unica che crea punti di corrispondenza fittizi. (maggiore è la differenza tra le distribuzioni degli indicatori, maggiore è la distorsione dei punti di corrispondenza.)

## SPIEGAZIONE DELLA GRIGLIA DELLA SLIDE SUCCESSIVA

Nel caso di normalizzazione, ogni metodo ha il suo concetto di bilanciamento, come : la distanza dalla media in termini di deviazioni standard nella standardizzazione. Nell' indicizzazione la distanza in percentuale dalla media. La normalizzazione mediante rescaling è l'unica che crea punti di corrispondenza fittizi. (maggiore è la differenza tra le distribuzioni degli indicatori, maggiore è la distorsione dei punti di corrispondenza.) L'indicizzazione considera bilanciato un insieme di valori quando sono "equidistanti" dalla media (la base) in termini percentuali. in rescaling (0.1,0.5,0.8), in indicizzazione anche quelli doppio la media

Table 7.1 – *Correspondence grid for different normalization methods*

Standardization				Rescaling				Indicization			
Scale	Exp	Nor	Bet	Scale	Exp	Nor	Bet	Scale	Exp	Nor	Bet
2.5	128.5	187.6	119.4	1.0	250.0	209.5	108.4	200	160.0	300.0	200.0
2.0	118.8	180.1	115.5	0.9	231.0	197.7	103.9	180	144.0	270.0	180.0
1.5	109.1	172.6	111.6	0.8	212.1	185.8	<b>99.4</b>	160	128.0	240.0	160.0
1.0	99.4	165.0	107.8	0.7	193.1	174.0	94.9	140	112.0	210.0	140.0
0.5	89.7	157.5	103.9	0.6	174.2	162.2	90.5	120	96.0	180.0	120.0
<b>0.0</b>	<b>80.0</b>	<b>150.0</b>	<b>100.0</b>	0.5	155.2	<b>150.3</b>	86.0	<b>100</b>	<b>80.0</b>	<b>150.0</b>	<b>100.0</b>
-0.5	70.3	142.5	96.1	0.4	136.3	138.5	81.5	80	64.0	120.0	80.0
-1.0	60.6	135.0	92.2	0.3	117.3	126.7	77.0	60	48.0	90.0	60.0
-1.5	50.9	127.4	88.4	0.2	98.4	114.8	72.5	40	32.0	60.0	40.0
-2.0	41.2	119.9	84.5	0.1	<b>79.4</b>	103.0	68.1	20	16.0	30.0	20.0
-2.5	31.5	112.4	80.6	0.0	60.5	91.2	63.6	0	0.0	0.0	0.0



```
{R}  
set.seed(22)
```

```
# Genera i dati
```

```
n <- 1000
```

```
exp_data <- rexp(n, rate = 1)
```

#La funzione `rexp(n, rate)` genera un campione di `n` numeri casuali estratti da una distribuzione esponenziale con parametro di scala `rate`. La distribuzione esponenziale è una distribuzione di probabilità continua che descrive il tempo tra gli eventi di un processo di Poisson

```
norm_data <- rnorm(n, mean = 5, sd = 2)
```

#La funzione `rnorm(n, mean, sd)` genera un campione di `n` numeri casuali estratti da una distribuzione normale con media `mean` e deviazione standard `sd`.

```
beta_data <- rbeta(n, shape1 = 2, shape2 = 5)
```

#La funzione `rbeta(n, shape1, shape2)` genera un campione di `n` numeri casuali estratti da una distribuzione beta con parametri di forma `shape1` e `shape2`. La distribuzione beta è una distribuzione di probabilità continua che descrive la variabilità di una proporzione tra 0 e 1.

```
data <- data.frame(exp_data, norm_data, beta_data)
```

```
...
```

con `rate=1` ci si aspetta che un evento si verifichi in media una volta ogni unità di tempo  
`shape1` e `shape2` sono spesso chiamati "alpha" e "beta" e rappresentano il numero di successi e insuccessi osservati in una serie di Bernoulli.



```
{r}
# Funzione per standardizzare
standardize <- function(x) {
  (x - mean(x)) / sd(x)
}
# calcola la media e la deviazione standard del vettore x e restituisce un vettore
in cui ogni elemento è standardizzato sottraendo la media e dividendo per la
deviazione standard. La standardizzazione trasforma i dati in modo che abbiano una
media di zero e una deviazione standard di uno

# Funzione per il rescaling
rescale <- function(x) {
  (x - min(x)) / (max(x) - min(x))
}
#calcola il minimo e il massimo del vettore x e restituisce un nuovo vettore in
cui ogni elemento è rescalato in modo che cada tra 0 e 1. La rescalatura trasforma
i dati in modo che abbiano la stessa scala di valori

# Funzione per l'indicizzazione
index <- function(x, base = 100) {
  (x / mean(x)) * base
}
#calcola la media del vettore x e restituisce un nuovo vettore in cui ogni
elemento è indicizzato rispetto alla media, moltiplicando ogni elemento per un
valore di base specificato. L'indicizzazione trasforma i dati in modo che siano
espressi come una deviazione rispetto alla media

#l'obiettivo è rendere comparabili i nostri dati con quelli di altre distribuzioni
...
```

```

# {r}
# Applica le funzioni di normalizzazione
data_std <- as.data.frame(apply(data, 2, standardize))
data_rescaled <- as.data.frame(apply(data, 2, rescale))
data_indexed <- as.data.frame(apply(data, 2, index))
...

```

Description: df [1,000 × 3]

exp_data <dbl>	norm_data <dbl>	beta_data <dbl>
-0.0767818690	0.5415405539	1.505479787
0.0783647021	as.data.frame(x, row.names = NULL, optional = FALSE, ...)	
-0.3182622505	-0.1045095721	-0.171049506
-0.5770509387	1.7298690369	-0.535689391
-0.8104391104	-2.3842322925	0.783681576
-0.5428308443	0.5072733695	0.815879027
0.4107838841	-0.4764971059	0.106100274
0.2182189275	-0.1620933338	-0.270664127
-0.7047013073	-1.0135256100	-0.595141965
-0.2495167320	-1.5005087693	1.388874885

