

**VERSO UNA STIMA
DEL BENESSERE DEI BAMBINI**

CER – Centro Europa Ricerche

Ottobre 2004

INDICE

Abstract

- 1 Introduzione**
- 2 Le caratteristiche ideali degli indicatori e gli indicatori realmente disponibili**
- 3 Indicatori analitici e indicatori di sintesi**
- 4 Il benessere del bambino nei paesi del Mediterraneo**
 - 4.1 L'ANALISI FATTORIALE**
 - 4.2 LE DETERMINANTI DELL'INDICE DI SVILUPPO**
 - 4.3 IL BENESSERE DEL BAMBINO NEI DIVERSI PAESI**

Abstract

In questo lavoro di analisi empirica il CER propone e sperimenta una nota metodologia statistica, l'analisi per componenti principali, per il calcolo di un indicatore di benessere del bambino. L'analisi si colloca nel solco della riflessione avviata da studiosi quali Jacques Van der Gaag e Erika Dunkelberg e tocca tre aspetti essenziali: le componenti essenziali che spiegano la variabilità del benessere del bambino nei diversi Paesi; il peso relativo delle misure di policy realizzabili nelle varie materie rilevanti nella determinazione del benessere; la classificazione dei diversi Paesi considerati e la loro dislocazione in diversi gruppi con caratteristiche omogenee al loro interno. Proprio i risultati della classificazione finale, non dissimili da quelli conseguiti da lavori di impianto più deliberatamente teorico, rappresentano secondo gli Autori una conferma della congruità della metodologia utilizzata e, quindi, spingono a valorizzarne le indicazioni più preziose, quelle riguardanti il mix di politiche più adeguato a migliorare la situazione.

1. Introduzione

E' a partire dagli anni Settanta che si assiste a una divaricazione crescente tra la dinamica del reddito e l'evoluzione della "qualità della vita". Del concetto di qualità della vita possono naturalmente accogliersi diverse definizioni, che includano, per esempio, la libertà personale di scegliere¹, la salubrità dell'ambiente, la qualità delle relazioni interpersonali, ma un fatto è certo: esso non può essere così angusto da comprendere il solo elemento del reddito prodotto e distribuito.

E' per questo che, negli anni Ottanta, una serie di studi empirici, dapprima negli Stati Uniti e poi in Europa, hanno cominciato a mettere a fuoco misure alternative del benessere: l'Indice di Benessere Economico Sostenibile (Cobb, Daly, 1989), l'Indice di Salute Sociale (Miringoff, 1999), il Genuine Progress Indicator (Redefining Progress, 2001).

Il PIL rimane un indicatore molto attraente per le sue caratteristiche di apparente oggettività e universalità che derivano dalla consolidata metodologia statistica con cui viene costruito. E' difficile riconoscere a misuratori più complessi del benessere queste stesse caratteristiche: inevitabilmente mano a mano che il concetto di qualità della vita si avvicina a quello di felicità cresce il grado di soggettività delle misurazioni. Ma la sfida deve essere accettata. Diversamente, il rischio è quello di mettere sullo stesso piano situazioni che il senso comune avverte come molto diverse.

Quanto detto vale anche, naturalmente, ove si voglia costruire un indice di benessere del bambino, indice che, a sua volta, può essere considerato tra quelli che contribuiscono alla definizione di un indicatore generale e sintetico di sviluppo umano e di benessere.

Per avere un'idea delle condizioni di vita dei minori il quadro economico deve essere integrato da altri indicatori, quali il tasso di mortalità infantile, l'aspettativa di vita dei più piccoli, il tasso di frequenza scolastica o il tasso di criminalità giovanile. Si tratta solo di esempi: il benessere è un concetto multiforme sul piano concettuale e non sempre direttamente misurabile sul piano empirico.

¹ Secondo il premio Nobel Sen, per esempio, la qualità della vita tende a coincidere con la possibilità da parte degli individui di svolgere liberamente, nel rispetto dei diritti altrui, attività coerenti con i valori in cui credono.

La fase di identificazione degli indicatori costituisce, quindi, un momento delicato nel processo di costruzione di un indice di benessere. L'esclusione di un elemento rilevante a spiegare il fenomeno determina l'introduzione di una distorsione nei risultati e sbilancia il peso relativo delle variabili considerate rispetto a quelle omesse. Nel processo di selezione degli indicatori il parere di esperti che abbiano già potuto identificare le aree tematiche rilevanti è fondamentale.

Nel nostro caso il punto di riferimento essenziale è stato il saggio di Jacques Van der Gaag e Erika Dunkelberg² "Verso un indice completo del benessere dell'infanzia", che, oltre a offrire un'ampia rassegna degli studi condotti in materia, consente di identificare gli indicatori rilevanti a rappresentare il fenomeno. In particolare, gli Autori individuano le motivazioni teoriche per considerare nella misura del benessere del bambino le tre dimensioni stimate come rilevanti per la valutazione del grado di sviluppo umano in generale. Tali dimensioni sono quella economica (stile di vita e prosperità), quella culturale (grado di istruzione, diffusione della formazione scolastica della popolazione), e quella relativa alla "salute e longevità" degli individui.

2. Le caratteristiche ideali degli indicatori e gli indicatori realmente disponibili

Abbiamo citato il nostro quadro di riferimento teorico. Questo punto di riferimento è indispensabile quando si affronta il problema preliminare di individuare una serie di indicatori. Questi, a loro volta, debbono poi avere alcune caratteristiche tecniche: la significatività, la disponibilità diffusa e l'omogeneità. La significatività fa riferimento alla capacità dell'indicatore di fornire una misura quantitativa di un aspetto specifico del fenomeno; la disponibilità è legata alla possibilità di rilevarlo rispetto a tutte le unità statistiche considerate; l'omogeneità dell'indicatore dipende dal grado di comunanza dei criteri di rilevazione adottati.

Per ciò che concerne in particolare la nostra analisi, la forte eterogeneità che caratterizza le unità statistiche di riferimento, ovvero i Paesi del Mediterraneo (un insieme che comprende sia i Paesi più avanzati come la Francia, che paesi meno sviluppati come l'Albania o l'Iraq) rendeva più difficile la selezione equilibrata degli indicatori per almeno due motivi. Non solo, infatti, i Paesi meno sviluppati possono patire qualche arretratezza anche sotto il profilo della disponibilità delle informazioni, ma, soprattutto, in presenza di

² "Measuring Child Well-Being in the Mediterranean Countries – toward a comprehensive Child Welfare index", Jacques van der Gaag, Amsterdam Institute for International Development e Erika Dunkelberg, Consultant, HDNED World Bank.

strutture economico-sociali fortemente differenziate, non sempre è corretto attribuire la medesima significatività al medesimo indicatore. Vedremo più avanti come abbiamo fronteggiato questo problema che, per molti versi, rimane ineliminabile.

