

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE
CORSO DI LAUREA TRIENNALE IN
STATISTICA PER L'ECONOMIA E L'IMPRESA



RELAZIONE FINALE

**Costruzione di un indicatore di turismo sostenibile
utilizzando il COIN Tool della commissione Europea**

Relatrice Prof.ssa Francesca Bassi
Dipartimento di Scienze Statistiche

Laureando *Thomas Bozzolan*
Matricola 2005389

Anno Accademico 2023/2024

Ai miei genitori e ai loro sacrifici.

Indice

Introduzione	1
1 Turismo sostenibile e COIN Tool	3
1.1 Definizione e concetto di turismo sostenibile	3
1.1.1 Principi della sostenibilità ambientale nel turismo	4
1.1.2 Principi della sostenibilità economica nel turismo	5
1.1.3 Principi della sostenibilità socio-culturale nel turismo	6
1.2 Ruolo degli indicatori nella misura della sostenibilità	7
1.3 COIN Tool della commissione Europea	8
1.3.1 Vantaggi e svantaggi degli indicatori compositi	10
2 Presentazione dei dati	11
2.1 Il sondaggio, gli errori campionari e non campionari	12
2.2 L'Eurobarometro	15
2.2.1 Flash Eurobarometer 499	16
2.3 Analisi esplorativa	17
3 Costruzione dell'indicatore composito e applicazione del COIN Tool	21
3.1 Costruzione dell'indicatore composito	21
3.1.1 Passaggi 1-3: definizione e selezione	23
3.1.2 Passaggi 4-6: preparazione e costruzione	25
3.1.3 Passaggi 7-10: verifica e presentazione	29
3.2 Applicazione del COIN Tool	31
Conclusioni	38
Bibliografia	41

Introduzione

Negli ultimi decenni, il turismo ha conosciuto una crescita senza precedenti, diventando uno dei principali motori monetari a livello globale e una risorsa fondamentale per molte regioni del mondo, in grado di generare significativi benefici economici, sociali e culturali.

Tuttavia, l'espansione del settore ha comportato anche una serie di preoccupazioni legate agli impatti negativi, in particolare sull'ambiente, sulle comunità locali e sui biosistemi, sollevando la necessità di sviluppare strategie di gestione che riducano tali effetti. La pressione esercitata dalle attività turistiche su ecosistemi fragili, la crescita incontrollata dell'urbanizzazione nelle aree ad alta densità turistica e l'aumento delle emissioni di gas serra sono solo alcune delle criticità che evidenziano l'urgenza di ripensare le modalità di sviluppo turistico in ottica sostenibile.

Il turismo sostenibile si è affermato come un modello alternativo, volto a bilanciare le esigenze economiche, ambientali e sociali dei territori turistici. Il suo compito è quello di preservare le risorse naturali e culturali per le generazioni future, promuovendo al contempo un'esperienza turistica responsabile e consapevole. Tuttavia, la sua implementazione richiede strumenti adeguati alla valutazione e alla gestione delle diverse dimensioni della sostenibilità.

In questo contesto, il *COIN (COMposite INDicator) Tool* emerge come uno strumento di particolare interesse. Si tratta di una metodologia avanzata per la costruzione di indicatori compositi che, attraverso l'aggregazione di variabili quantitative e qualitative, facilita la comprensione e il monitoraggio delle prestazioni di fenomeni complessi come la sostenibilità. Nel settore del turismo, il *COIN Tool* può essere utilizzato per valutare le *performance* ambientali, economiche e sociali delle destinazioni turistiche, offrendo una base solida per la definizione di politiche e strategie mirate a migliorare la sostenibilità del turismo.

Lo scopo dell'elaborato è quello di analizzare il ruolo del *COIN Tool* nel contesto del turismo sostenibile, esplorandone il potenziale come strumento di supporto decisionale

per la gestione sostenibile delle destinazioni turistiche e mettendolo in paragone ad altre tipologie di indicatori.

In particolare, il lavoro è così strutturato:

- nel Capitolo 1 verrà definito il concetto di sostenibilità, diviso anche nei tre ambiti, e di turismo sostenibile. Successivamente ci si soffermerà sul ruolo dei diversi indicatori, prima quelli tradizionali poi quelli di turismo sostenibile, nella valutazione della sostenibilità, facendo particolare attenzione al *COIN Tool* ed ai suoi numerosi vantaggi;
- il Capitolo 2 sarà dedicato alla presentazione dei dati. Si parlerà, dunque, del questionario somministrato alla popolazione statistica in questione e verrà effettuata una breve analisi esplorativa, utile per notare qualche dato rilevante o per fare delle previsioni sul risultato dell'indagine;
- il Capitolo 3, nonché l'ultimo, avrà come soggetto la costruzione vera e propria dell'indice composito, evidenziando ogni passaggio chiave e fornendo il corrispondente quadro teorico. Comprenderà, inoltre, l'elenco dei passaggi effettuati dal *COIN Tool*, riferendosi poi all'esempio analizzato.

Capitolo 1

Turismo sostenibile e COIN Tool

Prima di parlare di turismo sostenibile, bisogna dare il significato della parola “sostenibilità”. Secondo l’enciclopedia Treccani, la sostenibilità è “la condizione di uno sviluppo in grado di assicurare il soddisfacimento dei bisogni della generazione presente senza compromettere la possibilità delle generazioni future di realizzare i propri”, che, in altre parole, implica un benessere costante, preferibilmente crescente, e la prospettiva di lasciare alle generazioni future una qualità della vita non inferiore a quella attuale.

Di questo tema se ne occupa principalmente l’ONU (Organizzazione delle Nazioni Unite) che, nel settembre del 2015, ha avviato un piano d’azione globale diretto ai 193 Stati membri, chiamato “Agenda 2030”. Questo programma mira a promuovere lo sviluppo sostenibile in tutto il mondo e include diciassette Obiettivi di Sviluppo Sostenibile (SDGs), che rappresentano dei traguardi comuni su insieme di questioni importanti per il progresso come la povertà, le disuguaglianze, il cambiamento climatico, la salute, l’istruzione e la sostenibilità ambientale. Gli SDGs sono interconnessi e mirano a garantire che nessuno venga lasciato indietro, infatti vengono chiamati “obiettivi comuni” proprio perché riguardano tutti i paesi e tutti gli individui, il cui fine è quello di portare il mondo sulla strada della sostenibilità. Inoltre, l’Agenda 2030 identifica il turismo come incentivo primario per la crescita economica rivolta alle comunità.

1.1 Definizione e concetto di turismo sostenibile

Il turismo sostenibile è un tipo di turismo che si sviluppa nel rispetto dell’ambiente, delle culture locali e delle economie delle comunità ospitanti. Esso mira a ridurre l’impatto negativo delle attività turistiche, promuovendo pratiche che preservano le risorse naturali e culturali, garantendo al contempo benefici economici e sociali per le popolazioni locali.

In questo modo, il turismo sostenibile cerca di offrire esperienze autentiche ai viaggiatori, mentre contribuisce al benessere delle comunità e alla protezione dell'ambiente.

Come appena accennato, si possono riconoscere i tre pilastri fondamentali su cui si basa il turismo sostenibile: sostenibilità ambientale, sostenibilità economica e sostenibilità socio-culturale.

1.1.1 Principi della sostenibilità ambientale nel turismo

La sostenibilità ambientale e il turismo sostenibile sono strettamente connessi, poiché quest'ultimo si propone di ridurre l'impatto negativo delle attività turistiche sull'ambiente e promuovere pratiche che preservino le risorse naturali.

In particolare, si basa su questi principi di fondo:

- la minimizzazione dell'impatto ecologico, promuovendo pratiche che riducono l'inquinamento e il degrado ambientale. Un esempio può essere l'uso di trasporti ecologici, come biciclette e mezzi pubblici, che contribuisce a ridurre le emissioni di carbonio;
- la conservazione delle risorse naturali, con la gestione in modo sostenibile delle risorse vitali come l'acqua, l'energia e il suolo. Le strutture ricettive possono adottare tecnologie di risparmio energetico, sistemi di raccolta dell'acqua piovana e pratiche di riciclo, per citare qualche metodo;
- la protezione della biodiversità, favorendo la salvaguardia degli ecosistemi e delle specie locali, creando quindi aree protette e implementando pratiche turistiche che non disturbano gli habitat naturali;
- l'educazione ambientale, sensibilizzando i turisti sull'importanza della conservazione ambientale e delle pratiche sostenibili. Attraverso esperienze come tour guidati, laboratori e attività interattive, i turisti possono imparare a rispettare l'ambiente e a trattarlo nella maniera più corretta possibile;
- lo sviluppo di infrastrutture sostenibili, costruendo strutture turistiche che utilizzano materiali ecologici e rispettano il contesto ambientale. Le strutture in questo modo dovrebbero integrarsi armoniosamente nel paesaggio naturale;
- la promozione del consumo responsabile, incoraggiando i turisti a scegliere prodotti locali e sostenibili, riducendo così l'impatto del trasporto di beni e sostenendo le economie locali.

I vantaggi derivanti da una buona sostenibilità ambientale sono la conservazione dell'ambiente, custodendolo in modo sano per le generazioni future, e il mantenimento dell'attrattiva per i turisti, che sono alla ricerca di esperienze sempre più autentiche.

1.1.2 Principi della sostenibilità economica nel turismo

Il secondo ambito del turismo sostenibile, quello economico, si concentra sulla creazione di un sistema turistico che porti benefici economici alle comunità locali, garantendo al contempo la sostenibilità a lungo termine.

Gli aspetti chiave di questa tipologia di sostenibilità sono i seguenti:

- i benefici per le comunità locali. Risulta fondamentale che una parte significativa delle entrate generate dal turismo rimanga nelle comunità che ospitano i turisti. Questo può essere realizzato attraverso la promozione di aziende locali, come ristoranti, negozi e strutture ricettive. Inoltre, il turismo sostenibile contribuisce alla creazione di posti di lavoro in vari settori ed è altrettanto importante che questi siano dignitosi, ben retribuiti e offrano opportunità di formazione e sviluppo professionale;
- il sostegno alle piccole imprese, con l'incoraggiamento dei visitatori a spendere in attività e negozi locali. Ciò include la promozione di mercati locali, prodotti artigianali e servizi gestiti da residenti. Sono essenziali anche le iniziative imprenditoriali nelle comunità locali, grazie alle varie opportunità per avviare nuove attività legate al turismo, come guide turistiche, corsi di cucina, o esperienze culturali;
- lo sviluppo economico sostenibile, ad esempio investendo gli introiti in infrastrutture locali, come strade, trasporti pubblici e servizi, migliorando, quindi, la qualità della vita per i residenti e l'esperienza dei visitatori o diversificare le economie delle comunità, riducendo la dipendenza da settori economici tradizionali e vulnerabili;
- le pratiche commerciali etiche, tramite un vero e proprio turismo responsabile, incentivando le imprese turistiche a operare in modo etico, rispettando i diritti dei lavoratori e le normative locali, si contribuisce a creare un ambiente di lavoro uguale e giusto per tutti. Questo è possibile soprattutto con la collaborazione tra governi, organizzazioni non governative, operatori turistici e comunità locali, così da sviluppare strategie economiche sostenibili e promuovere il turismo responsabile;

- la misurazione e il monitoraggio, tramite l'utilizzo di indicatori economici per controllare gli impatti del turismo nelle comunità locali, come l'occupazione, i redditi e la spesa dei turisti e per pianificare strategie future.

Un turismo ben gestito può contribuire a stabilizzare l'economia locale, riducendo la vulnerabilità a crisi economiche globali, e a svilupparla, creando opportunità economiche e migliorando la qualità di vita.

1.1.3 Principi della sostenibilità socio-culturale nel turismo

L'ultimo, ma non per importanza, punto cardine del turismo sostenibile è quello socio-culturale. Questo aspetto mira a garantire che il turismo rispetti e valorizzi le culture locali, promuovendo al contempo la coesione sociale.

