

**SINTETIZZARE MISURE ELEMENTARI: UNA SPERIMENTAZIONE DI ALCUNI CRITERI PER LA
DEFINIZIONE DI UN INDICE COMPOSTO**

Salvatore F. ALLEGRA*

Alessandro LA ROCCA*

*Istituto Nazionale di Statistica

L'articolo è frutto del lavoro congiunto dei due Autori. Nondimeno, i paragrafi 1 e 3 sono stati scritti da S. F. Allegra; i paragrafi 2, 4 e 5 sono stati scritti da A. La Rocca.

Sommario

La quantificazione di fenomeni sociali complessi richiede una selezione accurata della tecnica statistica da utilizzare per sintetizzare le misure elementari scelte come indicatori.

Il sistema di indicatori sociali denominato *DemoS*, prodotto dall'Istituto Nazionale di Statistica, è la base empirica da cui gli Autori sono partiti per sperimentare le procedure statistiche per la sintesi degli indicatori riportate dalla letteratura sull'argomento. Le sette tecniche applicate, due *monovariate* e cinque *multivariate*, sono state confrontate attraverso gli effetti che l'indice sintetico ha prodotto sulla classificazione delle osservazioni: le 103 province italiane. Le procedure statistiche *monovariate* non attribuiscono alcun peso alle misure elementari; quelle *multivariate*, invece, assegnano pesi diversi. Le diverse posizioni delle province nelle singole graduatorie mostrano l'influenza delle tecniche sugli indici prodotti.

Infine, il coefficiente di correlazione di Kendall ha permesso di misurare la similitudine delle tecniche di sintesi adottate.

Abstract

A complex social phenomena quantification requires a careful selection of the statistical technique to be used in order to synthesise the elementary measures chosen as indicators.

The social indicator system named *DemoS*, which was produced by the Italian National Statistics Institute (Istat), is the empirical foundation for the authors who wanted to test the statistical procedures suitable to the indicator synthesis dealt with by the literature. Seven techniques which have been used, whose two were *univariate* and five *multivariate*, have been compared on the basis of the effects which the synthetic index produced on the classification of the observations, namely 103 Italian provinces. *Univariate* statistical procedures do not give any weight to the elementary measures while, on the contrary, *multivariate* ones assigne different weights. The different provinces position in each scale shows the influence of the techniques on the indexes produced.

Finally, Kendall's correlation coefficient allowed the authors to evaluate the similitudes of the synthesis techniques.

1. Introduzione

La consapevolezza dell'insufficienza delle categorie economiche nella lettura della complessità dei processi di modernizzazione delle società industrializzate e l'esigenza di comprendere la rilevanza sociale di tali processi davano origine, nel decennio successivo alla fine del secondo conflitto mondiale, a nuovi spunti di riflessione in relazione all'analisi dell'impatto dello sviluppo economico sugli scenari sociali. Bisogni prima inesplorati – derivanti da un accresciuto benessere – prendevano forma in quegli anni e con i bisogni nuove domande in cui trovavano spazio le preoccupazioni circa i costi – materiali e immateriali – dello sviluppo economico (cfr. Curatolo, 1972, pp. 25-30; Delvecchio, 1995, pp. 58-9).

L'interrogativo principale che molti studiosi si ponevano era il seguente: quali strumenti affiancare ai tradizionali strumenti di analisi al fine di enucleare le questioni concernenti le condizioni di vita delle persone? Naturalmente la risposta esigeva un cambiamento di prospettiva, non soltanto da un punto di vista operativo, ma anche da un punto di vista teorico. Ciò significava abbandonare l'idea che sviluppo economico fosse uguale a progresso sociale (Delvecchio, op. cit., p. 60) e, di fatto, ciò ridefiniva la capacità esplicativa degli strumenti tradizionali a favore di schemi di ricerca che tenessero conto delle *variabili sociali* e del loro patrimonio informativo. Si trattava, in altri termini, di individuare gli strumenti più adeguati all'interpretazione dei fenomeni sociali. L'approccio era certamente votato alla conoscenza e orientato all'azione; l'obiettivo d'altronde era ben definito: fornire utili *indicazioni* alla progettazione di politiche volte a migliorare le condizioni di vita delle persone.

Gli indicatori sociali nascevano in questo ambito e con queste finalità. Nella prima metà degli anni '60, negli Stati Uniti d'America il gruppo di studiosi che si riconosceva intorno al "Movimento per gli indicatori sociali" contribuì significativamente allo sviluppo e alla diffusione di tale tematica (Curatolo, op. cit., pp. 36-47; Miles, 1985, pp. 25-33).

Il riconoscimento dell'importanza degli indicatori sociali per lo studio e la rilevazione dei fenomeni fu immediato. Sulla scia dell'interesse suscitato dai temi dell'ambiente e del benessere sociale (temi legati al miglioramento della qualità della vita), che venivano posti per la prima volta in maniera pressante all'attenzione della collettività, fu possibile gettare le basi per un uso integrato di indicatori sociali ed economici. Questi ultimi con una tradizione consolidata e un *corpus* di conoscenze più consistente erano infatti già formalizzati e condivisi da tutti i Paesi a statistiche avanzate. L'obiettivo ambizioso di costruire un sistema integrato di contabilità sociale analogo a quello economico (*Sec*) non è stato però raggiunto.

Le ricerche nelle quali gli indicatori sociali, sin dall'inizio, hanno trovato applicazione sono dunque molteplici, specie nel campo della misurazione della qualità della vita (Delvecchio, op. cit., p. 89 e ss.), concetto che comprende – secondo una definizione recente –, oltre alla dimensione materiale del *welfare*, gli “aspetti immateriali quali la percezione dello stato di salute, delle relazioni sociali e della qualità dell'ambiente naturale, nonché altre caratteristiche, quali il benessere soggettivo dei cittadini” (Vitali, 2002, p. 2).

Oggi, a distanza di quasi mezzo secolo, è possibile affermare che l'apporto più importante della tematica degli indicatori sociali, anche in virtù del dibattito scientifico che essa ha prodotto, è avere posto le fondamenta per l'edificazione di sistemi di statistiche sociali in ciascuno dei Paesi industrializzati. Sono molti i settori della ricerca che ne riconoscono l'importanza e varie organizzazioni internazionali, come l'Onu, producono indici sintetici che sono diventati ormai istituzionali.