1-10 of 1,000 rows

Previous 1 2 3 4 5 6 ... 100 Next

```

library(gridExtra)
library(ggplot2)
library(tidyr)

plot_distributions <- function(data, title) {
  # Trasforma le colonne di dati in formato lungo
  data_long <- pivot_longer(data, cols = c("exp_data", "norm_data", "beta_data"),
                             names_to = "tipo", values_to = "valore")

  # Crea il grafico delle distribuzioni
  ggplot(data_long, aes(x = valore, color = tipo, group = tipo)) +
    geom_density() +
    labs(title = title, x = "valore", y = "Densità") +
    theme_minimal()
}

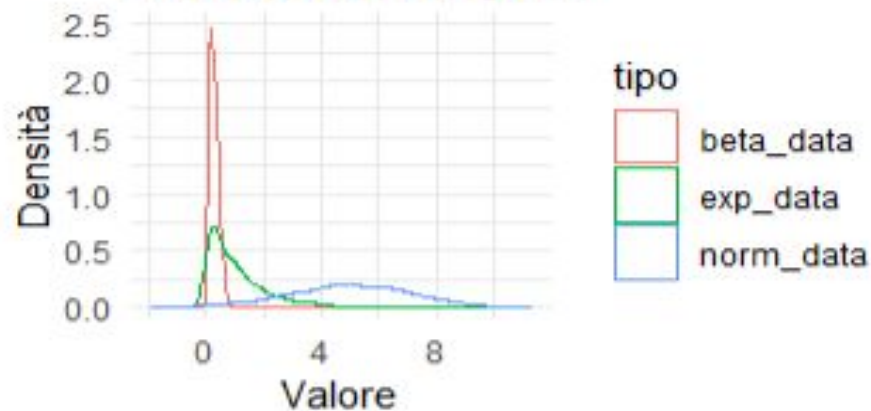
# Grafici delle distribuzioni originali e normalizzate
plot_original <- plot_distributions(data, "Distribuzioni originali")
plot_std <- plot_distributions(data_std, "Distribuzioni standardizzate")
plot_rescaled <- plot_distributions(data_rescaled, "Distribuzioni riscalate")
plot_indexed <- plot_distributions(data_indexed, "Distribuzioni indicizzate")

# visualizza i grafici
grid.arrange(plot_original, plot_std, plot_rescaled, plot_indexed, ncol = 2)

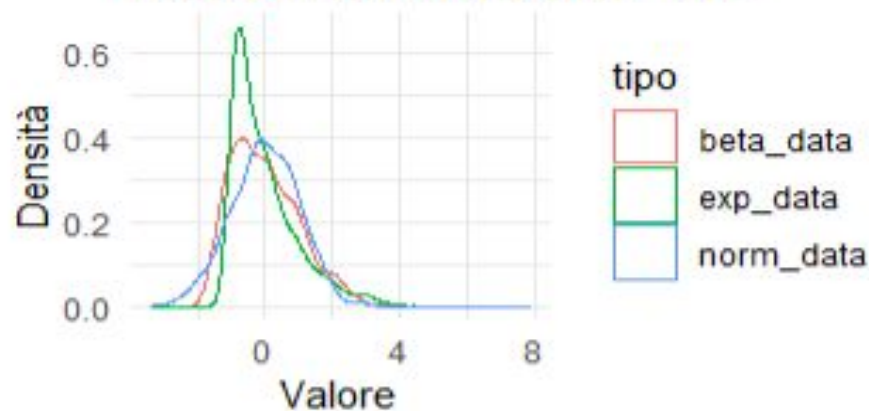
```



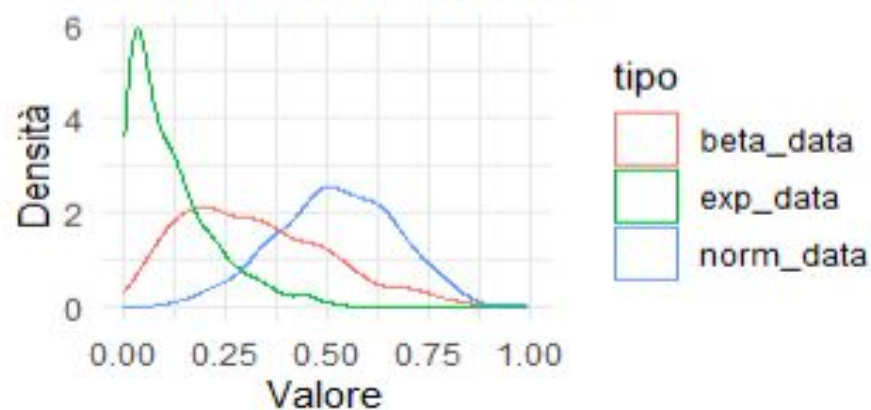
### Distribuzioni originali



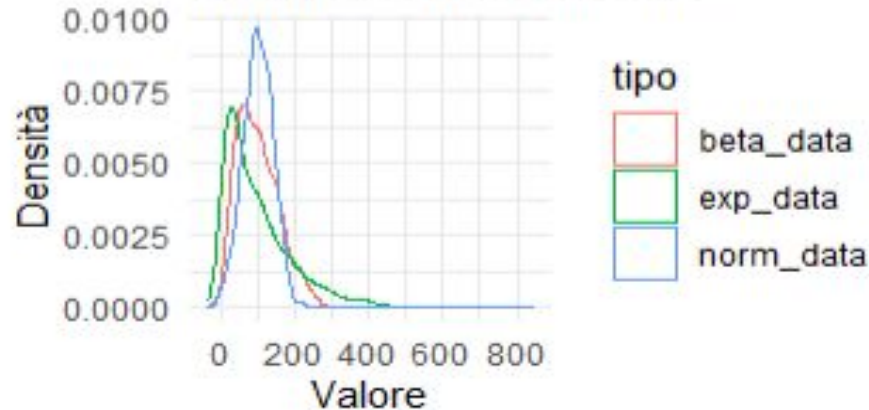
### Distribuzioni standardizzate



### Distribuzioni riscalate



### Distribuzioni indicizzate





## importanza di scegliere un metodo di normalizzazione appropriato

- La scelta del metodo di normalizzazione è fondamentale per evitare risultati distorti nell'aggregazione degli indicatori.
- Il metodo scelto dovrebbe essere appropriato per il tipo di dati e l'approccio di aggregazione adottato.
- La costruzione di un indicatore composito deve seguire un preciso paradigma di lavoro e il "modello di bilanciamento" deve essere definito con attenzione.

### SPIEGAZIONE DEI GRAFICI IN R DELLA SLIDE PRECEDENTE

La standardizzazione 'centra' le distribuzioni degli indicatori attorno all'origine (media = 0) e le 'allunga' o 'accorcia' in modo che abbiano la stessa variabilità (varianza = 1).

La rescaling rende le varianze più omogenee, portando tutti i valori in un intervallo comune. Tuttavia, le distribuzioni degli indicatori non sono 'centrate' e ciò comporta la perdita di un valore di riferimento comune, come la media. Ciò significa che valori normalizzati uguali possono corrispondere a valori originali molto diversi.

L'indicizzazione con la media come base impostata a 100 'centra' tutte le distribuzioni attorno alla media, ma non 'normalizza' la loro variabilità. Gli indicatori indicizzati hanno gli stessi coefficienti di variazione degli indicatori originali.

può sembrare che lo faccia perché la scala degli indicatori viene modificata in modo che la media sia sempre uguale a 100, il che può rendere le distribuzioni più confrontabili



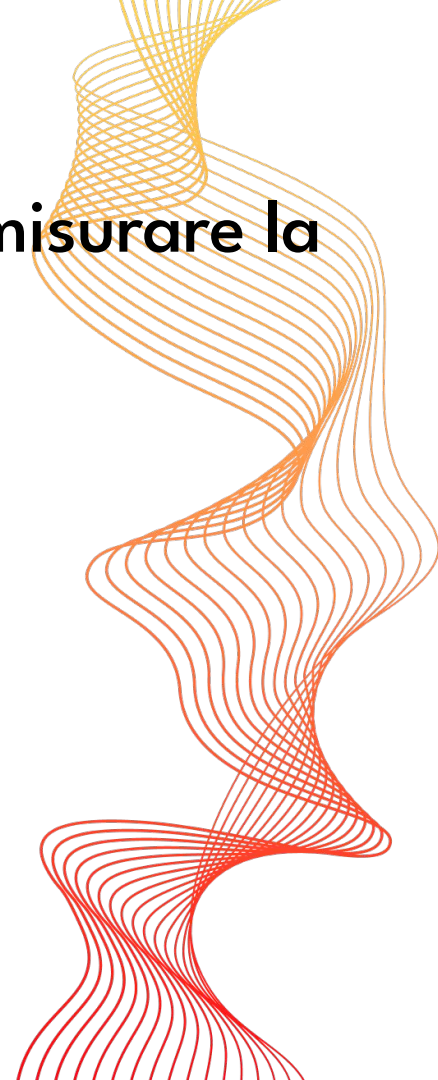
# Comprendere gli indicatori compositi e misurare la disuguaglianza

Una panoramica del metodo non compensativo o parzialmente compensativo degli indicatori compositi e della sua relazione con la misurazione della disuguaglianza.