Questi sono i punti più rilevanti a riguardo:

- il rispetto e la valorizzazione delle culture locali, poiché il turismo sostenibile incoraggia la conservazione dei patrimoni culturali, storici e artistici. Ciò può includere la salvaguardia di tradizioni, arti e mestieri locali, assicurando che le comunità possano tramandare le loro culture alle future generazioni. La promozione di esperienze autentiche incoraggia i turisti a vivere esperienze genuine, interagendo con le comunità locali, partecipando a eventi culturali e apprendendo le tradizioni locali. Questo non solo arricchisce l'esperienza del visitatore, ma rafforza anche il senso di identità culturale nelle comunità ospitanti;
- l'integrazione sociale tramite la partecipazione attiva delle comunità nella pianificazione e nella gestione del turismo. Questo coinvolgimento assicura che le loro esigenze, aspettative e preoccupazioni siano considerate, promuovendo un senso di appartenenza e responsabilità. Gioca un ruolo fondamentale anche la coesione sociale, grazie alla quale si può creare un ponte tra diverse comunità e culture. Inoltre, eventi e attività turistiche possono contribuire a costruire legami tra residenti e visitatori;
- lo sviluppo di capacità e la formazione. Il turismo sostenibile offre opportunità di formazione per le comunità locali, migliorando le loro competenze in vari ambiti, come l'ospitalità, la gestione delle risorse e il marketing. Ciò consente alle comunità di gestire meglio le attività turistiche e di trarne benefici. Non è da meno, anche in questo ambito, l'educazione alla sostenibilità, sensibilizzando i turisti e le comunità sulle pratiche sostenibili e sull'importanza della conservazione culturale. Ciò può includere *workshop*, programmi scolastici e attività di sensibilizzazione.

Con l'adempimento dei precedenti aspetti, verrà rafforzata l'identità culturale creando un senso di orgoglio e appartenenza tra i residenti. Inoltre, si sarà certi di rispettare non solo l'ambiente ospitante, ma anche le persone che ne fanno parte.

1.2 Ruolo degli indicatori nella misura della sostenibilità

Nell'ambito del turismo, l'uso degli indicatori tradizionali di sostenibilità è stato un tentativo di monitorare e valutare l'impatto delle attività turistiche. Questi strumenti, basati su metriche semplici, hanno fornito un'illusione di controllo e progresso, poiché non riescono a catturare la complessità delle interazioni tra turismo, comunità e ambiente.

Ad esempio, il numero di visitatori è spesso utilizzato per misurare il successo di una destinazione. Questo indicatore può sembrare vantaggioso, poiché un alto afflusso di turisti può tradursi in maggiori entrate per l'economia locale. Tuttavia, questo approccio ignora gli svantaggi, come il sovraffollamento, che può danneggiare le risorse naturali e ridurre la qualità dell'esperienza per i visitatori.

Un altro indicatore comune è l'occupazione alberghiera, che offre un'idea dell'attrattività di una destinazione. Sebbene un alto tasso di occupazione possa indicare una buona salute economica, non riflette l'impatto ambientale delle strutture ricettive, né la qualità dell'occupazione per i lavoratori del settore, né il vero grado di soddisfazione dei clienti. Gli alberghi possono essere pieni, ma questo non garantisce pratiche di gestione sostenibili o il benessere dei dipendenti.

La spesa media per turista rappresenta un ulteriore dato poco informativo. Monitorare quanto ogni visitatore spende può sembrare utile per stimare il profitto economico, ma una spesa elevata non implica necessariamente benefici per la comunità locale, soprattutto se le entrate vengono incanalate verso grandi catene internazionali piuttosto che verso le piccole imprese locali, e potrebbe incentivare forme di turismo di massa.

Un ultimo esempio sono gli indici di crescita annuali, che possono suggerire un aumento delle attività turistiche e, quindi, un segnale positivo. Invece no, perché una crescita continua, senza considerare gli effetti collaterali, può portare a un deterioramento delle risorse ambientali e culturali, minacciando a lungo termine la stessa attrattiva della destinazione.

Quanto appena descritto dimostra che gli indicatori tradizionali possono fornire dati utili, ma spesso mancano di una visione a 360° necessaria per una gestione veramente

sostenibile del turismo. È cruciale, quindi, spostare l'attenzione verso indicatori che considerino non solo la quantità, ma anche la qualità e l'impatto delle attività turistiche.

Proprio per questo motivo sono stati creati gli indicatori di turismo sostenibile, il cui compito è quello di fornire una panoramica più completa integrando i tre aspetti descritti nel precedente paragrafo. Devono, dunque, massimizzare i benefici sociali ed economici per la comunità locali, i benefici al patrimonio culturale e i benefici per l'ambiente, minimizzando i relativi impatti negativi .

1.3 COIN Tool della commissione Europea

La misurazione dei fenomeni multidimensionali ha suscitato un rinnovato interesse nella comunità scientifica negli ultimi anni. È ormai consapevolezza comune che numerosi fenomeni socio-economici non possano essere misurati attraverso un unico indicatore descrittivo, ma debbano invece essere rappresentati da una molteplicità di aspetti o dimensioni. Il *Composite Indicator Tool* (o più semplicemente *COIN Tool*) della Commissione Europea è uno strumento utilizzato per la costruzione, l'analisi e la modifica di indicatori compositi, che sintetizzano diversi dati in un unico valore per valutare eventi di qualsiasi tipo. Questo può contribuire a creare indicatori compositi di alta qualità che siano metodologicamente solidi e legittimi per gli utenti finali. Serve, allora, dare una spiegazione di questa tipologia di indicatori e del loro utilizzo.

In termini generali, un indicatore è una misura quantitativa o qualitativa derivata da una serie di fatti osservati che possono rivelare posizioni relative (ad esempio di un paese) in un'area specifica. Quando è valutato a intervalli regolari, un indicatore può evidenziare la direzione del cambiamento attraverso diverse unità e nel tempo.

Un indicatore composito è la combinazione matematica (o "aggregazione") di un insieme di indici che rappresentano le diverse dimensioni di un fenomeno da misurare (Saisana e Tarantola, 2002; Salzman, 2003; OECD, 2008). Si forma quando questi indicatori individuali vengono compilati in un unico indice sulla base di un modello sottostante. Esso dovrebbe idealmente misurare concetti complessi e multidimensionali che non possono essere catturati da un singolo indicatore come la competitività, l'innovazione, l'industrializzazione, lo sviluppo umano, le prestazioni ambientali, l'integrazione del mercato unico, e molto altro ancora. Per essere misurati, serve una "combinazione" di diverse dimensioni, considerate insieme come componenti del fenomeno (Mazziotta e Pareto, 2013). Inoltre, di solito, si utilizza anche una struttura gerarchica che suddivide il concetto in elementi noti come sotto-pilastri, pilastri e sotto-indici. Il risultato finale

è chiamato “indice” ed è utilizzato sia per creare una classifica che per riassumere i dati (Bandura, 2008).

Gli aspetti vengono misurati attraverso un numero di unità, che spesso sono paesi o regioni, ma possono essere anche altre entità, come università o aziende. I risultati degli indicatori compositi vengono solitamente utilizzati per confrontare o classificare le unità tra loro. Essi sono strumenti pratici potenti che possono aiutare i responsabili delle politiche a riassumere fenomeni complessi e interdipendenti, forniscono una visione d’insieme, sono facili da interpretare, semplici da comunicare e attraenti per il pubblico. Spesso, infatti, sembra più facile interpretare gli indicatori compositi piuttosto che identificare tendenze comuni tra molti indicatori separati (come quelli tradizionali visti nel paragrafo precedente).

Sono, inoltre, motori di comportamento e cambiamento, costringendo istituzioni e governi a mettere in discussione i propri standard. D’altra parte, è necessaria cautela per evitare situazioni in cui gli indicatori compositi possano inviare messaggi politici fuorvianti o parziali e i loro risultati possono invitare gli utenti a trarre conclusioni analitiche semplicistiche a causa di una scarsa costruzione o interpretazione. Dunque, gli indicatori compositi devono essere visti come un mezzo per avviare discussioni o stimolare l’interesse pubblico e la loro rilevanza dovrebbe essere valutata rispetto ai gruppi interessati dall’indice composito.

Si sottolinea, quindi, che gli sviluppatori di indicatori compositi provengono da una vasta gamma di ambiti, spaziando da scienziati, statistici e analisti dei dati a giuristi e responsabili politici. La costruzione appropriata di un indicatore composito è, infatti, un processo delicato e soggetto a critiche, data la sua natura soggettiva. È pertanto fondamentale garantire la legittimità e l’accuratezza dell’indicatore composito, attraverso un processo di costruzione quanto più meticoloso possibile, che includa un’analisi statistica rigorosa. Essi sono simili a modelli matematici o computazionali e, in quanto tali, la loro costruzione deve più all’abilità dell’analista che a regole scientifiche universalmente accettate per la codifica.

Proprio per questo nasce il *COIN Tool*. Esso guida l’utente attraverso una serie di passaggi semplici e intuitivi, mirati a costruire e visualizzare il proprio indicatore composito. Inoltre, consente di analizzare le relazioni tra gli indicatori, testare variazioni metodologiche e verificare la robustezza rispetto a determinate assunzioni. Questa capacità di esplorare e adattare gli indicatori in base ai risultati ottenuti rappresenta un passo fondamentale per sviluppare indicatori di alta qualità, metodologicamente solidi e utilizzabili da parte degli utenti finali. È importante riconoscere che le metodologie e

le pratiche in questo campo sono in continua evoluzione. Pertanto, è essenziale rimanere aggiornati sulle nuove scoperte e approcci, affinché gli indicatori compositi possano continuare a rispondere efficacemente alle sfide attuali.

1.3.1 Vantaggi e svantaggi degli indicatori compositi

La facilità d'interpretazione non è l'unico vantaggio di cui gode questa tipologia di indicatori. Essi possono, infatti, ridurre la dimensione visibile di un insieme di indicatori senza perdere la base informativa sottostante, così da rendere possibile l'inclusione di maggiori informazioni nel limite di dimensioni esistente. Un indicatore ben fatto permette di facilitare la comunicazione con il pubblico (cittadini, media, ...), esperto o meno che sia, e consente agli utilizzatori di confrontare efficacemente più dimensioni complesse.

Un dato è tanto semplice da interpretare quanto da manipolare. Gli indicatori compositi, infatti, soprattutto se usati nell'ambito politico, possono essere oggetto di discussioni e controversie non indifferenti. Questo è dovuto, in primo luogo, a chi utilizza questa metodologia in modo inappropriato, con un processo di costruzione non trasparente o mancante di soliti principi statistiche. In secondo luogo, o chi ignora le dimensioni della *performance* o non spiega chiaramente la selezione sia degli indicatori stessi sia dei pesi usati.

Capitolo 2

Presentazione dei dati

Nella letteratura di riferimento, così come nelle statistiche ufficiali, è possibile trovare numerose proposte di indicatori composti per misurare la sostenibilità e l'economia circolare anche nel settore turistico, tuttavia, tutte si basano su dati raccolti a livello territoriale. In questo elaborato si utilizzeranno dati raccolti a livello individuale, su un campione rappresentativo di cittadini europei. Il dataset fornito deriva da un'indagine intitolata *Flash Eurobarometer 499 - Attitudes of Europeans towards Tourism, Autumn 2021* e non è altro che un questionario somministrato ad un campione rappresentante di oltre venticinquemila cittadini, di età pari o superiore a quindici anni, residenti nei ventisette Stati membri dell'Unione Europea. L'indagine è stata effettuata tra il 18 e il 28 ottobre 2021 tramite interviste telefoniche, sia su linea fissa che su telefoni cellulari.