Gli indicatori sociali sono divenuti pertanto il punto di forza di qualunque attività di *report* sociale, nonostante le numerose difficoltà di natura concettuale e terminologica, che a volte ne hanno limitato la portata esplicativa (cfr. Curatolo, op. cit., pp. 36-47; Miles, op. cit., pp. 137-48).

Di notevole rilievo sono le iniziative adottate nei vari Paesi dagli organismi istituzionali deputati alla produzione e alla diffusione dell'informazione statistica. Per quanto concerne l'Italia, l'Istituto Nazionale di Statistica (Istat) ha provveduto in più occasioni a fornire un compendio di statistiche basato su indicatori sociali (Istat, 1975; 1981; 1990; 1993). I fenomeni di interesse, articolati in aree tematiche omogenee, danno modo di delineare un quadro della realtà italiana attraverso una selezione ampia di dati statistici caratterizzanti i molteplici aspetti della vita sociale. L'approccio è descrittivo, dato che l'intento è chiaramente divulgativo: oltre ad un'utenza specializzata, pubblicazioni così strutturate si rivolgono a un pubblico costituito da persone interessate ad avere una visione d'insieme – sintetica dunque, ma esaustiva – della situazione sociale del Paese (Istat, 1993, p. 13). Occorre precisare, a tale proposito, che ciascuna area tematica presenta delle misure di base (elementari) che sono disaggregate dal punto di vista territoriale.

Evidentemente informazioni organizzate in questo modo permettono di perseguire una duplice finalità: da un lato, è possibile confrontare il comportamento del medesimo indicatore elementare nelle diverse aree territoriali; dall'altro, è possibile definire un indice composto degli indicatori sociali elementari che sia in grado di fornire una misura sintetica – empiricamente valutabile – del fenomeno che si intende analizzare.

Obiettivo del presente lavoro sarà proprio quello di esaminare i risultati di diverse procedure di sintesi a cui sono stati sottoposti gli indicatori elementari che compongono il sistema di indicatori sociali denominato *DemoS*, prodotto dall'Istat e diffuso in forma di metadati.

2. La scelta degli indicatori nell'ambito del sistema *DemoS*

Tra i macrodati prodotti dall'Istat in campo sociale e demografico, alcuni hanno assunto la forma di *sistemi di indicatori* poiché misurano fenomeni appartenenti ad aree tematiche tra loro coerenti. Dal punto di vista statistico, questi sistemi di indicatori si avvalgono di indici di composizione per la misura dei suddetti fenomeni.

Il lavoro che segue fa riferimento al sistema di indicatori sociali denominato *DemoS*. Si tratta di un sistema informativo contenente dati socio-demografici organizzati e strutturati secondo modalità di diffusione di tipo *data warehouse*. Ciò allo scopo di consentire all'utente un accesso semplice e personalizzato, sulla base di determinati criteri da lui definiti.

DemoS si compone di 9 aree tematiche, per un totale di 80 indicatori, che si presentano disaggregati sia a livello regionale sia a livello provinciale. Il numero di indicatori elementari in ciascuna area è variabile (cfr. tabella 1).

Tab. 1 – Aree tematiche di *DemoS* e numero di indicatori per ciascuna area

<i>Aree tematiche</i>	<i>Numero di indicatori</i>
Mobilità territoriale, dinamica migratoria, stranieri	9
Struttura demografica	5
Mortalità, natalità, fecondità, comportamenti riproduttivi	18
Matrimoni, separazioni e divorzi	10
Istruzione, scuola, formazione professionale	3
Cultura, sport, turismo, attività ricreative	5
Occupazione e mercato del lavoro	6
Abitazioni, qualità abitativa, attività edilizia	9
Litigiosità, criminalità, criminalità economica	15

Le fonti da cui sono stati desunti gli indicatori elementari che appartengono alle aree suddette – eccetto che per l'area relativa alle abitazioni – sono le rilevazioni e le indagini condotte periodicamente dall'Istat; la fonte degli indicatori concernente le abitazioni è censuaria. Gli anni di riferimento non sono perciò omogenei: i dati che si ricavano da rilevazioni e indagini vanno dal 1996 al 1999; per quanto riguarda l'area delle abitazioni l'anno di riferimento è il 1991.

È possibile sottoporre questo *set* di indicatori sociali all'applicazione delle tecniche statistiche per la sintesi degli indicatori comunemente riportate dalla letteratura sull'argomento. In altre parole, partendo da misure elementari, si può giungere alla definizione di indici composti, che risulteranno diversi in quanto differenti sono i criteri chiamati in causa per la loro costruzione.

La sintesi di più misure in una composita implica tuttavia due ordini di problemi. Essi sono individuabili in momenti distinti del percorso che conduce alla realizzazione della misura sintetica.

Il primo problema riguarda la coerenza semantica tra gli elementi da sintetizzare (gli indicatori empirici) che rappresentano le dimensioni del concetto e il concetto stesso (cfr. Marradi, 1984, p. 39). Questo rapporto dovrebbe essere supportato da una teoria (cfr. Fraire, 1989, p. 245-7). Ad esempio, misurare il concetto di benessere significa dare origine ad un indice che, pur essendo una sintesi di singole misure, è il portato di una riflessione teorica che investe il processo di operazionalizzazione del concetto (cfr. Grasso, 2002).

Il secondo problema attiene alla scelta delle tecniche statistiche delle quali ci si intende avvalere allo scopo di ottenere un indice sintetico. È un aspetto propriamente operativo, che certamente segue la riflessione teorica, ma che ha, a ben vedere, un effetto determinante sull'indice prodotto e quindi sull'interpretazione del risultato.

Il lavoro che qui viene presentato non entra nel merito della questione relativa al legame semantico nella relazione tra concetto e indicatore. Si dà per scontato che tale legame esista. L'utilizzo di un *set* di indicatori di fonte ufficiale porta già di per sé a non occuparsi di problemi riguardanti la natura del rapporto di indicazione, considerato che in letteratura è ormai consolidata una tradizione metodologica e di ricerca che utilizza indicatori di natura socio-demografica al fine di misurare concetti come qualità della vita e benessere sociale.

