

**UNIVERSITA' DEGLI STUDI DEL MOLISE
CAMPOBASSO**



Dip. S.E.G.E.S.

*Dottorato di Ricerca in Teoria e Metodi Quantitativi per l'Analisi dello
Sviluppo
XXII CICLO*

***UN'ANALISI STATISTICA DELLA COMPETITIVITA'
TURISTICA***

Settore scientifico disciplinare: SECS-S/01

TESI DI DOTTORATO
DI
DANIELA LEONELLI

Matricola: 141169

TUTOR:
Prof.ssa Livia De Giovanni

COORDINATORE:
Prof. Alberto Franco Pozzolo

A Marika e Giuseppe

UN'ANALISI STATISTICA DELLA COMPETITIVITA' TURISTICA

Indice

<i>Ringraziamenti</i>	6
<i>Presentazione della ricerca</i>	9

CAPITOLO 1

IL CONTESTO DI RIFERIMENTO DELLA RICERCA: LO SCENARIO INTERNAZIONALE, EUROPEO, NAZIONALE.

<i>1.1 Il contesto internazionale</i>	14
<i>1.2 Il contesto europeo</i>	24
<i>1.3 Il contesto nazionale</i>	29

CAPITOLO 2

COMPETITIVITA' E SOSTENIBILITA' DELLE DESTINAZIONI TURISTICHE

<i>2.1. La destinazione turistica competitiva</i>	37
<i>2.2. Il modello di competitività della destinazione di Crouch e Ritchie</i>	42
<i>2.3. Altri modelli e indicatori sulla competitività delle destinazioni turistiche</i> .	46
<i>2.4. La sostenibilità come fattore chiave per la competitività</i>	53

CAPITOLO 3
LA METODOLOGIA STATISTICA

3.1. <i>La cluster analysis e la cluster fuzzy</i>	56
3.2. <i>Introduzione alla cluster analysis</i>	61
3.3. <i>La cluster analysis non gerarchica</i>	72
3.4. <i>L'approccio fuzzy alla cluster analysis</i>	74
3.5. <i>Fuzzy K-means</i>	77
3.6. <i>Criteri di cluster validity</i>	81
3.6.1. <i>Indici di fuzziness</i>	82
3.6.2. <i>Indice fuzzy silhouette</i>	83
3.7. <i>L'analisi in componenti principali</i>	85
3.8 <i>Il software R</i>	93

CAPITOLO 4
L'ANALISI ESPLORATIVA DELLA DESTINATION
COMPETITIVENESS DELLE REGIONI ITALIANE

4.1. <i>Obiettivi e fasi della ricerca</i>	96
4.2. <i>La progettazione e la creazione della matrice di dati</i>	99
4.3. <i>Le determinanti della competitività e gli indicatori</i>	100
4.3.1. <i>Territorio</i>	101
4.3.2. <i>Clima</i>	103
4.3.3. <i>Offerta</i>	104
4.3.4. <i>Flussi</i>	106
4.3.5. <i>Accessibilità</i>	107
4.3.6. <i>Profilo economico-turistico</i>	108
4.3.7. <i>Sostenibilità e qualità</i>	110

4.3.7. <i>Performance imprese turistiche</i>	115
4.4. <i>Applicazione e risultati dell'analisi in componenti principali</i>	118
4.4.1. <i>ACP Territorio</i>	118
4.4.2. <i>ACP Clima</i>	121
4.4.3. <i>ACP Offerta</i>	123
4.4.4. <i>ACP Flussi</i>	126
4.4.5. <i>ACP Accessibilità</i>	129
4.4.6. <i>ACP Sostenibilità e qualità</i>	131
4.4.7. <i>ACP Profilo economico-turistico</i>	135
4.4.8. <i>ACP Performance imprese turistiche</i>	138
4.4.8.1 <i>ACP Performance settore alberghi</i>	138
4.4.8.2 <i>ACP Performance settore agenzie di viaggio e T.O.</i>	140
4.5 <i>Applicazione e risultati della cluster analysis crisp</i>	145
4.5.1. <i>Applicazione e risultati cluster analysis gerarchica</i>	145
4.5.2. <i>Applicazione e risultati cluster analysis non gerarchica</i>	149
4.6. <i>Classificazione fuzzy delle regioni italiane</i>	150
4.7. <i>I cluster regionali</i>	160
4.7.1. <i>Gruppo 1: Destinazioni di prossimità</i>	163
4.7.2. <i>Gruppo 2: Destinazioni leader</i>	167
4.7.3. <i>Gruppo 3: Destinazioni mature</i>	171
4.7.4. <i>Gruppo 4: Destinazioni da sviluppare</i>	174
4.8. <i>Conclusioni</i>	180
<i>Appendice</i>	183
<i>Bibliografia</i>	187
<i>Siti Internet consultati</i>	193

Ringraziamenti

Quando si arriva alla conclusione di un progetto inevitabilmente il pensiero va a tutto il tempo trascorso, a tutte le energie impegnate per realizzarlo, al risultato ottenuto, alle speranze riposte in esso...ma soprattutto il pensiero va a tutte le persone fondamentali per raggiungere l'obiettivo, a quelle sempre presenti e pronte a sostenerti.

Il mio percorso è stato lungo, difficoltoso, imprevisto, ma è stato sicuramente un percorso di crescita professionale e umana. Il corso di dottorato ha impegnato gli anni più intensi della mia vita, gli anni in cui s'investe sulla propria persona, sulla propria preparazione per essere pronti ad affrontare le sfide del futuro. Nel mio caso, però, questo lungo periodo da "dottoranda" ha occupato anche gli anni della mia crescita umana e personale. In questi anni tante cose sono cambiate, vivo in un'altra città, ho un nuovo e imprevedibile lavoro, e soprattutto sono diventata moglie e mamma.

Questi avvenimenti inevitabilmente hanno influito sul mio percorso accademico, sicuramente rallentandone i tempi e forse anche modificandone i contenuti, ma non hanno in nessun modo intaccato le mie aspirazioni nonostante le crescenti difficoltà nel conciliare tutto! E' difficile oggi per una donna essere una mamma presente, una moglie attenta, avere un lavoro e continuare a studiare per migliorare se stessa e non rinunciare ai propri obiettivi!!!

Riuscire a fare tutto questo è stato possibile per me solo grazie a molte persone speciali che mi hanno aiutato ad affrontare ogni giorno i miei limiti, le mie paure, i momenti difficili in cui pensavo di mollare.

E' grazie a tanta forza di volontà e al loro aiuto che oggi sono qui ad aggiungere un nuovo e importante tassello per me stessa e per la mia vita. I ringraziamenti dunque, sono più che mai dovuti ma, senza dubbio, sono soprattutto sinceri.

Un “grazie” speciale va alla Prof.ssa Livia De Giovanni, è grazie a lei se ho intrapreso questo percorso, ed è anche grazie ai suoi preziosi consigli, alla sua disponibilità, alla sua cortesia, alla sua preparazione se oggi sono arrivata fino a qui. E’ stata una presenza costante ma mai invadente, una presenza discreta, lontana da me geograficamente ma sempre vicina al momento delle mie richieste di aiuto. Nei tanti anni che ho avuto modo di conoscerla ho potuto apprezzare anche il suo lato umano, scoprendo una persona davvero ricca interiormente. Inoltre è stata lei, anni fa, a scoprire il mio lato “quantitativo”!

Un ringraziamento particolare al Prof. Pierpaolo D’Urso per avermi “ereditato” senza esitazione e avermi mostrato la sua disponibilità e il suo prezioso aiuto in questo percorso. Devo a lui anche l’esperienza nella didattica accademica, mi ha dato l’opportunità di trasferire una piccola parte di ciò che avevo imparato ad altri studenti, mi ha dato l’opportunità di saltare dall’altra parte della barricata e accorgermi quanto sia difficile, delicato, ma anche entusiasmante il lavoro di formatore.

Con affetto e stima ringrazio il Prof. Antonio Minguzzi che ha sempre creduto in me, dandomi la possibilità di assaporare la bellezza del lavoro appassionante di ricercatore. Lo ringrazio per le tante occasioni in cui mi ha mostrato la sua stima e il suo apprezzamento.

Un grazie a Daniela Di Gregorio per i momenti di lavoro, studio e divertimento che mi ha fatto vivere all’interno della facoltà... oltre al nome, condivido con lei gli stessi valori e la stessa integrità morale.

Un abbraccio e un ringraziamento a Nadia per i tanti momenti di confronto da “dottorande” e di divertimento da “amiche”.

Con amore ringrazio mio marito e mia figlia.

Probabilmente appare scontato ma sono loro il motore di tutta la mia forza di volontà, la voglia di non deluderli e soprattutto la voglia di poter trasmettere un giorno a mia figlia la passione per lo studio, per la ricerca, per la conoscenza, il valore dei propri “sogni” e delle proprie aspirazioni.

Ringrazio Giuseppe, mio marito, inconsapevole vittima di tanti miei sfoghi e “finanziatore” di questo mio investimento lunghissimo. Mi rendo conto che per chi sta fuori dall’ambiente accademico non è così facile condividere questa scelta e ancor di più comprenderla, ma ringrazio Giuseppe per avermi fatto semplicemente percorrere la mia strada.

Non potrebbe mancare in questa lista mia suocera per tutti i pomeriggi e le mattinate da “nonna sitter”. Senza la sua presenza e il suo aiuto non avrei avuto il tempo per dedicarmi a tutto questo. Il ringraziamento per gli stessi motivi si estende alle mie cognate e a mio suocero che mi hanno sempre aiutato per il raggiungimento di questo obiettivo.

E infine, li cito per ultimi, ma non per ordine d’importanza, un grazie ai miei genitori per tutto quello che mi hanno insegnato, perché grazie a loro sono una persona libera, mi hanno insegnato che si può ottenere molto solo con le proprie forze, con la propria testa e con la propria volontà ed io continuo a crederci nonostante la realtà spesso mostri tutt’altro. Un grazie alle mie sorelle...per esserci semplicemente, so di poter contare su di loro e questo per me è importante.

Un grazie affettuoso anche a Tina e Antonella per i caffè condominiali utili per distrarmi e ricaricarmi!!!!

Un ringraziamento alla Blu Srl per l’aiuto “tecnico” nella parte finale di questo percorso.

E infine voglio ringraziare me stessa.....sebbene sia una persona molto umile guardandomi indietro posso essere orgogliosa, sarebbe stato molto più facile abbandonare questa strada e invece io l’ho percorsa mettendoci tutta me stessa e rinunciando a molto altro.

Ho anche tante cose da rimproverarmi ma il bilancio finale di questa esperienza è più che mai positivo.

E per concludere ringrazio tutti quelli che hanno, volontariamente o involontariamente, riempito questi anni con i loro sorrisi, i loro consigli, le pacche sulle spalle e le critiche costruttive.

...Siate Affamati! Siate folli!!

Steve Jobs

Presentazione della ricerca

Il lavoro che si va a presentare si colloca al termine del corso di dottorato in “Teoria e metodi quantitativi per l’analisi dello sviluppo”, e il turismo rappresenta sempre di più una leva strategica di sviluppo a livello globale in termini economici, culturali, sociali, territoriali.

Molti Paesi hanno investito e stanno impegnando grandi risorse per lo sviluppo del settore turistico che ha assunto sicuramente un ruolo decisivo nell’economia e nella globalizzazione mondiale. Il peso rilevante del turismo nello scenario economico mondiale è testimoniato da tassi di crescita costantemente superiori a quelli del Pil mondiale; la sua importanza, non solo economica, ma anche culturale e sociale è cresciuta a ritmi intensi dalla seconda metà del Novecento ed è destinata a una continua crescita nel futuro.

Attualmente il settore turistico costituisce circa il 9% del Pil mondiale e occupa oltre 200 milioni di persone. Secondo l’*UNWTO* nel 2008 si sono mossi oltre 922 milioni di turisti, che hanno speso oltre 642 milioni di Euro. Le previsioni per il futuro stimano il superamento del miliardo e mezzo di viaggiatori per il 2020.

L’Europa, la più importante meta turistica del mondo e la principale regione per il numero di presenze turistiche, tra il 1950 e il 2008 ha visto gli arrivi internazionali passare da 25,3 milioni a oltre 488 milioni, e si stima un aumento superiore ai 700 milioni di turisti nel 2020.

L’importanza e la rilevanza del turismo per l’economia mondiale si palesano anche nella capacità del settore di sostenere la difficile situazione economica attuale. Il turismo può e deve, indubbiamente, rappresentare un

volano strategico di sviluppo e occupazione in grado di contribuire al rilancio competitivo dei sistemi economici nazionali.

La rilevanza del fenomeno turistico non riguarda soltanto le ricadute economiche e occupazionali che esso può generare, ma anche l'opportunità di conciliare obiettivi sociali, culturali e ambientali in un'ottica di sostenibilità a lungo termine.

La concorrenza tra destinazioni turistiche è sempre più intensa e si gioca a livello internazionale in uno scenario del turismo mondiale sempre più globalizzato.

Ogni destinazione è potenzialmente raggiungibile in poche ore e si trova, quindi, a competere con nuovi concorrenti sempre più numerosi.

E' necessario quindi dotarsi, anche nell'ambito del settore turistico e della concorrenza tra territori, di strumenti decisionali simili a quelli utilizzati negli altri settori economici.

Negli anni si è affermata, perciò, una nuova disciplina, nota come *destination management*, che mira a estendere al turismo i modelli concettuali, i metodi e gli strumenti degli studi aziendali in tema di *marketing*.

Nella letteratura di *tourism management* è cresciuto l'interesse per il tema della "*destination competitiveness*". Numerosi autori e centri di ricerca si sono occupati dei fattori che risultano determinanti per la competitività di una destinazione e delle possibili tecniche di misurazione della stessa da poter impiegare, in modo da avere gli strumenti per adeguare la pianificazione strategica ai nuovi contesti concorrenziali.

Negli ultimi decenni numerose destinazioni *leader* hanno mostrato una perdita di competitività. Contemporaneamente, si è assistito allo sviluppo di destinazioni emergenti e al consolidamento di destinazioni già affermate che hanno saputo rinnovare profondamente la loro offerta turistica mantenendo o accrescendo la loro quota di mercato.

L'Italia registra da alcuni anni una perdita di competitività tale da rendere necessari interventi urgenti e incisivi che possano rilanciare i prodotti

turistici nazionali in modo da tornare a occupare quella posizione di leadership che il nostro Paese merita data la sua innegabile vocazione turistica.

La destinazione Italia ha assistito negli anni alla progressiva erosione della propria quota di mercato nel turismo mondiale conseguendo risultati inferiori sia nei confronti dei suoi concorrenti storici, in particolare della Spagna e della Francia, che delle nuove destinazioni emergenti (tra cui, nell'area dell'Europa del Mediterraneo Meridionale, soprattutto Turchia e Croazia).

L'Italia ha da sempre sottovalutato l'importanza strategica ed economica del turismo come motore di sviluppo territoriale. L'evidente 'ritardo' dell'Italia rispetto ai suoi concorrenti, tradizionali ed emergenti, e la necessità di reinventare e innovare il suo modello di offerta adattandolo alle dinamiche attuali della domanda turistica internazionale richiede, però, un'attenta analisi di come la minore competitività a livello nazionale si sia realmente manifestata a livello regionale.

Il nostro paese, infatti, come per altri settori economici, presenta a livello regionale sostanziali differenze nello sviluppo turistico e nelle performance competitive.

Il lavoro di ricerca s'inserisce nel dibattito sulla competitività delle destinazioni.

L'obiettivo principale della ricerca è di indagare e misurare, in una prospettiva statistico-territoriale, le possibili determinanti della competitività turistica nelle diverse regioni italiane. Il lavoro mira a individuare cluster regionali con un'evoluzione competitiva sostanzialmente omogenea evidenziando gli aspetti comuni. Contemporaneamente si analizzano le cause della differenziazione dei risultati conseguiti dai cluster, cercando di comprendere se ciò sia conseguenza della sola attrattività potenziale di tali aree (dotazione di risorse naturali, storico-culturali ecc.) o anche della maggiore o minore capacità di sfruttare le potenzialità dei territori.

L'analisi statistica della competitività delle Regioni Italiane è stata effettuata attraverso una strategia d'analisi che presuppone l'utilizzo di tecniche statistiche multivariate utilizzate in combinazione tra loro.

Per ciascun cluster individuato l'analisi statistica, ha consentito di indagare le potenzialità territoriali a livello di risorse turistiche di diversa natura, il grado di sviluppo dell'industria turistica, i fattori di supporto all'attività turistica, la domanda attratta, il profilo economico-turistico, il livello di sostenibilità, le performance economiche delle imprese turistiche operanti sul territorio.

L'analisi della competitività dell'Italia e delle Regioni è stata sviluppata su indicatori relativi ai *driver* di competitività turistica indicati dai principali modelli internazionali (UNWTO, WEF) e sulla base di alcuni modelli sviluppati nella letteratura di *management* turistico (Crouch e Ritchie, Dywer e Kim, Enright e Newton, Heat). L'analisi di alcune ricerche sul tema dei principali centri studi nazionali in ambito turistico (TCI, ISNART) e di società di consulenza (The European House-Ambrosetti) hanno influito sulla scelta delle variabili da considerare e inserire nella ricerca.

Si è costruita così una matrice di dati molto vasta che comprende numerose variabili che concorrono alla definizione della competitività regionale.

Su tale matrice si sono applicate tecniche statistiche di tipo esplorativo per la riduzione della dimensionalità dei dati e per la classificazione delle unità statistiche in gruppi omogenei. In particolare, in una prima fase della ricerca si è applicata l'analisi in componenti principali per la riduzione delle numerose variabili raccolte. La fase successiva ha riguardato l'applicazione di tecniche di *cluster analysis* di tipo *crisp* e *fuzzy* per l'individuazione di gruppi di regioni omogenee, sulla base degli indicatori di performance individuati dall'analisi in componenti principali.