Si nota subito che il periodo a cui ci si sta riferendo non è uno qualsiasi. L'anno precedente, il 2020, è stato segnato dalla diffusione della pandemia di COVID-19 (o Coronavirus), una malattia infettiva respiratoria che è stata contratta inizialmente da alcuni lavoratori del mercato umido di Wuhan, in Cina, per poi diffondersi a livello globale, causando gravi problemi non solo alle persone, ma anche alle economie dei vari stati. All'ondata di contagi molti paesi hanno cercato di porre rimedio con il *lockdown* nazionale, salvaguardando la salute delle persone fino alla creazione di una cura definitiva. A poco più di due mesi dall'identificazione della sequenza virale, hanno avuto inizio le prime sperimentazioni dei vaccini per COVID-19 e si è assistito, a giugno dello stesso anno, ad un calo dei decessi. Dopo un'estate dove in Europa i casi si erano ridotti al minimo, ha avuto inizio una risalita consistente del numero di contagi, dando origine alla seconda ondata. A metà dicembre viene finalmente approvato il primo vaccino contro il Coronavirus e iniziano le prime iniezioni. Il 2021, quindi, è stato il secondo anno di pandemia, contraddistinto dal successo dei vaccini, ma anche dallo sviluppo di varianti sempre più contagiose. Nonostante ciò, hanno iniziato a calare

considerevolmente i ricoveri e i decessi, confermando l'efficacia delle cure. Si capisce, dunque, che il questionario è volto soprattutto a osservare le diverse reazioni dei cittadini europei nel loro comportamento di viaggio e nel loro pensiero di turismo sostenibile dopo questo brutto periodo di pandemia.

I dati risultanti dalla parte interessata di questo elaborato sono risposte “Sì” o “No” (rispettivamente indicate con 0 e 1) a vari aspetti legati ad una specifica domanda, ossia “Sei pronto a cambiare le tue abitudini di viaggio e turismo per essere più sostenibile? Se sì, in quali dei seguenti modi?”. I modi selezionati sono i seguenti:

- Pagare di più per proteggere l'ambiente naturale
- Pagare di più a beneficio della comunità locale
- Prendere ferie al di fuori dell'alta stagione turistica
- Viaggiare verso destinazioni meno visitate
- Contribuire a attività di compensazione delle emissioni di carbonio (ad esempio, piantare alberi)
- Consumare prodotti locali in vacanza
- Scegliere opzioni di trasporto in base all'impatto ecologico
- Ridurre i rifiuti durante la vacanza
- Ridurre il consumo d'acqua in vacanza
- Non sono disposto a cambiare le mie abitudini
- Qualcos'altro
- Non lo so

2.1 Il sondaggio, gli errori campionari e non campionari

Prima della nascita dei sondaggi non c'era un mezzo diretto per l'ascolto dell'opinione pubblica, quindi politici, opinionisti e giornalisti non potevano essere sufficientemente informati. Nel 1927 lo statistico statunitense George Gallup inventò il cosiddetto “sondaggio scientifico” e, grazie a questo, nel 1936 riuscì a predire correttamente il risultato

delle elezioni presidenziali intervistando solamente cinquantamila persone, mentre l'allora molto quotato *Literary Digest* sbagliò clamorosamente la previsione pur avendo intervistato molte più persone (circa due milioni e mezzo). Questo per far capire quanto un sondaggio ben fatto possa risultare molto affidabile e comodo nella gestione di un ampio campione di rispondenti, consentendo di ottenere risultati che riflettono con precisione le opinioni della popolazione. Ma cos'è in sostanza un sondaggio?

Il sondaggio è un metodo statistico volto a valutare le proporzioni di diverse caratteristiche di una popolazione a partire dallo studio di una parte di essa, chiamata campione. Questo particolare strumento di ricerca può essere realizzato in diversi modi, ad esempio tramite questionari online, interviste faccia a faccia, sondaggi telefonici o postali, e così via.

In questo caso, la metodologia adottata è stata quella dell'intervista telefonica. Qui gioca un ruolo fondamentale l'intervistatore, che inevitabilmente condiziona l'esito della rilevazione interagendo con il rispondente con il suo modo di condurre il sondaggio. Gli intervistatori, quindi, devono essere dei professionisti, appositamente formati sia sui possibili meccanismi che si creano con gli intervistati sia sui temi specifici dell'indagine alla quale sono chiamati a collaborare. Essi devono essere in grado di dare chiarimenti agli intervistati e di indurli a rispondere nel modo più corretto e veritiero possibile (non affrettando la risposta per esempio). Tecnicamente parlando, l'intervista si segue nel seguente modo: gli intervistati sono raggiunti per telefono, l'intervistatore registra direttamente al computer le risposte e il *software* stesso agevererà il percorso controllando possibili risposte incongruenti. Una minoranza di studi, inoltre, specifica che si ha un maggiore tasso di risposta fra le intervistatrici giovani o consiglia il *matching* tra il rispondente e l'intervistatore (uomo-uomo, donna-donna, ...).

L'intervista telefonica poi gode di alcuni vantaggi, come la rapidità con la quale si riescono a raccogliere i dati, i costi ridotti (soprattutto rispetto all'intervista faccia a faccia), la semplicità organizzativa grazie alle interviste centralizzate, e il minor rischio di errore da parte del rilevatore.

Guardando l'altra parte della medaglia, però, possiamo notare anche alcuni difetti. Innanzitutto, la popolazione oggetto d'indagine spesso non coincide con la popolazione di cui si ha la lista, fornita dall'elenco telefonico o da altre liste. Questo può essere dovuto da molteplici fattori, ad esempio la mancata copertura, poiché non tutti hanno il telefono o parecchi non consentono la pubblicazione del loro numero, la presenza di doppioni nella lista (numero privato e numero d'ufficio di un libero professionista) o di unità non appartenenti alla popolazione da indagare (i negozi). Inoltre, le probabilità

di accettazione all'intervista non sono equi-distribuite perché alcune categorie di popolazione fanno più fatica a rispondere al telefono, come gli anziani o i malati. In caso le caratteristiche dei non rispondenti fossero correlate all'oggetto d'indagine, si otterrebbero risultati distorti. Infine, viene a meno il coinvolgimento del rispondente e ciò potrebbe portare a risposte più sbrigative e incorrette.

Andando nello specifico, ci si può riferire anche alle due tipologie di errori esistenti, ossia l'errore non campionario, legato all'accuratezza dell'indagine, e l'errore campionario, legato, invece, alla precisione della stessa.

Il primo è difficile da controllare perché si può manifestare in vari momenti del processo di campionamento. Può accadere, infatti, nella fase di progettazione della ricerca come errore di specificazione, relativamente alla popolazione oggetto d'indagine o alla lista delle unità di popolazione, oppure nella fase di rilevazione dei dati come errore di misura, dovuto agli strumenti di indagine, ad esempio il questionario, che può non essere correttamente formulato, o al rispondente che può distorcere deliberatamente un'informazione, o nella fase finale di elaborazione dei risultati, dove si possono commettere dei semplici errori di calcolo o può esserci un errore di memorizzazione informatica.

Per ovviare a questo problema occorre optare per un metodo di campionamento probabilistico, rispetto che ad uno non probabilistico. Quest'ultimo è definito come una selezione in maniera soggettiva del campione dal ricercatore in base alla sua idea di rappresentatività, mentre l'altro consiste in una selezione dove ogni unità della popolazione ha una probabilità positiva di essere inclusa nel campione (le unità vengono, quindi, scelte a caso). Dunque, i risultati ottenuti con un campione probabilistico possono essere estesi all'intera popolazione e, inoltre, l'errore di campionamento può essere facilmente calcolato. Non viene sempre utilizzato però, poiché spesso non si dispone della lista delle unità da campionare o non sempre tutti i soggetti della popolazione sono ugualmente reperibili, per non parlare del costo che si andrebbe inevitabilmente ad alzare.

Un altro controllo da effettuare, quindi, riguarda il tasso di non risposta, ossia la percentuale delle persone selezionate per l'indagine che non sono state effettivamente intervistate perché non reperite o hanno rifiutato di rispondere. Bisogna, dunque, selezionare adeguatamente il campione per tenere questo tasso al minimo, per ridurre il rischio di autoselezione, che porterebbe ad un'errata rappresentazione della realtà perché accettano di rispondere più facilmente persone con caratteristiche particolari, in generale diverse da coloro che non rispondono.

Risulta importante, infine, testare le domande del questionario. Esse devono essere poste in modo chiaro e non ambiguo, cosicché tutti siano in grado di rispondere alla

domanda e non possano aggiungere parole per specificare una domanda incompleta. Le risposte, inoltre, devono essere tutte previste in modo esaustivo e mutuamente esclusivo.

L'errore campionario, invece, si riferisce alla variazione nei risultati che può derivare dalla scelta di un campione piuttosto che di un'intera popolazione. Anche se il campione è selezionato in modo randomizzato, ci saranno sempre delle fluttuazioni naturali nei dati raccolti. Questo tipo di errore è influenzato da molteplici fattori: la dimensione del campione, infatti i campioni più piccoli tendono a presentare maggiori variazioni rispetto ai campioni più grandi (maggiore è il campione, minore è l'errore campionario); il metodo di campionamento, poiché, se il campione non è selezionato in modo adeguato e risulta quindi non casuale o non rappresentativo, l'errore campionario può aumentare; la variabilità della popolazione che, se fosse molto eterogenea, aumenterebbe la probabilità che ci siano errori campionari.

2.2 L'Eurometro

Iniziato nel 1974 all'interno della Commissione Europea, l'Eurometro è stato concepito inizialmente come un mezzo per "rivelare gli europei a se stessi". Da allora, si è evoluto e ampliato significativamente con diversi strumenti di sondaggio. Nel 2007, il Parlamento Europeo ha avviato la propria serie regolare di sondaggi Eurometro, concentrandosi su temi specifici del Parlamento Europeo, comprese le elezioni europee.

L'Eurometro è diventato, quindi, lo strumento ufficiale di sondaggio utilizzato dal Parlamento Europeo, dalla Commissione Europea e da altre istituzioni e agenzie dell'Unione Europea per realizzare sondaggi mirati a conoscere e comprendere gli atteggiamenti dei cittadini europei. Le inchieste e gli studi riguardano argomenti di primaria importanza per la cittadinanza europea, come l'allargamento dell'UE, la situazione sociale, la cultura, la salute, la cultura, la tecnologia, l'ambiente, l'Euro, la difesa e molto altro.

Il suo scopo è quello di monitorare regolarmente l'evoluzione dell'opinione pubblica, che offre una panoramica dei pensieri e delle opinioni dei cittadini europei, aiutando le istituzioni a comprendere le loro esigenze e aspettative. Questo è utile sia ai politici, che riescono ad utilizzare le informazioni ottenute per adattare le varie strategie, prendere decisioni e valutare il proprio lavoro, sia ai ricercatori e agli analisti, che ottengono un quadro importante di come cambiano le opinioni e i *trend* nel tempo. Inoltre, viene stimolato il dibattito pubblico, poiché si forniscono informazioni su temi rilevanti per l'UE, contribuendo così ad una maggiore consapevolezza civica.

La combinazione dell'ampia gamma di argomenti trattati in modo coerente nel tempo, della regolarità delle pubblicazioni e della copertura geografica rende l'Eurobarometro una fonte unica di conoscenza e informazione nell'Unione Europea.

2.2.1 Flash Eurobarometer 499

A nome della Commissione Europea, della Direzione Generale per il Mercato Interno, dell'Industria, dell'Imprenditorialità e delle PMI (Piccole e Medie Imprese), *Ipsos European Public Affairs* ha condotto un'indagine *Flash Eurobarometer* sull'atteggiamento verso il turismo sostenibile da parte dei cittadini europei. Gli argomenti esplorati includono:

- il comportamento di viaggio, comprese la frequenza dei viaggi, le preferenze di alloggio e l'impatto della pandemia di COVID-19 sul comportamento di viaggio;
- i canali di prenotazione preferiti e le fonti di informazione per organizzare i viaggi;
- le principali ragioni per la scelta di una destinazione;
- le opinioni e informazioni sul viaggio sostenibile, comprese la facilità di trovare informazioni relative alla sostenibilità o all'accessibilità e la disponibilità a modificare le abitudini di viaggio per essere più sostenibili.