Il lavoro deve essere inteso come un'applicazione attraverso la quale si sono sperimentati criteri diversi di sintesi (sia di natura *monovariata* che *multivariata*) del concetto di benessere sociale per combinare gli indicatori elementari selezionati dalle aree tematiche che sono state scelte. Sono stati perciò

presi in considerazione 25 indicatori e si è deciso di scendere al massimo livello di disaggregazione territoriale, operando sulle 103 province.

Si riporta per completezza l'elenco degli indicatori elementari con l'anno di riferimento per ognuna delle aree tematiche scelte.

1. Mobilità territoriale, dinamica migratoria, stranieri:

- 1) saldo migratorio – 1997;
- 2) stranieri iscritti all'anagrafe per 1.000 abitanti – 1998.

2. Struttura demografica:

- 1) indice di invecchiamento della popolazione – 1998;
- 2) indice di dipendenza totale – 1998.

3. Mortalità, natalità, fecondità, comportamenti riproduttivi:

- 1) numero medio di figli per donna in età feconda – 1996;
- 2) tasso di mortalità infantile – 1997;
- 3) tasso generico di mortalità – 1999.

4. Istruzione, scuola, formazione professionale:

- 1) tasso di iscrizione alle scuole superiori – 1997.

5. Cultura, sport, attività ricreative:

- 1) copie di quotidiani diffuse per 1.000 abitanti – 1998;
- 2) spesa media per abitante per rappresentazioni teatrali e musicali – 1998;
- 3) sale cinematografiche aperte al pubblico per 100.000 abitanti – 1998.

6. Occupazione e mercato del lavoro:

- 1) tasso di attività – 1999;
- 2) tasso di occupazione – 1999;
- 3) tasso di disoccupazione – 1999.

7. Abitazioni, qualità abitativa, attività edilizia:

- 1) popolazione in condizione di sovraffollamento – 1991;
- 2) abitazioni fornite di telefono per 100 abitazioni – 1991;
- 3) abitazioni senza acqua potabile e gabinetto per 100 abitazioni – 1991.

8. Litigiosità, criminalità, criminalità economica:

- 1) delitti denunciati per criminalità violenta per 100.000 abitanti in età 14 anni e più – 1998;

- 2) delitti denunciati per lesioni dolose per 100.000 abitanti in età 14 anni e più – 1998;
- 3) delitti denunciati per furti per 100.000 abitanti in età 14 anni e più – 1998;
- 4) delitti denunciati per borseggi per 100.000 abitanti in età 14 anni e più – 1998;
- 5) delitti denunciati per scippi per 100.000 abitanti in età 14 anni e più – 1998;
- 6) delitti denunciati per altri motivi per 100.000 abitanti in età 14 anni e più – 1998;
- 7) delitti denunciati per produzione e spaccio di stupefacenti per 100.000 abitanti in età 14 anni e più – 1998;
- 8) minorenni denunciati per 10.000 abitanti in età 14-17 anni – 1997.

La scelta degli indicatori ha cercato di includere tutte le aree tematiche di *Demos*. È stata esclusa l'area "Matrimoni, separazioni e divorzi" ritenuta non coerente con il significato che è possibile attribuire all'indice composto. Gli indicatori sottoposti a sintesi hanno tutti una attinenza con il concetto di benessere sociale; in quanto riguardano aspetti del vivere sociale al cui miglioramento tendono le società occidentali. Con una buona approssimazione, l'indice risultante consente di misurare il benessere sociale delle province italiane, come confermano i numerosi lavori sul tema presenti in letteratura (si veda per tutti Delvecchio, op. cit.).

Per quanto concerne il lavoro qui presentato, pertanto, è stato necessario effettuare delle valutazioni preliminari al fine di stabilire quali misure elementari del sistema di indicatori sociali *DemoS* dovessero costituire la base empirica di partenza per la definizione dell'indice composto. La scelta degli indicatori è stata dettata da criteri di parsimonia. Per di più, allo scopo di verificare il grado di attendibilità dell'indice ottenuto, ossia della consistenza interna del risultato in relazione agli indicatori selezionati, si è fatto ricorso reiteratamente al calcolo del coefficiente α di Cronbach, il quale varia tra 0 (minima attendibilità) e 1 (massima attendibilità). Ciò ha consentito di eliminare di volta in volta quegli indicatori che portavano il valore del coefficiente molto al di sotto della soglia considerata limite ($\alpha = 0,80$) (cfr. Fraire, op. cit., p 252). Nel caso in esame il valore di α è risultato pari a 0,70.

Inoltre, la decisione di scendere al massimo livello di disaggregazione territoriale (le province) deriva dal fatto che alcune procedure di sintesi multivariate forniscono risultati affidabili solo se la numerosità delle osservazioni del collettivo statistico considerato è sufficientemente ampia.

Anche i *mass media* impegnati periodicamente nella produzione di analisi ecologiche su tematiche di rilevanza socio-economica impiegano *set* di indicatori e utilizzano dati disaggregati a livello provinciale per stilare classifiche in relazione al grado di *performance* delle singole realtà territoriali (si veda, per esempio,

l'indagine *panel* de *Il Sole 24 Ore* sulla qualità della vita). I criteri di sintesi però sono spesso carenti di una attenta riflessione statistica e metodologica (Attanasio, Capursi, 1997; Giudici, Avrini, 2002, pp. 78-9), sebbene, per *correggere* le statistiche ottenute grazie agli indicatori tradizionali, le analisi comincino ad avvalersi di tecniche di *sentiment* (cfr. Noto, 2003, p. 2), ossia di indicatori soggettivi che dovrebbero considerare la reale percezione che gli individui hanno di un dato fenomeno.

3. La scelta delle tecniche di sintesi degli indicatori

La scelta tra le tecniche disponibili per sintetizzare indicatori elementari è stata indirizzata verso quelle che mostrano maggiormente le problematiche conseguenti alla costruzione di un indice composto. In altre parole, i punti critici delle differenti tecniche nella definizione di un indice composto si riverberano sulla classificazione delle osservazioni, che, nel caso in esame, sono le province.