3. Indicatori analitici e indicatori di sintesi

In un ideale percorso metodologico, una volta individuati gli indicatori rilevanti e verificata la loro effettiva disponibilità e confrontabilità, si tratta di scegliere tra una rappresentazione più analitica e una più sintetica del fenomeno oggetto d'interesse.

Al riguardo, non c'è dubbio che una prima rappresentazione del benessere del bambino sia fornita dal semplice accostamento di tante misure diverse. Naturalmente, ogni variabile dà voce a un singolo aspetto del fenomeno. Ma quanto sono importanti i singoli aspetti? La risposta non è facile. Mentre il nesso che lega i singoli indicatori al fenomeno da spiegare è definito con chiarezza, è difficile discriminare in modo obiettivo -fra diverse combinazioni- quella associata al maggior grado di benessere. In effetti, non c'è ambiguità sul fatto che l'accesso all'acqua potabile è un requisito di benessere, così come lo è un basso tasso di criminalità giovanile o una percentuale ridotta di adolescenti che hanno avuto interruzioni della gravidanza. Tuttavia, rimane difficile capire se le condizioni di vita per un bambino sono migliori in un paese dove l'accesso all'acqua è garantito, ma l'economia è in stagnazione, piuttosto che in un altro paese dove, anche se il PIL cresce a tassi relativamente dinamici, le risorse idriche non sono diffuse a tutto il territorio. Il tentativo di costruire un indicatore che tenga conto di alcuni trade-off tra le variabili prese in considerazione allontana certamente da una rappresentazione obiettiva del fenomeno ma può far guadagnare qualcosa sul piano dei suggerimenti di politica economica e sociale. E' per questo che può avere un senso il tentativo di costruire, accanto a quelli analitici, un indicatore di sintesi.

Per la costruzione di un indice sintetico si possono scegliere tecniche differenziate, con un livello di complessità più o meno elevato.

In primo luogo, si tratta di scegliere tra un indicatore non ponderato e un indicatore ponderato.

L'indicatore non ponderato risulta dal calcolo di una media semplice degli indicatori selezionati. Il solo requisito che deve valere è l'esistenza di un legame univoco e ben definito fra ogni indicatore numerico e il fenomeno da spiegare. Ad esempio deve essere chiaro che un maggior livello di reddito pro capite equivale a un grado relativamente

più alto di benessere, così come una minore incidenza di nascite sottopeso è correlata positivamente al benessere del bambino. Se tali legami valgono per tutti gli indicatori, allora sarà sufficiente riportare ogni informazione a una medesima unità di misura, standardizzando le variabili e traducendole in numeri puri. Operata tale trasformazione, la media aritmetica rappresenterà il valore sintetico che permetterà di classificare ogni paese in una graduatoria generale. L'indice così ottenuto sarà rappresentativo del benessere del bambino, posto che non vi sia alcuna grave omissione e che l'insieme delle variabili selezionate esaurisca tutta l'informazione necessaria a spiegare il fenomeno.

In definitiva quando si realizza una sintesi non ponderata, le graduatorie costruite sulle singole variabili vengono sommate o mediate. La semplicità del metodo, sia concettuale che elaborativa, rappresenta un vantaggio. Tuttavia, mancando di un criterio di ponderazione, non si stabiliscono relazioni tra le variabili, e, quindi, non si indagano quei trade off che possono essere così interessanti da conoscere in sede di definizione delle politiche.

Prendendo spunto da quest'ultima considerazione si può sviluppare un nuovo metodo aggregativo che somma le singole graduatorie "pesando" le diverse variabili in analisi.

Nella definizione dei pesi due sono le principali traiettorie metodologiche. Nella prima, la taratura avviene basandosi su scelte soggettive di esperti o in base a interpretazioni teoriche del fenomeno (una visione economico-centrica attribuirà un peso notevole alle variabili economiche, considerando le altre come una "sovrastuttura" o un portato dello sviluppo dell'economia). Si può ridurre il grado di soggettività, aiutandosi con analisi statistiche che consentano di distinguere tra variabili "guida" e variabili di "sovrastuttura" in ragione della capacità esplicativa del fenomeno. Alle prime può poi essere assegnato un peso massimo nella costruzione dell'indice ponderato.

Nella seconda traiettoria metodologica, la taratura avviene attraverso una metodologia statistica che individua direttamente i pesi da assegnare a ciascuna delle variabili considerate. La nostra scelta è stata proprio quella di privilegiare questa seconda strada che, come è evidente da quanto abbiamo detto, può essere di ausilio anche nell'arricchimento del primo percorso metodologico.

A questo scopo abbiamo fatto ricorso all'analisi fattoriale ovvero, per essere più precisi, all'analisi in componenti principali che ne costituisce la particolare applicazione a dati di natura quantitativa. Tale metodologia viene normalmente utilizzata nello studio di fenomeni complessi, definiti dall'operare in differenti direzioni di molteplici fattori, e

consente di raggiungere tre risultati fondamentali: uno studio preliminare e approfondito sulle relazioni tra le variabili, la costruzione di un indicatore sintetico e la graduazione delle unità di rilevazione (nel nostro caso i paesi) sulla base di quest'ultimo.

Il principale pregio delle analisi fattoriali è quello di riuscire a gestire basi dati molto complesse in maniera unitaria, sintetizzando l'informazione da esse rivenienti sulla base delle relazioni osservate tra le singole variabili. Il "benessere del bambino", rientra a pieno titolo tra i fenomeni adatti ad essere esplorati tramite queste metodologie, visto il gran numero di fattori che intervengono nella sua definizione.

Rinviando all'appendice metodologica per una più completa e precisa descrizione della tecnica in questione, possiamo cercare di sintetizzarne qui le principali proprietà. Posta come unica assunzione che l'insieme delle variabili "primarie" considerate siano esaustive del benessere del bambino nei paesi del Mediterraneo, la tecnica menzionata studia il complesso delle relazioni tra variabili, per analizzarne i "comovimenti" e costruire sulla base di questi, i cosiddetti "fattori", ottenuti come combinazione lineare delle variabili di partenza. Tali fattori altro non sono che nuove variabili che sintetizzano le informazioni originariamente contenute nelle variabili di base. L'analisi restituisce tanti fattori quante sono le variabili primarie incluse nel calcolo. Essi sono ordinabili in base alla quota della variabilità totale spiegata, che a sua volta, può essere considerata come una misura del loro potere riassuntivo, o meglio della loro capacità di sintetizzare, senza una rilevante perdita di informazione.