I primi tre non verranno analizzati a fondo nel corso di questo elaborato, ma sorgono alcuni risultati interessanti da nominare.

Riguardo il primo tema, risulta che otto intervistati su dieci affermano di farlo "diverse volte all'anno" (43%) o "una o due volte all'anno" (37%). Un ulteriore 9% dice di viaggiare "una volta ogni pochi anni", quando prima della pandemia di COVID-19, la maggior parte dei cittadini dell'Unione Europea viaggiava prevalentemente per svago o lavoro. Inoltre, quando pianificano una vacanza, i tipi di sistemazione che i cittadini dell'UE considerano più probabili sono le sistemazioni commerciali (hotel o B&B) con pensione inclusa (42%), le sistemazioni private (casa vacanze o appartamento) prenotate direttamente dal proprietario (33%) e soggiornare da un amico o un familiare (30%). Poi, una grande maggioranza dei cittadini europei si aspetta che la pandemia di COVID-19 avrà qualche tipo di effetto a lungo termine sul loro comportamento di viaggio, mentre il 21% non si aspetta alcun impatto. Gli intervistati sono particolarmente propensi a pensare che la pandemia li porterà a prestare maggiore attenzione alle misure di salute e sicurezza (49%), a trascorrere più vacanze nel proprio paese (38%) e a viaggiare meno in generale (34%).

Parlando della seconda questione, quando organizzano le loro attività di viaggio e turismo, i cittadini dell'UE si affidano più spesso a piattaforme online per servizi di alloggio professionale come hotel e B&B (31%), seguiti da piattaforme online che combinano servizi di viaggio come alloggio, noleggio auto, voli, ... (25%). Tuttavia, i metodi "offline" per organizzare viaggi e attività turistiche rimangono importanti: il 24% degli intervistati è più propenso a organizzare il proprio viaggio tramite qualcuno che conosce, mentre il 19% preferisce effettuare una prenotazione "di persona" in un'agenzia di viaggi o in una compagnia di trasporti. Inoltre, quando gli intervistati pianificano un viaggio, la loro fonte di informazione preferita è "raccomandazioni da amici, colleghi o familiari" (56%), seguita, a una certa distanza, da "esperienza personale" (37%) e "siti web che raccolgono recensioni e valutazioni dei viaggiatori" (34%).

Infine, valutando il terzo argomento, i cittadini dell'UE tendono a basare la loro decisione su una destinazione di viaggio su tre aspetti principali, ossia le "offerte culturali nella destinazione" (44%), il "prezzo complessivo del viaggio" e il "contesto naturale nella destinazione" (entrambi 43%). Il quarto e quinto motivo più importante per gli intervistati per scegliere una destinazione sono le "attività disponibili nella destinazione" (30%) e l'"accessibilità dei servizi e delle attività per tutti" (24%).

2.3 Analisi esplorativa

Soffermandosi sull'oggetto vero e proprio di questa tesi, ossia il turismo sostenibile, si nota che quando pianificano un viaggio e cercano informazioni relative alla sostenibilità o all'accessibilità, i cittadini dell'UE trovano spesso difficile ottenere informazioni affidabili sulla *carbon footprint* delle opzioni di trasporto. La "*carbon footprint*" è una misura che quantifica l'impatto delle attività umane sul cambiamento climatico, esprimendo la quantità totale di gas serra emessi direttamente o indirettamente da un individuo, un'organizzazione, un evento o un prodotto, solitamente espressa in termini di tonnellate di CO₂ equivalente. Risulta, quindi, che il 48% ritiene molto o piuttosto difficile trovare informazioni affidabili su questo aspetto, rispetto al 33% che lo considera molto o piuttosto facile. Gli intervistati trovano anche spesso difficile ottenere informazioni sugli impegni di sostenibilità delle destinazioni (43%, contro il 41%, che trova questa informazione facile), sull'accessibilità della destinazione per persone con disabilità o mobilità ridotta (39%, contro il 39% che trova questa informazione facile) e sulle certificazioni di sostenibilità delle strutture ricettive (38%, contro il 46% che trova questa informazione facile).

In particolare, nella tabella sottostante (Figura 2.1), si possono osservare le varie posizioni dei paesi trattati in questa indagine riguardo le domande poste all'inizio del capitolo.

Analizzando i dati contenuti nella tabella, si può vedere che una grande maggioranza dei cittadini dell'UE (82%) è disposta a cambiare almeno alcune delle proprie abitudini di viaggio e turismo per essere più sostenibile, il 15% degli intervistati, al contrario, afferma di non essere pronto a farlo. Le azioni che i cittadini dell'UE sono più disposti a intraprendere includono il consumo di prodotti locali in vacanza (55%), la riduzione dei rifiuti durante il soggiorno (48%), viaggiare al di fuori dell'alta stagione turistica (42%) e visitare destinazioni meno affollate (41%). Una percentuale più ridotta è pronta a rendere i propri viaggi e attività turistiche più sostenibili pagando di più, sia per proteggere l'ambiente naturale (35%) che a beneficio della comunità locale (33%).

Si possono notare ulteriori informazioni particolari. Innanzitutto, lo stato che sembra voler cambiare maggiormente le proprie abitudini per essere più sostenibile risulta essere la Romania, seguita dalla Svezia. A proposito si osserva che ben l'89% dei rispondenti della Romania è disposto a ridurre i rifiuti durante le loro vacanze e l'86% consumerebbe prodotti locali durante il soggiorno, percentuali che sfiorano la totalità. Al contrario, i paesi che sono meno propensi a questo cambio di mentalità sembrano essere Italia, Portogallo, Bulgaria, Malta e Lituania poiché non presentano valori elevati in nessun ambito selezionato. Mentre, il 28% degli abitanti del Cipro non sono proprio propensi a cambiare le loro consuetudini durante i viaggi.

Per avere qualche indicazione su come il COVID-19 ha influenzato i pareri delle persone, si dovrebbero confrontare i dati della tabella precedente con quelli di un anno precedente al 2020.

Flash Eurobarometer 499
Attitudes of Europeans towards tourism

Q5 Are you prepared to change your travel and tourism habits to be more sustainable? If so, in which of the following ways?

		Pay more to protect the natural environment	Pay more to the benefit of the local community	Take holidays outside of the high tourist season	Travel to less visited destinations	Contribute to carbon-offsetting activities (e.g. plant trees)	Consume locally sourced products on holiday	Choose transport options based on ecological impact	Reduce waste while on holiday	Reduce water usage on holiday	I am not prepared to change my habits	Something else	Don't know/ not applicable
EU27		35	33	42	41	34	55	36	48	35	15	1	4
BE		21	23	28	27	15	28	19	41	28	21	3	3
BG		12	9	21	18	21	34	15	36	8	23	3	1
CZ		42	40	55	50	50	69	36	65	47	12	2	2
DK		26	21	30	24	15	33	38	25	19	25	5	8
DE		47	41	37	43	33	54	39	45	39	18	1	3
EE		22	28	45	31	17	51	20	39	14	14	2	6
IE		30	27	37	29	22	34	24	31	22	5	4	5
EL		52	47	66	63	50	77	50	70	52	12	0	0
ES		41	35	53	52	52	71	46	68	52	11	1	2
FR		29	39	39	38	27	52	30	45	31	18	2	7
HR		21	16	41	33	32	48	24	52	28	14	1	2
IT		14	11	27	23	13	42	23	26	19	18	1	3
CY		17	13	41	33	22	43	27	42	29	28	2	1
LV		23	19	40	34	26	51	23	45	14	20	2	1
LT		22	13	36	28	18	36	22	30	7	24	2	4
LU		51	50	51	42	37	60	45	56	42	5	1	3
HU		26	20	39	34	33	58	23	58	32	11	1	2
MT		33	26	26	19	10	15	22	27	24	1	10	3
NL		38	30	45	43	32	44	37	44	29	15	2	5
AT		39	35	38	36	28	47	31	38	32	6	2	5
PL		27	26	49	47	36	61	36	53	40	10	1	3
PT		16	14	29	19	13	32	21	20	16	6	2	14
RO		74	62	74	67	82	86	66	89	41	6	0	1
SI		35	30	55	47	31	63	27	52	36	21	1	1
SK		57	54	69	58	54	75	51	75	57	8	1	2
FI		31	37	55	36	30	64	36	51	34	11	1	1
SE		70	67	66	64	69	82	66	75	63	8	1	1

Flash Eurobarometer 499 - Attitudes of Europeans towards tourism
Fieldwork: 18/10 - 28/10/2021 / %, Base: all respondents (n=25714)

FIGURA 2.1: Percentuale di cittadini che ha selezionato la specifica modalità, divisi nei vari paesi (nella prima riga si ha la percentuale totale, ossia in base a tutti i cittadini dell'Unione Europea)

Capitolo 3

Costruzione dell'indicatore composito e applicazione del COIN Tool

3.1 Costruzione dell'indicatore composito

Si possono riconoscere dieci passaggi principali per la costruzione e l'analisi di un indicatore composito. È importante sottolineare che questi passaggi vengono normalmente seguiti in modo iterativo, ad esempio cambiando la selezione degli indicatori e persino il quadro concettuale, sulla base dei risultati dell'indice o dell'analisi statistica (il *COIN Tool* affronta approssimativamente i passaggi 3-8).

1. Quadro teorico: fornisce la base per la selezione e la combinazione delle variabili in un indicatore composito significativo, seguendo il principio di idoneità allo scopo. Questo è importante per ottenere una chiara comprensione e definizione del fenomeno multidimensionale da misurare, degli obiettivi e degli utenti finali, per strutturare i vari sotto-gruppi del fenomeno (se necessario) e per compilare un elenco di criteri di selezione per le variabili sottostanti;
2. Selezione dei dati, delle variabili e degli indicatori: basata su criteri come la rilevanza, il valore aggiunto, la disponibilità dei dati e considerazioni statistiche. Viene fatta per verificare la qualità degli indicatori disponibili, discutendone i punti di forza e di debolezza e per creare una tabella riassuntiva sulle caratteristiche dei dati (ad esempio disponibilità, fonte, tipo, ecc. . .);

3. Imputazione dei dati mancanti: necessaria per fornire un dataset completo. Si vuole stimare i valori mancanti, fornire una misura dell'affidabilità di ciascun valore imputato, valutare l'impatto dell'imputazione sui risultati dell'indicatore composito e discutere la presenza di *outlier* nel dataset;
4. Analisi multivariata: utilizzate per studiare la struttura complessiva del dataset, valutarne l'idoneità e guidare le scelte metodologiche successive. Si verifica, quindi, la struttura statistica del dataset e la si confronta con il quadro teorico per discutere delle possibili differenze;
5. Normalizzazione: porta gli indicatori su una scala comune, rendendo le variabili comparabili. Ciò implica una selezione di procedure di normalizzazione adeguate che rispettino sia il quadro teorico sia le proprietà dei dati, una discussione della presenza di *outlier* nel dataset ed eventuali aggiustamenti di scala o trasformazioni di indicatori fortemente simmetrici;
6. Pesatura e aggregazione: effettuate in base al quadro teorico sottostante. Consente di adattare l'effetto o l'importanza di ciascun indicatore in base al concetto misurato, considerando in particolare se valori elevati di un certo indicatore debbano compensare valori bassi di un altro, sempre in relazione agli obiettivi dell'indice e dell'indagine;
7. Analisi di incertezza e sensibilità: condotte per valutare la robustezza dell'indicatore composito. Bisogna identificare tutte le possibili fonti di incertezza nello sviluppo dell'indicatore composito, accompagnando ai punteggi e alle classifiche composite i relativi limiti di incertezza;
8. Interpretazione dei dati: necessario per rivelare i principali fattori trainanti di una buona o cattiva *performance* complessiva, poiché la trasparenza risulta un fattore fondamentale per una buona analisi e una politica efficace. Comprende il controllo di correlazioni e casualità, se possibile;
9. Collegamenti ad altri indicatori: identifica eventuali collegamenti con altri indicatori (semplici o composti) attraverso regressioni. Occorre tenere conto dei risultati dell'analisi di sensibilità;
10. Visualizzazione dei risultati: può influenzare l'interpretabilità, con una buona visualizzazione si comunica efficacemente il messaggio. Perciò, serve identificare

un insieme coerente di strumenti di presentazione per il pubblico di riferimento, selezionare la tecnica di visualizzazione che comunica il maggior numero di informazioni che presenta i risultati nel modo più chiaro e accurato possibile.