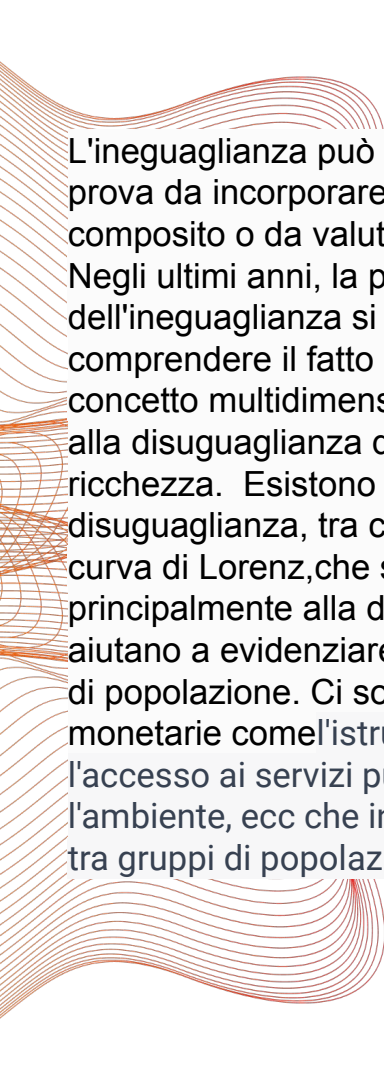


L'ineguaglianza, definita come *"la situazione ingiusta nella società quando alcune persone hanno più opportunità, denaro, ecc. rispetto ad altre persone"*

(CUP, 2008)








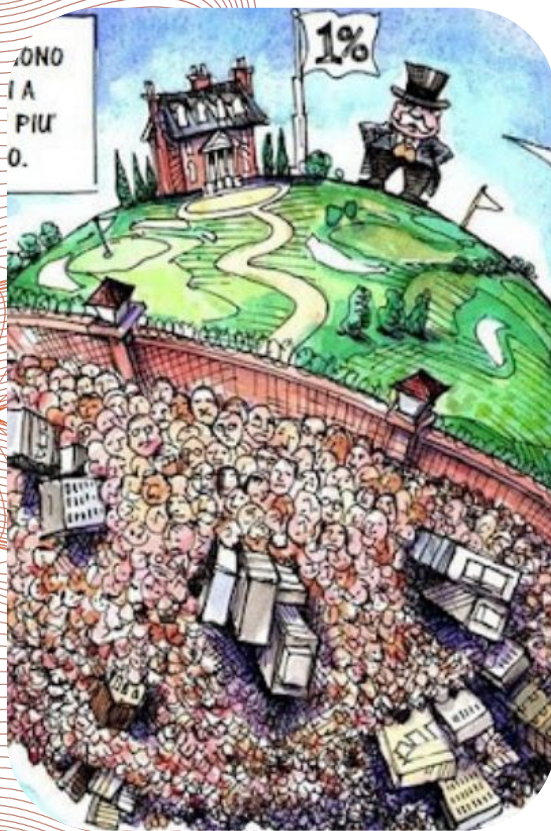
L'ineguaglianza può essere valutata come prova da incorporare in un indicatore composito o da valutare autonomamente. Negli ultimi anni, la percezione dell'ineguaglianza si è evoluta per comprendere il fatto che l'ineguaglianza è un concetto multidimensionale che non si limita alla disuguaglianza dei redditi o della ricchezza. Esistono diverse misure di disuguaglianza, tra cui l'indice di Gini e la curva di Lorenz, che si applicano principalmente alla distribuzione del reddito e aiutano a evidenziare le differenze tra i gruppi di popolazione. Ci sono inoltre risorse non monetarie come l'istruzione, la salute, il genere, l'accesso ai servizi pubblici, l'occupazione, l'ambiente, ecc che influenzano le differenze tra gruppi di popolazioni



## Misurare la disuguaglianza e il benessere

- La percezione della disuguaglianza è cambiata riconoscendo la multidimensionalità del concetto.
  - Può essere misurato come parte di un indicatore composito o come valutazione autonoma.
  - Le misure tradizionali della disuguaglianza sono l'indice di Gini e la curva di Lorenz, ma l'approccio si è esteso alle risorse e alle opportunità non monetarie.
- 

# Disuguaglianza Verticale e dispersione verticale



- La disuguaglianza verticale si riferisce alla differenza di posizione economica e sociale tra individui all'interno di una data società. Questa forma di disuguaglianza si basa sulla distribuzione del reddito, della ricchezza, del potere e delle opportunità.
- La dispersione verticale si riferisce alla variazione tra i valori degli indicatori compositi tra i diversi gruppi di individui o paesi che si trovano a diversi livelli di sviluppo



## Disuguaglianza orizzontale e dispersione orizzontale

- La disuguaglianza orizzontale si riferisce alle differenze tra individui in relazione alle loro caratteristiche personali come l'età, il genere, l'etnia, la religione, l'orientamento sessuale, la salute, l'istruzione e così via.
- La Dispersione orizzontale si riferisce alla variazione tra le modalità delle varie dimensioni che compongono l'indicatore composito.





## Atkinson Misure di disuguaglianza

- La misura di Atkinson confronta una media generale "sensibile alla base" e la media aritmetica "neutra".
- Tiene conto della distribuzione relativa dei singoli valori, riflettendo concentrazioni irregolari di dati e tenendo conto della loro variabilità.

$A(\varepsilon) = 0$  per la perfetta uguaglianza e  
 $A(\varepsilon) = 1$  per la massima disuguaglianza.

concentrazione relativa dei  
dati e la variabilità

$$A(\varepsilon) = 1 - [M(1-\varepsilon)/M1] \forall \varepsilon > 0$$

parametro  
di  
sensibilità,

media  
generalizzata  
di ordine  $q$

media  
aritmetica

Questo indice confronta una media generale "sensibile alla base" (è una media ponderata dei valori di una variabile) con una media aritmetica "neutra" (che corrisponde alla media dei valori della variabile se fossero distribuiti in modo uniforme) per valutare la distribuzione relativa dei valori individuali.

L'indice di sviluppo umano (HDI) è una misura dello sviluppo umano sviluppata dal Programma di sviluppo delle Nazioni Unite (UNDP) per valutare il benessere e la qualità della vita delle persone in diversi paesi. Viene calcolato utilizzando tre dimensioni chiave: una vita lunga e sana, l'accesso alla conoscenza e uno standard di vita dignitoso.

La prima dimensione, una vita lunga e sana, è misurata dall'aspettativa di vita alla nascita. Questo è un indicatore della salute generale di una popolazione, nonché della qualità dei servizi sanitari e dell'accesso all'acqua pulita e ai servizi igienico-sanitari.

La seconda dimensione, l'accesso alla conoscenza, è misurata da due indicatori: anni di scolarizzazione attesi e anni medi di scolarizzazione. Gli anni di scolarizzazione previsti si riferiscono al numero di anni che un bambino in età scolare può aspettarsi di ricevere un'istruzione, mentre gli anni medi di scolarizzazione si riferiscono al numero medio di anni di istruzione che gli adulti di età pari o superiore a 25 anni hanno ricevuto.

La terza dimensione, uno standard di vita dignitoso, è misurata dal reddito nazionale lordo (RNL) pro capite, aggiustato per la parità di potere d'acquisto (PPP). Questo indicatore riflette il benessere economico di un paese ed è calcolato dividendo il reddito totale di un paese per la sua popolazione, aggiustato per le differenze di potere d'acquisto tra i paesi.