Il lavoro di tesi è organizzato in quattro parti descritte di seguito.

Il primo capitolo del lavoro è dedicato all'analisi dell'ambiente di riferimento della ricerca attraverso l'analisi delle principali dinamiche della domanda turistica internazionale, con un *focus* sull'area dell'Europa del Mediterraneo Meridionale (la principale area turistica al mondo e quella in cui operano la gran parte dei concorrenti diretti dell'Italia), e il posizionamento competitivo del nostro Paese nello scacchiere mondiale.

Successivamente, nel secondo capitolo l'attenzione sarà rivolta al dibattito sulla competitività e sostenibilità delle destinazioni turistiche con una sintesi della letteratura in tema di *destination competitiveness* e dei modelli più noti cui la ricerca fa riferimento.

Il terzo capitolo sarà dedicato alla descrizione degli aspetti teorici delle metodologie statistiche utilizzate e del *software* utile all'implementazione dell'analisi.

Seguirà, nell'ultimo capitolo, l'ampia trattazione riguardante la descrizione della ricerca oggetto di tale tesi, evidenziandone gli obiettivi, le fasi e i risultati ottenuti, gli aspetti scientifici più rilevanti, le problematiche, i limiti e i possibili sviluppi futuri.

CAPITOLO 1

Il contesto di riferimento della ricerca: lo scenario internazionale, europeo, nazionale.

SOMMARIO: 1.1. Il contesto internazionale. - 1.2. Il contesto europeo. - 1.3. Il contesto nazionale.

1.1 Il contesto internazionale

Il 2008¹ è stato un anno particolarmente difficile, fortemente influenzato dall'esplosione della crisi e dalla sua propagazione dagli Usa verso l'Europa e gli altri continenti. Tra i fattori principali della crisi figurano, in particolare, l'aumento del prezzo del petrolio e delle materie prime in generale, l'instabilità dei tassi di cambio e la stretta creditizia delle banche. Tali fattori hanno minato la fiducia dei consumatori contribuendo all'innescarsi della recessione globale che, inevitabilmente, si è ripercossa anche sulla propensione ai consumi turistici. Si è registrato, quindi, a partire dal secondo semestre 2008, un brusco contenimento del *trend* espansivo della domanda turistica mondiale.

Secondo i dati² dell'UNWTO³, dopo un aumento del 5,4% verificatosi nella prima metà dell'anno, gli arrivi di turisti stranieri a livello mondiale, sono

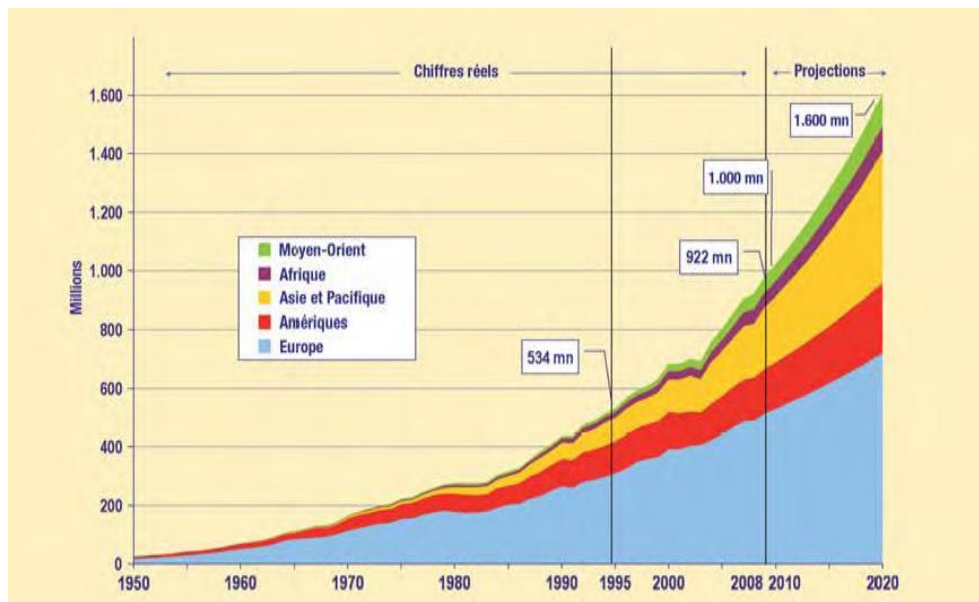
¹ Il 2008 è l'anno di riferimento per i dati raccolti e utilizzati nella ricerca. La scelta di riferirsi a tale periodo è dovuta a motivazioni di ordine sia pratico sia teorico. Le motivazioni di ordine pratico sono relative al fatto che, al momento della creazione del database, il 2008 rappresentava l'anno più recente di cui erano disponibili i dati, per molte delle variabili identificate. Inoltre tale scelta è stata perseguita anche perché il 2008 è risultato un anno particolare dal punto di vista dell'economia mondiale e del turismo.

² La fonte dei dati a livello internazionale è UNWTO World Tourism Barometer, Vol.7, N°2 June 2009.

³ L'**United Nations World Tourism Organisation** è stata fondata nel 1925 con la denominazione di *International Congress of Official Tourist Traffic Associations*. Nel 1975

diminuiti nel secondo semestre di circa un -1,3%, determinando un risultato annuale complessivo di 922 milioni di arrivi (fig.1), cioè solo 17 milioni in più rispetto all'anno precedente (+1,9%). Tale risultato per il 2008 appare certamente negativo rispetto al +6,9% del 2007 e, soprattutto, rispetto all'andamento medio dei quattro anni precedenti, che avevano fatto registrare un tasso di crescita medio del +7% tra il 2004 e il 2007.

Fig.1- Serie storica e proiezione arrivi dei turisti internazionali (1950-2020)



Fonte: Tourism Highlights 2009

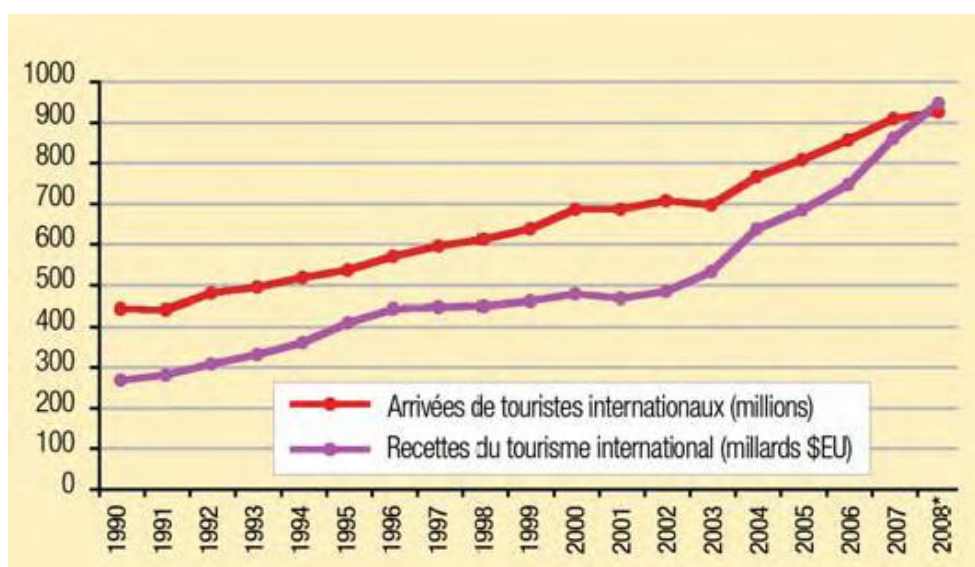
Le entrate del turismo internazionale⁴ hanno raggiunto i 642 miliardi di euro nel 2008, registrando un aumento dell'1,8% in valore reale rispetto all'anno precedente, con un incremento sostanzialmente parallelo a quello degli arrivi. A tali entrate vanno aggiunti gli introiti del trasporto

l'organizzazione ha assunto la denominazione attuale fissando la sede a Madrid. Attualmente aderiscono all'UNWTO più di 150 Stati membri. L'adesione dell'Italia risale al 1978.

⁴ *L'entrate del turismo internazionale per i Paesi di destinazione sono considerate come esportazioni e comprendono tutte le transazioni legate ai consumi dei visitatori sul territorio sia da parte dei turisti in senso stretto (che effettuano pernottamenti) sia da parte degli escursionisti (che non effettuano pernottamenti).*

internazionale⁵ dei passeggeri che, per il 2008, hanno raggiunto i 183 miliardi di dollari. Il totale delle entrate, quindi, per il 2008, supera il trilione di dollari (3 miliardi di dollari al giorno). Con questi numeri il turismo rappresenta, attualmente, il 30% delle esportazioni di servizi nel mondo e il 6% delle esportazioni totali di beni e servizi.

Fig. 2 - Serie storica degli arrivi e delle entrate del turismo internazionale (1990-2008)

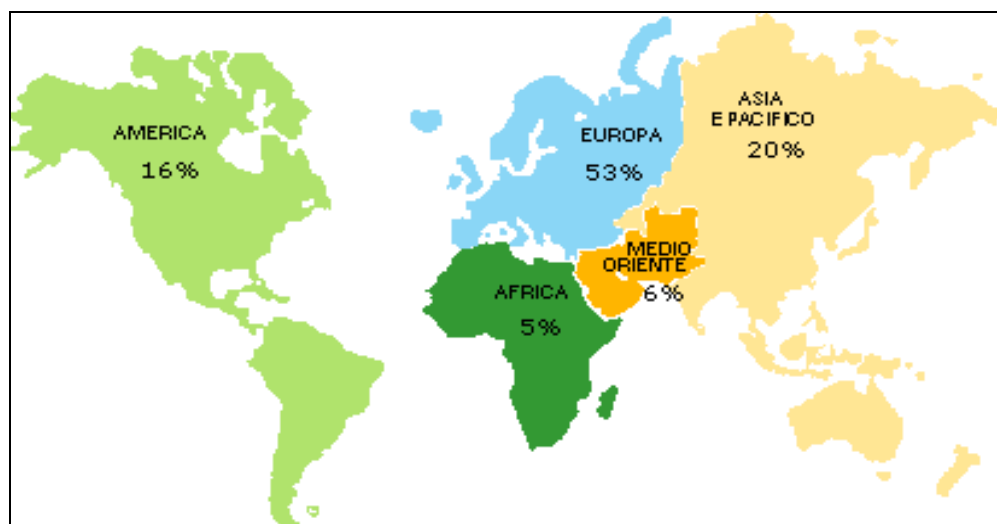


Fonte: *Tourism Highlights 2009- UNWTO*

In tale contesto l'Europa, nonostante sia la destinazione più matura, ha mantenuto la sua posizione di leadership mondiale con una quota di mercato per gli arrivi internazionali sostanzialmente stabile (53%), seguita a grande distanza dall'area Asia e Pacifico (20%), dalle Americhe (16%), dal Medio Oriente (6%) e dall'Africa (5%) (Fig.3).

⁵ Nella voce entrate valutarie dell'UNWTO non sono incluse le spese di trasporto sostenute dai turisti per raggiungere quella destinazione.

Fig. 3- Quota di mercato per arrivi internazionali (2008)



Fonte: ns. elaborazione su dati UNWTO

La crescita registrata, seppur positiva, è stata molto contenuta nelle macro-aree mondiali, ad eccezione del Medio Oriente che ha fatto registrare un tasso di crescita a due cifre sia in termini di arrivi (+18,2%) che di entrate (+16,8%), con 55 milioni di arrivi di turisti e 30,6 miliardi di euro di entrate.

L'area europea del Mediterraneo, la più grande destinazione turistica del mondo, pur non subendo riduzioni, è tra quelle che hanno conseguito il minor tasso di crescita degli arrivi internazionali (+ 0,6%) e una diminuzione dello 0,5% per le entrate. L'Europa⁶ nel complesso ha manifestato una stagnazione degli arrivi, con un esiguo incremento dello 0,1% rispetto al 2007.

L'andamento degli arrivi nel continente Asiatico e nel quadrante Pacifico appare particolarmente significativo in termini di brusca diminuzione. L'incremento del 10,5% del 2007 si riduce fortemente ad un esiguo +1,6% nel 2008, con un peggioramento maggiore nel secondo semestre dell'anno pari a un -2,8%.

⁶ Si veda il par.1.2 per un approfondimento delle dinamiche del turismo in Europa.

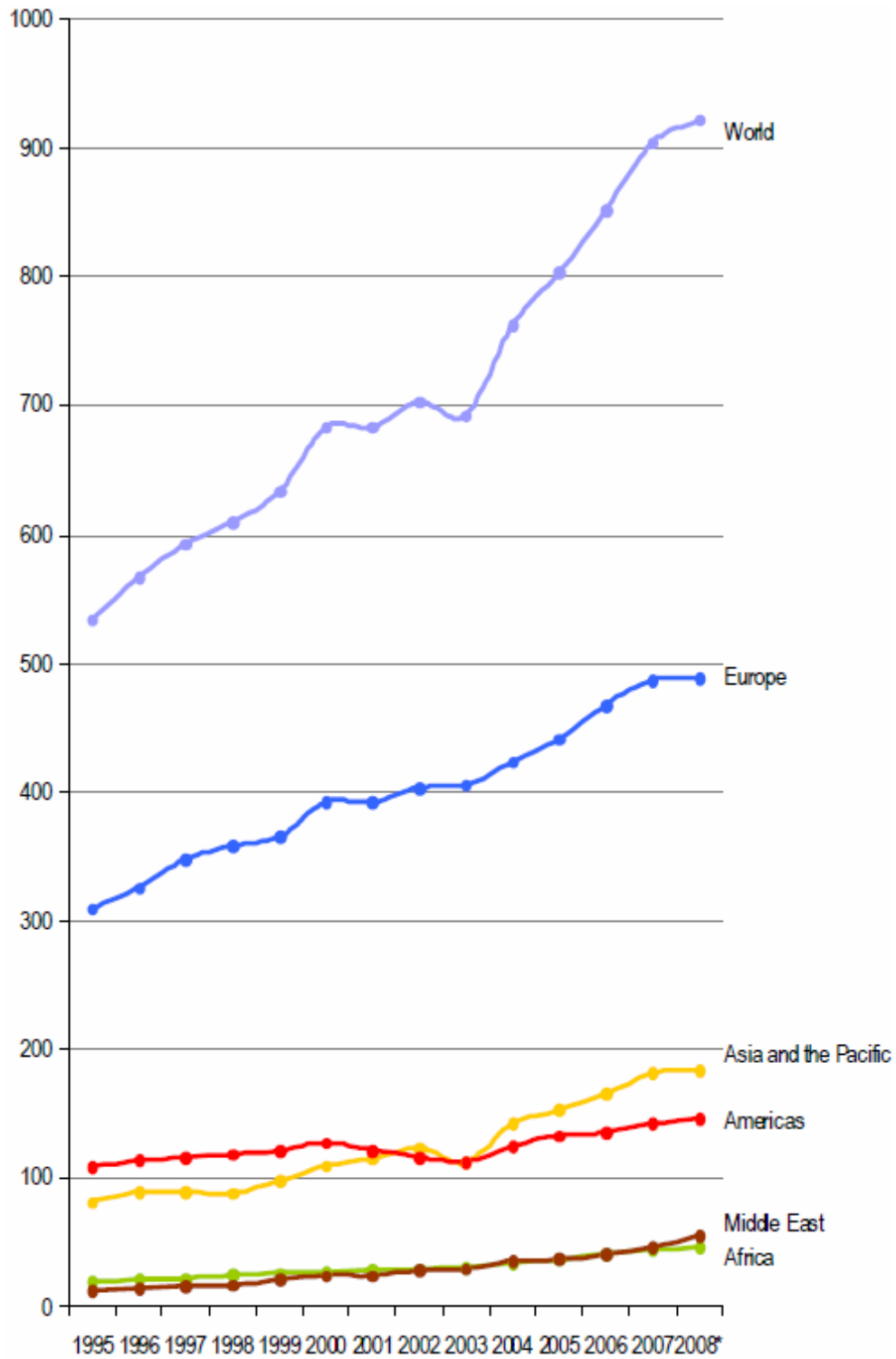
Il quadrante asiatico di Nord Est e l'Australia evidenziano le *performance* peggiori con una riduzione, rispettivamente, del 4,6% e del 3,3% del totale degli arrivi. Tali destinazioni sono mete privilegiate di un turismo a lungo raggio che risulta particolarmente penalizzato dall'aumento dei prezzi del carburante e del conseguente rialzo dei biglietti aerei, questo, inevitabilmente, incide sulla riduzione degli arrivi internazionali.

Anche il continente americano mostra i segni di un forte rallentamento dei flussi, che passano da un +6,8% del primo trimestre a +1,1% nella parte finale dell'anno con una netta differenziazione tra le sue macro-aree. L'area nordamericana registra un brusco rallentamento nella seconda parte dell'anno, la crescita degli arrivi passa dal + 7% d'inizio anno al +0,7% del secondo semestre. Per gli stati sudamericani il rallentamento è molto meno evidente e si conferma tutto sommato una crescita abbastanza sostenuta (+4,8% nel secondo semestre).

Ben più evidente appare la contrazione cui si assiste, tra il primo e il secondo semestre 2008 in Medio Oriente, si passa da +17,8% a + 5,2%.

Il continente africano, infine, appare il meno toccato dalla crisi in termini di variazione dei flussi turistici internazionali in entrata (dal +5,4% al +4,1% tra primo e secondo semestre 2008). Il dato appare ragionevole se si riflette sul relativo isolamento economico del continente dalle dinamiche più accentuate della globalizzazione.

Fig.4- Serie storica degli arrivi internazionali nelle macro-aree mondiali (1995-2008.)