3.1.1 Passaggi 1-3: definizione e selezione

I primi tre passaggi rappresentano la fase preliminare della costruzione dell'indice, che non è da sottovalutare poiché fornisce la base per tutto il lavoro successivo. Se gli obiettivi non coincidono con i risultati ottenuti e se la selezione dei dati non viene fatta in modo adeguato, non si avrà un'analisi efficace.

Un solido quadro teorico è il punto di partenza, deve definire chiaramente il fenomeno da misurare e i suoi sotto-componenti, selezionando indicatori individuali e pesi che riflettano la loro importanza relativa e le dimensioni dell'indicatore composito complessivo. Questo processo si basa su ciò che è desiderabile misurare e non su quali indicatori sono disponibili. Inoltre, i concetti multidimensionali possono essere suddivisi in diversi sotto-gruppi, che non devono essere (statisticamente) indipendenti l'uno dall'altro, e i legami esistenti devono essere descritti teoricamente o empiricamente nella misura massima possibile. Ciò potrebbe anche rendere più facile determinare i pesi relativi tra i diversi fattori. Si identificano poi i criteri di selezione per gli indicatori sottostanti, che dovrebbero fungere da guida per l'inclusione o meno nell'indice dell'indicatore composito complessivo. Deve essere il più preciso possibile e descrivere il fenomeno misurato, cioè input, output o processo. In questa fase, così come nella successiva, si dovrebbero coinvolgere esperti e parti interessate, al fine di tenere conto di molteplici punti di vista e aumentare la robustezza del quadro concettuale e del set di indicatori.

I punti di forza e di debolezza degli indicatori compositi derivano in gran parte dalla qualità delle variabili sottostanti. Idealmente, le variabili dovrebbero essere selezionate in base alla loro rilevanza, solidità analitica, tempestività, accessibilità, ecc. . . . Sebbene la scelta degli indicatori debba essere guidata dal quadro teorico, il processo di selezione dei dati potrebbe essere piuttosto soggettivo, poiché non esiste un insieme definitivo di indicatori. Una mancanza di dati pertinenti di questa specie può anche limitare la capacità del progettista di costruire indicatori compositi solidi. Vista, dunque, la scarsità di dati quantitativi internazionalmente comparabili, gli indicatori compositi spesso includono dati qualitativi provenienti da sondaggi o revisioni politiche. Le misure *proxy* possono essere utilizzate quando i dati desiderati non sono disponibili o quando la comparabilità tra i paesi è limitata. L'accuratezza delle misure dovrebbe essere verificata attraverso analisi di correlazione e sensibilità e, per avere un confronto obiettivo tra paesi piccoli e grandi, è necessaria la trasformazione di scala delle variabili in una misura di

dimensione appropriata. Infine, il tipo di variabili selezionate deve corrispondere alla definizione dell'indicatore composito previsto. Di conseguenza, la qualità e l'accuratezza degli indicatori compositi dovrebbero evolvere parallelamente ai miglioramenti nella raccolta dei dati e nello sviluppo degli indicatori. D'altra parte, non si può sostenere l'idea che utilizzare ciò che è disponibile sia necessariamente sufficiente ed è scontato dire che dati scadenti produrranno risultati scadenti. Da un punto di vista pragmatico, tuttavia, è necessario fare compromessi nella costruzione di un composito, ciò che si ritiene essenziale è la trasparenza di questi compromessi.

Bisogna fare i conti anche con i dati mancanti, che spesso ostacolano lo sviluppo di indicatori compositi robusti, poiché possono essere assenti in modo casuale o non casuale. I modelli di assenza possono essere:

- assenti completamente in modo casuale (MCAR): i valori mancanti non dipendono dalla variabile di interesse o da qualsiasi altra variabile osservata nel set di dati;
- assenti in modo casuale (MAR): i valori mancanti non dipendono dalla variabile di interesse, ma sono condizionati da altre variabili nel set di dati;
- non assenti in modo casuale (NMAR): i valori mancanti dipendono dai valori stessi.

Sfortunatamente, non esiste un test statistico per NMAR e spesso non c'è una base su cui giudicare se i dati sono mancanti in modo casuale o sistematico, mentre la maggior parte dei metodi che imputano valori mancanti richiede un meccanismo di assenza casuale, cioè MCAR o MAR. Quando ci sono motivi per assumere un modello di assenza non casuale (NMAR), il modello deve essere esplicitamente modellato e incluso nell'analisi. Dunque, può essere molto difficile e potrebbe implicare assunzioni *ad hoc* che probabilmente influenzano il risultato dell'intera indagine statistica.

Esistono tre metodi generali per affrontare i dati mancanti: la cancellazione dei casi, l'imputazione singola o l'imputazione multipla. Il primo, noto anche come analisi dei casi completi, omette semplicemente i record mancanti dall'analisi, ma questo approccio ignora possibili differenze sistematiche tra campioni completi e incompleti e produce stime non distorte solo se i record cancellati sono un sotto-campione casuale del campione originale (assunzione MCAR). Inoltre, gli errori standard saranno generalmente più grandi in un campione ridotto, dato che vengono utilizzate meno informazioni. Come regola generale, se una variabile ha più del 5% di valori mancanti, i casi non vengono cancellati (Little & Rubin, 2002). Gli altri due approcci considerano i dati mancanti come parte dell'analisi, con la possibilità di portare alla minimizzazione del *bias* e all'uso di dati "costosi da raccogliere" che altrimenti verrebbero scartati dalla cancellazione dei

casi. Tuttavia, può anche consentire ai dati di influenzare il tipo di imputazione (Dempster & Rubin, 1983). L'incertezza nei dati imputati dovrebbe essere riflessa da stime di varianza, rendendo possibile il conto degli effetti dell'imputazione nel corso dell'analisi. In particolare, però, l'imputazione singola è nota per sottovalutare la varianza, poiché riflette parzialmente l'incertezza dell'attribuzione. Il metodo di imputazione multipla, che, invece, fornisce diversi valori per ogni valore mancante, può rappresentare più efficacemente l'incertezza dovuta all'assegnazione. Nessun modello di imputazione è esente da assunzioni e quindi i relativi risultati dovrebbero essere verificati accuratamente per le loro proprietà statistiche, come le caratteristiche distributive, così come per la loro significatività.

3.1.2 Passaggi 4-6: preparazione e costruzione

Negli ultimi decenni, si è assistito ad un aumento del numero di indicatori compositi sviluppati da varie agenzie nazionali e internazionali. Gli indicatori individuali, però, vengono talvolta selezionati in modo arbitrario, prestando poca attenzione alle interrelazioni tra di essi. Ciò può portare a indici che sovraccaricano e confondono sia i decisori sia il pubblico in generale. La natura sottostante dei dati deve essere attentamente analizzata prima della costruzione di un indicatore composito, sarà poi utile per valutare l'adeguatezza del set di dati e fornire una comprensione delle implicazioni delle scelte metodologiche.

L'analista, quindi, deve prima decidere se la struttura nidificata dell'indicatore composito è ben definita e se il set di indicatori disponibili è sufficiente o appropriato per descrivere il fenomeno. La decisione può basarsi sul parere di esperti e sulla struttura statistica del set di dati. Diverse metodologie analitiche possono essere utilizzate per l'esplorazione se le dimensioni del fenomeno sono statisticamente ben bilanciate nell'indicatore composito. In caso contrario, potrebbe essere necessaria una revisione degli indicatori individuali.

Un esempio di metodologia può essere l'analisi delle componenti principali (PCA), il cui obiettivo è quello di rivelare come le diverse variabili cambiano in relazione l'una all'altra e come sono associate, trasformando le variabili correlate in un nuovo insieme di variabili non correlate utilizzando una matrice di covarianza o la sua forma standardizzata, ossia la matrice di correlazione. L'analisi fattoriale (FA) è simile alla PCA, ma si basa su un particolare modello statistico. Un modo alternativo per indagare il grado di correlazione tra un insieme di variabili è utilizzare il coefficiente alpha di Cronbach (c -alpha), che è la stima più comune della coerenza interna degli elementi in un modello

o in un sondaggio. Queste tecniche di analisi multivariata sono utili per ottenere informazioni sulla struttura del set di dati dell'indicatore composito. Tuttavia, è importante evitare di effettuare analisi multivariate se il campione è piccolo rispetto al numero di indicatori, poiché i risultati non avranno proprietà statistiche note.

L'analisi dei *cluster* è un altro strumento per classificare grandi quantità di informazioni in insiemi gestibili. È stata applicata a una vasta gamma di problemi di ricerca e settori, dalla medicina alla psichiatria e all'archeologia. Essa è utilizzata anche nello sviluppo di indicatori compositi per raggruppare informazioni sui paesi in base alla loro somiglianza su diversi indicatori individuali. Quando il numero di variabili è elevato o quando si ritiene che alcune di esse non contribuiscano a identificare la struttura di *clustering* nel set di dati, possono essere applicati modelli continui e discreti in sequenza. I ricercatori spesso eseguono prima una PCA e poi applicano un algoritmo di *clustering* sui punteggi oggetto delle prime componenti (analisi tandem). Tutti questi procedimenti richiedono cautela, poiché possono identificare dimensioni che non necessariamente aiutano a rivelare la struttura di *clustering* nei dati e potrebbero effettivamente mascherare le informazioni tassonomiche.

Seguendo con i passaggi, si arriva alla normalizzazione, fondamentale prima di qualsiasi aggregazione dei dati, poiché gli indicatori in un set di dati spesso presentano unità di misura diverse. Esistono diversi metodi di normalizzazione:

- la classificazione: la tecnica di normalizzazione più semplice. Questo metodo non è influenzato da valori anomali e consente di seguire nel tempo la *performance* dei paesi in termini di posizioni relative. Tuttavia, la *performance* dei paesi in termini assoluti non può essere valutata, poiché le informazioni sui livelli vengono perse;
- la standardizzazione: converte gli indicatori in una scala comune con una media pari a 0 e una deviazione standard pari a 1. Allora, gli indicatori con valori estremi hanno un effetto maggiore sull'indicatore composito, fatto che potrebbe non essere desiderabile se l'intenzione è quella di premiare comportamenti eccezionali, cioè se un risultato estremamente buono su alcuni indicatori è considerato migliore di molti punteggi medi. Questo effetto può essere corretto nella metodologia di aggregazione, ad esempio escludendo i punteggi degli indicatori individuali migliori e peggiori dall'inclusione nell'indice o assegnando pesi differenziali in base alla "desiderabilità" dei punteggi degli indicatori individuali;
- il metodo del minimo-massimo: normalizza gli indicatori per avere un intervallo identico $[0, 1]$ sottraendo il valore minimo e dividendo per l'intervallo dei valori dell'indicatore. Però, i valori estremi o anomali potrebbero distorcere l'indicatore

trasformato e la normalizzazione minimo-massimo potrebbe ampliare l'intervallo degli indicatori che si trovano all'interno di un piccolo intervallo, aumentando l'effetto sull'indicatore composito più della standardizzazione;

- l'utilizzo di una scala categoriale che assegna un punteggio, di tipo numerico o qualitativo, per ciascun indicatore. Spesso, i punteggi si basano sui percentili della distribuzione dell'indicatore tra i paesi. Ad esempio, il 5% superiore riceve un punteggio di 100, le unità tra l'85° e il 95° percentile ricevono 80 punti, i valori tra il 65° e l'85° percentile ricevono 60 punti, così via fino a 0 punti, premiando così i paesi con le migliori *performance* e penalizzando i peggiori. Tuttavia, è difficile seguire gli aumenti nel tempo e le scale categoriali escludono grandi quantità di informazioni sulla varianza degli indicatori trasformati. Inoltre, quando c'è poca variazione nei punteggi originali, le bande dei percentili costringono la categorizzazione dei dati, indipendentemente dalla distribuzione sottostante. Una possibile soluzione può essere quella di regolare le fasce dei percentili tra gli indicatori individuali per ottenere variabili categoriali trasformate con distribuzioni quasi normali;
- la trasformazione degli indicatori sopra o sotto la media in modo tale che i valori attorno alla media ricevano 0, mentre quelli superiori o sotto una certa soglia ricevono rispettivamente 1 e -1. Ciò nonostante, l'arbitrarietà del livello di soglia e l'omissione delle informazioni sui livelli assoluti sono spesso criticate. Ad esempio, se il valore di un dato indicatore per il paese A è tre volte (300%) sopra la media, e il valore per il paese B è il 25% sopra la media, entrambi verrebbero considerati "sopra la media" con una soglia del 20% attorno alla media;
- i metodi per indicatori ciclici: assegnano implicitamente meno peso alle serie più irregolari nel movimento ciclico dell'indicatore composito, a meno che non venga eseguita una levigatura *ad hoc* in precedenza.