Le tecniche adottate sono quelle riportate nei numerosi lavori sull'argomento (Fraire, op. cit., pp. 250-8; Slotje, 1991; Delvecchio, op. cit., pp. 117-82; Giudici, Avrini, op. cit.). Sono state escluse le tecniche che non sfruttano le proprietà della scala di misurazione in cui sono espressi i dati impiegati per il presente lavoro. Le diverse sintesi degli indicatori elementari sono state pertanto effettuate mantenendo il carattere cardinale delle misure. Non sempre, tuttavia, si conserva tale caratteristica; talvolta accade che misure cardinali siano retrocesse a misure di *ranking* e dunque trattate mediante strategie di sintesi di tipo ordinale, al fine di ridurre la soggettività che deriva dall'assegnare pesi specifici agli indicatori elementari (si veda, per esempio, il metodo di Borda, cit. in Grasso, op. cit., pp. 278-80).

Nell'ambito delle procedure che sfruttano le proprietà delle variabili cardinali, la scelta è tra criteri di sintesi che non attribuiscono alcun peso alle varie misure e criteri che, viceversa, assegnano a queste pesi diversi. Ciò conduce evidentemente alla distinzione tra tecniche *monovariate* e tecniche *multivariate* di aggregazione dei dati.

Nel lavoro qui presentato vengono mostrate 7 operazioni di sintesi di indicatori elementari: 2 di queste sono classificabili come *monovariate*; mentre, le rimanenti 5 rientrano tra le procedure *multivariate* di tipo fattoriale.

Si definisce indice sintetico (o indice composto) una misura S_i , ottenuta attraverso una opportuna combinazione delle misure elementari per ciascuna unità statistica del collettivo. Se si indica con $\mathbf{P}_{N,M}$ la matrice dei dati di ordine $(N \times M)$, in cui N rappresenta il numero delle unità osservate (le 103 province)

e M il numero degli indicatori elementari selezionati (25, nel caso in esame), allora $p_{i,j}$ è la determinazione assunta dalla i -ma provincia in relazione al j -mo indicatore, per ($i = 1, 2, \dots, N$) e ($j = 1, 2, \dots, M$).

Prima di procedere, è necessario effettuare un'operazione di *ri-orientamento* delle misure elementari di segno negativo. Esse sono trasformate in misure di segno positivo, affinché tutte siano rivolte nella medesima direzione. Formalizzando, il ribaltamento degli indicatori di segno negativo può essere espresso nel modo seguente: $d_{i,j} = y - p_{i,j}$, dove y è l'unità di riferimento per il valore assunto dalla i -ma provincia in relazione al j -mo indicatore.

Passando dalla matrice $\mathbf{P}_{N,M}$ alla sua equivalente in forma standardizzata $\mathbf{Z}_{N,M}$, in cui i valori originali sono espressi come scarti dalla media relativizzati allo scarto quadratico medio, al fine di sganciare gli indicatori elementari dalle rispettive unità di misura, è possibile costruire l'indice sintetico S_i .

Più esattamente, dalla matrice $\mathbf{Z}_{N,M}$ si giunge alla matrice \mathbf{S}_N di ordine $(N \times 1)$ – cioè a un vettore colonna –, nel caso di calcolo di un unico indice S_i . Nel caso di calcolo di più indici S_i , dalla matrice $\mathbf{Z}_{N,M}$ si perviene alla matrice $\mathbf{S}_{N,C}$ di ordine $(N \times C)$, con $C \ll M$.

Poiché l'obiettivo da raggiungere è quello di ridurre attraverso opportune tecniche statistiche la matrice iniziale lungo le colonne, si possono ottenere i risultati seguenti: *a*) un vettore colonna quando la misura ottenuta è unica (l'efficacia della procedura di sintesi sarà allora massima); *b*) una nuova matrice, ma di ordine inferiore a quello della matrice iniziale, quando gli indicatori elementari non si prestano ad essere sintetizzati con un solo indice.

Come già detto, le tecniche di costruzione degli indici sintetici si differenziano in relazione alla possibilità di introdurre fattori di ponderazione.

Nel lavoro che segue, le 2 procedure *monovariate* utilizzate per definire gli indici sono: 1) la media aritmetica degli M indicatori elementari, ${}^1S_i = (1/M) \sum_{j=1}^M z_{i,j}$; 2) la tecnica che, nel cosiddetto modello di Hellewig, determina la *misura dello sviluppo*, ${}^2S_i = D_{i,0} / D_0$.

Il primo indice composto 1S_i si ricava dalla media aritmetica semplice degli M indicatori convertiti in scarti standardizzati $z_{i,j}$ nella matrice $\mathbf{Z}_{N,M}$ mediante una trasformazione lineare di ciascuno degli

indicatori di media \overline{X}_j e scarto quadratico medio σ_j . I valori $z_{i,j}$ esprimono, infatti, i valori originali di ogni indicatore come scarti dalla media relativizzati allo scarto quadratico medio e danno origine a una distribuzione con media uguale a 0 e scarto quadratico medio pari a 1. È così possibile disporre di numeri puri, ossia di valori liberati dalle unità di misura originarie. L'indice sintetico ${}^1S_i = (1/M) \sum_{j=1}^M z_{i,j}$, associato all' i -ma unità statistica del collettivo sotto osservazione, sarà dunque il risultato della media aritmetica dei valori assunti dagli indicatori per quella unità statistica.

Il secondo indice composto 2S_i si ottiene dalla sintesi degli M indicatori mediante la distanza da un valore *ideale*, che, per ciascun indicatore, è il valore massimo raggiunto. In altre parole, le misure elementari, convertite – ancora una volta – in scarti standardizzati $z_{i,j}$ nella matrice $\mathbf{Z}_{N,M}$, per ogni unità statistica del collettivo, vengono poste a confronto con i singoli valori *ideali* attraverso il calcolo della distanza euclidea. Il risultato è un valore $D_{i,0}$, che diverrà il numeratore dell'indice 2S_i . Considerando la media aritmetica \overline{D}_0 delle distanze euclidee delle singole unità statistiche dal valore *ideale* e il loro scarto quadratico medio σ_0 , è possibile determinare il denominatore D_0 dell'indice 2S_i . Tale denominatore contiene poco meno della totalità delle distanze, in quanto $D_0 = \overline{D}_0 + 2\sigma_0$. L'indice sintetico ${}^2S_i = D_{i,0} / D_0$, associato all' i -ma unità statistica del collettivo preso in esame, indicherà il suo livello di sviluppo, che sarà tanto più elevato quanto più 2S_i sarà vicino a 0.