Ma quale è il legame tra le nuove variabili e le variabili originarie? Per ciascuna variabile "primaria", considerata nell'analisi, è presente un coefficiente (coordinata sul fattore) il quale ne misura il peso nella determinazione di una generica variabile "secondaria". Quest'ultima, quindi, risulta caratterizzata in misura maggiore da quelle variabili cui corrispondono i coefficienti più grandi in valore assoluto: sono proprio queste variabili a conferire significato al fattore e a individuare come i singoli aspetti si leghino nella definizione del benessere del bambino. Dunque analizzando tali coordinate per i diversi fattori individuati, si possono comprendere i legami tra variabili originarie, e tramite lo studio di questi ultimi, cogliere il significato stesso della variabile "secondaria". Nella terminologia dell'analisi in questione si dice che in tal modo si "battezza" il fattore considerato.

Ma l'individuazione di tutti gli assi fattoriali non esaurisce il compito di elaborazione. Lo sforzo del ricercatore è volto a questo punto ad individuare, fra tutti, proprio i fattori rappresentativi del benessere.

Infatti anche se un fattore sintetizza una quota rilevante di variabilità totale, potrebbe essere il caso che esso dia un peso “preponderante” a indicatori e relazioni giudicati dal ricercatore “marginali”, o addirittura spurie, rispetto al fenomeno da sintetizzare. La sensibilità del ricercatore deve intervenire a identificare tra tutte le relazioni presenti nei fattori solo quelle rilevanti per la rappresentazione multi-dimensionale del fenomeno.

Una volta selezionati gli assi rilevanti, l’analisi restituisce anche le coordinate sui fattori per ciascuna delle unità di rilevazione (i Paesi), che sono utilizzabili, direttamente, o tramite ulteriori elaborazioni, ai fini della graduatoria.

Il metodo più immediato per elaborare una graduatoria a partire da un solo asse fattoriale consiste nel considerare come valore dell’indice proprio la coordinata di ogni Paese su tale fattore. Chiarita la “direzione” dell’asse rispetto al fenomeno che si vuole spiegare (ad esempio una coordinata positiva equivale a maggior benessere relativo e viceversa) si tratterà solamente di normalizzare i valori delle coordinate per assicurare che ognuna si muova in un intervallo di variazione definito di cui sia noto l’ordine di grandezza.

L’analisi in componenti principali non consente solamente di giungere a una graduatoria generale del benessere del bambino, ma permette anche di estendere l’indagine in altre direzioni. In particolare, è possibile scomporre i risultati ottenuti identificando, per ciascun paese, aree di sofferenza e aree di successo.

La metodologia impiegata individua da un lato il mix di caratteristiche necessarie per conseguire con successo il benessere del bambino e dall’altro restituisce il valore delle coordinate dei paesi su ogni fattore rilevante. Osservando la performance di un paese rispetto agli elementi necessari a garantire il successo in termini di benessere generale è possibile discernere le caratteristiche che collocano quel paese in un dato posto della graduatoria. Inoltre, mantenendo la distinzione degli indicatori in aree tematiche, è possibile verificare se un risultato deludente dal punto di vista del benessere generale debba essere attribuito a carenze concentrate in una specifica area tematica, si tratti di progresso in termini di istruzione, piuttosto che di condizioni economiche o altro. Sulla base di queste informazioni, si può quindi sfruttare l’analisi per trarre indicazioni di tipo normativo, ovvero identificare i percorsi da seguire per guadagnare terreno in classifica e quindi progredire in termini di benessere.

L’analisi in componenti principali identifica le caratteristiche di un fenomeno che si presenti per tutti le unità di rilevazione con le medesime caratteristiche. Sottesa all’analisi,

vi è l'ipotesi che esiste un'unica nozione di benessere del bambino e che questa sia comune a tutti i paesi. Nonostante ciò, un possibile approfondimento dello studio consiste nell'indagare il *data set* per individuare diversi modelli di benessere per il bambino. L'idea è quella di verificare se, nell'ambito delle variabili che lo condizionano, i vari paesi hanno adottato modelli diversi, sulla base di una diversa matrice culturale, dell'indirizzo politico o della mancanza di risorse. Tale analisi può portare a definire tante classifiche per quante sono le tipologie di paesi individuati. E' questo il modo in cui abbiamo ritenuto di fronteggiare il problema enunciato infine del paragrafo 2.

4. Il benessere del bambino nei paesi del Mediterraneo

Abbiamo utilizzato la metodologia descritta per identificare un indicatore di benessere dei bambini nei paesi del Mediterraneo.

Il data set di riferimento, costituito principalmente di dati di fonte ONU, è stato fornito da Lynkeus e contempla complessivamente 33 paesi appartenenti a cinque diverse aree geografiche: Paesi arabi, Medio oriente, Nord Africa, Europa dell'est e Europa del Mediterraneo. Le variabili complete di cui sono disponibili rilevazioni per ciascuno dei Paesi sono in tutto 39, distinte in sei aree tematiche: demografia, nutrizione, salute, istruzione, economia e indicatori sociali. L'elenco completo dei paesi e delle variabili disponibili è riportato nelle tavole 1a e 1b.

Tavola 1a. I paesi del Mediterraneo

Paesi arabi	1) Bahrain 2) Iran 3) Iraq 4) Kuwait 5) Oman 6) Qatar 7) Arabia Saudita 8) Emirati arabi uniti 9) Yemen	Medio oriente	1) Israele 2) Giordania 3) Libano 4) Siria	Nord Africa	1) Algeria 2) Egitto 3) Libia 4) Marocco 5) Tunisia
Europa dell'Est	1) Albania 2) Bosnia 3) Bulgaria 4) Croazia 5) Romania 6) Serbia e Montenegro 7) Slovenia 8) Macedonia 9) Turchia	Europa del mediterraneo	1) Cipro 2) Francia 3) Grecia 4) Italia 5) Malta 6) Spagna		

Prima di passare in rassegna i risultati è bene precisare che il data set originario è stato adattato alle esigenze dell'analisi sotto due aspetti. Innanzi tutto ognuna delle variabili osservate è stata opportunamente trasformata e ricondotta ad assumere intervalli di variazione il più possibile omogenei. Tale trasformazione è necessaria per evitare che le diverse unità di misura con cui le variabili sono state rilevate distorcano i risultati dell'analisi. Lo scopo è di non attribuire rilevanza alle variabili solo in quanto caratterizzate da un ordine di grandezza relativamente maggiore, come è il caso ad esempio del livello del prodotto pro capite e del tasso di mortalità infantile: il primo infatti segna una media di circa 10 mila dollari, valore decisamente sproporzionato rispetto al secondo che, essendo una percentuale, varia necessariamente fra 0 e 100.