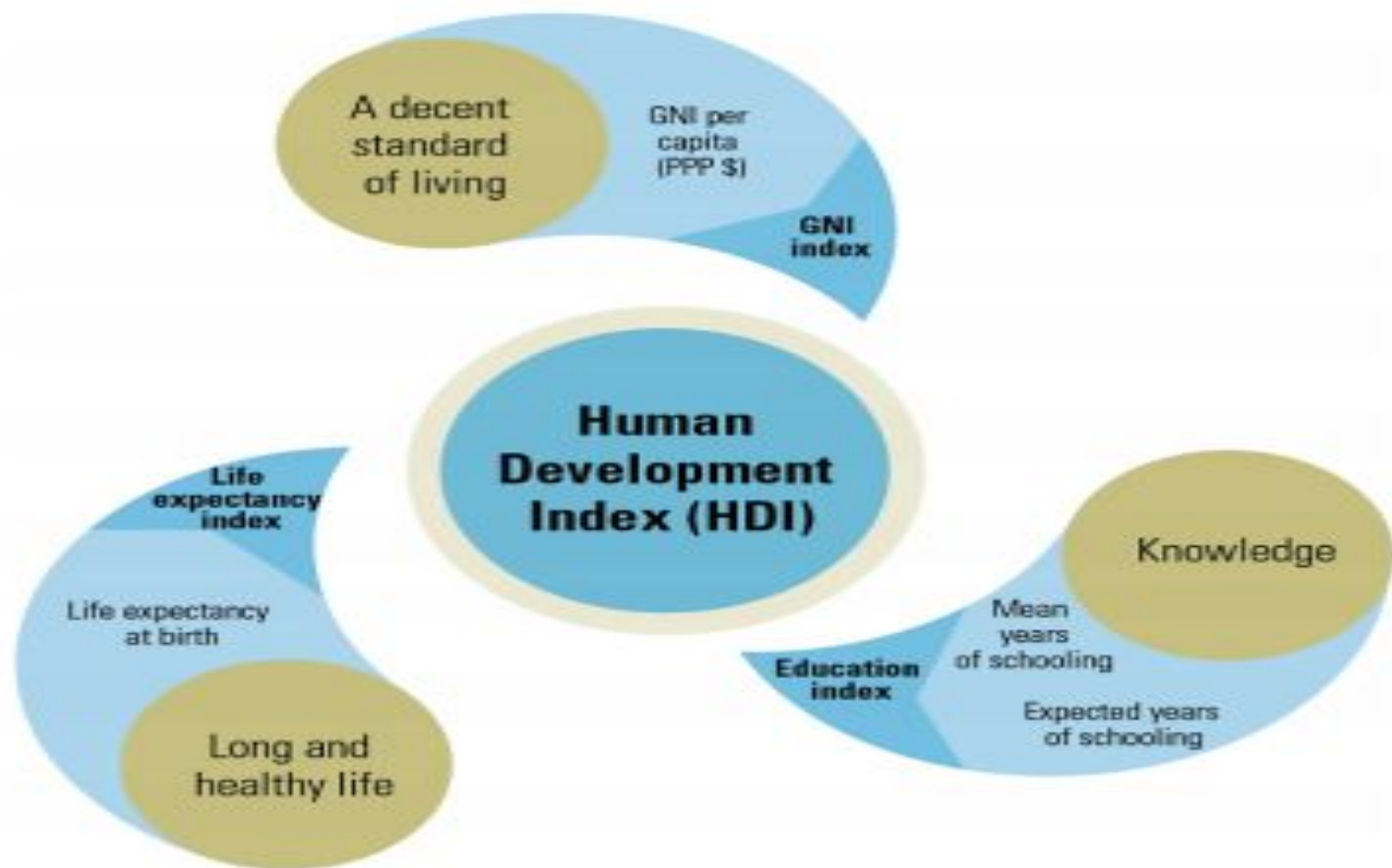
L'ISU è calcolato prendendo la media geometrica di queste tre dimensioni, ciascuna delle quali ha lo stesso peso. L'indice risultante varia da 0 a 1, con un valore più alto che indica livelli più elevati di sviluppo umano. L'UNDP calcola anche i sottoindici per ciascuna dimensione, nonché un HDI aggiustato per la disuguaglianza (IHDI) che tiene conto delle disparità nello sviluppo umano all'interno dei paesi.

## Disuguaglianza e benessere

- La disuguaglianza svolge un ruolo significativo nella misurazione del benessere.
- L'indice di sviluppo umano (HDI) è un indicatore composito che incorpora la disuguaglianza nella sua metodologia.
- I tre pilastri principali dell'indice sono l'indice di aspettativa di vita (LE), l'indice di istruzione (EI) e l'Indice del reddito nazionale pro capite lordo (LI).



*The three pillars of the Human Development Index (UNDP, 2020)*







## Miglioramenti all'HDI

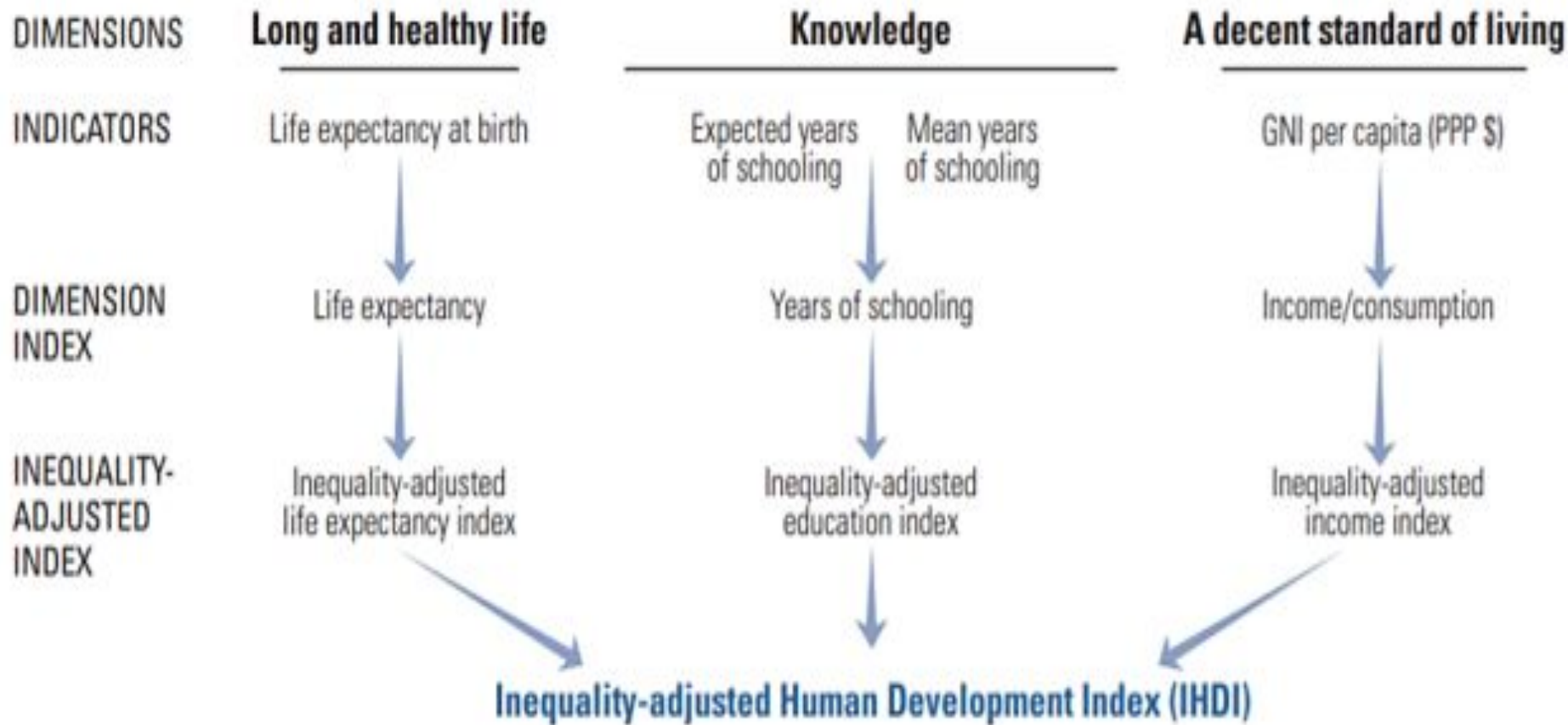
- (IHDI) è stato calcolato aggiustando l'HDI originale con il valore dell'Income Gini Index (GI).