L'adozione di procedure *multivariate* – le quali attribuiscono pesi diversi w_j , per $(j = 1, 2, \dots, K)$, a ciascuna delle misure elementari selezionate – porta a distinguere i seguenti due casi: I) il caso di un solo indice sintetico per gli indicatori considerati, in corrispondenza dell' i -ma unità statistica del collettivo; II) il caso di due o più indici sintetici, in corrispondenza dell' i -ma unità statistica.

Tra le tecniche *multivariate* sono state scelte quelle di fattorizzazione, proposte nelle seguenti 5 varianti, in relazione al metodo di estrazione e di rotazione degli assi fattoriali: 1) analisi delle componenti principali con un solo fattore estratto (3S_i); 2) analisi delle componenti principali con due fattori estratti e considerati ortogonali, ma con l'adozione del primo fattore (${}^{4a}S_i$); 3) analisi delle componenti principali con somma dei punteggi fattoriali dei due fattori ortogonali (5S_i); 4) analisi delle componenti principali con

media dei punteggi fattoriali dei due fattori ortogonali (6S_i); 5) analisi dei fattori principali con un fattore estratto (7S_i).

L'ipotesi di base, comune a tutte le 5 varianti, è che la correlazione tra le misure elementari è spiegata da fattori non osservabili (*latenti*). Punto di partenza è la matrice di correlazione $\mathbf{R}_{M,M}$ delle misure elementari standardizzate. Dalla matrice dei coefficienti fattoriali, che si ricava dalla matrice dei pesi fattoriali, è possibile ottenere uno o più indici sintetici, a seconda del numero di fattori selezionati.

La somma dei prodotti tra il generico valore $z_{i,j}$ rappresentato dal fattore e il relativo punteggio fattoriale (*factor score*) dà luogo al seguente indice sintetico: ${}^a S_i = \sum_{j=1}^M z_{i,j} \cdot {}^a f_j$, dove $z_{i,j}$ è la determinazione standardizzata assunta dalla i -ma unità osservata per il j -mo indicatore e f_j è il punteggio fattoriale per il j -mo indicatore. Fissando un unico fattore si ottiene un solo indice composto (3S_i), che, tuttavia, può non essere in grado di rappresentare tutti gli indicatori elementari.

Per prendere in considerazione altre misure non rappresentate dall'unico asse fattoriale è necessario utilizzare un secondo asse. Si ottiene in questo modo un secondo indice analogo al primo: ${}^b S_i = \sum_{j=1}^M z_{i,j} \cdot {}^b f_j$; pertanto per ogni unità si avranno due indici composti, ${}^a S_i$ e ${}^b S_i$, ma, attraverso la rotazione degli assi, muteranno i contributi delle misure elementari e il punteggio fattoriale del primo fattore si modificherà in virtù della presenza di un secondo asse (${}^{4a}S_i$).

Inoltre, l'esigenza di avere un unico indice sintetico può comportare che ${}^a S_i$ e ${}^b S_i$ possano essere accorpati attraverso un'operazione di somma dei punteggi fattoriali dei due fattori ortogonali (5S_i) oppure mediante la media dei punteggi (6S_i). In altri termini, i criteri di sintesi adottati – per questi due casi – considerano i primi due assi fattoriali estratti dall'analisi delle componenti principali. Si dà conto così di una quantità maggiore di varianza spiegata, rappresentando gli indicatori *sfuggiti* alla prima componente.

Infine, è possibile ottenere un altro criterio di sintesi, mutando il metodo di estrazione dei fattori. Attraverso l'analisi dei fattori principali (7S_i) si analizza solo la varianza attribuibile ai fattori comuni (la *comunalità*).

4. Dalla sintesi degli indicatori alla classificazione delle osservazioni

La selezione delle proprietà di un concetto, ovvero delle misure elementari da sintetizzare in un indice, qualora si intenda misurare un fenomeno complesso – nel caso in esame, il benessere sociale – richiede un attento processo logico-concettuale che descriva il passaggio dalla definizione nominale (astratta) del concetto agli indicatori elementari. Tale processo è indispensabile poiché è finalizzato a una corretta misurazione empirica.

In realtà, per l'analisi e la misurazione di fenomeni sociali (si pensi agli studi sulla qualità della vita e sul benessere sociale) è possibile rifarsi ad una tradizione di ricerca ormai consolidata, che utilizza come strumenti di misura *set* di indicatori socio-demografici.

La sintesi degli indicatori elementari in un indice di benessere sociale nel lavoro qui presentato ha prodotto diversi risultati a seconda delle tecniche adottate

Il primo indice sintetico 1S_i è quello che si ottiene attraverso una procedura di sintesi *monovariata*, ovvero dalla media aritmetica degli indicatori. Dal punto di vista del linguaggio delle matrici, si è di fronte a un vettore colonna \mathbf{S}_N che si compone di 103 indici sintetici, tanti quante le unità del collettivo (le province); essi potranno dunque essere graduati e posti a confronto. Occorre tuttavia tenere presente due limiti sostanziali: *a)* tale procedura, essendo *monovariata*, non assegna alcun peso agli indicatori elementari; *b)* la media, non essendo un indice robusto, risente notevolmente dei valori che nella distribuzione occupano posizioni estreme.

Ciò premesso, la provincia con la migliore *performance* è quella di Bolzano; seguono quella di Bergamo e la provincia di Trento. Viceversa, i risultati peggiori sono quelli mostrati dalla province di Napoli, Enna e dalla provincia di Crotone (cfr. tabella 2).

La seconda tecnica *monovariata* conduce all'indice sintetico 2S_i , che è la *misura dello sviluppo* nel modello di Hellewig. Come nel caso precedente, il risultato è un vettore colonna \mathbf{S}_N che presenta 103 indici sintetici. L'ordine delle province sarà dunque derivato dal loro livello di sviluppo, misurato dal valore dell'indice: più tale valore sarà prossimo allo 0, maggiore sarà il grado di sviluppo indicato. In prima posizione, nella graduatoria così formata, figura la provincia di Verona; al secondo posto si colloca la provincia di Bolzano, seguita dalla provincia di Trento. Enna è, invece, la provincia che ottiene il risultato peggiore; seguono le province di Napoli e Crotone (cfr. tabella 2).