Fonte: Tourism Highlights 2009- UNWTO

Anche la classifica dei Paesi sia rispetto alle entrate sia rispetto agli arrivi, nel 2008, risulta leggermente modificata (tab. 1 e tab. 2).

La Francia resta in testa alla graduatoria per il numero di arrivi, mentre a livello di entrate la *leadership* è degli Stati Uniti. In entrambe le graduatorie rientrano otto delle dieci grandi destinazioni inserite. Il nostro Paese mantiene la quinta posizione nella classifica degli arrivi internazionali con 42,7 milioni di turisti e il quarto posto per ciò che concerne le entrate, con 45,7 miliardi di dollari.

Le entrate del turismo internazionale realizzate dai primi dieci paesi in classifica rappresentano il 49% del totale mondiale, leggermente inferiore la percentuale per ciò che riguarda gli arrivi, per i quali, le prime dieci destinazioni al mondo rappresentano il 45% del totale mondiale.

Tab. 1- Top Ten destinazioni per numero di arrivi (in milioni)

Posizione		2007 (milioni)	2008 (milioni)	08/07 %
↔	Francia	81,9	79,3	-3,2%
↑	Stati Uniti	56	58	3,6%
↓	Spagna	58,7	57,3	-2,3%
↔	Cina	54,7	53	-3,1%
↔	Italia	43,7	42,7	-2,1%
↔	Regno Unito	30,9	30,2	-2,2%
↑	Ucraina	23,1	25,4	9,8%
↑	Turchia	22,2	25	12,3%
↓	Germania	24,4	24,9	1,9%
↔	Messico	21,4	22,6	5,9%
	Mondo	904	922	1,9%

Fonte: Elaborazione su dati UNWTO 2008

Tab.2- Top Ten destinazioni per entrate turistiche (in miliardi di dollari)

Posizione		2007 (miliardi)	2008 (miliardi)	08/07 %
↔	Stati Uniti	96,7	110,1	13,8
↔	Spagna	57,6	61,6	-0,4
↔	Francia	54,3	55,6	-4,6
↔	Italia	42,7	45,7	-0,1
↑	Cina	37,2	40,8	9,7
↑	Germania	36,0	40,0	3,5
↓	Regno Unito	38,6	36	1,6
↔	Australia	22,3	24,7	10,3
↑	Turchia	18,5	22,0	18,7
↓	Austria	18,9	21,8	7,5
	Mondo	857	944	1,8

Fonte: Elaborazione su dati UNWTO 2008

Le performance delle singole “destinazioni-Paese”, evidenziano che i Paesi che hanno manifestato gli incrementi maggiori per numero di arrivi sono l’Arabia Saudita (+28%), L’Egitto (+15,9%), l’Indonesia (+13,2%), la Turchia (+12,3%), la Bulgaria (+12,2%).

Analizzando il contesto internazionale dal punto di vista dei mercati di origine dei flussi turistici, si può evidenziare che, accanto ai mercati tradizionali dell’Europa, dell’America e dell’Asia e Pacifico, si affiancano, ormai da alcuni anni, molti Paesi emergenti come Brasile, Russia, India e Cina, (BRIC). Questi Paesi grazie all’aumento del reddito disponibile generano una percentuale sempre crescente di flussi *outgoing* diventando mercati d’origine importanti per il turismo internazionale e per l’Europa.

L’Europa occupa attualmente la posizione di leadership anche per ciò che concerne i flussi di turisti in uscita, avendo generato nel 2008 il 55% degli arrivi dei turisti internazionali. Seguono l’Asia e Pacifico con il 20% e l’America con il 16%.

La crisi in atto si è ripercossa anche a livello di mercati d’origine e la crescita è stata anche in questo caso molto contenuta.

Il Medio Oriente, tuttavia, ha ottenuto una buona *performance* con un tasso di crescita annuale del 16%. Anche i flussi in partenza dall’Africa sono rimasti abbastanza positivi con un tasso di crescita del 5%.

Gli effetti della crisi economica si sono manifestati in misura ancora maggiore nel corso del 2009. Gli arrivi internazionali hanno segnalato un calo del 4,2%, fermando il dato a quota 880 milioni. Anche i ricavi turistici sono diminuiti tra il 2008 e il 2009 del 5,7%, registrando entrate per 850 miliardi di dollari.

Altre crisi economiche hanno colpito l'industria del turismo in passato, questa crisi, tuttavia, mostra alcune caratteristiche specifiche che la rendono diversa dalle precedenti. La crisi attuale è di natura più globale, colpisce, infatti, sia le destinazioni emergenti sia quelle ormai mature. Inoltre, secondo gli esperti, gli effetti della crisi sono destinati a perdurare nel tempo, ci si aspetta una contrazione rapida e solo una lenta ripresa per il futuro.

A conclusione dell'analisi sull'evoluzione delle tendenze del contesto competitivo internazionale si ritiene opportuno approfondire i diversi fattori esogeni⁷ che hanno influito sullo sviluppo del settore (Liguori, 2008):

- la recessione economica in atto ha portato alla sostanziale riduzione dei flussi in partenza dagli USA, con conseguenze importanti sul turismo, data la rilevanza degli Stati Uniti come mercato d'origine dei flussi.
- Il massiccio impiego delle nuove tecnologie dell'informazione e della comunicazione, legate in particolare all'uso di internet ha influenzato notevolmente il turismo sia dal lato della domanda che da quello dell'offerta.
- La profonda e sostanziale riorganizzazione del trasporto aereo internazionale, la crisi di compagnie di bandiera e le numerose operazioni di fusione e acquisizione tra compagnie continentali e intercontinentali, lo sviluppo del trasporto low cost, l'impiego e la diffusione della rete come strumento di prenotazione e acquisto della biglietteria aerea, hanno radicalmente modificato le abitudini di milioni

⁷ In proposito si veda Liguori M., "La competitività dell'Italia e delle Regioni Italiane" in E.B.I.T., *Turismo, Prospettive e Governance*, 2008, pp 43-44.

di turisti, rendendo accessibili destinazioni un tempo esclusivo e favorendo al contempo la diffusione degli short-break.

- Lo sviluppo del turismo culturale e la scoperta di nuovi prodotti turistici, il rinnovamento di prodotti tradizionali come il turismo crocieristico, le crescenti difficoltà di prodotti maturi come il balneare che sono di rilievo strategico per numerose destinazioni tra cui l'Italia.
- L'ingresso di paesi *new comers* come protagonisti nella competizione internazionale.
- L'aggressività delle strategie di marketing turistico e territoriale dei principali *competitors* dell'area Mediterraneo sia per il prodotto mare che per quello culturale (Spagna, Turchia, Grecia, Croazia, Egitto, Tunisia).
- L'incremento esponenziale del numero di turisti (*leisure e business*) da e per il cosiddetto "BRIC" (Brasile, Russia, India e Cina) come conseguenza della eccezionale crescita registrata da tali paesi nell'ultimo decennio.
- Il profondo mutamento dei comportamenti di acquisto e di consumo dei turisti sempre più evoluti, attenti alla qualità della vita, alla sostenibilità ambientale, alla dimensione esperienziale e culturale dei viaggi; propensi a cumulare più *short breaks* durante l'anno che un'unica vacanza con una lunga permanenza presso la medesima destinazione.
- Il rallentamento dell'economia in numerosi paesi europei aggravata dalla crisi finanziaria globale nata negli USA, da cui deriva, soprattutto negli ultimi anni, una riduzione del reddito disponibile per le spese turistiche per una consistente percentuale di popolazione.
- Le condizioni e i cambiamenti climatici che con i loro effetti influenzano una molteplicità di variabili chiave per l'attrattività e la sostenibilità delle destinazioni turistiche.

L'impatto di tali fenomeni sul turismo ha così contribuito a delineare nuovi scenari per l'Europa, per il Mediterraneo, per l'Italia e le sue singole destinazioni. Sostanziali cambiamenti sono avvenuti sia dal lato della domanda (i viaggiatori provengono da grandi paesi che fino a qualche anno fa erano esclusi dal mercato turistico mondiale, mutano le aspettative e i comportamenti dei turisti, cresce la richiesta di viaggi che trasmettano nuove emozioni), sia del lato dell'offerta con la nascita di nuove destinazioni, determinando nuovo modo di fare di turismo, trasformandosi da semplice svago a vero e proprio bisogno.

1.2. Il contesto europeo

L'Europa è la principale destinazione turistica mondiale con una quota di mercato di oltre il 50% degli arrivi e delle entrate valutarie. I flussi turistici nel 2008 hanno raggiunto i 488 milioni di persone che hanno generato entrate per 322 miliardi di euro.

La rilevanza economica del fenomeno turistico nell'ambito dell'Unione Europea è misurabile attraverso il contributo del settore al Pil e all'occupazione. Il turismo genera più del 5% del PIL⁸, tale contributo risulta ancora più elevato se si considerano anche i settori attinenti, contribuendo a più del 10% del PIL dell'Unione europea e circa il 12% dell'occupazione totale.

A tale proposito, se si osserva la tendenza degli ultimi dieci anni, il tasso di creazione di occupazione nel settore del turismo dell'Unione europea è stato sopra la media osservata per l'economia UE nel suo insieme.

Il turismo è, dunque, un'attività economica in grado di creare crescita e occupazione, contribuendo nello stesso tempo allo sviluppo e all'integrazione economica e sociale. Nell'ambito UE il turismo rappresenta la terza maggiore attività socio-economica dopo il settore del commercio e della distribuzione e

⁸ Fonte dei dati: *Study on the Competitiveness of the EU tourism industry, settembre 2009.*

quello della costruzione, con circa 1,8 milioni di imprese, principalmente piccole e medie imprese, che occupano il 5,2% circa della manodopera totale.⁹

Nel corso del 2008 l'Europa ha evidenziato una stagnazione degli arrivi internazionali registrando una crescita pari allo 0,1%. Nonostante il Vecchio Continente sia ormai una destinazione matura e mostri già da qualche tempo una crescita rallentata, la stasi del 2008 appare come un risultato estremamente negativo se paragonato al +4,2% del 2007.

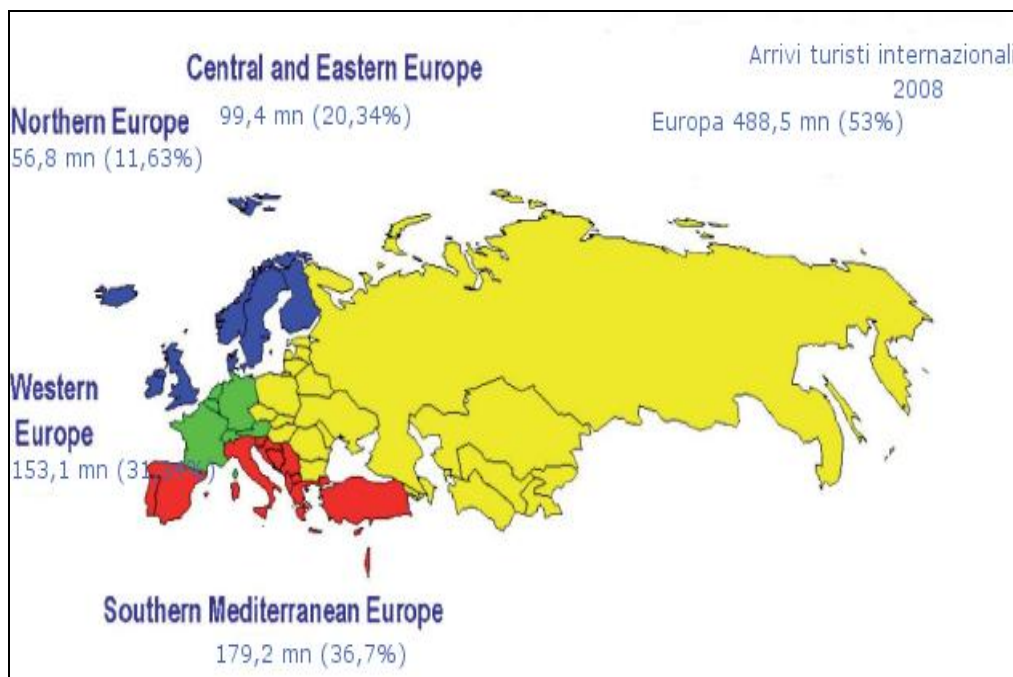
Analizzando le performance a livello di macro aree¹⁰, i paesi dell'area nord occidentale segnano riduzioni consistenti, intorno ai due punti percentuali, mentre l'Europa mediterranea, all'interno della quale si colloca il nostro paese, registra un andamento leggermente migliore (+0,6%). L'area centro-orientale registra la performance migliore con un incremento degli arrivi pari al +2,5% su tutto l'anno 2008.

L'Europa meridionale è la destinazione preferita dai turisti internazionali attirando il 36,7% dei flussi diretti in Europa, leggermente inferiore la quota dell'Area Occidentale (31,3%), seguono l'Area Centro-Orientale con il 20,34% degli arrivi e l'Europa settentrionale con l'11,63% (fig.5).

⁹ Approssimativamente 9,7 milioni di posti di lavoro, di cui una quota considerevole è rappresentata da donne e giovani.

¹⁰ Le macroaree in cui è suddivisa l'Europa sono: Europa del Nord (Danimarca, Finlandia, Islanda, Irlanda, Norvegia, Svezia, Regno Unito), Europa Occidentale (Austria, Belgio, Francia, Germania, Liechtenstein, Lussemburgo, Paesi Bassi e Svizzera), Europa Centro-orientale (Bulgaria, Rep. Ceca, Estonia, Ungheria, Lettonia, Lituania, Polonia, Romania, Fed. Russa, Slovacchia, Ucraina), Europa del sud/mediterranea (Albania, Andorra, Bosnia Erzegovina, Croazia, Cipro, Ex Repubblica della Macedonia, Grecia, Israele, Italia, Malta, Montenegro, Portogallo, San Marino, Serbia, Slovenia, Spagna, Turchia).

Fig. 5 – Quote di mercato per numero di arrivi delle macro-aree europee



Fonte: ns. elaborazione su dati UNWTO 2008

Prendendo in considerazione l'evoluzione dei flussi turistici nel periodo 2000-2008, si può rilevare in primo luogo che l'area centro-orientale del vecchio continente è cresciuta più delle altre (+43%). La crescita dell'area mediterranea si è attestata sulla media continentale (+22%), mentre, l'area occidentale, principale meta turistica al mondo prima del 2000, ha fatto registrare una crescita nel periodo 2000-2008 del +9,6%.

L'area nordeuropea ha conseguito, nel periodo considerato, un incremento maggiore dell'area mediterranea (+30,47%), qualificandosi però come la meno attrattiva in termini di arrivi effettivi di turisti stranieri nell'anno 2008.

A livello di singole destinazioni, la Spagna è cresciuta in maniera significativa dal 2000 al 2008 (+23%) a dispetto di Italia e Francia, entrambi +6%. Ucraina, Turchia, Bulgaria e Croazia hanno registrato tassi di crescita nel

periodo considerato nettamente superiori alla media e a quelli dei paesi leader, conquistando così quote di mercato sempre più significative.

Destinazioni mature quali Grecia (+33%), Regno Unito (+30,06%) e Germania (+31%) hanno evidenziato tassi di crescita nettamente superiori a quelli dell'Italia. Polonia e Portogallo sono le destinazioni che hanno registrato le *performance* peggiori negli anni considerati.

La crisi economica e finanziaria ha avuto effetti non trascurabili sulla domanda di prestazioni turistiche, i cittadini europei hanno adattato le loro abitudini alle circostanze, preferendo destinazioni più vicine e riducendo la durata dei loro soggiorni o le spese sul posto¹¹.

Nonostante la situazione di crisi generale, alcuni Paesi dell'Area Mediterranea hanno evidenziato dei tassi di crescita consistenti, come la Turchia (+13%), Israele (+24%), e alcuni Paesi dell'Area occidentale dei Balcani.

Tab. 3-Top Ten delle destinazioni in Europa per numero di arrivi

Posizione		2007 (milioni)	2008 (milioni)	08/07 %
↔	Francia	81,9	79,3	-3,2%
↔	Spagna	58,7	57,3	-2,3%
↔	Italia	43,7	42,7	-2,1%
↔	Regno Unito	30,9	30,2	-2,2%
↑	Ucraina	23,1	25,4	+9,8%
↑	Turchia	22,2	25	+12,3%
↓	Germania	24,4	24,9	+1,9%
↔	Austria	20,8	21,9	+5,6%
↔	Fed.Russa	20,6	21,6	+4,7%
↔	Grecia	16,2	15,9	-1,4%
	Europa	487,3	487,9	+0,1%

Fonte: ns. elaborazione su dati UNWTO 2008

¹¹ In Europa, nel complesso, gli arrivi di turisti internazionali sono diminuiti del 5,6% circa nel 2009, ma alcune regioni, soprattutto dell'Europa orientale o settentrionale, hanno registrato un calo fino all'8% (Barometro OMT del turismo mondiale, volume 8, gennaio 2010).

L'importanza strategica del turismo in termini di sviluppo socio-economico dell'Europa, ha portato le istituzioni comunitarie a programmare ed attuare una nuova politica turistica al fine di migliorare la competitività del settore e creare nuovi posti di lavoro.

Nel corso degli anni l'Unione europea ha posto le basi di una politica europea del turismo puntando sui fattori che ne determinano la competitività e tenendo conto, allo stesso tempo, degli imperativi dello sviluppo sostenibile.

L'entrata in vigore del trattato di Lisbona riconosce formalmente l'importanza del turismo a livello europeo. L'Unione europea ha acquisito la competenza di sostenere, coordinare e completare l'intervento degli Stati membri in questo settore permettendo di stabilire un quadro d'azione coerente.

Gli elementi prioritari della nuova politica europea del turismo sono: qualità dei servizi turistici, diversificazione delle attività e dei prodotti, sviluppo del turismo sostenibile, promozione dell'Europa come destinazione unica.