Parlando, invece, dei pesi, occorre dire che essi possono avere un effetto significativo sull'indicatore composito complessivo e sulle classifiche dei paesi. Esistono diverse tecniche di pesatura, alcune sono derivate da modelli statistici, altri da metodi partecipativi. Indipendentemente dal metodo utilizzato, sono essenzialmente giudizi di valore. Mentre alcuni analisti possono scegliere i pesi basandosi esclusivamente su metodi statistici, altri potrebbero premiare (o penalizzare) componenti ritenuti più (o meno) influenti, a seconda dell'opinione degli esperti, per riflettere meglio le priorità politiche o i fattori teorici.

La maggior parte degli indicatori compositi si basa su una pesatura uniforme, ovvero tutte le variabili hanno lo stesso peso. Questo vuol dire sostanzialmente che tutte le variabili hanno lo stesso valore nel composito, ma in ogni caso, la pesatura uniforme non significa “nessun peso”, implica, anzi, che i pesi siano uguali. Inoltre, se le variabili sono raggruppate in dimensioni e queste sono ulteriormente aggregate nel composito, applicare una pesatura uniforme alle variabili potrebbe comportare una pesatura diseguale della dimensione (le dimensioni che raggruppano un numero maggiore di variabili avranno un peso maggiore). Ciò potrebbe portare a una struttura sbilanciata nell'indice composito.

I pesi possono essere scelti anche per riflettere la qualità statistica dei dati. Dunque i pesi più elevati potrebbero essere assegnati ai dati statisticamente affidabili con una copertura ampia. Lo svantaggio di questa metodologia ricade sullo sbilanciamento a favore degli indicatori facilmente disponibili, penalizzando le informazioni che sono statisticamente più problematiche da identificare e misurare. Va notato che ci sarà quasi sempre una certa correlazione positiva tra misure diverse dello stesso aggregato. I modelli statistici come l'analisi delle componenti principali o l'analisi fattoriale possono essere utilizzati per raggruppare indicatori individuali in base al loro grado di correlazione. Tuttavia, i pesi non possono essere stimati con questi metodi se non esiste correlazione tra gli indicatori.

In alternativa, possono essere utilizzati metodi partecipativi che incorporano vari *stakeholder* (esperti, cittadini e politici) per assegnare i pesi. Questo approccio è fattibile quando c'è una base ben definita per una politica nazionale (Munda, 2005, 2007). Un esempio è l'allocatione del budget, in cui agli esperti viene assegnato un “budget” di N punti, da distribuire su un certo numero di indicatori individuali, “pagando” di più per quegli indicatori la cui importanza si vuole enfatizzare (Jesinghaus, 1997). Questo metodo è ottimale nel caso si volessero utilizzare una decina di indicatori perché, se ne fossero coinvolti di più, si potrebbe indurre un serio stress cognitivo negli esperti che devono allocare il budget.

Anche i metodi di aggregazione variano. Mentre il metodo di aggregazione lineare è utile quando tutti gli indicatori individuali hanno la stessa unità di misura, a condizione che vengano rispettate alcune proprietà matematiche, le aggregazioni geometriche sono più adatte se il modellatore desidera un certo grado di non compensabilità tra indicatori o dimensioni individuali. Inoltre, le aggregazioni lineari premiano gli indicatori di base in proporzione ai pesi, mentre le aggregazioni geometriche premiano i paesi con punteggi più elevati. In entrambe le aggregazioni, i pesi esprimono sostanzialmente dei compromessi tra indicatori, poiché un *deficit* in una dimensione può essere compensato da un

surplus in un'altra. Ciò implica un'incoerenza tra il modo in cui i pesi sono concepiti e il significato effettivo quando si utilizzano aggregazioni geometriche o lineari. In un'aggregazione lineare, la compensabilità è costante, mentre con le aggregazioni geometriche la compensabilità è inferiore per gli indicatori compositi con valori bassi.

Per garantire che i pesi rimangano una misura di importanza, dovrebbero essere utilizzati altri metodi di aggregazione, in particolare metodi che non consentano la compensabilità. Inoltre, se obiettivi diversi sono ugualmente legittimi e importanti, potrebbe essere necessaria una logica non compensativa. Questo è solitamente il caso in cui dimensioni altamente diverse sono aggregate nel composito, come nel caso degli indici ambientali che includono dati fisici, sociali ed economici. Se l'analista decide che un aumento delle prestazioni economiche non può compensare una perdita di coesione sociale o un peggioramento della sostenibilità ambientale, né l'aggregazione lineare né quella geometrica sono adatte. Un approccio multi-criterio non compensativo (MCA) potrebbe garantire non compensabilità trovando un compromesso tra due o più obiettivi legittimi. Nella sua forma base, questo approccio non premia i valori anomali, poiché conserva solo informazioni ordinarie, ossia quei paesi che hanno un vantaggio (o svantaggio) maggiore in indicatori individuali. Tuttavia, questo metodo potrebbe risultare computazionalmente costoso quando il numero di paesi è elevato, poiché il numero di permutazioni da calcolare aumenta esponenzialmente (Munda & Nardo, 2007).

Per quanto riguarda l'elemento temporale, mantenere i pesi invariati nel tempo potrebbe essere giustificato se il ricercatore intende analizzare l'evoluzione di un certo numero di variabili. I pesi non cambiano con l'MCA, essendo associati al valore intrinseco degli indicatori utilizzati per spiegare il fenomeno. Se, invece, l'obiettivo dell'analisi è definire le migliori pratiche o stabilire priorità, allora i pesi dovrebbero necessariamente cambiare nel tempo.

L'assenza di un modo "obiettivo" per determinare pesi e metodi di aggregazione non porta necessariamente a rifiutare la validità degli indicatori compositi, a patto che l'intero processo sia trasparente. Gli obiettivi del modellatore devono essere chiaramente dichiarati fin dall'inizio e il modello scelto deve essere testato per vedere in quale misura soddisfa gli obiettivi del modellatore.

3.1.3 Passaggi 7-10: verifica e presentazione

La robustezza degli indicatori compositi e dei messaggi politici sottostanti può essere messa in discussione da tutte le procedure precedentemente citate, pertanto dovrebbero essere affrontate tutte le potenziali fonti di incertezza. Una combinazione di analisi di

incertezza e analisi di sensibilità può aiutare a valutare la robustezza dell'indicatore composito e migliorare la trasparenza.

L'analisi di incertezza si concentra su come l'incertezza nei fattori in ingresso si propaga attraverso la struttura dell'indicatore composito e influisce sui valori dell'indicatore stesso. Essa valuta il contributo delle singole fonti di incertezza alla varianza dell'output e, sebbene sia utilizzata più frequentemente rispetto all'analisi di sensibilità e sia quasi sempre trattata separatamente, l'uso iterativo di entrambe durante lo sviluppo di un indicatore composito potrebbe migliorare la sua struttura (Saisana 2005; Tarantola, 2000; Gall, 2007). I risultati dell'analisi di robustezza sono generalmente riportati come classifiche dei paesi con i relativi intervalli di incertezza, che derivano dalle incertezze in gioco. Ciò, quindi, consente di comunicare all'utente l'intervallo plausibile dei valori dell'indicatore composito per ciascun paese.

Le misure di sensibilità, invece, rappresentano quanto l'incertezza nell'indicatore composito per un paese sarebbe ridotta se quella particolare fonte di incertezza venisse rimossa. I risultati di un'analisi di sensibilità sono spesso mostrati anche come diagrammi a dispersione, con i valori dell'indicatore composito per un paese sull'asse verticale e ciascuna fonte di incertezza in ingresso sull'asse orizzontale, poiché aiutano a rivelare schemi nelle relazioni input-output.

Tuttavia, la valutazione della robustezza non è sufficiente per garantire un indicatore composito sensato, perché l'analisi statistica dovrebbe aiutare a riflettere sul quadro teorico inizialmente utilizzato. Il modellatore, quindi, deve essere in grado di rispondere a domande riguardanti soprattutto la coerenza dell'indice, dato che ottenere il modello teorico corretto è la principale sfida di un indicatore composito.

Un'osservazione utile a questo punto riguarda l'analisi della correlazione, che non deve essere confusa con quella della causalità. La correlazione indica semplicemente che la variazione nei due set di dati è simile. Un cambiamento nell'indicatore individuale non porta necessariamente a un cambiamento nell'indicatore composito e viceversa. Invece, la causalità rimane poco chiara nell'analisi di correlazione. Va notato, inoltre, che gli indicatori compositi spesso includono alcuni degli indicatori con cui sono correlati, portando a una doppia contabilizzazione.

La modalità di presentazione degli indicatori compositi non è un aspetto secondario. Questi indicatori devono essere in grado di comunicare rapidamente e con precisione una storia a decisori e altri utenti finali. Le tabelle, sebbene forniscano informazioni complete, possono a volte offuscare questioni sensibili che emergono immediatamente con una rappresentazione grafica. Pertanto, chi presenta deve decidere, in ogni situazione, se includere una tabella, un grafico o entrambi. Un formato tabellare è la presentazione

più semplice, in cui l'indicatore composito è mostrato per ogni paese come una tabella di valori. Di solito, i paesi sono visualizzati in ordine decrescente di rango e le conseguenti classifiche possono essere utilizzate per monitorare i cambiamenti nelle *performance* degli stessi nel tempo. Sebbene le tabelle siano un approccio completo per visualizzare i risultati, possono sembrare troppo dettagliate e poco attraenti visivamente. Tuttavia, possono essere adattate per mostrare informazioni mirate per gruppi di paesi raggruppati per una determinata caratteristica (localizzazione geografica, PIL, ecc. . .).

Gli indicatori compositi possono essere visualizzati anche attraverso un basilare diagramma a barre. I paesi sono sull'asse verticale e i valori dell'indicatore composito sull'asse orizzontale. La barra superiore indica la *performance* media di tutti i paesi (inoltre, può essere considerata come un obiettivo da raggiungere) e consente al lettore di identificare come un paese si sta comportando rispetto alla media. Gli indicatori individuali sottostanti possono essere ugualmente mostrati in un diagramma a barre. Inoltre, mediante l'uso dei colori può rendere il grafico più attraente e mettere in evidenza i paesi che performano bene o male, che stanno crescendo o meno.

Infine, ci sono i grafici a linee, che possono essere utilizzati per illustrare invece i cambiamenti di un indicatore composito nel tempo.