In secondo luogo il numero delle variabili considerate è stato ulteriormente ridotto di tutta una serie di indicatori considerati ridondanti. Si tratta di indicatori che colgono solo aspetti specifici di uno stesso fenomeno, e pertanto non aggiungono informazione rilevante alle nostre finalità: è il caso ad esempio di alcuni indicatori sociali quali il numero di televisioni, di radio e di telefoni. Tutti e tre danno una indicazione del grado di diffusione delle telecomunicazioni, presentano un elevato grado di correlazione e non si differenziano sostanzialmente fra i paesi. Per questi motivi sono stati aggregati in un unico indicatore, chiamato appunto diffusione delle telecomunicazioni (TLC). Sulla base di analoghe considerazioni altri indicatori sono stati aggregati, trasformati o eliminati dall'analisi; nella tavola 2 è riportato l'elenco definitivo delle 19 variabili complessivamente incluse nelle elaborazioni.

Il data set così predisposto ha costituito la base per l'applicazione della analisi in componenti principali (Acp).

Come abbiamo detto, questa analisi consente di fornire una rappresentazione sintetica del fenomeno, attraverso l'estrazione degli assi fattoriali. Il primo fattore raccoglie la quota principale di varianza; il secondo, ortogonale al primo, fornisce una rappresentazione della varianza residua, e così via. E' chiaro perciò che le relazioni descritte dai primi assi fattoriali sono quelle che sintetizzano la quota maggiore di informazione. La capacità sintetica dell'analisi, quindi, dipende da quanto è grande la quota di variabilità raccolta dal primo, o al massimo dai primi due assi fattoriali.

Tavola 1b. Elenco delle variabili complete

INDICATORI DEMOGRAFICI
1) Popolazione 0-4 anni
2) Popolazione 0-17 anni
3) Popolazione 5-9 anni
4) Popolazione 10-14 anni
5) Popolazione 15-19 anni
6) Popolazione totale
7) Child dependency ratio
8) Tasso di natalità
9) Tasso totale di fertilità
10) Tasso di mortalità infantile
11) Aspettativa di vita alla nascita (Totale)
12) Popolazione urbana
13) Crescita della popolazione urbana
INDICATORI DELL'ALIMENTAZIONE
1) Bambini nati sotto peso
INDICATORI DELLA SALUTE
1) Mortalità delle madri per parto
2) Bambini di 1 anno vaccinati contro il DPT
3) Bambini di 1 anno vaccinati contro la poliomielite
4) Bambini di 1 anno vaccinati contro il morbillo
5) Numero dei medici
6) Spesa sanitaria totale
7) Spesa sanitaria procapite
8) Spesa sanitaria privata
9) Spesa sanitaria pubblica
10) Spesa sanitaria out-of-pocket
INDICATORI DELL'ISTRUZIONE
1) Iscrizione a all'istruzione e alla cura della prima infanzia (Totale)
2) Tasso di iscrizione alla scuola secondaria (lordo)
3) Tasso di iscrizione all'asilo (lordo)
4) Rapporto studenti insegnanti nella scuola primaria
5) Indicatore di educazione
INDICATORI ECONOMICI
1) Reddito nazionale lordo pro-capite
INDICATORI SOCIALI
1) Numero di radio
2) Numero di televisioni
3) Linee telefoniche
4) Contratti di telefonia mobile
5) Numero di computer in uso
6) Utilizzatori di internet
7) Indicatore di sviluppo umano
8) Indicatore di sviluppo umano
9) Popolazione con accesso a servizi sanitari adeguati

L'ACP applicata alle 19 variabili riportate in tavola 2 ha permesso di estrarre un primo asse fattoriale che spiega ben il 49 per cento, ovvero poco meno della metà, della variabilità complessiva. Poiché la capacità sintetica del primo asse è risultata decisamente elevata, le relazioni così descritte si sono rivelate già sufficienti a fornire una descrizione del fenomeno latente che cerchiamo di spiegare. Ne riportiamo nella figura 1 la rappresentazione grafica.

Tavola 2. Elenco delle variabili incluse nell'analisi

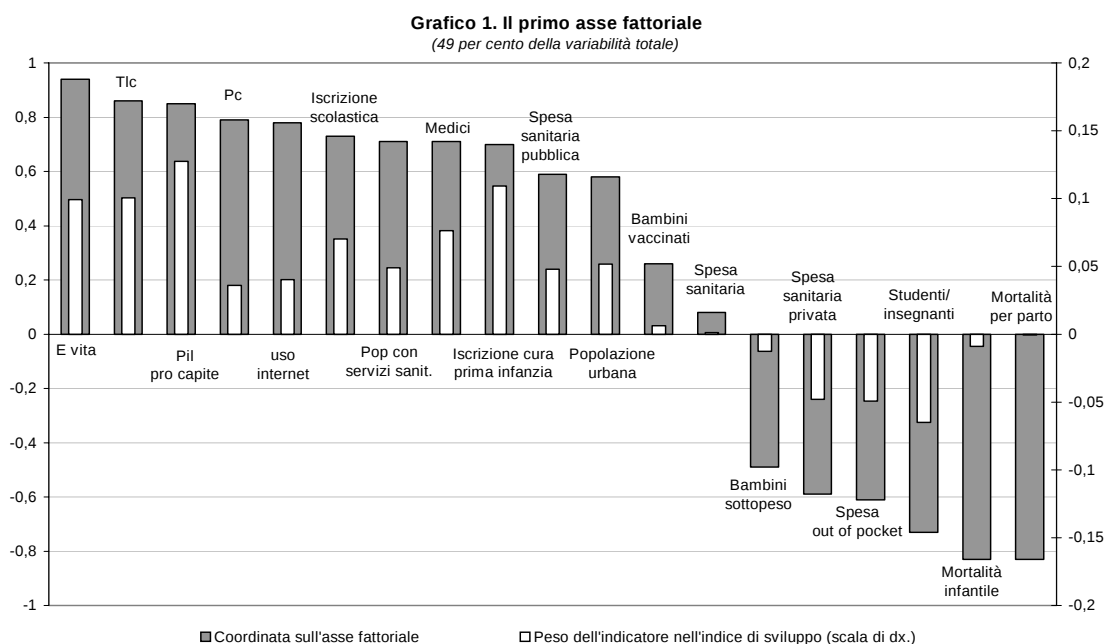
INDICATORI DEMOGRAFICI
1) Tasso di mortalità infantile
2) Aspettativa di vita alla nascita (Totale)
3) Popolazione urbana
INDICATORI DELL'ALIMENTAZIONE
1) Bambini nati sotto peso
INDICATORI DELLA SALUTE
1) Mortalità delle madri per parto
2) Bambini vaccinati
3) Numero dei medici
4) Spesa sanitaria in % del PIL
5) Spesa sanitaria privata (in % del totale)
6) Spesa sanitaria pubblica (in % del totale)
7) Spesa sanitaria out-of-pocket (in % del totale)
INDICATORI DELL'ISTRUZIONE
1) Rapporto studenti insegnanti nella scuola primaria
2) Iscrizione a all'istruzione e alla cura della prima infanzia (Totale)
3) Tasso di iscrizione alla scuola secondaria (lordo)
INDICATORI ECONOMICI
1) Reddito nazionale lordo pro-capite (indice)
INDICATORI SOCIALI
1) Diffusione delle TLC
2) Numero di computer in uso
3) Utilizzatori di internet
4) Popolazione con accesso a servizi sanitari adeguati