GI è interpretata come la perdita del benessere in percentuale del massimo benessere raggiungibile, causato dalla disparità di reddito

- IHDI corregge la disuguaglianza all'interno di ciascuno dei tre pilastri dell'IHDI, aspettativa di vita, istruzione e reddito, utilizzando i dati granulari disponibili per ciascuna dimensione.

Per calcolare l'IHDI, l'HDI viene corretto utilizzando il valore dell'indice di Gini del reddito, che è una misura della disuguaglianza di reddito in una società. In particolare, il valore dell'indice di Gini viene utilizzato per calcolare la perdita di benessere in percentuale del massimo benessere raggiungibile, causato dalla disuguaglianza di reddito. corregge anche la disuguaglianza all'interno di ciascuno dei tre pilastri dell'IHDI, ovvero l'aspettativa di vita, l'istruzione e il reddito, utilizzando i dati disponibili per ciascuna dimensione. L'IHDI riflette non solo il livello di sviluppo umano di una società, ma anche il livello di disuguaglianza all'interno della società

*The structure of the IHDI (UNDP, 2019)*





## SPIEGAZIONE SLIDE PRECEDENTE

L'IHDI (Indice di sviluppo umano aggiustato per la disuguaglianza) è una misura dello sviluppo umano che tiene conto sia dei risultati che delle disuguaglianze nella distribuzione di tali risultati. Viene calcolato aggiustando l'indice di sviluppo umano (HDI) per la disuguaglianza in ciascuna delle sue tre dimensioni: vita lunga e sana, istruzione e tenore di vita.

La dimensione della vita lunga e sana è misurata dall'aspettativa di vita alla nascita. La dimensione dell'istruzione è misurata da due indicatori: anni di scolarizzazione attesi e anni medi di scolarizzazione. La dimensione del tenore di vita è misurata dal reddito nazionale lordo (RNL) pro capite (PPP \$), che viene aggiustato per le differenze di potere d'acquisto tra i paesi.

L'IHDI viene calcolato aggiustando ciascuno di questi indicatori per la disuguaglianza utilizzando una formula che tiene conto della distribuzione di ciascun indicatore tra la popolazione. Gli indici corretti per la disuguaglianza risultanti vengono quindi combinati per formare l'IHDI, che varia da 0 a 1, con valori più alti che indicano un livello più elevato di sviluppo umano con meno disuguaglianza.

L'IHDI fornisce una visione più sfumata dello sviluppo umano rispetto al solo HDI, tenendo conto della distribuzione dei risultati dello sviluppo umano all'interno di una società. Evidenziando le disuguaglianze, può aiutare i responsabili politici a identificare dove sono necessari interventi per garantire che tutti i membri della società traggano beneficio dal progresso dello sviluppo umano



## Limitazioni di IHDl

- IHDl non è sensibile all'associazione, in quanto non cattura le disuguaglianze che si sovrappongono tra le sue dimensioni.
- Gli indicatori alternativi corretti per la disuguaglianza, come la media ponderata basata su Gini(GW) e l'indice di felicità mondiale, hanno anche loro i loro svantaggi.

La GW(Gini Weight) è influenzata dalla scelta delle unità di misura e dalla rappresentatività del campione.

Il WHI si concentra sulla soddisfazione della vita e non tiene conto di altre dimensioni del benessere, come la salute e l'istruzione

### World Happiness Index:

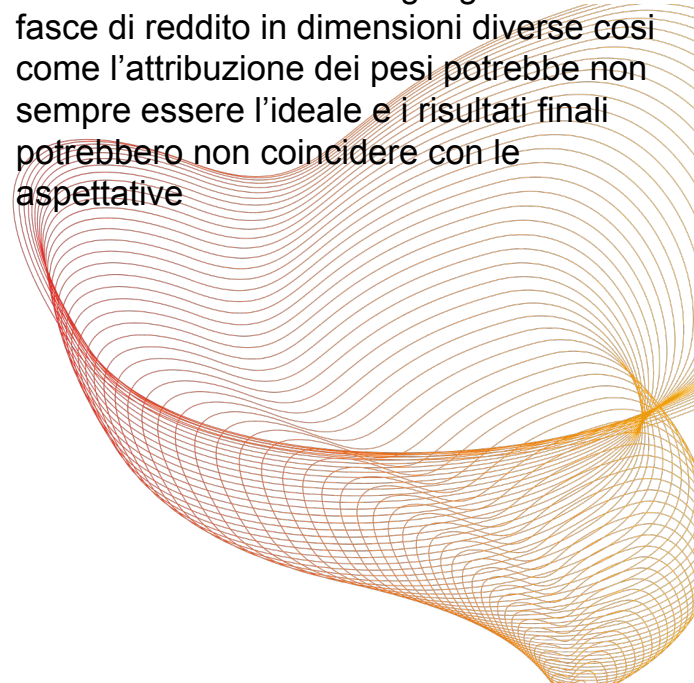
$$QL = GDP\ rank + GINI\ rank + LIFE\ rank + SUIC\ rank + AIR\ rank$$




## sfide nella misurazione della disuguaglianza multidimensionale

- Sebbene l'integrazione della disuguaglianza negli indicatori compositi sia stata un'area promettente, permangono sfide nel catturare la disuguaglianza multidimensionale in modo completo.

è impossibile calcolare la disuguaglianza su tutte le dimensioni perchè è possibile concentrarsi solo su un determinato numero limitato di dimensioni o variabili di interesse. Ci possono essere problemi nel confrontare diverse aree geografiche o fasce di reddito in dimensioni diverse così come l'attribuzione dei pesi potrebbe non sempre essere l'ideale e i risultati finali potrebbero non coincidere con le aspettative





## sviluppi futuri delle misure di disuguaglianza

- È necessario sviluppare e perfezionare ulteriormente gli indicatori per catturare in modo completo la disuguaglianza multidimensionale.





# metodi per affrontare la disuguaglianza tra le dimensioni

- Sono stati proposti diversi metodi per affrontare la disuguaglianza tra le dimensioni. PMI(Povertà Multidimensionale)
- Approccio di conteggio che combina gli elementi di realizzazione a livello individuale.
- Metodo di Alkire-Foster e Mazziotta-Pareto basato su doppi cutoff per costruire un indice di realizzazione multidimensionale o un indice di deprivazione multidimensionale(povertà).
- Metodo Mazziotta e Pareto basato sulla penalizzazione delle unità che hanno maggiore variabilità tra le diverse dimensioni dell'indicatore composito.