L'industria del turismo deve far fronte a una concorrenza mondiale sempre più vivace, con i paesi emergenti o in via di sviluppo che attirano un numero crescente di turisti. Per affrontare questa concorrenza, l'Europa deve proporre un'offerta turistica sostenibile e di qualità valorizzando i propri punti di forza, in particolare la diversità dei suoi paesaggi e del suo straordinario patrimonio culturale. Deve inoltre rafforzare la cooperazione con quei paesi che possono diventare una fonte di visitatori per le destinazioni europee.

Nei paragrafi precedenti si è delineato lo scenario di fondo della ricerca analizzando le dinamiche evolutive del turismo a livello internazionale ed europeo. In tale contesto risultano evidenti le performance non brillanti della destinazione Italia. Ciò è l'immediata conseguenza del ritardo competitivo accumulato dall'Italia sia rispetto ai suoi principali *competitors* nell'area mediterranea e mondiale, sia nei confronti dei paesi *new comers*. Nel giro di un trentennio, il nostro Paese, ha perso la *leadership* mondiale sugli arrivi

internazionali, scivolando tra la quinta e la sesta posizione nel *ranking* mondiale, con previsioni di ulteriori perdite nei prossimi anni.

Alla luce di queste considerazioni introduttive, il paragrafo successivo sarà dedicato a mostrare il quadro dell'evoluzione della domanda turistica nel contesto nazionale, al fine di completare il quadro di riferimento in cui si muove l'intera ricerca.

1.3 Il contesto nazionale

Dallo scenario internazionale appena descritto è evidente che l'Italia mostra, ormai da qualche anno, una perdita di competitività rispetto ai suoi diretti concorrenti e alle altre destinazioni mondiali. Nel periodo di forte crescita mondiale degli arrivi e delle presenze internazionali, l'Italia ha perso terreno, crescendo nettamente meno rispetto ai suoi diretti *competitor* e, attualmente, soffre maggiormente la crisi economica internazionale in atto.

L'Italia è sempre stata una destinazione *leader* nel panorama turistico internazionale come meta di viaggiatori stranieri.¹² Dalla metà degli anni Novanta e, in particolare, dal 2001 in poi, il nostro paese ha registrato performance inferiori, sia nei confronti di alcuni concorrenti storici, sia rispetto a quei paesi che, solo recentemente, si sono introdotti nella competizione internazionale come destinazioni capaci di attrarre flussi turistici consistenti.

Nei paesi diretti concorrenti dell'Italia, con storia e livello di sviluppo simili, l'importanza complessiva del settore turistico ha continuato a crescere, la Spagna ad esempio ha incrementato la sua quota di mercato in termini di arrivi dal 7,9% del 1995 all'11,74% del 2008, mentre l'Italia è passata dal 7,04% all'8,75% nello stesso periodo. Sia la Spagna sia la Francia hanno inoltre guadagnato quote nel tempo: la Francia ha conquistato la leadership

¹²E' sufficiente ricordare che la parola "turismo" ha origine dai Grand Tour settecenteschi in cui l'Italia era considerata come destinazione fondamentale.

mondiale dal 1990 e la Spagna, a partire dal 2002, si è attestata al secondo posto.

L'Italia, invece, ha perso la posizione di leader, classificandosi nel 2008 al quinto posto come importanza per arrivi internazionali.

Questo risultato, tuttavia, non corrisponde all'indiscutibile propensione turistica del nostro Paese che offre una ricchezza unica dovuta alle molteplici alternative disponibili. L'Italia, tuttavia, soffre la mancanza di un'efficace presa di coscienza del ruolo cruciale del settore e del suo valore economico e imprenditoriale. Manca la capacità di valorizzare l'immenso potenziale turistico nazionale attraverso una visione strategica d'insieme e di condizioni territoriali che siano orientate complessivamente all'attività turistica. Ciò ha comportato il ritardo, ormai decennale, dell'Italia rispetto ai suoi concorrenti in termini di competitività del comparto turistico.

Appaiono interessanti al riguardo alcune considerazioni formulate dal Touring Club Italiano sulla competitività della destinazione Italia che evidenzia quali fattori di criticità del nostro Sistema-Paese:

- una *governance*, frammentata in un complesso di strutture e di soggetti, incapace di ottimizzare le risorse finanziarie, complessivamente pari a quelle dei nostri diretti concorrenti, in politiche di promozione efficaci ed efficienti rivolte ai mercati potenziali;
- un sistema di offerta turistica dominato da attori di piccole e piccolissime dimensioni, spesso estranei a logiche di *networking* e di innovazioni di prodotto e di processo, nonché tagliati fuori dai circuiti internazionali di distribuzione e promozione sia tradizionali che innovativi;
- un'eccessiva varietà e variabilità della qualità dei servizi erogati da tutti gli attori dell'offerta turistica e più in generale dai territori. Ciò comporta una qualità complessiva dell'offerta italiana inferiore alle

altre destinazioni internazionali, con un conseguente peggioramento del rapporto qualità/prezzo dell'industria turistica italiana;

- una concorrenza internazionale sempre più manageriale e creativa, capace di conseguire un chiaro posizionamento della propria offerta e di recuperare i *gap* infrastrutturali, organizzativi e commerciali dei decenni precedenti con una velocità sconosciuta alla maggioranza dei territori del nostro paese.

Il turismo italiano ha bisogno dunque di essere ripensato e riorganizzato affinché torni ad occupare il ruolo da protagonista che ci si aspetta da un Paese come l'Italia. Interessante è il piano di azione proposto da Ambrosetti nel Rapporto 2008 – Sistema turismo Italia, in cui si ipotizzano cinque tappe per tornare ad essere vincenti.

Nel rapporto si evidenzia la necessità di promuovere il Paese in maniera unitaria ottimizzando le risorse e recuperando l'efficacia e l'efficienza dell'attività promozionale; aumentare l'offerta turistica italiana con nuovi prodotti turistici coerenti con le risorse del territorio che riescano a ridurre la stagionalità e fidelizzano la clientela; diffondere qualità e standard internazionali aumentando la penetrazione delle catene alberghiere e favorendo l'ammodernamento del comparto; aumentare l'orgoglio di appartenenza al settore facendo diventare il turismo come il settore professionale più attrattivo del Paese, rendendo l'occupazione nel settore più sostenibile e destagionalizzata; valorizzare i turisti diffondendo tra la popolazione la cultura pro-turismo.

E' inoltre evidenziata la necessità di dotarsi di un sistema di rilevazione statistica che sia in grado di fornire, in tempi brevi, dati completi, uniformi e certi, necessari alle scelte strategiche degli operatori che operano nel turismo e delle istituzioni che hanno responsabilità di governo del settore.

In questo quadro generale di problematiche, interventi e azioni necessarie, bisogna però riconoscere una significativa disparità tra le varie Regioni

Italiane. Alcune di esse, infatti, registrano tassi di crescita consistenti mentre altre accusano ulteriormente la distanza dai leader nazionali.

La differenza di *performance* delle varie regioni, come si avrà modo di evidenziare anche nelle pagine seguenti e nei risultati della ricerca, è fondamentalmente indipendente dall'attrattività potenziale delle risorse culturali, ambientali e turistiche dei vari territori, a testimonianza di una differente capacità di risposta alle sfide del nuovo scenario competitivo da parte dell'imprenditoria e dei *policy maker* locali.

Nel corso del 2008 l'Italia, pur collocandosi al quinto posto nel ranking mondiale per gli arrivi, ha sofferto la crisi economica mondiale con una rilevante riduzione dei flussi di turisti, sia italiani sia stranieri, che si rafforza nel corso dell'anno passando da una riduzione del 2.6% del primo trimestre al -5,1% nell'ultimo¹³.

La flessione (-2,1%) degli arrivi dei turisti internazionali desta preoccupazione soprattutto se confrontata con il dato in controtendenza, relativo all'Europa mediterranea (+0,6 %). La crisi sembra evidenziare le fragilità della nostra penisola, sia in termini di rapporto qualità-prezzo dei servizi, sia relativamente alla varietà, qualità e capacità d'innovazione dell'offerta a livello di destinazioni turistiche.

Anche a livello di entrate l'ultimo trimestre si rileva particolarmente disastroso per l'Italia con una riduzione del 7,7%.

Anche i dati Istat¹⁴ per il 2008 segnalano una riduzione dei flussi turistici.

Le presenze turistiche registrate nelle strutture ricettive italiane sono state circa 374 milioni contro i 376,4 milioni dell'anno precedente mostrando una sostanziale fase di stagnazione per il settore (-0,8%). Per quanto riguarda gli arrivi nel 2008 hanno raggiunto i 95,5 milioni contro i 96,2 milioni del 2007 (-0,6%).

¹³ Fonte Dati UNWTO, *Barometer*, vol.7, n°2, Giugno 2009.

¹⁴ Le rilevazioni ISTAT sugli arrivi in Italia differiscono, in termini percentuali, rispetto ai dati UNWTO.

La flessione delle presenze nel corso del 2008 riguarda in modo particolare la componente straniera (-1,02%) e in misura minore la componente italiana (-0,6%). In termini di arrivi la situazione è sostanzialmente simile, con una netta flessione della componente straniera (-2,5%) e una leggera crescita degli arrivi domestici.

La componente straniera della domanda ha ampliato la propria quota nel tempo, rappresentando con 161.8 milioni di presenze il 43% delle presenze in Italia nel 2008. La componente nazionale fa registrare il 57% delle presenze turistiche negli esercizi ricettivi (sia alberghieri sia extralberghieri) pari a oltre 53,7 milioni di turisti ed un totale di 211,9 milioni di presenze nel corso del 2008.

La permanenza media resta invariata (3,9) rispetto allo scorso anno, comunque in linea con la tendenza, riscontrata negli ultimi anni, di soggiornare per periodi più brevi nonostante si viaggi più spesso.

Tab. 4 Presenze italiani e stranieri per regione e graduatoria delle Regioni Italiane

Presenze italiani	Posizione	REGIONI	Posizione	Presenze stranieri
7.471.502	13	Piemonte	10	4.086.828
2.043.497	18	Valle D'Aosta	17	1.069.843
13.474.140	5	Lombardia	5	14.829.365
19.483.287	4	Trentino	2	23.089.172
24.930.656	2	Veneto	1	35.676.417
5.106.266	16	Friuli-Venezia Giulia	12	3.772.661
9.984.799	9	Liguria	9	4.145.715
29.322.847	1	Emilia-Romagna	6	9.038.550
21.528.480	3	Toscana	4	19.733.476
3.926.657	17	Umbria	13	2.084.669
9.806.766	10	Marche	15	1.671.596
10.557.835	7	Lazio	3	21.118.292
6.539.833	15	Abruzzo	18	1.020.643
609.550	20	Molise	20	49.655
11.114.279	6	Campania	7	7.608.107
10.469.631	8	Puglia	14	1.713.745
1.681.069	19	Basilicata	19	181.304
7.024.711	14	Calabria	16	1.468.628
8.381.095	12	Sicilia	8	5.557.224
8.412.378	11	Sardegna	11	3.881.544

Fonte: ns. elaborazione su dati ISTAT - Capacità e movimento negli esercizi ricettivi, 2008

Un segnale evidente della crisi in atto proviene dalla suddivisione delle presenze all'interno delle diverse tipologie di esercizi ricettivi. La diminuzione riguarda in modo particolare le strutture alberghiere (-1,04%) mentre la flessione registrata per le strutture extra - alberghiere è più contenuta, intorno al - 0,7%.

Emergono, quindi, due comportamenti diversi per la componente nazionale e straniera. Gli stranieri arrivano meno numerosi in Italia, ma restano tendenzialmente più a lungo, probabilmente per ammortizzare i crescenti costi dello spostamento dovuti al caro petrolio. Inoltre scelgono maggiormente strutture ricettive extra - alberghiere, che permettono di contenere il costo complessivo della vacanza, o, a parità di spesa, di allungare leggermente il periodo di soggiorno.

Gli italiani, al contrario, sembrano aver reagito alla diminuzione, reale o attesa, del proprio reddito disponibile spostandosi all'interno della penisola e per periodi più brevi, scegliendo anch'essi, al pari degli stranieri, strutture ricettive extra - alberghiere, ma anche sfruttando di più la rete delle conoscenze e dei familiari per abbattere i costi del soggiorno senza dover necessariamente rinunciare alla vacanza¹⁵.

Per ciò che riguarda le preferenze sui "prodotti" turistici si manifestano sostanziali differenze nelle scelte tra italiani e stranieri.

Con riferimento al totale della domanda (misurata in termini di presenza presso le strutture ricettive), il prodotto "mare" si colloca al primo posto, assorbendo circa un terzo del turismo in Italia; al secondo posto le "città" con una quota del 29%; la "montagna" segue distanziata al quarto posto con un 16% della domanda complessiva. Gli altri prodotti (lacuale 8%, campagna/collina/termale, ecc.) seguono con quote intorno al 5%.

¹⁵ Fonte: UNIONCAMERE, "Competitività del sistema turistico Italiano", 2010.

Tab. 5 – Quota di mercato per tipo di località.

	Totale		%
	Arrivi	Presenze	
LOCALITA' DI INTERESSE TURISTICO			
Città d'interesse storico e artistico	33.025.977	91.097.369	29.3
Località montane	9.797.222	48.912.171	15.7
Località lacuali	5.453.991	24.985.578	8.0
Località marine	21.443.994	117.986.626	38.0
Località termali	3.433.510	13.631.656	4.3
Località collinari e d'interesse vario	3.881.559	13.790.523	4.4
Totale	77.036.253	310.403.923	100

Fonte: ns. elaborazione su dati ISTAT

Nel mercato estero, tuttavia, il prodotto preferito è rappresentato dalle “città” che con il 38% del totale della domanda estera supera notevolmente il “balneare”; più debole è il turismo montano che assorbe una quota di domanda analoga a quello “lacuale”. Nel mercato interno, invece, prevale il balneare, poco meno del 40% dei turisti italiani s’indirizza verso questo prodotto; al contrario è molto più limitata l’attenzione verso le “città” che hanno un peso non molto superiore a quello del “montano”. Del tutto marginali sono il turismo lacuale e campagna e collina.

Completando l’analisi con i dati sulla spesa turistica¹⁶ e in particolare sulla bilancia turistica possiamo rilevare che a fronte di una spesa verso l’estero di € 20.922.000, la spesa degli stranieri in Italia si attesta nel 2008 a 31.090.000 con una bilancia turistica in attivo per €10.168.000. Queste cifre testimoniano la rilevanza di un settore che, in questo momento di crisi economica e di recessione, può rappresentare sicuramente una fonte “vitale” di entrate valutarie. E’ importante quindi che le istituzioni, a tutti i livelli amministrativi, prendano coscienza e sappiano coordinare e gestire le immense risorse turistiche del nostro Paese per renderlo maggiormente attrattivo per i flussi turistici.

¹⁶ Fonte: Dati UIC, Turismo Internazionale.

A questo punto, è necessario analizzare le nuove tendenze nel modo di fare turismo.

Lo scenario turistico della domanda, infatti, non cambia solo perché si aggiungono nuovi mercati, ma anche perché si modificano i comportamenti nel modo di fare vacanza. Il turista moderno ricerca nella vacanza la possibilità di vivere un'esperienza unica, lontana dalla banale categorizzazione "balneare" o "culturale", che lo allontani dal quotidiano e lo porti a scoprire nuove sensazioni.

Il turista di oggi si dirige sempre di più verso i luoghi autentici, i borghi, le tradizioni, le comunità, tutte quelle risorse che nella letteratura sono definite come risorse immateriali o heritage tourism. Si tratta di un patrimonio di risorse culturali ben più ampio di quello che in genere è promosso e commercializzato come prodotto turistico - culturale.

Lo scopo del viaggio è la ricerca di esperienze e di emozioni che possono offrire le persone, le comunità più che i monumenti in sé. Le sagre, le feste, gli eventi legati alle tradizioni e anche alla religiosità nel nostro Paese, tutti quegli elementi tangibili e intangibili di carattere culturale, linguistico, storico, di modi di vita, di tradizioni enogastronomiche e artigianali, che una data comunità ha ricevuto dal suo passato rappresentano un patrimonio in grado di motivare la domanda turistica più attenta e colta.

Il turismo non deve limitarsi a vendere camere d'albergo o a erogare servizi, il turismo deve saper comunicare e vendere i territori in tutte le loro complessità e peculiarità.

Le nuove forme di turismo tematico - rurale, naturalistico, spirituale, enogastronomico, sportivo, escursionistico, di avventura – rappresentano una grande opportunità di diversificazione e di sviluppo economico, in virtù del mutato rapporto dei turisti con la natura e la riscoperta delle radici del territorio.

CAPITOLO 2

COMPETITIVITA' E SOSTENIBILITA' DELLE DESTINAZIONI TURISTICHE

SOMMARIO: 2.1. La destinazione turistica competitiva. - 2.2. Il modello Crouch e Ritchie. – 2.3. Altri modelli e indicatori sulla competitività delle destinazioni turistiche. – 2.4. La sostenibilità come fattore chiave per la competitività.

2.1. La destinazione turistica competitiva.

L'obiettivo principale del lavoro è misurare e confrontare il livello di competitività raggiunto dalle regioni italiane sotto il profilo turistico raggruppando in cluster le regioni con vantaggi competitivi omogenei ed evidenziando le relative diversità nel modo di valorizzare i fattori d'attrattiva determinando esternalità positive per i territori e le comunità.

Il tema della competitività della destinazione turistica e la ricerca di strumenti e indicatori utili alla sua misurazione sono da tempo oggetto di dibattito nella letteratura turistica.

La crescente concorrenza nazionale e internazionale rende sempre più evidente come il confronto competitivo si giochi innanzitutto a livello di destinazioni e, solo successivamente, a livello di singole imprese (Sainaghi, 2008).