4.1 *L'analisi fattoriale*

Nel grafico sono riprodotte con l'istogramma scuro le coordinate delle singole variabili sul primo asse (o fattore). Ogni coordinata va letta secondo due diversi criteri. Il primo è legato alla dimensione in valore assoluto della coordinata stessa (ovvero in termini grafici alla “grandezza” dell'istogramma): essa dà conto dell'intensità del legame fra variabile sottostante e il fenomeno da sintetizzare; perciò coordinate più elevate (ad esempio l'aspettativa di vita, che ha coordinata pari a circa 0,9) sono relativamente più correlate con il benessere del bambino rispetto a coordinate prossime a zero .

Il secondo criterio riguarda invece la “vicinanza” o la “distanza” delle coordinate stesse fra di loro. Coordinate positive e “prossime” stanno a significare che le due variabili tendono ad avere un elevato grado di correlazione fra di loro; al contrario coordinate che hanno segni opposti, e sono quindi “distanti”, indicano che le variabili sottostanti sono caratterizzate da correlazione negativa. Ad esempio, si riscontra nei dati che laddove l'aspettativa di vita è elevata (coordinata positiva e grande di dimensione) si assiste anche a un forte grado di diffusione delle telecomunicazioni (TLC, coordinata anch'essa positiva e grande di dimensione, dunque “vicina”); di converso, negli stessi paesi, si riscontra pure un tasso decisamente basso di mortalità infantile (variabile con coordinata decisamente negativa, che è appunto in “opposizione” alle precedenti in quanto raggiunge la “massima distanza” da quelle positive).

Per agevolare la lettura dei risultati, abbiamo ordinato le coordinate del primo fattore dalla più grande alla più piccola. Così nel grafico 1 procedendo da sinistra verso destra, incontriamo dapprima le coordinate positive e poi quelle negative.



La prima evidenza che emerge immediatamente con chiarezza è che le relazioni fra variabili selezionate e il fenomeno latente sono coerenti con le più intuitive aspettative a priori, confermando che il primo asse costituisce una sintesi credibile del benessere del bambino.

L'analisi infatti ribadisce l'esistenza di una relazione positiva fra benessere del bambino e livelli relativamente più elevati dell'aspettativa di vita. Così pure è positiva la correlazione fra il benessere e il livello del pil pro-capite, indicatore per eccellenza delle condizioni economiche e almeno in parte delle condizioni di vita. Il benessere è anche legato a una serie di indicatori di sviluppo sociale, quali il grado di diffusione delle telecomunicazioni, il possesso dei personal computer e l'uso di internet.

In misura minore (le coordinate vanno via via assottigliandosi man mano che si procede verso destra) osserviamo che caratterizzano il fenomeno anche gli elevati tassi di partecipazione scolastica, la presenza di servizi sanitari adeguati e la congrua disponibilità di assistenza medica.

Passando ora all'estremo opposto del grafico, troviamo che sono correlati negativamente con il benessere del bambino (le coordinate sono infatti negative) un insieme di altri indicatori. Fra le variabili legate all'istruzione ad esempio, osserviamo che un basso rapporto fra studenti e insegnanti è indice di benessere. Così pure, per ciò che concerne gli aspetti demografici, lo sono bassi tassi di mortalità infantile o bassi tassi di mortalità delle madri per parto.

Per quanto riguarda invece gli indicatori della sanità, il risultato appare diversificato. Innanzi tutto emerge che la spesa sanitaria complessiva (in rapporto al PIL) non caratterizza in

modo rilevante questo asse fattoriale. La sua coordinata è fra tutte quella più vicina allo zero. Tuttavia si osserva che la sua componente pubblica si associa positivamente ai fattori di sviluppo, mentre le sue componenti private sono connotate da coordinate negative. Tale risultato riflette il fatto che il campione di Paesi di riferimento è in qualche misura sbilanciato. I casi evidenti di successo in termini di benessere del bambino, come la Francia o l'Italia, sono tutti paesi in cui è consolidato da tempo un sistema sanitario a base prevalentemente pubblica. I paesi dove il benessere è meno diffuso sono invece Paesi dove il sistema sanitario è ancora poco presente o comunque alimentato solo dai capitali privati. Lo stesso esercizio ripetuto a livello mondiale potrebbe non restituire lo stesso risultato e in questo senso va interpretato con cautela.

Nonostante, nel loro complesso, le relazioni individuate siano piuttosto intuitive, il benessere del bambino si conferma come un fenomeno multidimensionale; ce lo dice la presenza sul primo fattore di diverse variabili caratterizzate da coordinate (in valore assoluto) elevate. La multidimensionalità del benessere viene adeguatamente rappresentata sul primo asse, la cui potenza sintetica è come abbiamo detto molto forte, e non vi è dunque alcuna necessità di rintracciare negli altri fattori ulteriori elementi che servano a spiegare il fenomeno.

4.2 *Le determinanti dell'indice di sviluppo*

Per determinare la rilevanza di ciascuna variabile nello spiegare il benessere del bambino abbiamo effettuato ulteriori elaborazioni completando l'informazione che viene dalla coordinata della variabile sul fattore con informazioni sulla dimensione stessa della variabile. A tale scopo, era importante valutare quanto azioni di policy su determinate aree di intervento influenzassero il benessere del bambino: tradotto in termini statistici ciò significava valutare quanto impulsi, misurati come deviazioni standard delle variabili introdotte nell'analisi, avessero effetto sull'indice complessivo. La misura di tale contributo è anch'esso riportato nel grafico 1 con l'istogramma bianco.

Visto in tale modo, il quadro cambia sostanzialmente rispetto alla visione dei valori delle coordinate. Come era lecito attendersi il peso più rilevante nel modificare il benessere dei bambini viene giocato dal reddito pro-capite. Ma tale supremazia è tutt'altro che schiacciante:

dato un totale di variazioni, il reddito ne determina “solo” il 12,8%. Questo dato ci allontana da una visione meramente economicistica.