$$MPI_i = Mzi - Szi \cdot CVi$$

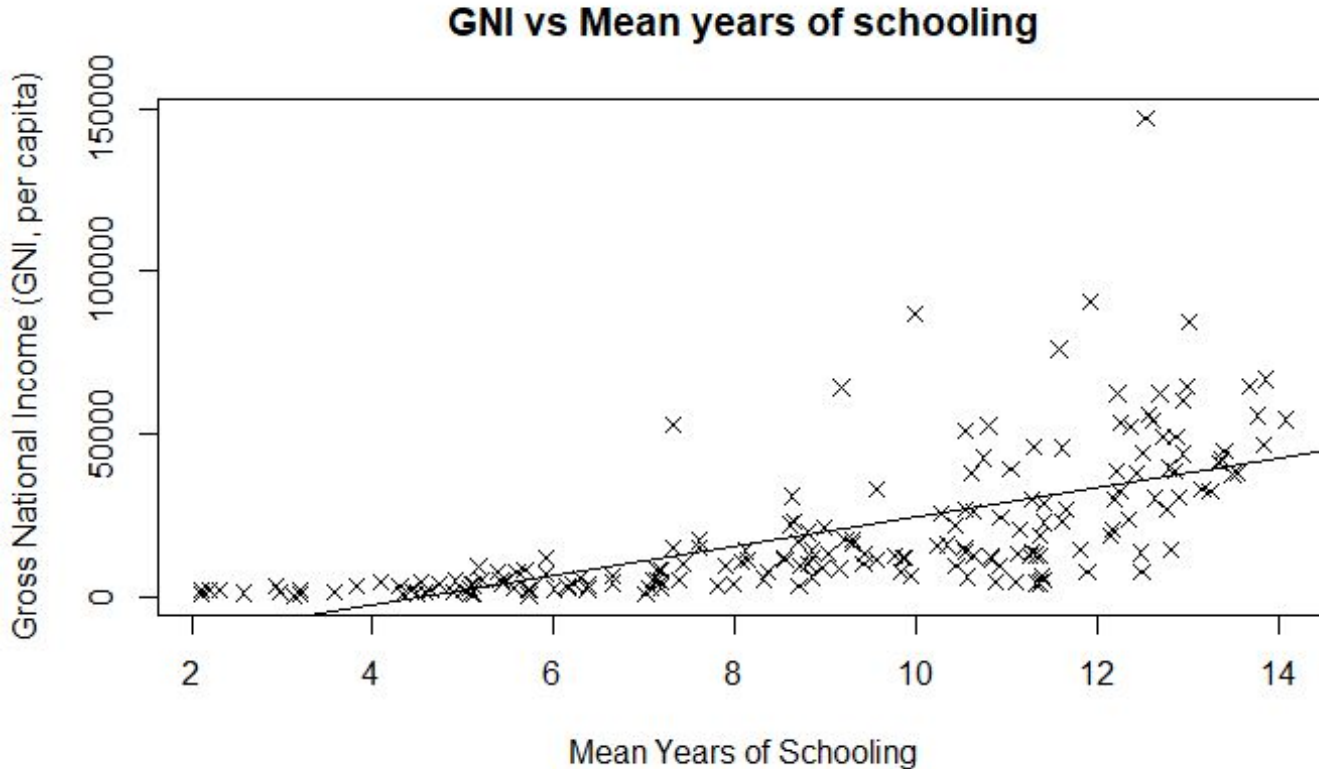
si utilizza una lista di elementi di realizzazione (come avere accesso all'acqua potabile, all'elettricità, all'istruzione, ecc.) per ogni individuo e si verifica quali di questi elementi sono presenti o assenti per ciascun individuo. Si considera che un individuo è povero se ha un numero minimo di elementi di realizzazione mancanti.

Le unità statistiche che non soddisfano la soglia di non-deprivazione per almeno una dimensione sono considerate deprivate. si utilizza una funzione di aggregazione per combinare le informazioni delle diverse dimensioni dell'indicatore composito e calcolare l'indice di realizzazione  
FORMULA MPI: i singoli indicatori vengono convertiti in una scala comune con media 100 e deviazione standard 10. valore medio, la deviazione standard e il coefficiente di variazione per l'unità  $i$ . dove  $s$  e  $C$  sono penalità. L'obiettivo è quello di premiare le unità che presentano un maggiore equilibrio tra i valori degli indicatori.

HDI rank	Country	Human Development Index (HDI)	Life expectancy at birth	Expected years of schooling	Mean years of schooling	Gross national income (GNI) per capita
1	Switzerland	0,962	84,0	16,5	13,9	66.933
2	Norway	0,961	83,2	18,2	13,0	64.660
3	Iceland	0,959	82,7	19,2	13,8	55.782
4	Hong Kong, China (SAR)	0,952	85,5	17,3	12,2	62.607
5	Australia	0,951	84,5	21,1	12,7	49.238
6	Denmark	0,948	81,4	18,7	13,0	60.365
7	Sweden	0,947	83,0	19,4	12,6	54.489
8	Ireland	0,945	82,0	18,9	11,6	76.169
9	Germany	0,942	80,6	17,0	14,1	54.534
10	Netherlands	0,941	81,7	18,7	12,6	55.979
11	Finland	0,940	82,0	19,1	12,9	49.452
12	Singapore	0,939	82,8	16,5	11,9	90.919
13	Belgium	0,937	81,9	19,6	12,4	52.293
13	New Zealand	0,937	82,5	20,3	12,9	44.057
15	Canada	0,936	82,7	16,4	13,8	46.808
16	Liechtenstein	0,935	83,3	15,2	12,5	146.830
17	Luxembourg	0,930	82,6	14,4	13,0	84.649
18	United Kingdom	0,929	80,7	17,3	13,4	45.225
19	Japan	0,925	84,8	15,2	13,4	42.274
19	Korea (Republic of)	0,925	83,7	16,5	12,5	44.501
21	United States	0,921	77,2	16,3	13,7	64.765
22	Israel	0,919	82,3	16,1	13,3	41.524



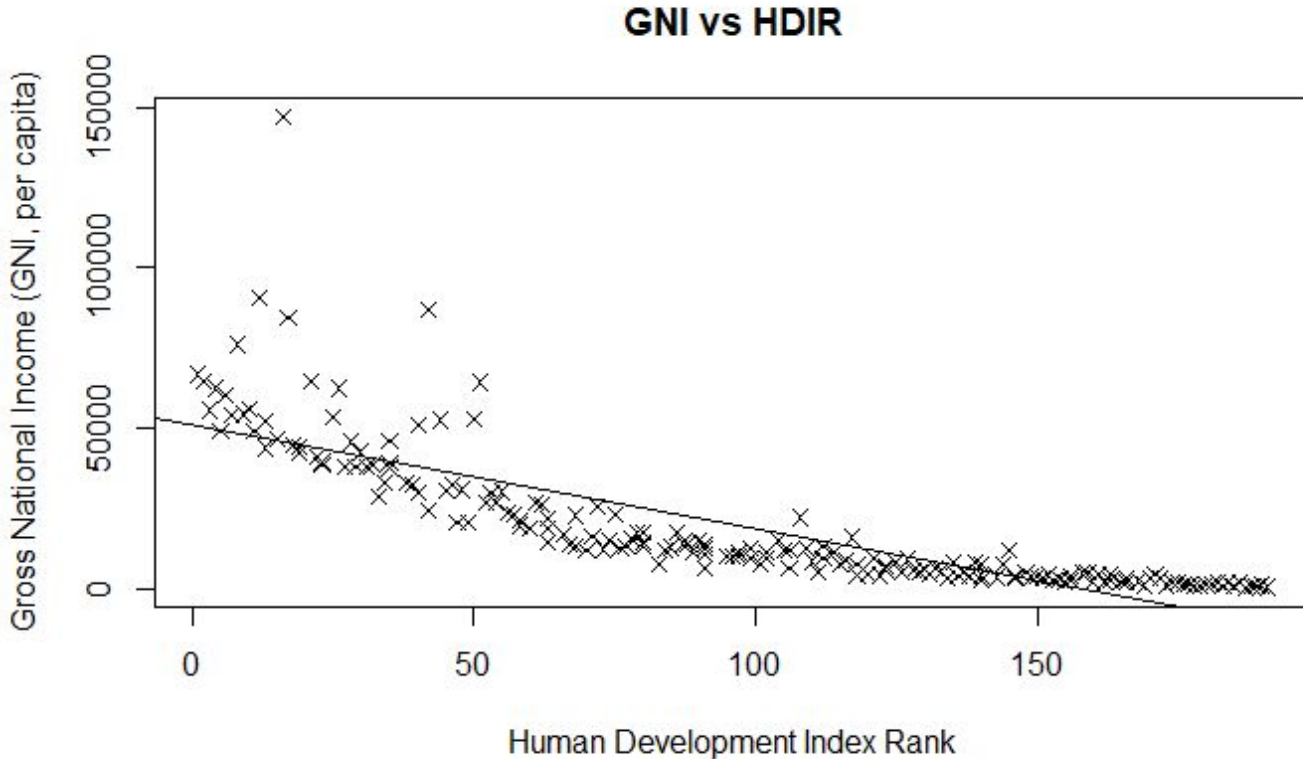
## Esiste una relazione tra anni medi di scolarizzazione e reddito nazionale lordo pro capite?