La centralità delle destinazioni nella competizione turistica è riconosciuta da molti autori che le definiscono come soggetti competitivi. (Ritchie e Crouch, 2000; Enright e Newton, 2004).

Efficace è l'affermazione proposta da Ritchie e Crouch: “*il prodotto primario nel turismo è l'esperienza proposta dalle destinazioni. La competizione, pertanto, trova il suo baricentro nelle destinazioni*”.

In definitiva sono le destinazioni a competere sui mercati turistici internazionali proponendo un'offerta armonica di beni immateriali e di fattori strumentali che consentano la più efficiente fruizione di tali beni. I territori, quindi, sono chiamati ad agire in una logica imprenditoriale, cercando di creare condizioni favorevoli per attrarre risorse scarse che producano ricchezza e favoriscano lo sviluppo di quelle esistenti.

La gestione strategica del territorio deve essere finalizzata alla valorizzazione del potenziale di risorse presenti nello specifico sistema locale secondo un'ottica integrata di interrelazioni solidali tra tutti i suoi fattori e componenti. Ciò implica, necessariamente, anche il coinvolgimento diretto degli operatori turistici che non possono più limitarsi a gestire le imprese solo in funzione dell'ottimizzazione dei loro fattori interni di produzione e di strategie di marketing orientate solo alla *mission* aziendale. Il successo delle imprese è strettamente condizionato e interdipendente da quello della destinazione in cui operano.

Nella prospettiva dell'integrazione dell'offerta turistica locale risulta sempre più importante il ruolo dei *policy maker* nei processi di coordinamento delle attività gestite dai singoli operatori turistici.

Il *destination management* è sempre più interessato al tema della competitività dei territori e delle destinazioni, la sua efficacia, infatti, si misura attraverso la capacità di innescare, congiuntamente con le imprese e le organizzazioni pubbliche e private locali, un percorso di sviluppo turistico capace di generare un benessere sostenibile ai residenti lungo la dimensione economica, sociale, politica, culturale ed ecologica. (Crouch e Ritchie, 1999).

Il dibattito nella comunità scientifica si è focalizzato, anche, sul concetto di destinazione turistica, che è stato analizzato secondo una duplice prospettiva.

Alcuni autori hanno studiato la destinazione dalla prospettiva della domanda, ovvero ritengono che sia il turista a definire la destinazione attraverso la scelta degli elementi che andranno a comporre la sua vacanza e la sua esperienza di viaggio. Essi definiscono la destinazione turistica come un set di prodotti, servizi e attrazioni naturali e/o artificiali in grado di attrarre turisti verso uno specifico luogo.

Altri ricercatori hanno definito la destinazione turistica come sistema di offerta connesso con uno specifico territorio. Secondo questa prospettiva la destinazione coincide con il concetto di località, inteso come insieme di prodotti e operatori che assumono una determinata struttura in virtù delle relazioni, più o meno intenzionali, tra i vari *stakeholder*. La struttura più “forte” identifica la destinazione come un vero e proprio sistema nel quale le relazioni tra tutte le parti consentono una gestione comune delle risorse e delle attività. Al contrario la struttura “più debole” configura un sistema d’offerta con relazioni non necessariamente intenzionali tra le diverse parti.

Sebbene ci sia condivisione nel riconoscimento che la destinazione rappresenta un amalgama di prodotti, servizi, elementi naturali ed artificiali, in grado di attrarre un certo numero di visitatori all’interno di un luogo geografico, ciò che differenzia le due prospettive è il “collante” di tale amalgama (Sainaghi, 2008). Per gli studiosi che assumono la prospettiva della domanda, è il turista attraverso la sua esperienza di fruizione a trasformare la destinazione in un prodotto unitario, per coloro che assumono il punto di vista dell’offerta, tale unitarietà è il risultato dell’agire congiunto degli operatori presenti in quel determinato territorio.

Il prodotto “destinazione turistica” è un sistema di offerta complesso composto dai fattori d’attrattiva e dai servizi turistici presenti in un territorio. Il prodotto turistico diventa sempre più articolato in relazione alle aspettative del turista che orienta la sua attenzione verso la destinazione turistica nel suo complesso e, quindi, verso un’offerta integrata piuttosto che verso i singoli elementi che la compongono. La destinazione deve riuscire ad assumere tratti

di unicità e distintività rispetto alle destinazioni concorrenti che saranno il fondamento del suo vantaggio competitivo.

Sebbene in letteratura siano state proposte varie definizioni di competitività e analizzati possibili strumenti di misurazione, la competitività rimane tuttavia un concetto non univocamente e facilmente definibile. Dwyer e Kim (2003) definiscono la competitività come un concetto relativo e multidimensionale, perché alla sua definizione concorre un complesso *range* di fattori. La competitività è sempre relativa poiché suggerisce una logica comparativa, una nazione o un'impresa è competitiva poiché è capace di attrarre le preferenze di alcuni target di clientela più di quanto non siano in grado di fare rispettivamente altri territori o altre imprese concorrenti. Nel loro lavoro anche Enright e Newton (2004) evidenziano l'importanza di definire il contesto concorrenziale per una destinazione: *“a given location is competitive or uncompetitive in an industry, not in the abstract, but against relevant competing location”*.

Le definizioni di competitività in letteratura definiscono tale concetto sia in un'ottica macro (competitività del sistema regionale, nazionale, ecc) che in un'ottica micro (competitività delle imprese).

A livello macro può essere definita competitiva, in via generale, una regione o nazione¹⁷ con un tasso di crescita nel lungo periodo equiparabile a quello delle nazioni più simili a lei dal punto di vista economico. In questa prospettiva, la competitività è un concetto molto ampio che comprende tutte le variabili sociali, culturali, e economiche che influenzano le prestazioni di una nazione nel mercato internazionale. (Dwyer e Kim, 2003)

A livello di impresa, il termine competizione può assumere due significati diversi : il primo relativo alla competizione all'interno del proprio

¹⁷ I più noti indici di competitività delle nazioni sono quelli calcolati e diffusi dall'IMD (Institute for Management Development) e dal WEF (World Economic Forum). L'IMD pubblica annualmente il *World Competitiveness Yearbook* che è uno dei più importanti rapporti sulla competitività delle nazioni/regioni. Sull'indice di competitività del WEF si veda il paragrafo 2.3.

mercato di riferimento per ottenere i migliori risultati economici e finanziari e il secondo relativo alla competizione per acquisire le migliori risorse per trasformarle in competenze distintive che possano garantire il vantaggio competitivo nel tempo.

Sugli effetti che la competitività deve produrre, a qualsiasi livello di analisi, vi è larga convergenza: essa deve essere in grado di migliorare la qualità della vita dei residenti e il loro reddito (*economic prosperity*). (Dwyer e Kim, 2003, p. 375; Crouch e Ritchie, 1999, p. 138). Stessa convergenza di opinioni tra i vari autori sulla necessità che la competitività sia un obiettivo a lungo termine e orientata ai principi della sostenibilità. Il successo di una destinazione turistica sta nella sua capacità di attrarre turisti, soddisfare le loro esigenze di esperienza turistica realizzando profitto e accrescendo il benessere dei residenti in maniera sostenibile.

Circa le variabili da analizzare per valutare, misurare, gestire o sviluppare la competitività di una destinazione quasi tutti gli autori concordano nel considerare sia dimensioni tradizionalmente sviluppate dal filone della competitività e principalmente riconducibili alle quattro dimensioni del “diamante” di Porter (1990), sia di introdurre alcune variabili specifiche per le peculiarità del prodotto turistico.

Il posizionamento competitivo della destinazione turistica e le variabili che ne influenzano la competitività sono oggetto di numerosi studi in letteratura (Ritchie e Crouch, Enright e Newton, Kozak, Mihalic, Heat, Dwyer e Kim), che considerano principalmente gli elementi dell’offerta, risorse e attrattive della destinazione (clima, cultura e storia, attività, eventi, ricettività, trasporti). Molti di questi autori si soffermano sulla necessità di indicatori sia quantitativi che qualitativi per poter valutare la competitività, su come la competitività della *destination* può essere sostenuta ed accresciuta nel lungo periodo difendendo la posizione di mercato rispetto alle destinazioni concorrenti, sulle cause determinanti, sui fattori ambientali e sulle strategie che assumono un ruolo determinate per la competitività della destinazione.

2.2. Il modello di competitività della destinazione di Crouch e Ritchie

Il dibattito sulla competitività turistica di una destinazione si è sviluppato nell'ambito della letteratura di tourism management a partire dai primi anni '90 grazie al contributo di Crouch e Ritchie.

Le loro ricerche (dal 1999 al 2003) hanno introdotto un modello concettuale sulla competitività delle destinazioni che è basato sulla teoria del vantaggio comparato degli economisti Ricardo e Smith, e sulla teoria del vantaggio competitivo di Porter. Il loro modello è riconosciuto a livello internazionale come il più dettagliato modello di analisi della competitività della destinazione turistica. Essi affermano che una *“destinazione è competitiva se dispone di un appeal capace di generare un'esperienza turistica superiore a quella offerta da territori alternativi e quindi concorrenti”*.

Tuttavia essi sostengono che, allo stesso tempo, una destinazione competitiva è quella che riesce a conciliare maggior successo e maggior benessere per i suoi residenti nella piena applicazione dei principi della sostenibilità.

Secondo il modello di Crouch e Ritchie (fig.6) la competitività di una destinazione turistica è basata su:

- il patrimonio di risorse culturali, storiche, paesaggistiche, sul patrimonio di risorse umane, fattori endogeni che una destinazione ha in dotazione o in eredità (vantaggio comparato);
- la capacità di valorizzare le risorse esistenti o di creare risorse in un'ottica di lungo periodo (vantaggio competitivo).

Mentre i vantaggi comparati derivano dal set di risorse non riproducibili in dotazione della destinazione (clima, paesaggio, flora, fauna,

ecc) il vantaggio competitivo è legato all'abilità del *destination management* di utilizzare queste risorse efficacemente nel lungo periodo.

L'obiettivo principale del management di una destinazione è quello di realizzare e mantenere un vantaggio competitivo sostenibile attraverso un processo continuo di miglioramento e sviluppo strategico della stessa.

Nel loro modello gli autori prendendo spunto dalla teoria di Porter sulla competitività delle nazioni sostengono che la competitività di una destinazione turistica dipende dall'interazione tra le forze del macro-ambiente (economiche, sociali, ambientali ecc.), che non sono controllabili dai *destination manager*, e le forze del micro-ambiente, cioè l'insieme di soggetti interni ed esterni al territorio che influenzano direttamente e che possono essere influenzati da chi governa la destinazione turistica.

Ritchie e Crouch hanno cercato di adattare alla specificità delle destinazioni turistiche il modello aziendale dell'ambiente competitivo.

Il micro-ambiente è parte integrante del sistema turistico poiché riguarda azioni e attività direttamente legate ai membri del sistema. E' fatto di organizzazioni, influenze e forze che risiedono all'interno dell'arena competitiva e delle attività turistiche. Per una destinazione il micro-ambiente competitivo include le imprese turistiche che forniscono i beni e servizi necessari al turista, i fornitori, i consumatori, il canale dell'intermediazione. I fornitori sono l'insieme delle imprese e delle organizzazioni che forniscono all'industria turistica tutti quegli elementi necessari alla creazione del prodotto turistico (lavoro, formazione, energia, ecc...). Fondamentale è anche il canale dell'intermediazione che garantisce il collegamento tra l'industria dell'ospitalità e i turisti consentendo alla destinazione di raggiungere più facilmente il mercato. A livello micro le destinazioni si trovano quindi a competere con le altre ogni qualvolta i singoli elementi dell'ambiente competitivo vengono utilizzati da più di una destinazione. I turisti ad esempio sono potenziali o effettivi clienti di più destinazioni, gli attori dell'intermediazione servono gli operatori turistici di differenti località.

La competizione si sposta poi a livello macro. Il macro-ambiente si compone di una vasta gamma di fenomeni che hanno molteplici effetti su tutte le attività e quindi non sono limitati al settore turistico. Si tratta di quei fenomeni (politici, economici, demografici, ambientali, socio-culturali, tecnologici) che riescono ad influenzare direttamente l'attrattività di determinate destinazioni pur essendo apparentemente scollegate e distanti da esse. Si pensi all'introduzione di nuovi competitor o al rallentamento della spesa da parte di mercati importanti per effetto ad esempio di crisi economiche.

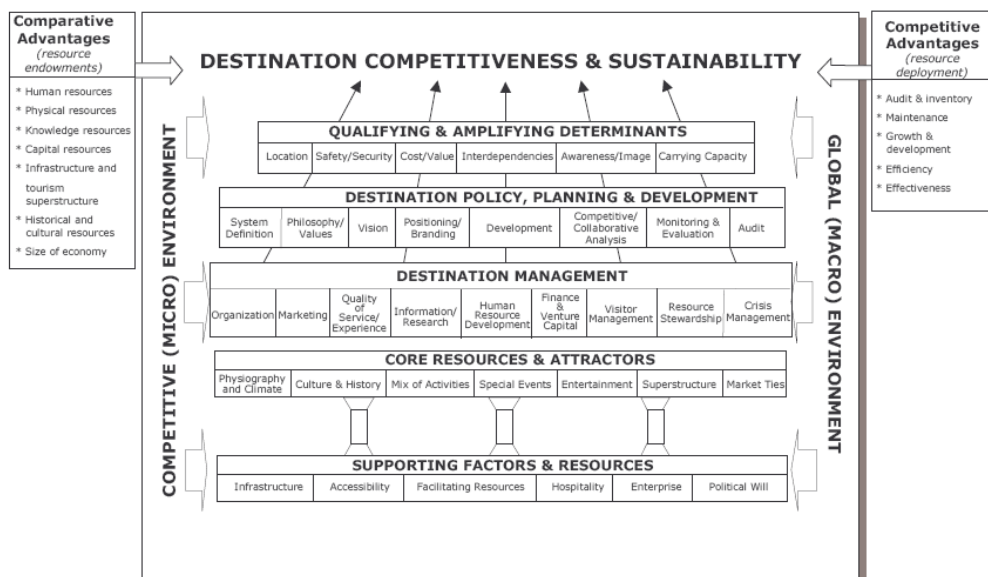
Il modello di Crouch e Ritchie sulla competitività della destinazione si basa su quattro componenti fondamentali:

- le “risorse e gli attrattori core” (fisiche, culturali e storiche, legami di mercato, attività offerte, eventi speciali, intrattenimento e strutture della ricettività e dell'industria turistica), sono i fattori motivanti della visita, definite come “*key motivators*” ovvero le ragioni fondamentali per cui un visitatore sceglie una destinazione rispetto ad un'altra;
- le “risorse ed i fattori di supporto” (infrastrutture, accessibilità, ospitalità, risorse e fattori di *facilitating*, imprenditorialità), rappresentano il fondamento sul quale deve basarsi un'industria turistica di successo e possono contribuire notevolmente alla competitività del sistema turistico locale;
- il “*destination management*”, rappresenta tutte quelle attività di gestione strategica e responsabile delle risorse del territorio, strategie di marketing, finanza e venture capital, organizzazione, sviluppo delle risorse umane, informazione e ricerca, rafforzamento della qualità dei servizi offerti;
- le “*determinanti che qualificano la competitività*” di un territorio (localizzazione, interdipendenze, sicurezza, notorietà/immagine, rapporto costi-valore, capacità di carico). Questi fattori determinano ricadute positive e negative sulla competitività della specifica

destinazione turistica. Quest'ultimo gruppo di fattori, chiamati anche *situational conditioners*, hanno le potenzialità per moderare o ampliare la competitività di una destinazione filtrando le influenze degli altri tre gruppi di fattori.

In estrema sintesi, secondo gli autori, la competitività di una destinazione turistica dipende dall'attento monitoraggio e gestione di queste quattro componenti, alla luce dei cambiamenti del micro-ambiente e del macro-ambiente competitivo.

Fig.6 Modello di competitività della destinazione di Crouch e Ritchie



Fonte: Ritchie e Crouch (1999)

2.3. Altri modelli e indicatori sulla competitività della destinazione turistica

Il modello di Crouch e Ritchie cerca di identificare tutti i principali fattori che influenzano la competitività di una destinazione.

L'ampiezza e complessità del tema ha portato altri autori all'approfondimento del modello di Crouch e Ritchie e all'applicazione empirica dello stesso, in modo da avvalorare l'attendibilità della misurazione della competitività delle destinazioni turistiche, sia nel caso di singole aree di un paese che di intere nazioni.

Enright e Newton (2004) si sono concentrati sulla creazione di un modello di competitività che riuscisse ad integrare i due aspetti della competitività della destinazione in generale e della competitività delle imprese che operano all'interno della destinazione. Essi ritengono che, per una corretta comprensione della competitività della destinazione, non sia possibile escludere quei fattori che influenzano la competitività delle imprese o di altre organizzazioni che in essa operano per la creazione del prodotto turistico.

Gli autori hanno proposto, inoltre, un approccio quantitativo alla misurazione della competitività testando empiricamente il loro modello per la destinazione urbana di Hong Kong.