E, in effetti, a ben vedere, sono tutti gli altri elementi che nel loro complesso determinano in modo preponderante l'indice di benessere del bambino. Vi è subito da rilevare l'importante ruolo giocato dalle variabili di istruzione. L'indicatore di istruzione e cura della prima infanzia riveste una importanza particolare con una potenza esplicativa pari all'11 per cento. Quando considerate nel loro complesso, le variabili legate all'istruzione rappresentano una delle più importanti aree di intervento, con un peso complessivo nell'indice di benessere del bambino pari al 24,4 per cento.

Anche l'insieme di indicatori legati alla sanità e alla nutrizione determinano un ruolo importante nella definizione dell'indice. Complessivamente il loro peso ammonta al 24,3 per cento del totale, una misura sostanzialmente equivalente a quella dell'istruzione. E' il numero dei medici, con un peso nell'indice del 7,6 per cento, ad avere un ruolo dominante. Fra gli altri elementi contemplati nell'area legata alle condizioni di salute, vi è la spesa sanitaria pubblica, che determina il 4,8 per cento dell'indice. Mentre per la componente privata della spesa sembra valere il contrario, per cui una sua riduzione sembra associarsi positivamente allo sviluppo. Fra le variabili legate alla nutrizione infine sottolineiamo che la percentuale di bambini nati sotto peso ha un ruolo molto limitato nella determinazione dell'indice (appena l'1,5 per cento).

I fattori di sviluppo sociale contano per un altro 22,6 per cento del totale. La diffusione delle telecomunicazioni, con un contributo nell'indice del 10 per cento, è la determinante principale. Seguono altri due indicatori di sviluppo tecnologico, quali l'uso del computer e la diffusione di internet, che insieme ammontano ad un contributo pari al 7,6 per cento. Quasi il cinque per cento dell'indice è invece influenzato dalla diffusione e dalla qualità dei servizi sanitari.

La parte rimanente dell'indice, ovvero circa il 16 per cento, è da attribuire ai fattori demografici: l'aspettativa di vita è il più rilevante e spiega il 10 per cento delle variazioni dell'indice. A seguire, osserviamo che il fenomeno dell'inurbamento, sintetizzato da una percentuale crescente di popolazione che vive negli agglomerati urbani, spiega il benessere per il 5,2 per cento. Poco meno dell'uno per cento è invece il peso nell'indice della mortalità infantile.

In definitiva, il contributo del PIL è rilevante di per sé, ma non preponderante rispetto ad altri elementi nella determinazione del benessere del bambino. In generale, il risultato ottenuto dà pienamente ragione dei tentativi di definire indici di benessere estesi che

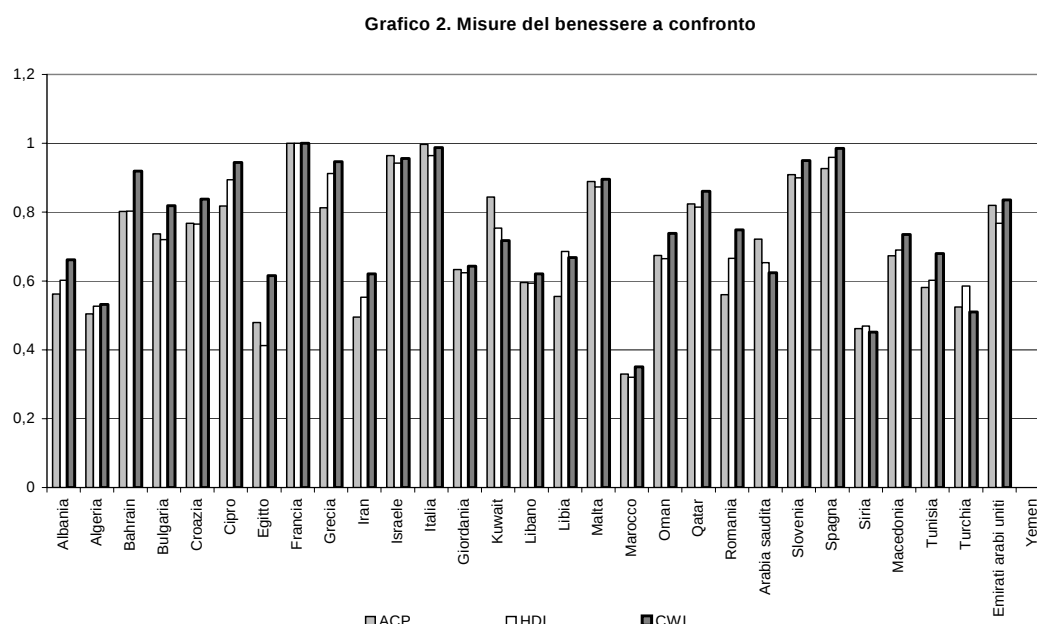
considerino, accanto ai fattori di natura strettamente economica, anche ulteriori determinanti dello sviluppo.

In termini di indicazioni di policy, ne scaturisce l'esigenza di interventi che ambiscano non solo a stimolare la dinamica del prodotto, ma anche a garantire alla popolazione maggiore istruzione e maggiore sicurezza sanitaria, con l'obiettivo di affiancare alla crescita economica la creazione di una società cui non manchino gli elementi basilari per una sopravvivenza dignitosa, e che sia pronta a raccogliere le sfide dello sviluppo tecnologico. Sul piano statistico, l'analisi suggerisce le proporzioni di tali interventi.

4.3 *Il benessere del bambino nei diversi paesi*

Abbiamo, infine, utilizzato le coordinate e i pesi delle variabili per costruire una classifica dei vari paesi rispetto al benessere del bambino.

Nel grafico 2 riportiamo il valore dell'indice di benessere ottenuto sulla base delle componenti principali a confronto con il valore dell'indice di sviluppo umano e dell'indice di benessere del bambino (così come presentati nell'articolo di Van der Gaag, rispettivamente alle pagine 63 e 65).³



³ Per effettuare il confronto abbiamo standardizzato i valori dei tre indici, facendoli muovere in un comune campo di variazione.

L'indice da noi elaborato a partire dall'ACP si muove in linea con gli altri due indici di benessere. Il grado di correlazione con l'indice di sviluppo umano è di oltre il 97 per cento, del 95 per cento circa con l'indicatore di benessere dei bambini.

Le graduatorie redatte a partire da questi valori sono riportate nella tavola 3: si osserva che l'indice basato sulle componenti principali individua gli stessi casi di successo (ad esempio la Francia e l'Italia sono sempre al primo e al secondo posto rispettivamente).