Gradient of  
the line is:  
0.6507158183  
74473



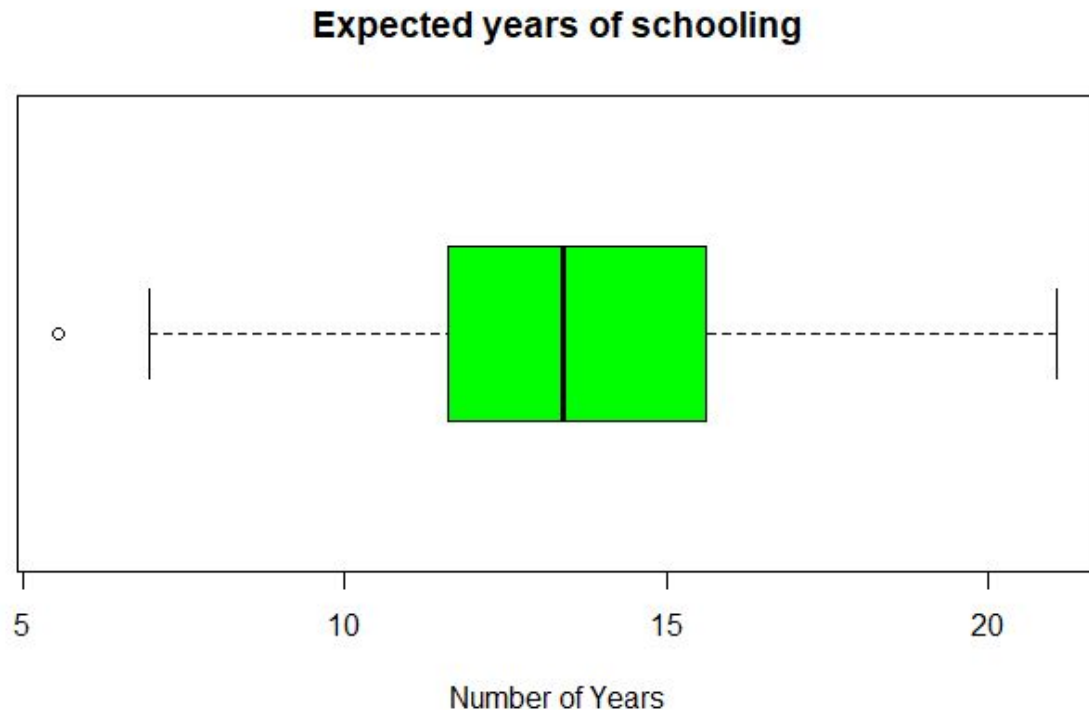
## Esiste una relazione tra HDI rank e reddito nazionale lordo pro capite?



Gradient of  
the line is:  
 $-0.816698747$   
813642



## ## Quanta differenza c'è negli anni previsti di scolarizzazione?

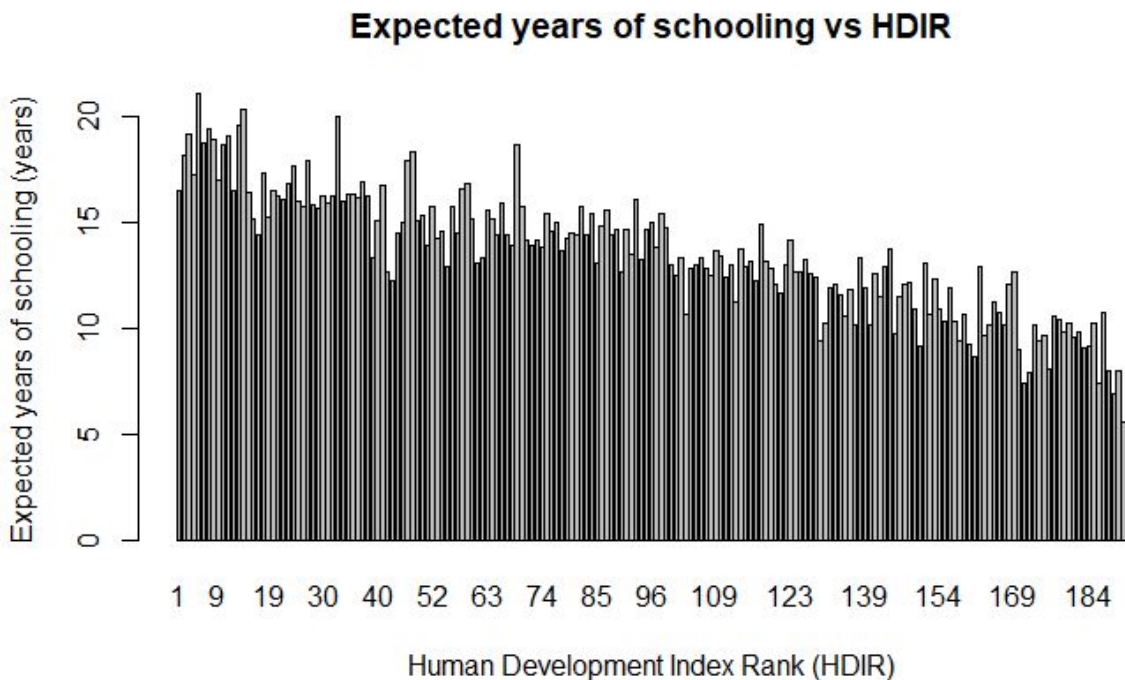


The standard deviation is:  
2.9239114851838  
3

Min. 1st Qu.  
5.543 11.601  
Median Mean  
13.405 13.535  
3rd Qu. Max.  
15.624 21.055



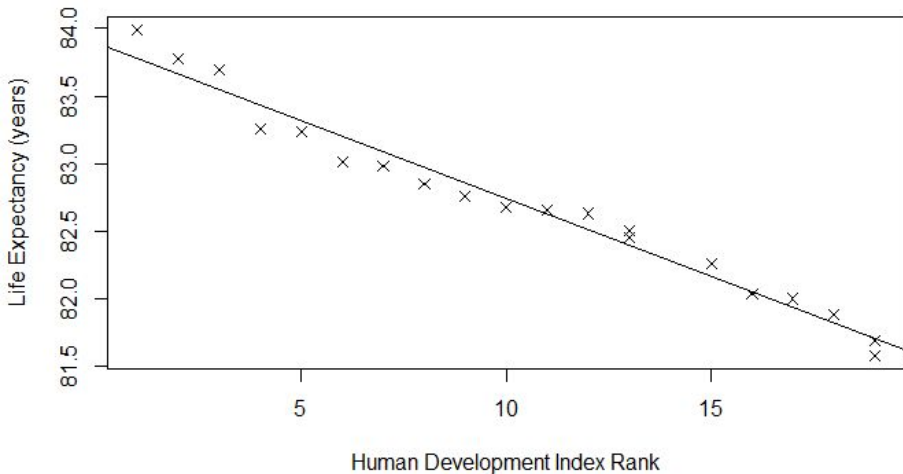
## ## Quanta differenza c'è negli anni previsti di scolarizzazione?