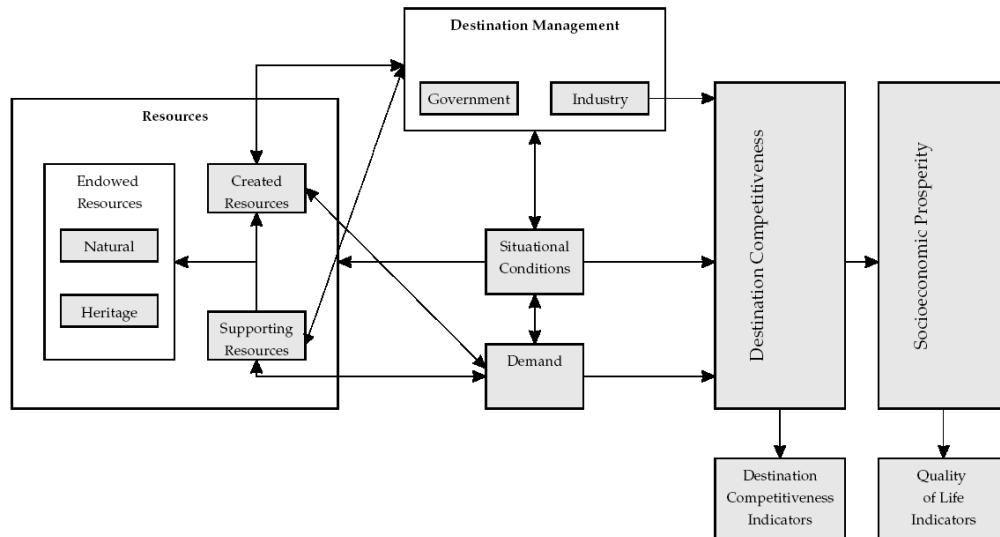
Enright e Newton (2004) utilizzano le sei tipologie di attrattori (*attractors*) già identificate da Crouch e Ritchie (elementi fisici e ambientali, cultura e storia, legami con il mercato, attività, eventi, fattori di supporto), insieme ad altre 31 variabili che sono definite come fattori business (*business factors*) come il livello qualitativo delle strutture, l'organizzazione e la strategia delle imprese, la competizione e cooperazione tra le stesse.

Dwyer e Kim (2003) hanno sollevato alcune critiche al modello di Crouch e Ritchie e hanno proposto un'ottimizzazione del *framework* creando uno schema (fig.7) dove la competitività dipende principalmente dalle seguenti categorie di variabili:

- la *sezione delle risorse o attrattori*, che rappresentano la ragione principale alla base dell'elevato o ridotto *appeal* della destinazione rendendola potenzialmente interessante per certi segmenti di clientela;
- la *sezione del destination management* che permette di valorizzare le risorse presenti e di trasformare la loro potenzialità in prodotti turistici, attraverso azioni di marketing, pianificazione e sviluppo, gestione, sviluppo delle risorse umane, implementazione di piani ambientali;
- le *condizioni contingenti*, si tratta di un insieme piuttosto ampio di fattori tipicamente esogeni, non controllabili da parte del management di una destinazione, ma che possono avere profonde conseguenze sulla sua competitività. Alcuni esempi sono rappresentati dall'ubicazione della destinazione, dalle caratteristiche dell'ambiente competitivo locale e globale, dalla sicurezza e dalla competitività dei prezzi;
- le *condizioni della domanda*. Questa sezione comprende i principali segmenti serviti dalla destinazione, le dimensioni della domanda nazionale, il grado di internazionalizzazione, il livello qualitativo dell'attuale domanda, i trend in atto, le percezioni e le preferenze dei turisti che possono influenzare l'esperienza nella destinazione.

Dwyer e Kim dimostrano quindi che la determinazione della competitività di un territorio è complessa, perché influenzata da numerose condizioni difficilmente controllabili da un *destination manager*. Inoltre, evidenziano come il fondamento della competitività sta nella capacità di trasformare le risorse o gli attrattori della destinazione in prodotti turistici che soddisfino i target di domanda.

Fig. 7 Il modello di Dwyer e Kim



Fonte: Dwyer, L., and C. Kim (2003).

Nel loro modello Dwyer e Kim distinguono le risorse in:

- *risorse endogene* che comprendono i principali attrattori naturali (clima, flora, fauna, panorama e altri *asset* fisici) e culturali (*heritage*), quali monumenti, tradizioni, enogastronomia, storia. Questo tipo di risorse è in grado di attrarre specifici *target* di clientela, ed è per sua natura non riproducibile;
- *risorse artificiali* che comprendono le risorse create dall'uomo, quali le infrastrutture turistiche, gli eventi, le attività che possono essere svolte all'interno di una destinazione (sportive e ricreative), i servizi di entertainment, lo *shopping*.
- *risorse di supporto* che comprendono le infrastrutture generali della destinazione (strade, aeroporti, treni, bus, telecomunicazioni), la qualità dei servizi offerti (che permettono di mantenere nel tempo gli standard qualitativi e di assicurare un'esperienza al cliente finale), l'accessibilità della destinazione (sistema dei trasporti e vincoli all'ingresso del paese), l'ospitalità (grado di accoglienza della popolazione locale), l'esistenza di legami con specifici

mercati (tipicamente legami etnici che si possono generare in seguito a fenomeni di emigrazione o grazie alla presenza di affinità linguistiche, culturali, religiose, ecc.).

In parte ispirato al modello di Crouch e Ritchie anche l'indice di competitività dell'industria dei viaggi e del turismo, "*The Travel & Tourism Competitiveness Index*", sviluppato dal *World Economic Forum* (WEF) nel 2007 ed ampliato nel 2008 (fig.8). Tale indice occupa sicuramente un posto di rilievo nel panorama dei modelli di analisi della competitività turistica.

Si tratta di un complesso e sintetico indicatore messo a punto con la collaborazione di una molteplicità di enti in vari paesi, con l'obiettivo di fornire uno strumento in grado di valutare i fattori e le *policy* che consentono uno sviluppo competitivo del turismo migliorando l'attrattiva turistica di un Paese.

Attraverso questo indicatore si valuta la competitività di una destinazione calcolando un indice complessivo che deriva dalla combinazione di una serie di parametri raggruppati in tre macroindicatori tematici:

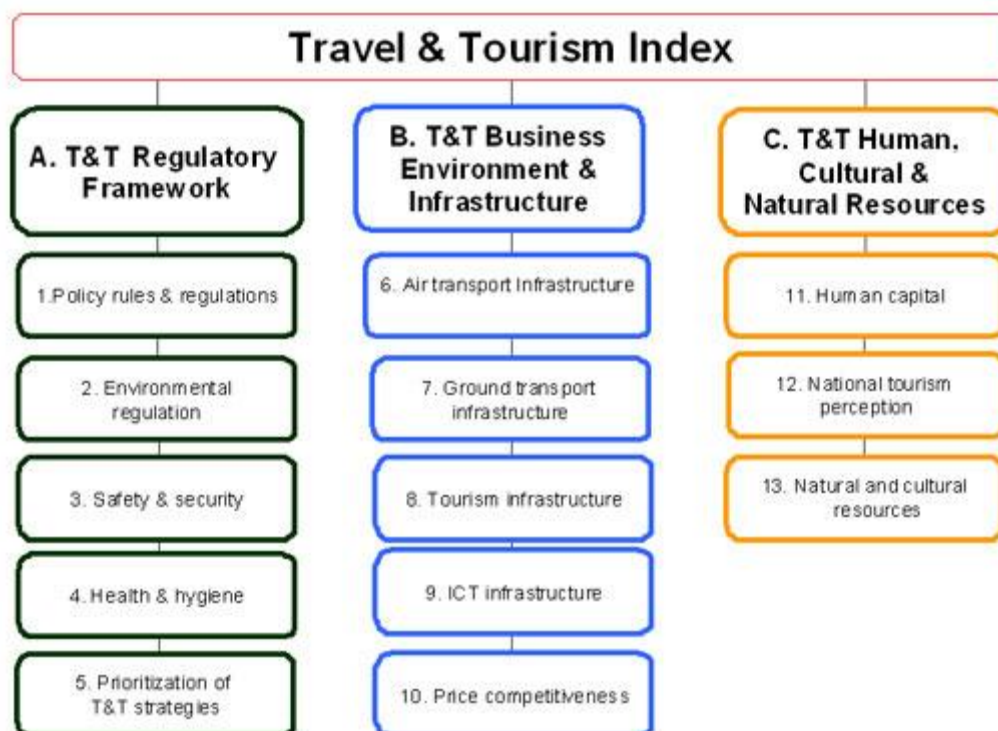
- il quadro normativo e regolamentare. Questo macroindicatore tenta di valutare l'adeguatezza della legislazione per lo sviluppo turistico e l'importanza del settore nella programmazione di governo;
- l'ambiente imprenditoriale e le infrastrutture del turismo. Esprime la vitalità delle imprese del settore e la competitività della dotazione di infrastrutture turistiche, economiche e tecnologiche di un paese e la relativa competitività dei prezzi dell'offerta turistica;
- le risorse umane, culturali e naturali. Si tratta di indicatori volti a misurare le varie componenti della competitività di un territorio quali le risorse culturali, l'apertura verso i turisti stranieri, ecc.

Ciascuno di questi fattori è a sua volta sintesi di ulteriori indici volti a misurare i seguenti “pilastri” della competitività:

- 1. Policy rules and regulations*
- 2. Environmental sustainability*
- 3. Safety and security*
- 4. Health and hygiene*
- 5. Prioritization of Travel & Tourism*
- 6. Air transport infrastructure*
- 7. Ground transport infrastructure*
- 8. Tourism infrastructure*
- 9. ICT infrastructure*
- 10. Price competitiveness in the T&T industry*
- 11. Human resources*
- 12. Affinity for Travel & Tourism*
- 13. Natural resources*
- 14. Cultural resources.*

L'indice complessivo è calcolato sulla base dei tre sottoindici precedentemente illustrati.

Fig.8 The Travel & Tourism competitiveness Index

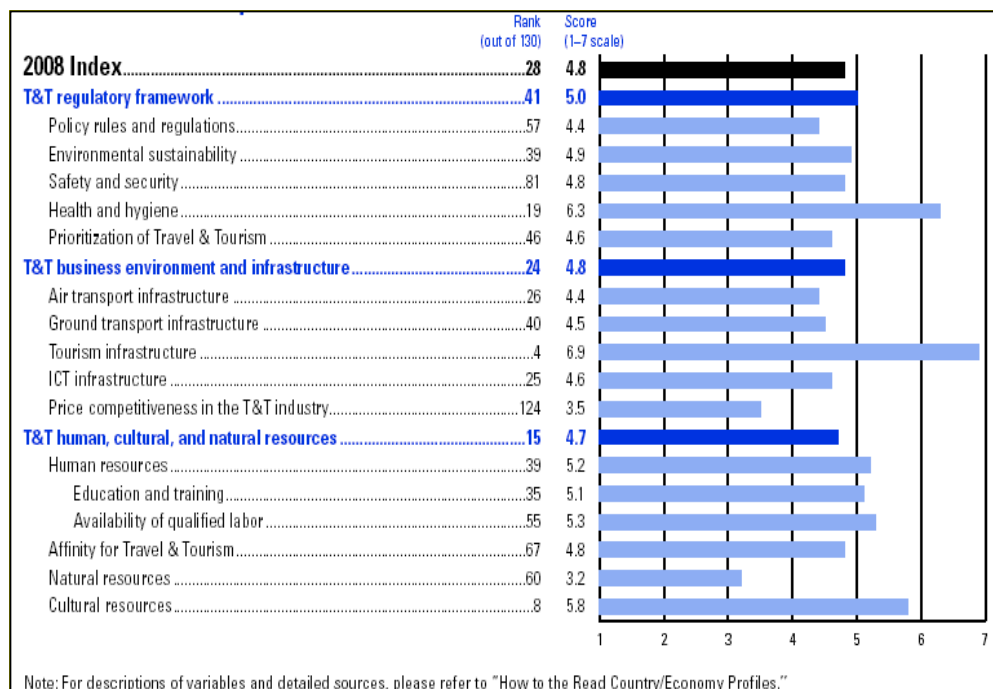


Fonte: WEF 2008

Secondo il WEF L'Italia nel 2008 si posiziona al 28° posto su una graduatoria di 130 Paesi, all'ultimo posto della classifica dei Paesi EU15.

Gli aspetti per i quali l'Italia occupa una buona posizione sono le risorse culturali per le quali occuperebbe l'8° posto, le infrastrutture turistiche, salute e igiene (fig.9). Gli aspetti per i quali il nostro Paese risulta penalizzato sono il quadro normativo e l'ambiente imprenditoriale, la competitività di prezzo e la sicurezza.

Fig.9 Indice TCI per l'Italia



Fonte: WEF 2008

Il rapporto sulla competitività del WEF va visto come uno strumento di dialogo tra settore privato e governi per riuscire a capire su quali leve agire per incentivare lo sviluppo del settore a livello delle singole nazioni.

Le aree di criticità del nostro Paese derivano dalla frammentazione dell'offerta alberghiera e della governance del sistema turistico che producono effetti negativi nelle aree *policy*, prezzi, sostenibilità dello sviluppo ed immagine. Al contrario i punti di forza evidenziati dall'indice segnalano che l'Italia vive ancora di una rendita di posizione nei suoi tipici elementi distintivi quali patrimonio culturale e quantità dell'offerta turistica.

2.4. La sostenibilità come fattore chiave per la competitività

La leva principale di competitività per il turismo risiede nei fattori di attrattività ambientale, artistica e culturale, si tratta di risorse spesso non riproducibili che giocano un ruolo fondamentale nel determinare i flussi turistici verso una determinata località o destinazione.

Le strategie di promozione e sviluppo del settore turistico devono, pertanto, essere finalizzate a tutelare e valorizzare tali risorse che potrebbero essere minacciate da un incontrollato sviluppo della stessa attività turistica.

Il turismo, infatti, come attività che interagisce fortemente con il territorio comporta esternalità negative che sono all'origine della possibile diminuzione della identità sociale e culturale dell'area ospitante, dell'aumento della produzione dei rifiuti, dell'aumento del consumo di beni primari e risorse (acqua, energia ecc.), della modificazione e distruzione degli ecosistemi montani, lacustri, costieri, marini, della perdita di biodiversità, degli impatti estetici e visivi, dell'inquinamento del suolo e dell'acqua, della congestione e dell'inquinamento acustico, la concentrazione dei benefici in poche aziende di elevate dimensioni e/o estere, l'aumento della domanda di mobilità, il lavoro nero e/o minorile. (Andriola, Manente, 2005).

Nasce quindi l'esigenza di conciliare competitività del turismo e rispetto del patrimonio locale nella piena attuazione dei principi della sostenibilità in un'ottica a lungo termine. Ritchie e Crouch, sottolineano la centralità della sostenibilità, ricordando, però, come quest'ultima non possa essere intesa solo in senso economico ed ecologico, ma anche sociale, culturale e politico (Ritchie e Crouch 2000). Data la rilevanza della sostenibilità per la competitività i due autori suggeriscono di riconoscere una connessione strutturale tra i due concetti e di aggiungere indissolubilmente al termine competitività il termine sostenibilità.

La domanda turistica attuale è sempre più domanda di esperienze in un territorio nel quale si ricercano valori ambientali e culturali quali clima, natura, tradizioni, risorse storiche ed artistiche; *"il futuro del turismo europeo si basa sulla qualità dell'esperienza dei turisti"* si legge nell'Agenda per un turismo sostenibile e competitivo della Commissione Europea, ed è proprio per restare concorrenziale che il settore turistico deve tendere verso la sostenibilità.

L'obiettivo del turismo, più di ogni altra attività economica, deve essere quello di tutelare i beni ambientali ed in generale incrementare la qualità ambientale.

In occasione della prima Conferenza mondiale sul Turismo Sostenibile a Lanzarote, si è evidenziata la necessità di ripensare le politiche di sviluppo turistico e le politiche territoriali ad esse associate e di dotare le località turistiche di strumenti idonei a coniugare crescita e qualità.

Tale intento ha guidato anche la politica europea del turismo che ha ribadito la centralità della gestione ambientale come opportunità per riqualificare il nostro tessuto produttivo e rendere più competitivi imprese e territori.

La politica europea del turismo mira a stimolare la competitività senza trascurare lo stretto legame che sussiste tra sviluppo del settore e sostenibilità, l'obiettivo principale e quello di contribuire a "migliorare la concorrenzialità dell'industria europea del turismo e creare più posti di lavoro e di qualità migliore grazie alla crescita sostenibile del turismo in Europa e a livello mondiale".

Un elemento fondamentale per inquadrare il fenomeno turistico e le sue complesse relazioni è la "capacità di carico" di cui si riporta la definizione dell' Organizzazione Mondiale del Turismo (WTO, 1999) : *"numero massimo di persone che visita, nello stesso periodo, una determinata località senza compromettere le sue caratteristiche ambientali, fisiche, economiche e socioculturali e senza ridurre la soddisfazione dei turisti"*.

Per ogni destinazione turistica si può definire una *capacità di carico fisica o ecologica* come il numero di visitatori oltre il quale le risorse ambientali o culturali della destinazione risultano danneggiate; una *capacità di carico economica*, cioè il limite oltre il quale si riduce drasticamente la qualità dell'esperienza turistica percepita e si determina una riduzione dei benefici derivanti dal turismo per cui l'impatto economico, dapprima positivo, crolla; una *capacità di carico sociale* che rappresenta il limite oltre il quale le funzioni non turistiche dell'area risultano danneggiate o ostacolate, con conseguente degrado nella qualità della vita della popolazione ospitante o danno sulle altre attività produttive.

Secondo l'Organizzazione Mondiale del Turismo è sostenibile quel turismo capace di "*gestire le risorse in modo che le necessità economiche, sociali ed estetiche possano venire soddisfatte mantenendo al contempo l'integrità culturale, le caratteristiche fondamentali degli ecosistemi, la biodiversità e lo stile di vita delle popolazioni locali*".

La sostenibilità del turismo coinvolge, pertanto, numerosi aspetti quali l'utilizzo responsabile delle risorse naturali, la considerazione dell'impatto ambientale delle attività, l'impiego di energie pulite, la protezione del patrimonio e la salvaguardia dell'integrità naturale e culturale delle destinazioni turistiche, la qualità dell'accoglienza.