Tavola 3. Classifiche a confronto

	Indice di benessere basato sull'ACP		Indice di sviluppo umano (HDI)		Indice di benessere del bambino (CWI)	
	Valore	Rango	Valore	Rango	Valore	Rango
Francia	1,000	1	1,000	1	1,000	1
Italia	0,997	2	0,965	2	0,987	2
Israele	0,964	3	0,943	4	0,956	4
Spagna	0,927	4	0,959	3	0,986	3
Slovenia	0,909	5	0,900	6	0,949	5
Malta	0,889	6	0,873	8	0,896	9
Kuwait	0,844	7	0,754	13	0,717	17
Qatar	0,824	8	0,814	9	0,861	10
Emirati arabi uniti	0,820	9	0,767	11	0,836	12
Cipro	0,818	10	0,894	7	0,945	7
Grecia	0,813	11	0,912	5	0,946	6
Bahrain	0,802	12	0,804	10	0,919	8
Croatia	0,767	13	0,766	12	0,837	11
Bulgaria	0,737	14	0,721	14	0,818	13
Arabia Saudita	0,722	15	0,654	19	0,624	22
Oman	0,674	16	0,665	18	0,738	15
Macedonia	0,673	17	0,690	15	0,735	16
Giordania	0,633	18	0,625	20	0,643	21
Libano	0,596	19	0,593	23	0,621	23
Tunisia	0,581	20	0,602	21	0,679	18
Albania	0,563	21	0,602	22	0,662	20
Romania	0,560	22	0,666	17	0,749	14
Libia	0,555	23	0,686	16	0,668	19
Turchia	0,525	24	0,586	24	0,510	27
Algeria	0,504	25	0,527	26	0,532	26
Iran	0,495	26	0,553	25	0,621	24
Egitto	0,479	27	0,413	28	0,616	25
Siria	0,462	28	0,470	27	0,452	28
Marocco	0,330	29	0,321	29	0,351	29
Yemen	0,000	30	0,000	30	0,000	30

Verifichiamo che, anche per i casi di insuccesso, le tre graduatorie collocano gli stessi paesi al penultimo e all'ultimo posto (infatti si tratta sempre di Marocco e Yemen).

Il forte grado di analogia con gli altri indicatori sintetici ci induce a reputare plausibili i risultati ottenuti e a valorizzare, quindi, anche le conclusioni di policy che analisi come la

nostra possono contribuire a mettere meglio a fuoco grazie all'attribuzione di pesi relativi alle aree critiche in cui l'azione risulterebbe più efficace.

Appendice Metodologica

L'Analisi Fattoriale

Sotto la definizione di analisi fattoriale si raccolgono un'insieme di tecniche statistiche di tipo multidimensionale che tentano di sintetizzare le informazioni di una larga massa di dati interconnessi tra di loro perseguendo l'obiettivo di mantenere il massimo delle potenzialità esplicative della molteplicità delle informazioni sotto la condizione di semplificare e rendere più accessibili i contenuti delle stesse.

L'impostazione dell'analisi fattoriale che fa riferimento alla scuola francese dell'*Analyse des données*, si basa su una formalizzazione della matrice dei dati su di uno spazio non probabilizzato (in particolare, su di uno spazio vettoriale euclideo); ed è riassumibile da un modello del tipo:

$$(\underline{D}, R^m, (R^n), \underline{M})$$

con R^m spazio vettoriale delle unità (R^n spazio vettoriale delle variabili) e $\underline{M}$ una metrica su tale spazio.

Lo scopo di tale modello è l'analisi simultanea di una molteplicità di variabili rappresentando la matrice $\underline{D}$ come una nuvola di punti nello spazio R^m o nello spazio R^n .

Per m o n piuttosto grandi (come nel caso in esame) risulta difficile dare un'interpretazione dei dati leggendo direttamente le posizioni dei punti in spazi di tali dimensioni: un modello statistico così impostato deve necessariamente far ricorso a *sottospazi* di dimensione inferiore su cui proiettare i punti, in modo da poter interpretare le distanze, nonché il quadro delle posizioni reciproche e di associazione, tra i punti stessi.

Il problema chiave diviene, quindi, quello di determinare questi sottospazi di proiezione in modo che la riduzione della dimensione conservi il massimo dell'informazione originaria possibile e in modo da ridurre l'effetto distorsivo di *schacciamento* che l'operazione di

proiezione induce sulle reali distanze tra i punti. Si tratta, in altri termini, di individuare il sottospazio che minimizza tale distorsione e che, tra tutti i sottospazi paralleli ad esso, meglio si adatta alla situazione osservata.

La soluzione del problema⁽⁴⁾ è data dall'individuazione degli *assi o dimensioni fattoriali*: essi rappresentano nuove variabili statistiche (ottenute come combinazione lineare delle variabili originarie) che, riducendo e sintetizzando la matrice $\underline{D}$, consentono di leggere in modo semplificato, ma possibilmente più ricco di significato, l'informazione originaria.

Tra le principali caratteristiche dell'approccio descritto, oltre alla creazione di dimensioni ridotte attraverso le quali *leggere* in modo semplice e significativo l'informazione contenuta in matrici di dati complesse, assumono un ruolo essenziale:

- l'esplicitazione della dualità tra unità e variabili (che consente la valorizzazione della dimensione delle unità statistiche);
- la ricerca di forme e strutture difficilmente individuabili a priori;
- l'uso sistematico di rappresentazioni grafiche, molto utili e semplici dal punto di vista interpretativo e comunicativo.

Gli assi fattoriali, essendo per costruzione ortogonali, risultano *informativi*, ovvero ogni asse spiega una parte della variabilità complessiva del fenomeno non spiegata dagli altri assi.

Inoltre, tali assi sono individuati in ordine gerarchico decrescente, in modo che la variabilità spiegata dal primo sia superiore a quella spiegata dal secondo che, a sua volta, è superiore a quella del terzo e così via.

Infine, l'interpretazione dei risultati dell'analisi si basa su diversi indici della qualità di rappresentazione, differenti a seconda della tecnica specifica utilizzata, ma costruiti sempre a partire dagli assi fattoriali.

L'Analisi in componenti principali

⁴Per ulteriori approfondimenti su tale impostazione, dovuta principalmente alla scuola francese e meglio nota come 'Analyse des données', cfr.:

- J.P.Benzecri, "L'Analyse des Données, Tome 1: La Taxinomie, Tome 2: L'Analyse des correspondances", Dunod, Paris, 1973;
- R. Coppi, "Appunti di statistica metodologica: analisi lineare dei dati" Dipartimento di Statistica e Calcolo delle Probabilità, Università degli Studi "La Sapienza", Roma 1986;
- L.Lebart, A.Morineau, J.P.Fenelon, "Traitement des données statistiques", Dunod, Paris, 1982

L'analisi in componenti principali è una tecnica statistica multivariata che trova impiego nell'ambito della sintesi dei dati e nello studio delle variabili latenti..

Accade spesso che per il fenomeno sotto studio si disponga di un elevato numero di variabili ognuna delle quali rappresenta uno o più aspetti del fenomeno stesso. Sotto tali condizioni risulta evidente la necessità di rappresentare il fenomeno con un numero inferiore di variabili derivanti da quelle originariamente osservate che conservino quanta più informazione possibile.