Anche questo secondo indice mostra un limite fondamentale: non facendo uso di alcun criterio di ponderazione, la tecnica utilizzata annulla di fatto il peso specifico di ciascuna misura elementare.

Quando si passa dalle tecniche *monovariate* a quelle *multivariate* entrano in gioco i punteggi fattoriali (*scores*), i quali rappresentano le coordinate delle unità statistiche del collettivo rispetto al nuovo sistema di riferimento. Se a questo sistema si attribuisce il significato di *variabile* di sintesi, allora è possibile ordinare le osservazioni in funzione di tale dimensione (fattore) (cfr. Zani, 2000, p. 79).

Il primo criterio di ordinamento utilizza il metodo di estrazione delle componenti principali, la cui caratteristica fondamentale è spiegare il massimo della varianza possibile tra le misure elementari originali identificando una serie di combinazioni lineari ortogonali (cioè non correlate).

Una sola componente (3S_i) ha consentito di spiegare il 34% della varianza presente tra gli indicatori, garantendo quindi una discreta rappresentazione della variabilità della matrice dei dati iniziali. Ordinare le unità lungo questa dimensione ha in realtà significato graduare le province in relazione alle misure che più hanno contribuito a formarla. Da un punto di vista operativo, ciò è stato possibile attraverso l'analisi delle correlazioni degli indicatori originali con l'unica componente estratta. La buona rappresentazione delle misure elementari sintetizzate tramite questa dimensione emerge anche dal fatto che per 14 indicatori su 25 è stata spiegata più del 30% di varianza totale.

La sintesi in un solo fattore ha comportato inevitabilmente una perdita di informazioni, come risulta dalla lettura dei valori caratteristici dell'analisi delle componenti principali. Se si confronta il risultato – in termini di posizioni delle province nella graduatoria – con quello ottenuto attraverso le tecniche *monovariate* viste precedentemente, emergono sia delle similitudini sia delle evidenti discordanze, specie per le posizioni di testa. I tre criteri di sintesi concordano, in modo particolare, nel classificare le province di Enna, Crotone, Caltanissetta, Foggia e Reggio Calabria; tutte geograficamente omogenee e posizionate in coda alla graduatoria. Gli scostamenti più marcati si sono registrati per le province di Trieste e Rimini, rispettivamente, al secondo e al quarto posto (cfr. tabella 2). Tali differenze sono da attribuire al fatto che la tecnica di sintesi, sfruttando una sola dimensione, non rappresenta tutti gli indicatori e quindi sottostima o sovrastima le *performance* delle province. Nel caso specifico della province di Trieste e Rimini, gli indicatori non rappresentati adeguatamente sull'unica dimensione sono proprio quelli che mostrano i valori più bassi. Questi stessi indicatori hanno contribuito a porre le due osservazioni in posizioni arretrate nelle graduatorie formate tramite le procedure *monovariate*.

Allo scopo di semplificare la struttura dei fattori, consentendone la separazione, si sono estratti due fattori e si è operata – mediante il criterio *varimax* – una rotazione ortogonale degli assi. Si sono poi ordinate le unità secondo lo *score* ottenuto sulla prima dimensione (4S_i). Il risultato è una graduatoria delle province diversa dalla precedente, per il fatto che il criterio *varimax* tende a produrre correlazioni con gli assi contenendo il più possibile quelle con valori intermedi.

Ciò implica che alcune province – per esempio, quella di Rimini, Milano e Roma – occupino posizioni molto lontane rispetto a quelle scaturite dall’analisi delle componenti principali senza alcuna rotazione. Altre province – innanzitutto quelle di Pistoia, Taranto e Vibo Valentia, ma anche le province di Aosta, Avellino, Pescara, Ragusa, Trapani – non risentono invece per nulla o quasi della rotazione (cfr. tabella 2). Evidentemente le misure che *entrano* nelle componenti o *escono* da queste per effetto della rotazione assumono valori bassi su queste province, incidendo quindi minimamente sul punteggio finale.

I seguenti due criteri di sintesi prendono in considerazione le prime due dimensioni estratte dall’analisi delle componenti principali e ruotate ortogonalmente. Il valore dell’indice è dato dalla somma dei punteggi fattoriali (uno per dimensione), in un caso (5S_i); dalla media, nell’altro caso (6S_i).

È stato possibile in questo modo sfruttare una quota maggiore di varianza spiegata, pari a circa il 52%. Ciò ha permesso di rappresentare le misure elementari non rivelate dalla prima componente. È evidente che, in virtù della presenza di questi indicatori, molte province occupano delle posizioni diverse rispetto alle graduatorie precedentemente formate. Inoltre, è importante notare che, per alcune province (quelle di Siena, Grosseto, Rovigo, Sassari e Napoli), la somma dei punteggi produce risultati perfettamente coincidenti con la media ponderata (cfr. tabella 2).

Infine, se si cambia il metodo di estrazione dei fattori a favore dell’analisi dei fattori principali, si ricava un ultimo indice sintetico (7S_i). Come già evidenziato, mediante tale criterio di sintesi, si analizza esclusivamente la *comunalità*, vale a dire solo la varianza riconducibile ai fattori comuni.

Le differenze – in termini di posizione delle unità osservate – sono interamente attribuibili al diverso contributo che le misure elementari danno alla formazione dei fattori e alla rappresentazione che ne consegue. I risultati sono molto simili a quelli ottenuti applicando la procedura di ordinamento basata sull’analisi delle componenti principali senza alcuna rotazione. In particolare, la provincia di Bologna è prima in entrambe le graduatorie; le province di Rimini e Trieste, rispettivamente, al secondo e al terzo posto, occupano la quarta e la seconda posizione nella graduatoria che si ottiene dall’analisi delle

componenti principali. In coda, la provincia di Enna è ultima in ambedue le graduatorie; seguono le province di Caserta e Caltanissetta, che, nella graduatoria scaturita dall'analisi delle componenti principali senza rotazione, occupano posizioni molto prossime (cfr. tabella 2).