CAPITOLO 3

LA METODOLOGIA STATISTICA

SOMMARIO: 3.1. La cluster analysis e la cluster fuzzy.- 3.2. Introduzione alla cluster analysis.- 3.3. La cluster analysis non gerarchica.- 3.4. L'approccio fuzzy alla cluster analysis.- 3.5. Fuzzy K-means.- 3.6. Criteri di cluster validity.- 3.6.1. Indici di fuzziness.- 3.6.2. Fuzzy Silhouette.- 3.7. L'analisi in componenti principali.- 3.8. Il software R.

3.1 La cluster analysis e la cluster fuzzy

Nello studio della competitività turistica delle regioni si è scelto di utilizzare una combinazione di metodi fattoriali e di *clustering* che applicati allo stesso insieme di dati costituiscono una strategia di analisi con l'obiettivo di "misurare" le *performance* di competitività turistica regionali sulla base di una serie di indicatori.

Questo capitolo è dedicato agli aspetti metodologici delle tecniche statistiche utilizzate quali la *cluster analysis*, ed in particolare, la *cluster analysis fuzzy* e l'analisi in componenti principali.

La *Cluster Analysis* è una tecnica statistica multivariata di tipo "esplorativo" per la classificazione delle unità statistiche in gruppi omogenei: la metodologia ricerca nelle n osservazioni p -dimensionali gruppi di unità tra loro simili, ignorando a priori l'esistenza reale di tali gruppi omogenei nel *dataset*. In questo senso la tecnica ha un ruolo esplorativo di ricerca di strutture latenti, si delega alla procedura statistica il compito di desumere dai dati la partizione ottimale in gruppi che siano caratterizzati da un elevato grado di omogeneità interna e allo stesso tempo da un'elevata disomogeneità tra gli stessi.

I metodi di *clustering* erano già noti alla fine del XIX secolo, ma l'interesse da parte degli statistici è cresciuto attorno agli anni 60. Si valuta che gli algoritmi di *clustering* elaborati fino ad oggi siano circa un migliaio.

La versatilità delle tecniche di *cluster analysis* nella loro possibile applicazione in diverse discipline, dalle scienze fisiche (fisica, medicina, biologia), a quelle sociali (economia, sociologia, psicologia), ha suscitato negli anni un crescente interesse da parte della comunità scientifica per gli aspetti metodologici e applicativi. E' soprattutto grazie all'evoluzione di numerosi *software* statistici che si è potuta gestire sempre più agevolmente la complessità computazionale di molti dei metodi di classificazione contribuendo in maniera decisiva al loro sviluppo con una vasta produzione di algoritmi sempre più complessi e sempre più efficienti nella classificazione dei dati.

Un processo di *clustering* può essere definito in maniera diversa; come una partizione di un insieme di unità elementari in modo che la suddivisione risultante goda di alcune proprietà considerate desiderabili; come raggruppamento di unità molto simili tra loro in gruppi che abbiano la caratteristica di essere il più possibile distinti tra loro.

Indipendentemente dalla definizione, i fattori che caratterizzano un metodo di classificazione sono essenzialmente due:

- a) la misura del grado di diversità tra le coppie di unità statistiche;
- b) l'algoritmo con cui si ricercano i *clusters*.

Le molteplici combinazioni di questi due elementi danno vita ad una gran quantità di metodi diversi di classificazione che possono distinguersi per il tipo di algoritmo utilizzato o per il tipo di risultato fornito.

La distinzione più nota è quella, basata sul tipo di algoritmo, che distingue tra metodi gerarchici e metodi non gerarchici.

I metodi di *clustering* gerarchico attraverso operazioni di tipo "*bottomup*" (aggregativo) o "*top down*" (scissorio) realizzano fusioni o divisioni successive dei dati. Si tratta di procedure iterative che considerano

tutti i livelli di distanza e i gruppi che si ottengono a un certo livello di distanza sono contenuti nei gruppi ottenuti a un livello di distanza inferiore. Nel caso dei metodi aggregativi le n unità iniziali sono fuse in gruppi sempre più ampi fino alla situazione finale estrema di unico gruppo; nel caso dei metodi “scissori” sono definite partizioni sempre più dettagliate dell’insieme iniziale fino alla scomposizione finale delle unità in n *clusters* contenenti ciascuno un elemento. La caratteristica distintiva rispetto ai metodi non gerarchici è che l’assegnazione di un’unità ad un *cluster* è definitiva, per cui, una volta che l’unità è attribuita ad un determinato *cluster*, non sarà più riallocata in altri gruppi.

I metodi di *clustering* non gerarchico individuano un’unica partizione procedendo a riallocazioni successive delle unità tra i gruppi definiti a priori, fino alla partizione giudicata “ottima” sulla base di un predeterminato criterio.

Un’ulteriore distinzione tra i metodi è basata sulla classificazione finale che essi producono. Si distingue tra metodi *crisp*, metodi *fuzzy* e metodi di classificazione sovrapposta. Con metodi di classificazione *crisp*, s’intendono tutti quei metodi che forniscono una partizione classica, cioè una suddivisione delle unità in gruppi tra loro disgiunti e tali che la loro unione fornisca l’insieme di tutte le unità. I metodi di classificazione sovrapposta forniscono una suddivisione delle unità in gruppi non disgiunti, cioè tali che una medesima unità possa appartenere a più di un gruppo (ricoprimento classico dell’insieme delle unità). I metodi di classificazione sfocata (*fuzzy*) suddividono l’insieme delle unità in modo che un’unità può appartenere solo in parte a un gruppo e, quindi, per la parte rimanente appartiene ad altri gruppi.

Un’ultima categoria di metodi che si può definire come classificazione sovrapposta sfocata, la quale fornisce dei ricoprimenti sfocati può scaturire dall’unione della classificazione sovrapposta e di quella *fuzzy*. Naturalmente, all’interno di queste distinzioni, valgono ancora quelle fatte in precedenza tra metodi gerarchici e non gerarchici, esistono, infatti, tecniche gerarchiche e

tecniche non gerarchiche di classificazione classica, sovrapposta, sfocata e sovrapposta sfocata.

Il numero di algoritmi a disposizione è piuttosto ridotto per la classificazione di tipo *fuzzy* e per quella sovrapposta rispetto agli algoritmi prodotti per i metodi di classificazione crisp che sono i più utilizzati.

Tuttavia i metodi di *cluster analysis* di tipo *fuzzy* riescono a trattare con efficacia l'imprecisione che spesso caratterizza i dati, poiché non offrono classificazioni nel senso proprio del termine, ma attribuiscono ogni unità a ciascun gruppo con un diverso gradi di appartenenza.

Per rappresentare i risultati ottenibili da una classificazione si può utilizzare una matrice di dimensioni $n \times c$, con un numero di righe n pari al numero delle unità e un numero di colonne c pari al numero dei gruppi, che contiene i valori di una funzione di appartenenza.

In particolare, si tratta di una funzione a c valori (dove c è il numero di gruppi della partizione o del ricoprimento) che associa a ogni unità n , c numeri, ognuno dei quali esprime il grado di appartenenza dell'unità i -esima al c -esimo gruppo (con $i=1,2,\dots, n$ e $c=1,2,\dots, C$).

Per i metodi classici tale funzione sarà definita nell'insieme $\{0,1\}$, cioè può assumere solo i due valori 1 e 0, che indicano, rispettivamente, se un'unità appartiene (1) o non appartiene (0) all'insieme. Per i metodi sfocati l'insieme di definizione della funzione è l'intervallo $[0,1]$ e quindi la funzione di appartenenza assume tutti i valori compresi tra zero e uno esprimendo il grado con cui un'unità appartiene a un gruppo. Nella tabella 6 si riportano le regole empiriche con le quali assegnare le unità ai gruppi in corrispondenza dei valori assunti dalla funzione di appartenenza.

Tab.6 Gradi di appartenenza e assegnazione dell'unità ai gruppi

Valore grado di appartenenza u_{ik} dell'unità i al gruppo k	Grado di assegnazione dell'unità i al gruppo k
$u_{ik} > 0,90$	Alto
$0,70 < u_{ik} < 0,90$	Medio
$0,50 < u_{ik} < 0,70$	Basso
$u_{ik} < 0,50$	Non assegnabile al gruppo

Un'altra differenziazione può essere fatta tra partizioni e ricoprimenti.

Le partizioni sono i raggruppamenti che devono rispettare il vincolo:

$$S = \sum_{i=1}^c u_{ic} = 1$$

ovvero la somma dei gradi di appartenenza è necessariamente uguale all'insieme unità; producono partizioni i metodi di classificazione classici e i metodi di classificazione sfocati.

I ricoprimenti sono raggruppamenti che devono rispettare il vincolo:

$$S = \sum_{i=1}^c u_{ic} \geq 1$$

per cui la somma dei gradi di appartenenza può anche superare il valore unitario; producono ricoprimenti i metodi di classificazione sovrapposta e sovrapposta sfocata

Nelle pagine seguenti saranno analizzati in primo luogo i metodi di *clustering* classici (*crisp*) gerarchici e non gerarchici e poi si analizzeranno le

tecniche di *cluster analysis fuzzy* utilizzate nell'ambito della ricerca sulla competitività turistica delle regioni italiane.

3.2. Introduzione alla Cluster Analysis

In ogni procedimento di clustering la definizione esatta dello scopo della classificazione delle unità rappresenta la base di tutte le scelte successive che il ricercatore dovrà compiere. Di seguito sono elencate le fasi di un processo di *clustering*:

1. selezione degli elementi del collettivo da sottoporre ad analisi;
2. scelta delle variabili e loro eventuale trasformazione per tenere conto di aspetti riconducibili alla scala e all'unità di misura (standardizzazione);
3. selezione di un criterio per valutare la dissomiglianza esistente tra gli elementi osservati;
4. scelta di un algoritmo di raggruppamento delle unità;
5. determinazione del numero dei gruppi che si formano tra gli elementi del collettivo in esame, ovvero individuazione della cosiddetta partizione ottimale;
6. verifica della congruenza dei risultati e loro interpretazione.

La scelta delle variabili dipende naturalmente dalle finalità assegnate alla classificazione e rappresenta una fase cruciale che può condizionare fortemente i risultati della classificazione. La scelta delle variabili è un'operazione di natura molto soggettiva poiché rispecchia essenzialmente le conoscenze, le convinzioni e le idee del ricercatore. La selezione dovrebbe fondarsi su tutti gli aspetti che si ritengono importanti per gli scopi prefissati tenendo conto che l'inclusione di variabili poco discriminanti o poco rilevanti al fine della

caratterizzazione dei gruppi esercita un ruolo negativo sulla qualità dei risultati.

Una buona classificazione dovrebbe mostrarsi essenzialmente stabile rispetto a piccoli cambiamenti nell'insieme delle variabili considerate.

Un problema particolare sorge quando le variabili presentano unità di misura diverse o hanno ordini di grandezza diversi, infatti, l'importanza relativa di ciascuna variabile nella formazione di gruppi è direttamente collegata alla varianza delle diverse variabili. Pertanto, variabili caratterizzate da un più elevato grado di dispersione hanno un maggiore impatto sulla misura di distanza. Un possibile rimedio consiste nella standardizzazione delle variabili, mediante sottrazione della media e divisione per la deviazione standard.

Per orientarsi nella scelta delle variabili può essere opportuno osservare la correlazione tra le variabili considerate per comprendere il reale contributo informativo di ciascuna di queste.¹⁸

Le scelte effettuate nelle prime due fasi del procedimento di *clustering*, ci portano ad avere la disponibilità di un collettivo statistico di n elementi ciascuno rappresentato da p variabili. I dati sono raccolti in una matrice come la seguente:

¹⁸ *L'applicazione dell'analisi in componenti principali a tutte le variabili disponibili può essere d'aiuto nella scelta di quelle da utilizzare nella cluster analysis. Se k componenti principali tengono conto di una percentuale elevata della varianza totale, il ricercatore può eseguire la classificazione direttamente sugli scores di tali CP, che costituiscono il segnale degli aspetti rilevanti, mentre le restanti componenti rappresentano i residui, cioè il rumore (noise). In alternativa il ricercatore potrebbe applicare la cluster analysis a un sottoinsieme delle variabili di partenza e precisamente solo a quelle più fortemente correlate con le prime k componenti principali, essendo le restanti poco connesse con gli aspetti fondamentali dell'indagine.*

$$X = \begin{pmatrix}
 & \begin{matrix} x_{11} & \dots & x_{1s} & \dots & x_{1p} \end{matrix} \\
 \begin{matrix} x_{i1} & \dots & x_{is} & \dots & x_{ip} \end{matrix} & & & & \\
 & \begin{matrix} x_{n1} & \dots & x_{ns} & \dots & x_{np} \end{matrix} \\
 \end{pmatrix}$$

Vettore - unità

Vettore - variabile

ove x_{is} rappresenta la determinazione della s -esima variabile quantitativa osservata sull' i -esima unità statistica ($i=1, \dots, n; s=1, \dots, p$). I vettori riga rappresentano quindi le n unità statistiche, ciascuna unità statistica è quindi un vettore di p -elementi, contenenti i valori da essa assunti sulla prima, la seconda, la j -esima e la p -esima variabile.

Per ripartire l'insieme degli n elementi in c gruppi, si sottopone ad una serie di elaborazioni successive il complesso delle informazioni contenute nella matrice.

Si trasforma tale matrice di partenza in una matrice di dissimilarità o di distanze tra le n coppie di osservazioni.

Si sceglie poi un algoritmo che definisca le regole su come raggruppare le unità in sottogruppi sulla base delle loro similarità.

Lo scopo della classificazione è quello di identificare il minor numero di gruppi in modo che gli elementi appartenenti ad un gruppo siano il più possibile simili tra loro e il più possibile diversi dagli elementi appartenenti ad altri gruppi.

Il punto di partenza fondamentale è la definizione di una misura di similarità o di distanza tra gli oggetti, cioè tra le righe della matrice dei dati. Per la misura della dissomiglianza sono disponibili numerosi indicatori che ovviamente se vengono applicati agli stessi insiemi di dati forniscono in genere risultati diversi. Le misure impiegate con maggiore frequenza sono gli indici di similarità e le distanze, in particolare quelle appartenenti alla cosiddetta metrica di *Minkowski*.

La scelta del ricercatore tra indici di dissimilarità e metrica è legata al tipo di dati che si hanno a disposizione.

Per dati quantitativi possiamo utilizzare delle misure di distanza, cioè delle metriche.

Per dati di tipo qualitativo bisogna utilizzare misure “*matching-type*”, cioè di associazione (similarità o dissimilarità).

Effettuata la scelta della misura di diversità da utilizzare, l’altro punto fondamentale è la regola in base alla quale si formano i gruppi.

Dopo aver utilizzato misure di dissomiglianza per sintetizzare la diversità dei profili tra coppie di unità statistiche si perviene alla determinazione di una matrice simmetrica detta matrice delle distanze (o degli indici di similarità) dalla forma:

$$D = \begin{pmatrix} 0 & d_{12} & \dots & d_{1n} \\ & 0 & \dots & d_{jn} \\ & & & 0 \end{pmatrix}$$

dove i termini della diagonale principale sono ovviamente nulli, poiché stanno ad indicare la distanza di un unità con se stessa. Di conseguenza è sufficiente considerare la matrice al di sopra (o al di sotto) della diagonale principale.

A questo punto è necessario procedere alla scelta di un idoneo metodo di raggruppamento. I più diffusi metodi gerarchici aggregativi si basano sull’impiego delle matrice delle distanze (metodo del legame singolo, metodo del legame completo, metodo del legame medio) e si differenziano unicamente per il diverso criterio utilizzato per calcolare le distanze tra i gruppi ai fini delle successive aggregazioni. L’individuazione dei gruppi si ottiene partendo dalla matrice delle distanze e aggregando le due unità più vicine, ovvero con

distanza minima, in un *cluster*. Si ricalcolano le nuove distanze del gruppo dalle altre unità ottenendo una nuova matrice delle distanze.

Al passo successivo una terza unità entra a far parte del *cluster* trovato al passo precedente, oppure, due unità vengono fuse per formare un diverso *cluster*.

Si continua a procedere in questo modo fino alla formazione di un unico *cluster* contenente tutte le unità.

Il procedimento descritto si basa sulla scelta del criterio di assegnazione delle unità ai cluster che corrisponde a diversi algoritmi aggregativi.

Nel caso del metodo del legame singolo, la distanza tra i gruppi è misurata dalla più piccola delle distanze esistenti tra gli elementi appartenenti ad un gruppo e quelli appartenenti ad un altro; in quello del legame completo, alla maggiore di tali distanze. Nel metodo del legame medio tra i gruppi, tale distanza è definita come la media aritmetica di tutte le distanze tra gli elementi dei gruppi; una variante a questo metodo è il metodo del legame medio nei gruppi nel quale si considerano anche le distanze tra le unità appartenenti al medesimo gruppo.

I metodi gerarchici appena descritti operano sulla matrice delle distanze, esistono altri metodi gerarchici che utilizzano anche la matrice dei dati di partenza.

Il metodo del centroide si applica solo a variabili quantitative e lavora sui singoli vettori di osservazioni, ad ogni passo ricalcola la matrice delle distanze partendo non dalle distanze precedenti ma dai centroidi di ciascun cluster.

Per ogni gruppo si calcola il centroide¹⁹ (o individuo medio, cioè un elemento che come modalità delle diverse variabili, presenta le modalità medie del gruppo). La distanza tra un'unità e un gruppo o tra due gruppi è calcolata come distanza tra i rispettivi centroidi.

¹⁹ Il centroide di un gruppo è il vettore p -dimensionale i cui elementi sono costituiti dalla media dei corrispondenti elementi dei vettori associati alle unità appartenenti al gruppo.