3.2 Applicazione del COIN Tool

Come per i dieci passaggi degli indicatori compositi, il *COIN Tool* può essere utilizzato in modo iterativo. Il caso analizzato in questo elaborato, inoltre, riguarda una situazione particolare perché il dataset è composto da dieci risposte “Sì” (indicate con 1) e “No” (indicate con 0) per ogni rispondente, rappresentando un insieme di dieci variabili dicotomiche. Il procedimento da adottare è quindi molto più semplice di quanto descritto in precedenza, poiché verranno considerati solo alcuni passaggi.

Lo strumento *COIN* è costruito in Excel per essere accessibile al maggior numero possibile di utenti e mira anche a offrire la massima flessibilità nell'accomodare molteplici possibili strutture di indicatori compositi e approcci metodologici. Questa combinazione porta alla creazione di un file Excel di grandi dimensioni, che può risultare lento da usare a causa delle sue dimensioni e del modo in cui Excel funziona.

Per affrontare parzialmente questo problema, lo strumento *COIN* è disponibile in due versioni: una versione “Completa” (*Full*) e una versione “Leggera” (*Lite*). La versione Lite è identica alla versione Completa, ma ha una serie di schede rimosse, che riduce le dimensioni del file di circa la metà, rendendolo più veloce da eseguire. La prima è sufficiente per costruire un indicatore composito e analizzarne le correlazioni, mentre se

si desidera indagare sulle incertezze e sull'aggregazione non compensativa, è consigliabile usare la versione Completa. Dunque, per questo elaborato verrà usata la versione Lite.

Si inizia inserendo i dati grezzi nell'apposito spazio e nel formato utilizzato *dal COIN Tool*, ma, visto che esso lavora con un massimo di 300 unità, il lavoro sarà svolto con i primi 300 rispondenti registrati del dataset fornito, ossia i primi 300 cittadini dell'Unione Europea. Inoltre, i dati mancanti devono essere indicati con "n/a", ma in questo caso non ne sono presenti. Per lo studio di questo elaborato verrà, quindi, considerato questo "sotto-campione" dei dati.

Ogni unità deve essere assegnata con un nome completo, in questo caso ci saranno solo dei numeri ad identificare le persone e un codice abbreviato (rispettivamente "Person001" e "Unit.001").

Poi si dovrà specificare la struttura dell'indice. Qui, gli indici individuali sono dieci, ognuno riguardante una modalità relativa alla domanda del sondaggio (specificate all'inizio del secondo paragrafo), che verranno aggregati nella definizione di un unico indicatore composito. I valori degli indicatori devono essere disposti in una tabella, dove ogni riga corrisponde a un'unità e ogni colonna corrisponde a un indicatore. Poi, si inseriscono i pesi per ogni indicatore, che verranno utilizzati come tali quando gli indicatori saranno aggregati al livello del sotto-pilastro. I pesi sono relativi e non è necessario che la somma totale sia pari a 1. Ad esempio, impostare tutti i pesi a 1 è equivalente a dare lo stesso peso a tutti gli indicatori, mentre un peso pari a 0 implica che l'indicatore non viene considerato nell'aggregazione. Per questo passaggio è fondamentale il lavoro del costruttore, perché il peso da attribuire a ciascun indicatore è prevalentemente soggettivo (si vedano i metodi di aggregazione nel paragrafo precedente). Quindi, si possono attribuire importanze diverse per le varie modalità imposte, anche in base a cosa si vuole risaltare nei risultati.

Per ogni indicatore va inserita anche una direzione, dove un valore di 1 significa che valori più alti dell'indicatore sono associati a valori più alti dell'indice o del concetto e un valore di -1 significa che valori più alti dell'indicatore sono associati a valori più bassi dell'indice o del concetto. Nell'esempio analizzato, si hanno tutte le variabili nella stessa direzione, eccetto una, ossia l'ultima, cioè quella relativa alla modalità "Non sono disposto a cambiare le mie abitudini". Questo perché tutte le altre riguardano degli atteggiamenti positivi da seguire e, quindi, un valore alto (1) denota un miglioramento positivo della sostenibilità (e il valore basso, ossia 0, non denota nessun miglioramento). Mentre, al contrario, un valore alto (sempre pari a 1) nella modalità sopra citata, esprime un andamento negativo riguardante il turismo sostenibile.

Si deve, in seguito, specificare la struttura dell'indice, quindi l'indice stesso, il sotto-indice, il pilastro e il sotto-pilastro a cui ogni indicatore appartiene, e si inserisce un nome per ciascun indicatore. In questo esempio, si utilizzano delle parole chiave riguardanti le modalità specificate da ogni indicatore, ad esempio per il primo comportamento ambientale si può utilizzare "Protezione ambientale", per il secondo "Beneficio comunità", e così via. . . . Si possono costruire diverse strutture, in base al pensiero soggettivo del costruttore o in base a caratteristiche predefinite del dataset. In questo caso le dieci modalità di comportamento sono state divise nel seguente modo: le prime due modalità (rappresentano i sotto-pilastri) creano il primo pilastro, denominato "Spendere di più", poiché fanno entrambe riferimento ad una spesa maggiore per far ottenere un beneficio alle comunità locali; le due seguenti vengono unite in "Altra scelta", in quanto si parla di scegliere un altro periodo e destinazione per la vacanza; tutte le altre modalità, eccetto l'ultima, vengono raggruppate nel terzo pilastro, chiamato "Riduzione inquinamento", perché si fanno riferimento ad azioni che si possono fare tutti i giorni per ridurre l'inquinamento; infine, l'ultima modalità viene lasciata da sola, poiché rappresenta l'unica azione negativa, quindi, risulta separata dalle altre. Quest'ultima rimane unica anche nella denominazione dei due sotto-indici, il suo ha nome "Azioni negative", mentre l'altro, composto dalle altre nove modalità (unendo i tre pilastri appena definiti), viene nominato "Azioni positive".

Dopo aver fatto questo, ci si occupa dell'inserimento dei dettagli metodologici. Si seleziona il metodo di normalizzazione più adeguato, nel caso analizzato non ce n'è bisogno, avendo a che fare con tutte variabili di tipo dicotomico. Si mettono i punteggi di minimo e massimo, che rappresentano gli intervalli ai quali ogni indicatore verrà scalato per i metodi Min-Max, Goalposts e DataMax (i valori tipici e predefiniti sono 0 e 100, rispettivamente). Si includono la media e la deviazione standard. (i valori tipici e predefiniti per la media e la deviazione standard sono 50 e 10, rispettivamente). Infine, si introduce il tipo di aggregazione per ogni livello dell'indice. Gli indicatori possono essere aggregati ai sotto-pilastri solo tramite la media aritmetica, mentre tutte le altre aggregazioni possono essere selezionate come media aritmetica o geometrica.

La media aritmetica è la ben nota media ponderata dei punteggi e il punteggio dei sotto-pilastri è definito come

$$s_c = \frac{1}{w_i} \sum_{i=1}^n w_i x_i \quad (3.1)$$

dove n è il numero degli indicatori nel sotto-pilastro.

Mentre la media geometrica è calcolata come segue

$$s_c = \left[\prod_{i=1}^n x_i^{w_i} \right]^{\frac{1}{\sum w_i}} \quad (3.2)$$

cioè i pesi vengono utilizzati come esponenti e gli indicatori vengono moltiplicati anziché sommati.

Come già accennato nel paragrafo precedente, le medie aritmetiche sono spesso descritte come compensative, il che significa che punteggi elevati in un indicatore possono compensare punteggi bassi in un altro. Al contrario, le medie geometriche premiano le unità con profili equilibrati, in modo che punteggi elevati in un indicatore forniscano solo una piccola compensazione per punteggi bassi in un altro. Si nota, inoltre, che le medie geometriche richiedono valori di pilastri non zero. Se si verificano punteggi di pilastri pari a zero, è necessario impostare i parametri di normalizzazione in modo che gli indicatori normalizzati non incontrino valori zero.

Mentre, per le tipologie di normalizzazione sono disponibili cinque metodi di normalizzazione nel *COIN Tool*:

1. Normalizzazione Min-Max: l'approccio più comune. Per un indicatore grezzo X i punteggi normalizzati x sono calcolati come:

$$x = \frac{X - \min(X)}{\max(X) - \min(X)} \times 100 \quad (3.3)$$

Questa è una trasformazione lineare, che scala i valori nell'intervallo $[0,100]$ e il *COIN Tool* consente all'utente di regolare questo intervallo.

2. Normalizzazione DataZ (Z-score): simile alla Min-Max

$$x = \frac{S(X - \text{mean}(X))}{\text{std}(X)} + M \quad (3.4)$$

ma anziché limitare l'indicatore a un intervallo fisso, lo vincola ad avere una media M e una deviazione standard S . Mettendo M pari a 0 e S pari a 1 si ottiene uno Z-score standard, ma questo risulterà in valori negativi. I valori consigliati sono M pari a 50 e S pari a 10, anche se possono essere regolati nel *COIN Tool*.

3. Normalizzazione Datamax: un'altra trasformazione lineare che utilizza il valore massimo dell'indicatore come punto di riferimento

$$x = \frac{X}{\max(X)} \quad (3.5)$$

Questo comporta valori inferiori a 1. Il *COIN Tool* consente di scalare questi valori su un intervallo specificato tramite i parametri minimo e massimo.

4. Normalizzazione Goalpost: normalizza gli indicatori rispetto ai valori “teorici” minimo e massimo (i limiti del goalpost), che solitamente coprono un intervallo più ampio rispetto ai valori degli indicatori. Viene calcolato come

$$x = \frac{X - L(X)}{U(X) - L(X)} \quad (3.6)$$

dove $L(X)$ e $U(X)$ sono i limiti inferiori e superiori, specificati singolarmente per ciascun indicatore nella parte precedente. Questi punteggi vengono quindi scalati sull'intervallo definito dai punteggi minimo e massimo.

5. Normalizzazione Percentili (Percentranks): sostituisce semplicemente i punteggi di ciascun indicatore con i loro ranghi percentuali. I ranghi percentuali sono semplicemente i ranghi, espressi come valori percentuali. Poiché si basano sui ranghi, questa è una trasformazione non lineare, poiché le distanze relative tra le unità non vengono mantenute. L'implicazione è che i valori anomali (*outliers*) vengono automaticamente trattati e tutti gli indicatori assumeranno una distribuzione uniforme. Questo può essere desiderabile quando gli indicatori hanno distribuzioni problematiche e solo i ranghi sono importanti. Tuttavia, se i valori anomali devono essere conservati, questa opzione non dovrebbe essere utilizzata.

Si arriva alla scheda *Framework*, dove viene fornito un riassunto della struttura dell'indice. Il *COIN Tool* riassume automaticamente la struttura dell'indice fornita dall'utente nella scheda iniziale. Qui, sono richiesti alcuni dettagli aggiuntivi prima che inizi l'analisi dei dati, come l'assegnazione dei pesi e dei nomi a ciascun livello di aggregazione (sotto-pilastri, pilastri, sotto-indici e indice).

Il *COIN Tool* dedica cinque schede per illustrare le proprietà dei dati e offrire possibilità di trattare i valori anomali (*outliers*). Infatti, la scheda *Statistics* fornisce una panoramica delle proprietà statistiche dei dati, prima di qualsiasi trattamento. Gli utenti possono verificare i dati mancanti, i valori minimi e massimi, la media e la deviazione standard, e gli eventuali valori anomali.