Tab. 2 – Collocazione delle province per ciascun indice sintetico

Province	¹ S	² S	³ S	⁴ S	⁵ S	⁶ S	⁷ S
Agrigento	92	99	99	88	58	78	99
Alessandria	62	71	47	7	4	5	48
Ancona	29	27	56	38	29	32	49
Aosta	36	33	53	52	48	46	44
Arezzo	37	42	35	11	14	10	41
Ascoli Piceno	42	40	61	50	37	38	60
Asti	83	83	30	17	31	22	35
Avellino	54	64	78	77	45	66	85
Bari	81	70	82	95	96	99	79
Belluno	43	46	59	41	18	30	47
Benevento	68	77	80	74	33	57	84
Bergamo	2	4	38	42	49	42	42
Biella	53	55	51	36	24	27	31
Bologna	66	59	1	30	94	69	1
Bolzano	1	2	32	51	62	54	38
Brescia	11	10	37	53	64	56	33
Brindisi	93	89	87	92	88	92	80
Cagliari	47	47	71	78	90	89	72
Caltanissetta	100	100	100	96	84	93	101
Campobasso	77	82	76	71	32	53	75
Caserta	87	88	97	101	98	101	102
Catania	98	91	93	100	97	100	93
Catanzaro	91	93	101	93	79	91	98
Chieti	52	57	72	64	27	44	68
Como	21	24	40	47	57	48	39
Cosenza	84	87	94	87	63	80	91
Cremona	8	11	42	20	17	18	40
Crotone	101	101	102	99	93	98	100
Cuneo	38	36	58	33	15	20	59
Enna	102	103	103	94	54	84	103
Ferrara	41	41	21	8	28	16	14
Firenze	71	60	3	40	91	67	4
Foggia	95	96	96	98	87	97	94
Forlì	16	15	25	12	25	19	23
Frosinone	63	65	75	72	41	60	77
Genova	88	75	17	58	83	68	10
Gorizia	65	86	5	25	76	52	6
Grosseto	40	48	48	10	6	6	46
Imperia	89	74	9	43	77	61	12
Isernia	61	79	74	66	16	41	74
L'Aquila	70	66	34	60	21	39	36
La Spezia	55	56	68	19	22	21	69
Latina	48	45	66	76	82	75	67
Lecce	73	73	86	82	66	77	83
Lecco	19	28	54	57	56	51	53
Livorno	57	52	23	37	55	43	27
Lodi	20	18	44	45	50	45	51

<i>Province</i>	¹ <i>S</i>	² <i>S</i>	³ <i>S</i>	⁴ <i>S</i>	⁵ <i>S</i>	⁶ <i>S</i>	⁷ <i>S</i>
Lucca	31	19	50	32	20	25	50
Macerata	34	29	46	34	36	33	55
Mantova	4	16	14	2	3	2	15
Massa-Carrara	78	63	63	59	52	55	62
Matera	75	76	85	81	60	73	88
Messina	97	95	88	86	71	82	87
Milano	33	30	6	63	100	86	5
Modena	10	7	8	6	34	12	11
Napoli	103	102	98	103	103	103	95
Novara	15	17	24	9	19	13	29
Nuoro	72	78	77	79	69	76	76
Oristano	58	62	73	70	51	63	73
Padova	25	23	39	56	68	58	34
Palermo	99	97	91	102	101	102	92
Parma	14	12	7	1	8	4	7
Pavia	35	38	49	18	10	11	43
Perugia	22	21	55	29	12	15	58
Pesaro e Urbino	17	14	45	15	9	8	56
Pescara	50	37	64	65	67	65	65
Piacenza	23	31	22	4	2	3	24
Pisa	26	22	28	24	38	31	28
Pistoia	30	32	16	16	39	28	22
Pordenone	13	13	43	28	30	26	37
Potenza	94	98	92	85	61	79	90
Prato	60	54	27	61	85	70	17
Ragusa	51	58	79	80	65	74	78
Ravenna	28	20	15	13	42	29	9
Reggio di Calabria	96	94	95	97	89	96	97
Reggio nell'Emilia	5	5	10	5	23	9	8
Rieti	69	84	70	54	5	24	70
Rimini	86	92	4	69	102	95	2
Roma	39	34	11	67	99	88	21
Rovigo	24	35	57	21	7	7	57
Salerno	64	61	81	84	74	83	82
Sassari	44	43	69	75	72	72	71
Savona	67	68	12	14	44	34	13
Siena	18	25	20	3	1	1	16
Siracusa	90	80	83	91	92	94	86
Sondrio	32	39	65	62	40	49	63
Taranto	80	85	90	90	81	90	89
Teramo	79	69	67	73	73	71	64
Terni	46	49	60	35	11	17	61
Torino	76	51	26	68	95	85	20
Trapani	74	72	84	83	70	81	81
Trento	3	3	36	22	26	23	54
Treviso	9	8	33	39	47	40	32
Trieste	82	81	2	23	86	59	3
Udine	27	26	29	31	43	35	26
Varese	12	9	31	44	59	50	30
Venezia	45	44	18	55	80	64	18
Verbano-Cusio-Ossola	56	53	41	46	53	47	52
Vercelli	49	50	52	26	13	14	45
Verona	7	1	13	49	78	62	19
Vibo Valentia	85	90	89	89	75	87	96

<i>Province</i>	1S	2S	3S	4S	5S	6S	7S
Vicenza	6	6	19	27	46	37	25
Viterbo	59	67	62	48	35	36	66

Attraverso il calcolo del coefficiente di correlazione di Kendall è stato possibile pervenire ad una misura delle similitudini, dal punto di vista dei risultati prodotti, delle tecniche di sintesi sperimentate nel presente lavoro (cfr. tabella 3).