Min.	1st Qu.
52.53	65.75
Median	Mean
71.69	71.31
3rd Qu.	Max.
76.70	85.47

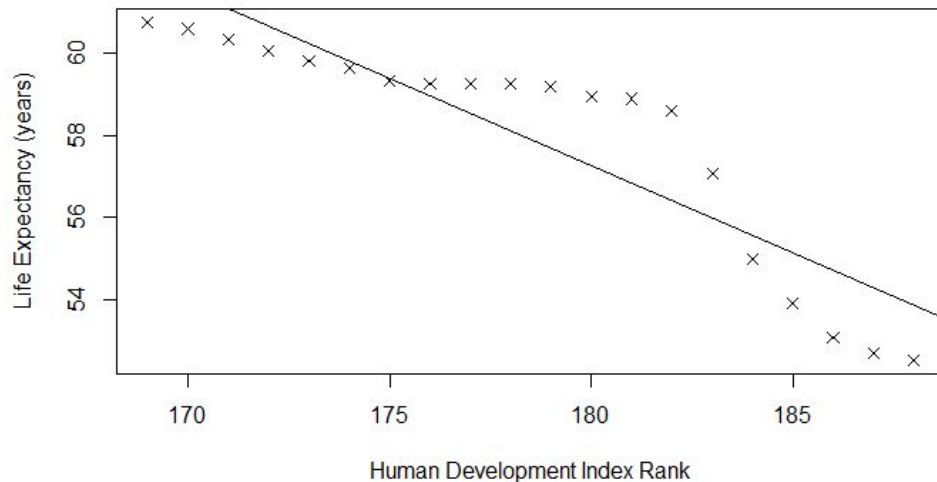


Life expectancy (top 20)



Correlation between life expectancy and HDIR for top 20 countries is:  
 $-0.985495270102488$

Life expectancy (bottom 20)



Correlation between life expectancy and HDIR for bottom 20 countries is:  
 $-0.895627132601915$

$$1. \text{ Life Expectancy Index (LEI)} = \frac{LE - 20}{85 - 20}$$

LE: Life expectancy at birth

$$2. \text{ Education Index (EI)} = \frac{MYSI + EYSI}{2}$$

MYSI:

$$\text{Mean Years of Schooling Index} = \frac{\text{MYS (Mean Years of Schooling)}}{15}$$

EYSI:

$$\text{Expected Years of Schooling Index} = \frac{\text{EYS (Expected Years of Schooling)}}{18}$$

$$3. \text{ Income Index (II)} = \frac{\ln(\text{GNIpc}) - \ln(100)}{\ln(75,000) - \ln(100)}$$

GNIpc: Gross national income at purchasing power parity per capita

$$\text{HDI} = \sqrt[3]{\text{LEI} \cdot \text{EI} \cdot \text{II}}$$

```

LE_min=20
LE_max=85
LEI <- (HDI$`Life expectancy at birth` - LE_min)/(LE_max - LE_min)
MYSI <- HDI$`Mean years of schooling`/15
EYSI <- HDI$`Expected years of schooling`/18
ED <- (MYSI+EYSI)/2
II <- (log(HDI$`Gross national income (GNI) per capita`) - log(100))/(log(75000) - log(100))
HDI_calcolato <- (LEI*ED*II)^(1/3)
as.data.frame(HDI_calcolato)
HDI_con_calcolo <- cbind(HDI$`Human Development Index (HDI)`, HDI_calcolato)
colnames(HDI_con_calcolo)[1] <- "HDI_originale"
as.data.frame(HDI_con_calcolo)

```

HDI_originale <dbl>	HDI_calcolato <dbl>	HDI_originale <dbl>	HDI_calcolato <dbl>
0.921	0.9211530	0.890	0.8901996
0.919	0.9192901	0.889	0.8892535
0.918	0.9178023	0.887	0.9052655
0.918	0.9182485	0.876	0.8762231
0.916	0.9155126	0.875	0.8751386
0.911	0.9106944	0.875	0.8745662
0.905	0.9051371	0.875	0.8749183
0.903	0.9027100	0.866	0.8656444
0.896	0.8955325	0.863	0.8632489
0.895	0.8945539	0.858	0.8578045

## SPIEGAZIONE DEL RAPPORTO SULL'HDI

Questo rapporto utilizza i dati dell'indice di sviluppo umano (HDI) e dei suoi componenti, forniti dal Programma di sviluppo delle Nazioni Unite (UNDP), per valutare i risultati. I dati HDI prendono in considerazione l'aspettativa di vita alla nascita, gli anni previsti e medi di scolarizzazione e il reddito nazionale lordo (GNI) per ciascun paese. Una volta calcolato, fornisce un valore numerico compreso tra 0 (il più basso) e 1 (il più alto) e anche una classifica (ordine crescente), dandoci 1 variabile dipendente (HDI) e 5 variabili indipendenti.

Il gradiente della linea è 0,65, il che lo rende positivo. Ciò significa che all'aumentare degli anni medi di scolarizzazione, aumenta anche l'GNI di quella nazione. La maggior parte dei paesi è strettamente legata alla linea di miglior adattamento, tuttavia sono stati rilevati alcuni valori anomali. Il gradiente della linea per l'GNI rispetto all'HDIR è -0,817, il che significa che al diminuire dell'GNI esiste una relazione lineare con quella dell'HDIR che anch'essa diminuirà. La maggior parte dei paesi è vicina alla linea di best fit, con meno valori anomali in questo confronto rispetto al precedente.

Non c'è molta differenza per gli anni previsti di scolarizzazione in tutto il mondo. Il numero previsto di anni di scolarizzazione è abbastanza simile in tutto il mondo, con una mediana di 13,405 anni di scolarizzazione mentre la media è appena inferiore a 13,535 anni. Inoltre, la deviazione standard è solo di 2,924, il che significa che la maggior parte dei paesi si trova in prossimità l'uno dell'altro. Sebbene l'intervallo sia ampio, 15,512 anni, non ci sono valori anomali perché Q1 è di 11,601 anni e Q3 è di 15,624 anni

Il grafico a barre mostra una diminuzione netta al diminuire dell'HDIR, a dimostrazione del fatto che un numero inferiore di anni previsti di scolarizzazione si traduce in un HDIR inferiore  
entrambi i grafici mostrano un andamento lineare e la correlazione tra i primi 20 e gli ultimi 20 era piuttosto alta, dimostrando così che esiste una relazione definita tra l'aspettativa di vita e la classifica HDI di quel paese. Il gradiente per i primi 20 è stato di -0,95 e per gli ultimi 20 è stato di -0,895. Ciò si correla correttamente ai dati poiché i paesi con un'aspettativa di vita inferiore tendevano anche a posizionarsi più in basso in termini di HDI rispetto ai paesi con un'aspettativa di vita elevata

In entrambi i grafici vediamo una regressione lineare negativa che indica una chiara correlazione tra le variabili per cui la diminuzione del grado HDI si traduce in una diminuzione dell'aspettativa di vita alla nascita.

Vengono poi riportate le formule per calcolare gli indici necessari per il calcolare l'HDI e un esempio concreto del calcolo dell'HDI che come attraverso la serie geometrica che come vediamo ci da gli stessi valori dell'HDI reale calcolato dalle Nazioni Unite



**Grazie per l'attenzione**

**Antonio Belmonte**

**Aldo Scaglione**

**Luca Mascia**

**Emanuele Iaccarino**