Il metodo delle componenti principali risolve questo problema costruendo un insieme di p variabili "artificiali" $(\xi_1, \xi_2, \dots, \xi_p)$, ottenute come combinazione lineare delle k variabili originarie $(X_1, X_2, \dots, X_k)$ con $p < k$, che siano incorrelate e tali che ognuna di esse abbia varianza massima .

Ciò accade perché una variabile statistica con una elevata variabilità tende ad assumere modalità molto differenti tra loro, se quindi si fa riferimento ad un collettivo di individui su cui la variabile statistica è rilevata la sua alta variabilità implica che gli individui tendono a differenziarsi fortemente in termini della variabile in questione.

In altre parole è necessario assicurarsi che la variabile artificiale sia in grado di fornire un contributo informativo rilevante o meglio che le componenti principali $\xi_1, \xi_2, \dots, \xi_p$ siano in grado di riassumere o sintetizzare il fenomeno in esame senza che ciò comporti una rilevante perdita di informazione.

Scopo dell'analisi in componenti principali, nell'ottica geometrica precedentemente considerata, è quello di rappresentare le unità e le variabili in sottospazi di bassa dimensione (contenuti rispettivamente in R^m e R^n) che conservino il massimo dell'informazione statistica presentata dalle due nuvole di punti presenti in tali spazi.

Si consideri una variabile statistica $X = (X_1, X_2, \dots, X_k)'$, e si indichino rispettivamente con μ e Σ il suo vettore di medie e la sua matrice di varianze e covarianze.

Come abbiamo precedentemente affermato il nostro scopo è quello di costruire una variabile artificiale ξ_1 , combinazione lineare delle variabili originariamente osservate:

$$\xi_1 = \alpha_{11}X_1 + \alpha_{12}X_2 + \dots + \alpha_{1k}X_k = \alpha_1' X$$

tale che la sua varianza risulti massima.

Poiché $E(\xi_1) = \alpha_1' \mu$ e $Var(\xi_1) = \alpha_1' \Sigma \alpha_1$, è evidente che per individuare la prima componente principale sarà necessario determinare il valore del vettore a k componenti α_1 .

Al fine di ottenere una soluzione univocamente determinata, occorre porre il seguente vincolo sulle componenti del vettore α_1 :

$$\alpha_1' \alpha_1 = 1$$

Essenzialmente si tratta di risolvere il seguente problema di minimo vincolato:

$$\begin{cases} \max & \alpha_1' \Sigma \alpha_1 \\ & \alpha_1' \alpha_1 = 1 \end{cases}$$

utilizzando il metodo dei Moltiplicatori di Lagrange si ricava la condizione

$$\Sigma \alpha_1 = \lambda \alpha_1$$

dove λ è il moltiplicatore di Lagrange.

La precedente è un'equazione agli autovalori ed evidenzia come λ , sia un autovalore della matrice Σ , mentre α_1 è il corrispondente autovettore di norma unitaria.

D'altra parte l'uguaglianza $\Sigma \alpha_1 = \lambda \alpha_1$ e il vincolo $\alpha_1' \alpha_1 = 1$ implicano che:

$$\text{Var}(\xi_1) = \alpha_1' \Sigma \alpha_1 = \lambda \alpha_1' \alpha_1 = \lambda$$

di conseguenza λ non può che essere il massimo autovalore della matrice Σ , che indicheremo con λ_1 , e α_1 è l'autovettore di norma unitaria ad esso corrispondente.

Resta in tal modo definita la variabile

$$\xi_1 = \alpha_1' X$$

che denomineremo prima componente principale.

In maniera analoga si possono calcolare le successive componenti principali. Ad esempio la seconda componente principale ξ_2 è ancora definita come combinazione lineare delle variabili originarie

$$\xi_2 = \alpha_{21} X_1 + \alpha_{22} X_2 + \dots + \alpha_{2k} X_k = \alpha_2' X$$

ed è caratterizzata dalle seguenti proprietà:

1. è incorrelata con ξ_1 ;
2. ha varianza massima tra tutte le variabili incorrelate con ξ_1 .

La seconda componente principale si calcola risolvendo il problema di massimo vincolato:

$$\begin{cases} \max & \alpha_2' \Sigma \alpha_2 \\ & \alpha_2' \alpha_2 = 1 \\ & \alpha_2' \Sigma \alpha_1 = 0 \end{cases}$$

dove il secondo vincolo deriva dall'annullarsi della covarianza tra ξ_1 e ξ_2 . Come in precedenza facendo ricorso al metodo dei moltiplicatori di Lagrange si ricava l'equazione agli autovalori

$$\Sigma \alpha_2 = \lambda \alpha_2$$

quindi analogamente a quanto detto precedentemente λ non può che essere il secondo più grande autovalore della matrice Σ , mentre α_2 è il corrispondente autovettore di norma unitaria

.

In modo analogo si possono definire le successive componenti principali, un'osservazione importante è che se le variabili $X_1, X_2, \dots, X_k$ sono incorrelate (la matrice delle varianze e covarianze è diagonale), esse sono esattamente uguali alle componenti principali.

Se, indubbiamente, nell'interpretazione delle componenti principali giocano un ruolo preponderante l'esperienza e la sensibilità del ricercatore è d'altra parte evidente che il generico coefficiente α_{jh} della combinazione lineare corrispondente alla variabile X_h misura il peso assunto da tale variabile nella determinazione della componente principale ξ_j .

Ciò significa che la componente principale ξ_j sarà maggiormente caratterizzata dalle variabili X_h a cui corrispondono i coefficienti α_{jh} più grandi in valore assoluto, e in tal modo sono proprio tali variabili a conferire un significato alla componente principale.

Come già accennato l'analisi in componenti principali viene spesso utilizzata per lo studio delle variabili "latenti", ossia di quelle variabili non direttamente osservabili.

In tale senso le variabili artificiali a cui dà luogo l'analisi rappresentano un tentativo di osservare, di misurare delle variabili nascoste non esplorabili direttamente.

In ogni caso, qualunque sia la ragione per cui si ricorre all'analisi in componenti principali, uno dei problemi da affrontare riguarda il numero di componenti principali da considerare.

A tale riguardo il criterio più utilizzato è quello di :

1. fissare una soglia $I^* = \text{"frazione di varianza totale che si desidera sia spiegata"};$
2. scegliere il più piccolo numero p di componenti principali tali che $I_p \geq I^*$.

E' evidente che la frazione della varianza totale spiegata dalle prime p componenti principali, può essere considerata come una misura del loro "potere riassuntivo": un valore elevato di I_p indica che le prime p componenti principali ricostruiscono un'alta frazione della varianza totale.