Tab. 3 – Confronto tra le tecniche di sintesi

<i>Tecniche di sintesi</i>	1S	2S	3S	4S	5S	6S	7S
1S		0,798211	0,366077	0,39882	0,324957	0,396155	0,355416
2S			0,468113	0,403389	0,209214	0,33676	0,448315
3S				0,639063	0,179136	0,414049	0,816486
4S					0,540072	0,774986	0,620027
5S						0,765087	0,193604
6S							0,417856
7S							

Dal confronto non si evince una distinzione netta, in termini di similitudine, tra gli indici ottenuti mediante le procedure *monovariate* (ossia 1S_i e 2S_i) e quelli scaturiti da criteri di sintesi di natura *multivariata* (vale a dire 3S_i , 4S_i , 5S_i , 6S_i e 7S_i). È però possibile affermare che le correlazioni più elevate si registrano fra indici della stessa *specie*: tra 1S_i e 2S_i la correlazione è pari a 0,80; tra 3S_i e 7S_i è 0,82.

Se è vero, da un lato, che 1S_i e 2S_i mostrano valori molto più bassi quando sono correlati con i rimanenti cinque (1S_i non va oltre lo 0,40 e 2S_i non supera lo 0,47), dall'altro, è vero pure che le

correlazioni tra gli indici di tipo *multivariato* non sono tutte rilevanti. Al contrario, le correlazioni più basse sono quelle tra 3S_i e 5S_i (0,18) e tra 5S_i e 7S_i (0,19).

In realtà, per interpretare in maniera corretta le differenze che emergono dal confronto, occorre anche tenere presente – soprattutto per le procedure *multivariate* – i criteri di costruzione dell'indice sintetico. In questo modo, per esempio, la spiccata concordanza fra 3S_i e 7S_i si giustifica con il fatto che ambedue le tecniche utilizzano un solo asse fattoriale. Laddove il metodo di estrazione degli assi non è il medesimo, la concordanza declina.

5. Una nota conclusiva

Le conclusioni di questo lavoro non contengono indicazioni o linee guida su qual è la tecnica più efficace per sintetizzare misure elementari, dal momento che si ritiene che non esista una regola univoca che consenta di ordinare le osservazioni in modo definitivo. Occorre, invece, valutare i risultati in funzione delle scelte effettuate.

Esse definiscono l'aspetto propriamente operativo nel processo di costruzione dell'indice composto, tuttavia sono il portato di considerazioni teoriche. Pertanto, le decisioni circa la trasformazione degli indicatori elementari in valori standardizzati oppure quelle riguardanti l'importanza da attribuire a ciascuna misura – tramite criteri di ponderazione – devono essere sempre lette con riferimento a un costrutto teorico dal quale prendere le mosse (nel caso in esame, il concetto di benessere sociale).

Similmente, la scelta tra procedure *monovariate* e tecniche *multivariate* è consequenziale alla riflessione circa il numero di dimensioni ritenuto sufficiente per la sintesi delle misure elementari. Diventa dunque necessario riflettere sul fatto che alcuni indicatori possono non essere rappresentati su un unico asse fattoriale, in quanto appartenenti ad aree tematiche diverse.

Le tecniche di sintesi *multivariate* sono da prediligere quando il numero di misure elementari da sintetizzare è rilevante e quando si ritenga che il contributo di ogni indicatore debba essere diverso.

Fra le tecniche *monovariate* la sintesi del modello di Hellewig, comporta un confronto di ciascuna osservazione con una provincia considerata ideale. Il limite consiste nell'impossibilità di rinvenire nella realtà le caratteristiche che fanno dell'osservazione un caso ideale.

Un esame congiunto delle sette graduatorie evidenziate in questo lavoro suggerisce di attribuire una posizione certa soltanto a quelle osservazioni (le province) che occupano posizioni prossime in ciascuna di

esse. Affidarsi, viceversa, ad un solo criterio di sintesi comporta inevitabilmente qualche problema. Il rischio è, da una parte, quello di *allineare* indicatori molto diversi per significato e valenza esplicativa, dall'altra, di sovrastimare o sottostimare le misure da sintetizzare. Considerazioni in merito alla *performance* di una provincia rispetto ad un determinato fenomeno possono così risultare arbitrarie.

Del pari, è bene procedere con molta cautela nel formulare considerazioni riguardo alle province che occupano posizioni molto diverse in ciascuna graduatoria. Le ragioni possono essere molteplici, ma tutte devono poter essere ricondotte al processo tramite il quale un fenomeno è concettualizzato e quindi tradotto in misure empiriche.

Riferimenti bibliografici

Attanasio M., Capursi V., *Graduatorie sulla qualità della vita: prime analisi di sensibilità delle tecniche adottate*, relazione presentata al Convegno della SIEDS, Alghero, 1997.

Curatolo R., "Indicatori sociali", *Atti della XXVII Riunione scientifica della Società Italiana di Statistica*, Volume I, Palermo, 1972.

Delvecchio F., *Scale di misura e indicatori sociali*, Cacucci Editore, Bari, 1995.

Fraire M., "Problemi e metodologie statistiche di misurazione di fenomeni complessi tramite indicatori e indici sintetici", *Statistica*, Numero 2, 1989.

Giudici P., Avrini F., "Modelli statistici per la costruzione di indicatori della qualità della vita: aspetti metodologici", *Rivista di statistica ufficiale*, Numero 1, 2002.

Grasso M., "Una misurazione del benessere nelle regioni italiane", *Politica economica*, Numero 2, 2002.

Istat, *Statistiche sociali*, Volume I, Roma, 1975.

Istat, *Statistiche sociali*, Volume II, Roma, 1981.

Istat, *Sintesi della vita sociale italiana*, Roma, 1990.

Istat, *Statistiche e indicatori sociali*, Roma, 1993.

Marradi A., *Concetti e metodi per la ricerca sociale*, La Giuntina, Firenze, 1984.

Miles I., *Social Indicators for Human Development*, Frances Pinter (Publishers), London, 1985.

Noto A., "L'opinione 'corregge' la statistica", *Il Sole 24 Ore – Qualità della vita. Dossier sull'Italia del 2003*, 22 dicembre 2003.

Slottje D.J., “Measuring the Quality of Life Across Countries”, *The Review of Economics and Statistics*, Number 4, 1991.

Vitali O., *Qualità della vita e sviluppo sostenibile. Relazione introduttiva*, relazione presentata alla VI Conferenza nazionale di statistica, Roma, 2002.

Zani S., *Analisi dei dati statistici*, Volume II, Giuffrè, Milano, 2000.