Nello specifico vengono forniti per ciascun indicatore: i valori mancanti come percentuale e numero rispetto al numero totale di unità, i valori massimi e minimi, i momenti statistici (media, deviazione standard, asimmetria (*skewness*) e curtosi (*kurtosis*)), la presenza di *outliers*, e se sono alti o bassi, le informazioni sui quartili e sulla mediana, il peso, la direzione, le informazioni strutturali e il nome dell'indicatore. E per ciascuna

unità vengono mostrati il numero di indicatori con dati disponibili, la copertura dei dati, espressa come percentuale (le unità con una copertura degli indicatori inferiore al parametro di copertura minima vengono evidenziate in rosso) e il nome dell'unità in questione. Infine, vengono visualizzati i valori dei dati stessi, con valori pari a zero, dati mancanti e valori negativi evidenziati. Inoltre, gli *outliers* sono evidenziati utilizzando informazioni basate sui quartili

L'intenzione di tutte queste informazioni è fornire quante più informazioni possibili sugli indicatori, sulle unità e sui singoli punti dati. Questo consente di verificare, ad esempio, che i valori massimi e minimi siano all'interno degli intervalli previsti, che i dati mancanti per ogni indicatore e unità siano all'interno dei limiti accettabili. Una regola generale potrebbe essere avere almeno il 65% di copertura degli indicatori per ogni unità e dimensione (questa soglia dipende dal caso specifico e può dipendere dal grado di correlazione tra gli indicatori all'interno di una dimensione, tra le altre cose), che i valori zero e negativi non siano dovuti a errori. Inoltre, in base alle informazioni presenti nella scheda *Statistics*, potresti voler verificare e regolare manualmente i tuoi dati, per poi reinserirli nella scheda *Database*, oppure procedere alla scheda *Winsorisation* per il trattamento degli *outliers*.

La scheda *Winsorisation* tenta di "correggere" gli indicatori che hanno alloro interno dei valori anomali, definiti come quegli indicatori che superano le soglie di asimmetria e curtosi. Opera sostituendo degli *outliers* con il valore più vicino che non è un *outlier*. Il metodo è solitamente utilizzato in presenza di pochi valori anomali (circa il 5% delle unità). Per ogni indicatore, il *COIN Tool* esegue iterativamente la *Winsorisation* dei valori anomali, fino a un numero massimo specificato (cinque valori come predefinito), finché l'asimmetria e la curtosi rientrano nelle soglie specificate. Se non porta l'asimmetria e la curtosi dentro le soglie, il *COIN Tool* raccomanda una trasformazione Box-Cox. Quest'ultima è un metodo per trasformare una distribuzione di probabilità non normale in una distribuzione che sia (più vicina a) normale. Risulta essere adatto nei casi di dati molto asimmetrici. Sebbene la trasformazione Box-Cox sia in realtà una famiglia parametrizzata di trasformazioni, solo alcune di esse sono implementate nel *COIN Tool*, ma dovrebbero coprire la maggior parte dei casi.

Si noti che l'operazione agisce sugli indicatori dove è possibile, ma si ha la possibilità di utilizzare il dataset originale nel calcolo dell'indice, se lo si desidera. Si potrebbe anche scegliere di trattare manualmente gli *outliers* al di fuori del *COIN Tool* e reinserire i dati trattati nella scheda iniziale.

Sono disponibili due grafici per visualizzare tutte le trasformazioni Box-Cox, insieme alla trasformazione selezionata, e gli indicatori originali. L'idea è mostrare l'effetto delle

varie possibili trasformazioni Box-Cox e della trasformazione selezionata, in termini di quanto i dati trattati si discostano dai dati originali.

La tabella *TreatedData* riassume i trattamenti applicati nelle due schede precedenti e illustra le proprietà del set di dati trattato. Questo set di dati viene utilizzato per tutti i calcoli successivi nel *COIN Tool*, inclusi i valori dell'indice, l'analisi delle correlazioni e così via. Nel caso non si volessero trattare i dati, c'è la possibilità di selezionare il set di dati originale o già sono stati trattati prima di inserirli nel *COIN Tool*.

La scheda *IndCorrel* fornisce informazioni sulle correlazioni tra gli indicatori, mentre la scheda *Rebalancing* mostra le correlazioni tra i diversi livelli di aggregazione. Quest'ultima offre l'opportunità di regolare i pesi in base alle correlazioni. La prima introduce una matrice dei valori di correlazione tra gli indicatori. I valori sono i coefficienti di correlazione di Pearson, che misurano l'associazione lineare tra ogni coppia di variabili, tenendo conto della direzione degli effetti. Le correlazioni sono calcolate sul set di dati finale selezionato nella scheda *TreatedData*. Questi valori non dipendono dai metodi di normalizzazione e aggregazione. La matrice di correlazione è colorata per facilitare l'analisi, le alte correlazioni sono colorate in verde e le correlazioni molto alte sono evidenziate in verde scuro, invece le correlazioni negative sono colorate in rosso. La matrice, quindi, aiuta a individuare problemi con il modo in cui gli indicatori sono associati. In un caso ideale, gli indicatori sarebbero tutti positivamente correlati, ma non così fortemente correlati da essere collineari.

Mentre la scheda *IndCorrel* mostra le correlazioni tra gli indicatori, la scheda *Rebalancing* mostra le correlazioni tra ciascun indicatore, sotto-pilastro, pilastro, sub-indice e i livelli di aggregazione superiori. Ogni tabella mostra anche i pesi iniziali degli indicatori e include una riga con i "pesi adattati" che possono essere inseriti dall'utente. Successivamente, vengono mostrate le correlazioni basate sui pesi iniziali, così come quelle basate sui pesi adattati.

Le parti successive *Scores*, *Heatmap* e *Rankings* presentano i punteggi e le classifiche dell'indice basati sui dati forniti e sulle scelte metodologiche selezionate. La prima mostra nuovamente i punteggi di ogni unità, ma aggiunge anche i punteggi a livello di indicatore. Le unità sono ordinate per classifica in base ai pesi adattati, ma viene anche riportata la classifica utilizzando i pesi iniziali, e la differenza di classifica. Quest'ultima può essere utilizzata come una metrica per valutare l'impatto dell'adattamento dei pesi. Tipicamente, alcune unità sono più sensibili di altre agli aggiustamenti dei pesi. La seconda, invece, fornisce una panoramica iniziale dei punteggi a ciascun livello di aggregazione e delle classifiche dell'indice complessivo. Le unità sono ordinate in base ai punteggi dell'indice, dal più alto al più basso e valori dei punteggi sono riportati anche

a livello di indice, sub-indice, pilastro e sub-pilastro. Sono codificati a colori, in modo che i punteggi più alti siano evidenziati in verde e i più bassi in rosso. Questo permette di vedere a colpo d'occhio quali unità performano bene in dimensioni particolari. La classifica può essere utilizzata per confermare (o meno) quali unità hanno una buona *performance* complessiva e quali no. La terza mostra le classifiche di ogni unità a ogni livello, dall'indicatore al livello dell'indice. Essa fornisce semplicemente informazioni aggiuntive su come le unità si confrontano tra loro in tutte le dimensioni dell'indice.

L'ultima scheda che compare nella versione *Lite* è la *ScoreCorrel*, che contiene una matrice completa delle correlazioni tra tutti gli indicatori, sotto-pilastri, pilastri, sotto-indici e l'indice complessivo (è una combinazione delle schede *IndCorrel* e *Rebalancing*).

Conclusioni

La stesura di questo elaborato ha permesso di mettere in luce l'utilizzo vantaggioso del *COIN Tool* della Commissione Europea. Esso ha permesso di esplorare e misurare in maniera approfondita le varie modalità trattate, riassumendo l'indagine sottostante e le dieci variabili dicotomiche considerate in un unico indicatore composito. Questa piattaforma ha fornito una panoramica completa, dettagliata e quantificabile delle dinamiche che insorgono nel comportamento sostenibile durante il turismo.

Se si volesse aggiungere un'ulteriore informazione, si potrebbe utilizzare un metodo aggregativo molto semplice, che somma tutte le risposte ottenute (si sommano dunque le righe del dataset), che equivale ad un'aggregazione mettendo i pesi pari a 1 per ogni domanda. In questo caso sarebbe necessario non contare l'ultima modalità o contarla come valore negativo. Poi si sceglie una soglia k , pari ad esempio a 6, che starebbe a significare la sufficienza di comportamento sostenibile. I valori dell'indicatore per ogni persona che superano la suddetta soglia stanno a significare un comportamento ben positivo di quel determinato cittadino dell'Unione Europea. Al contrario, un valore al di sotto di tale soglia dimostrerebbe un comportamento non sostenibile. Ponendo anche un livello ancora più basso, quindi pari a 3, si potrebbero valutare dei comportamenti negativi e comportamenti intermedi, rispettivamente al di sotto di 3 e compresi tra 3 e 6.

L'uso del *COIN Tool* ha, quindi, permesso di visualizzare chiaramente come la correlazione tra indicatori a diversi livelli (da quelli più granulari come gli indicatori individuali, fino agli aggregati finali come l'indice complessivo) influisca sulle decisioni di ponderazione. Il modello di ponderazione (comprese le direzioni adottate) utilizzato, ha aiutato a ottenere un equilibrio tra gli indicatori, evitando la distorsione che deriverebbe dall'attribuire un peso eccessivo ad un singolo indicatore.

I risultati ottenuti suggeriscono diverse implicazioni politiche e aree di intervento per i paesi che desiderano migliorare la loro competitività globale. In particolare, i paesi con bassa *performance* in settori come la logistica, le connessioni digitali e l'integrazione nei mercati internazionali potrebbero considerare politiche per migliorare le proprie

infrastrutture fisiche e digitali, puntando sulla riduzione delle barriere commerciali e sul miglioramento della connettività interna ed esterna.

Dal punto di vista della ricerca futura, il *COIN Tool* offre numerose opportunità per espandere l'analisi a nuovi indicatori e settori, migliorando ulteriormente la precisione e la granularità delle misure. Un aspetto interessante potrebbe essere l'inclusione di indicatori legati, appunto, alla sostenibilità ambientale e all'innovazione tecnologica, che stanno diventando sempre più centrali nelle politiche di sviluppo a lungo termine. Inoltre, sarebbe utile sviluppare metodi avanzati di visualizzazione dei dati per facilitare l'interpretazione dei risultati, soprattutto per i *policymaker* e i decisori.

In conclusione, l'analisi condotta con il *COIN Tool* ha fornito una solida base di dati e *insights* per comprendere i diversi comportamenti sostenibili dei cittadini europei e per identificare le aree critiche su cui concentrarsi per migliorare il turismo sostenibile in generale. La flessibilità del *COIN Tool* nell'adattarsi alle esigenze di normalizzazione e ponderazione ha consentito di affrontare efficacemente le sfide legate all'analisi di un grande set di dati. Nonostante le difficoltà incontrate nella scelta dei metodi più adatti per la ponderazione, l'approccio utilizzato si è dimostrato valido per ottenere una valutazione equilibrata e informata, utile per le politiche ambientali, economiche e sociali dei paesi analizzati.

Bibliografia

- ASSEMBLEA GENERALE DELL'ORGANIZZAZIONE DELLE NAZIONI UNITE (2015). *Trasformare il nostro mondo: l'Agenda 2030 per lo Sviluppo Sostenibile*. pp. 13–27.
- BECKER, W., BENAVENTE, D., DOMINGUEZ TORREIRO, M., TACAO MOURA, C., FRAGOSO NEVES, A., SAISANA, M. & VERTESY, D. (2019). *COIN Tool User Guide*. Publications Office of the European Union.
- CASTELLINI, A. (2024). *Misurare ciò che conta di più: i nuovi standard globali adottati dall'ONU per misurare la sostenibilità del turismo*.
- IPSOS EUROPEAN PUBLIC AFFAIRS (2021). *Flash Eurobarometer 499 – Attitudes of European towards tourism*. Directorate-General for Communication.
- JESINGHAUS, J., SREBOTNJAK, T., CHERCHYE, L., VAN PUYENBROECK, T., RIVIÈRE, P., GRIFFIN, T. & LATVALA, A. (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing.
- TERZI, S., OTOIU, A., GRIMACCIA, E., MAZZIOTTA, M. & PARETO, A. (2021). *Open Issues in Composite Indicators*. Università degli Studi Roma Tre - Dipartimento di Economia.