Nel metodo di *Ward* si definisce esplicitamente una funzione obiettivo, in particolare si considera la scomposizione della devianza totale (T) delle p variabili in Devianza nei gruppi (W=within) e devianza tra i gruppi (B=Between).

$$T = W + B$$

$$T = \sum_{s=1}^p \sum_{i=1}^n (x_{is} - \bar{x}_s)^2$$

Ad ogni passo della procedura gerarchica si aggregano tra loro i gruppi che comportano il minor incremento della devianza nei gruppi cioè che assicurano la massima coesione interna possibile.

E' importante sottolineare che, partendo da una stessa matrice delle distanze, l' algoritmo scelto per la formazione dei gruppi influenza notevolmente i risultati della classificazione gerarchica. Il problema della scelta della classificazione migliore può essere impostato in termini di proprietà desiderabili per il metodo di formazione dei gruppi e di verifica di tali proprietà per i vari algoritmi di classificazione. Partendo dalla matrice dei dati iniziali otteniamo la matrice delle distanze effettuando delle scelte "soggettive" (standardizzazione delle variabili, ponderazione delle variabili, tipo di metrica ecc..) si pone quindi il problema del grado di oggettività dei gruppi ottenuti. Un criterio è quello di richiedere che la massima distanza all'interno dei gruppi sia minore della minima distanza tra i gruppi in modo tale che i gruppi appartengano evidentemente a sottoinsiemi omogenei al loro interno e non esistono criteri convenzionali per assegnare un elemento ad un gruppo.

Il criterio appena enunciato si riferisce alla partizione ben strutturata minimale che si definisce come la partizione ben strutturata con minor numero di gruppi. Una partizione $P = \{C_1, C_2, C_3, \dots, C_g\}$ di n elementi u_i dice ben strutturata se $\max (d_{ij}) < \min (d_{rs})$, per ogni u_i, u_j appartenenti allo stesso gruppo e u_r, u_s appartenenti a gruppi diversi.

E' da rilevare inoltre che per ogni matrice delle distanze esiste una ed una sola partizione ben strutturata minimale. Tutti i metodi gerarchici ad un certo punto della classificazione individuano la partizione ben strutturata minimale e quindi sono tutti criteri accettabili per la classificazione in gruppi, ma il problema nasce dal punto di vista pratico dal momento che la partizione ben strutturata minimale è comunque molto dispersa cioè costituita da un numero eccessivo di gruppi e quindi è necessario individuare altri criteri aggiuntivi per giungere ad una classificazione oggettiva che sia soddisfacente e che riesca a soddisfare la duplice esigenza di sintesi dell'insieme dei dati e di correttezza nell'individuazione dei gruppi.

Si può considerare quindi un'altra proprietà richiesta ai metodi gerarchici che è quella di essere invarianti per trasformazioni monotone crescenti delle distanze.

Tale criterio è soddisfatto sia dal metodo del legame singolo che dal metodo del legame completo, non risulta invece valido per il metodo del legame medio. Tuttavia tali metodi consentono di individuare gruppi con caratteristiche diverse; il metodo del legame singolo può riunire in un gruppo anche elementi molto distanti quando tra essi esiste una successione di punti intermedi (effetto catena), il metodo del legame completo individua gruppi compatti al loro interno ma di forma circolare, sferica o ipersferica, assegnando eventuali punti intermedi ai gruppi principali individuati.

Tutti gli algoritmi gerarchici descritti possono essere ricondotti ad una formulazione unitaria proposta da Lance and Williams nel 1967:

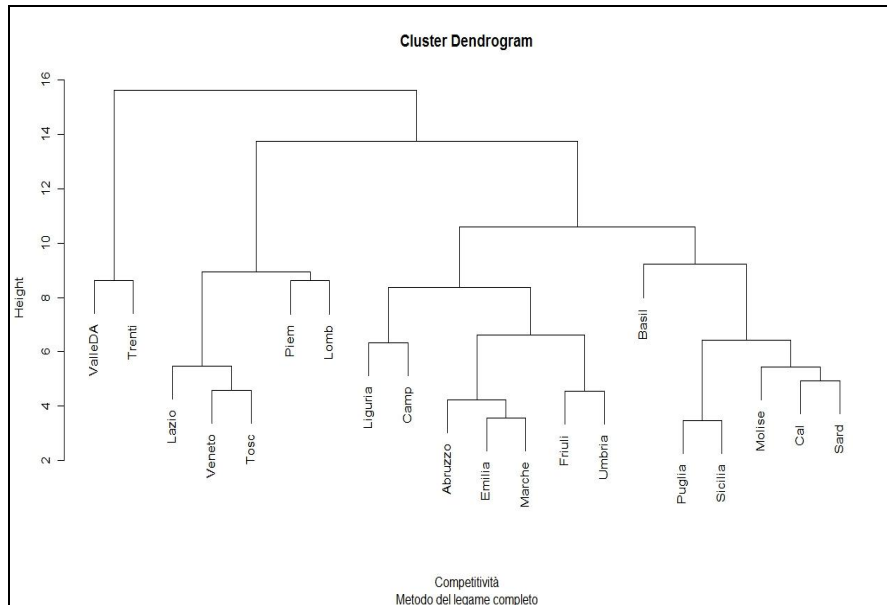
$$D [C_3, (C_1, C_2)] = \alpha_1 d_{31} + \alpha_2 d_{32} + \beta d_{12} + \gamma | d_{31} d_{32} |$$

Con α_1 , α_2 , β , γ come parametri fissati dal ricercatore. Per particolari valori di tali parametri ci si riconduce alle espressioni di distanza adottate dai principali metodi gerarchici già menzionati.

La mappa delle successive aggregazioni delle unità statistiche da n gruppi ad un solo gruppo può essere rappresentata graficamente mediante un diagramma a forma di albero (dendrogramma), che riproduce in ordinata i

livelli di distanza che caratterizzano le aggregazioni delle diverse partizioni (fig.10)

Fig.10 - Esempio di dendrogramma



Fonte : ns. elaborazione

Dal momento che i metodi gerarchici aggregativi generano una famiglia di partizioni delle n unità statistiche si pone il problema di scegliere quale sia la partizione ottimale ai fini dell'interpretazione dei risultati. Per la scelta del numero dei gruppi non esiste un criterio generale applicabile, ma si può far ricorso ad alcune procedure ormai consolidate.

L'obiettivo di fondo della cluster nel ricercare gruppi con massima coesione interna e separazione esterna ci porta ad individuare un valido criterio di giudizio basato sulla scomposizione della devianza totale delle p variabili. Infatti una buona classificazione è caratterizzata da una ridotta quota di devianza *within* e da un'elevata quota di devianza *between*.

Si può considerare quindi l'indice:

$$R^2 = 1 - W/T = B/T$$

che assume valori tra 0 e 1. In particolare una classificazione con R^2 prossimo ad 1 è caratterizzata da buona omogeneità nei gruppi e discreta separazione tra gli stessi. Naturalmente la semplice massimizzazione di tale indice porterebbe a considerare come partizione migliore quella in cui tutte le unità formano un gruppo, ma l'obiettivo è sempre quello di compendiare le esigenze di omogeneità interna e sintesi della classificazione.

Un indice alternativo è il *RMSSTD*, acronimo di *Root Mean Square Standard Deviation*

$$RMSSTD = \sqrt{\frac{W_h}{p(n_h - 1)}}$$

Dove W_h rappresenta la devianza delle p variabili nel cluster costituitosi al passo h della procedura e n_h la relativa numerosità del cluster.

Gli indici R^2 e *RMSSTD* sono utili per valutare il grado di coesione interna dei gruppi che si costituiscono ad ogni passo della procedura gerarchica e scegliere quindi la partizione giudicata più soddisfacente tra quelle individuate dai metodi gerarchici. Considerando due passi consecutivi del processo di aggregazione caratterizzati da $g+1$ e g gruppi, se nel passaggio dalla partizione meno aggregata con $g+1$ gruppi alla partizione più aggregata con g gruppi si manifesta un salto rilevante nei valori degli indici di bontà della partizione (R^2 e *RMSSTD*) si considera come soddisfacente la partizione meno aggregata, cioè quella del passo precedente al salto nei valori (con $g+1$ gruppi)²⁰.

²⁰ Gli indici R^2 e *RMSSTD* possono essere correttamente impiegati solo nel caso in cui le variabili considerate siano tra loro comparabili, oppure siano state opportunamente trasformate per renderle tali ad esempio attraverso la standardizzazione.

Si può inoltre costruire una rappresentazione grafica che ha in ordinata il numero dei gruppi via via formati ed in ascissa i corrispondenti valori della misura di dissomiglianza relativa ai due gruppi che si uniscono per generare la partizione in quel numero di gruppi. Si prende quindi in esame la spezzata che si ottiene unendo i punti così individuati, e se ne osserva l'andamento dall'alto verso il basso e da sinistra verso destra. Il numero di gruppi in corrispondenza del quale si evidenzia un forte appiattimento della spezzata identifica la ripartizione ottimale.

Una procedura alternativa consiste nell'esaminare i valori assunti dalla misura di dissomiglianza tra i due gruppi che si fondono ad ogni passo successivo. Si cerca di individuare in questo caso un salto consistente nella serie dei valori di tale misura. Si può immaginare infatti che là dove si registrano i salti maggiori le tappe di aggregazione corrispondenti comportino un cambiamento rilevante nell'omogeneità interna dei gruppi e quindi si può decidere di arrestare il processo di raggruppamento immediatamente prima di quel salto.

Un ulteriore criterio consiste nell'ispezione diretta del dendrogramma e si effettua un "taglio" in corrispondenza di un salto nei livelli di distanza in cui è avvenuta l'aggregazione. In questo modo si individuano in corrispondenza del taglio i gruppi che corrispondono alla partizione ottimale.

Spesso può essere utile valutare il grado di somiglianza tra due differenti partizioni individuate attraverso la cluster analysis. Il confronto può avvenire sia tra partizioni con il medesimo numero di gruppi, sia tra partizioni con un differente numero di cluster individuati. Per valutare numericamente la stabilità di una classificazione al variare delle opzioni scelte per ottenerla (metodo e tipo di distanza) sono stati proposti alcuni indici. Uno di questi indici, utile allo scopo del confronto tra due partizioni è l'indice di Rand che si basa sul confronto tra coppie di unità statistiche. Date due partizioni P e P^* , costituite rispettivamente di g e g^* gruppi, la formula dell'indice di Rand è la seguente:

$$R_{P;P^*} = \frac{c_{11} + c_{00}}{\binom{n}{2}}$$

Nell'espressione c_{11} indica il numero di coppie di unità che appartengono allo stesso cluster in entrambe le partizioni P e P^* , c_{00} è il numero di coppie di unità che appartengono a gruppi diversi sia in P che in P^* . Il denominatore indica naturalmente le possibili combinazioni a due a due delle n unità.

L'indice varia tra 0 e 1, assume valore 0 se tutte le coppie di unità appartenenti ad un gruppo in una partizione sono assegnate a gruppi diversi nell'altra partizione, vale 1 se le partizioni sono praticamente identiche.

L'indice di Rand può essere calcolato, in maniera più immediata, confrontando il numero di unità statistiche classificate simultaneamente nel gruppo r -esimo di P e nel gruppo c -esimo di P^* . In questo modo non è necessario effettuare i confronti a coppie ma soltanto conoscere il cluster di appartenenza di ciascuna unità nelle diverse partizioni.

3.3. *La cluster analysis non gerarchica*

Nei metodi gerarchici l'algoritmo cerca, ad ogni passo, la migliore scissione o aggregazione tra *clusters*, nel caso dei metodi non gerarchici l'algoritmo mira a ripartire le n unità in un numero predefinito di gruppi fornendo come risultato finale un'unica partizione in c gruppi basandosi sulla ottimizzazione di un criterio predeterminato. A differenza dei metodi gerarchici, l'assegnazione di un oggetto ad un *cluster* non è irrevocabile quindi le unità possono essere riassegnate ad un diverso *cluster* se l'allocazione iniziale risulta inappropriata. Le classi ottenute ad ogni iterazione intermedia vengono infatti cancellate e il processo di aggregazione ricomincia a partire dai nuovi centri ricalcolati.

L'inizializzazione del processo di classificazione avviene indicando c centri di partenza intorno a cui aggregare le unità e procedendo a spostamenti successivi delle unità tra i gruppi definiti a priori fino alla determinazione di una partizione che meglio risponde ai concetti di omogeneità interna ai gruppi e di eterogeneità tra gli stessi.

I diversi algoritmi non gerarchici differiscono tra loro nei seguenti aspetti:

1. come sono inizializzati i centri di partenza;
2. come gli elementi vengono assegnati ai diversi centri;
3. come alcune o tutte le unità vengono eventualmente riassegnate ad un diverso gruppo.

Il metodo più comune è il metodo delle k medie. Data una prima partizione ottenuta a priori, il metodo procede a riallocare le unità al gruppo con centroide più vicino, fino a che per nessuna unità si verifica che sia minima la distanza rispetto al centroide di un gruppo diverso da quello cui essa appartiene. Gli algoritmi più comunemente utilizzati sono *K-means* (o la sua variante *Fuzzy K-means*) e *Expectation-Maximization (EM)*.

La procedura iterativa può essere sintetizzata nelle seguenti fasi:

1. Scelta di c poli iniziali ($g_1, g_2, \dots, g_h, \dots, g_c$) che costituiscono i centroidi dei cluster della partizione iniziale;
2. Si costruisce quindi la partizione iniziale costituita da c gruppi raggruppando le unità intorno ai c centri allocando ciascuna unità al cluster il cui centroide risulta più vicino;
3. Calcolo dei centroidi dei c gruppi così ottenuti;
4. Calcolo della distanza di ogni elemento da ogni centroide; se la distanza minima non è ottenuta in corrispondenza del centroide del gruppo di appartenenza, allora l'unità è riallocata al gruppo che corrisponde al centroide più vicino;
5. Si ricalcolano nuovamente i centroidi;
6. Si ripetono le fasi 4 e 5 fino a che i centri risultano stabili ovvero non subiscono ulteriori modifiche rispetto alla iterazione precedente.

Come misura di distanza tra l'unità i ed il centroide viene normalmente utilizzata la distanza euclidea in quanto garantisce la convergenza dell'algoritmo in pochi passi (<10). Ciò consente di suddividere le unità statistiche in gruppi non sovrapposti, minimizzando la devianza interna ai gruppi stessi, e quindi massimizzando la loro omogeneità.

Il metodo delle k -medie è il metodo di classificazione non gerarchica più semplice computazionalmente e altrettanto semplice da implementare nei principali software statistici, fornisce buoni risultati a patto di fornire una ragionevole soluzione di partenza e un numero adeguato di *cluster*.

Il metodo delle k medie identifica cluster iper-sferici nel caso in cui venga utilizzata la distanza euclidea come misura di distanza o cluster iperellissoidali nel caso di distanza di *Mahalanobis*.

3.4. L'approccio fuzzy alla cluster analysis

La *fuzzy logic* può essere definita come una logica polivalente ovvero come un'estensione della logica tradizionale bivalente. Nella logica *fuzzy* una proposizione non è vera (=1) o falsa (=0) ma il suo grado di verità è compreso tra 0 e 1.

Fu Lofti A. Zadeh nel 1965 a sancire la nascita della logica *fuzzy* definendo un insieme *fuzzy* come una "classe di oggetti con un continuum di gradi di appartenenza". Gli insiemi *fuzzy* sono caratterizzati dalla funzione di appartenenza che assegna ad ogni oggetto un grado di appartenenza compreso tra 0 e 1. La caratteristica distintiva di questo approccio è la ridefinizione del concetto di appartenenza ad un insieme. Tutte le proprietà di inclusione, unione, intersezione, complemento, relazione, convessità, ecc. sono estese a tali insiemi. Nella logica *fuzzy*, in altre parole, il concetto di appartenenza è ridefinito in maniera quantitativa, associando ad ogni elemento il grado di appartenenza a quella classe. Inizialmente tale logica fu investita dalle dure critiche della comunità scientifica accademica tuttavia moltissimi studiosi e scienziati nei campi più svariati divennero sostenitori di tale logica. A partire dagli anni 80 la logica *fuzzy* ha investito diverse applicazioni industriali. La vicinanza della *fuzzy logic* al modo di ragionare del pensiero umano fanno di questo approccio un ottimo strumento per diversi ambiti disciplinari.

La *cluster analysis* è stata la prima tecnica statistica multivariata cui è stata applicata la teoria degli insiemi *fuzzy*.

I metodi di *fuzzy clustering*, tuttavia, sono stati poco sviluppati, rispetto a quelli classici detti *crisp*, infatti gli algoritmi a disposizione risultano abbastanza ridotti rispetto a quelli a disposizione nel *clustering* classico. I

metodi *fuzzy* fanno ricorso alla teoria degli insiemi *fuzzy*, e permettono di associare una unità ai gruppi con un certo grado di appartenenza. L'interesse per questi metodi nasce dalla consapevolezza che esiste un certo grado di imprecisione nei dati e che essi quindi siano in grado di rappresentarli più di quanto possa fare un metodo *crisp*.

L'incertezza nella *cluster analysis* può riguardare sia l'informazione empirica, quindi i dati, che l'informazione teorica, quindi il modello.

L'incertezza teorica è legata al processo di assegnazione delle unità ai cluster mentre l'incertezza empirica è dovuta all'imprecisione dei dati di partenza. Nella ricerca sulla competitività turistica i dati sono non sfocati quindi l'informazione empirica (i dati) è di tipo *crisp*, mentre è l'informazione teorica (il modello) ad essere di tipo *fuzzy*.

I metodi di *fuzzy clustering* sono, inoltre, più ricchi di informazioni in quanto forniscono il grado di coerenza di una unità con ciascun *cluster*, consentendo di stabilire una gerarchia di gruppi data dal diverso grado di appartenenza dell'unità ai gruppi a cui può appartenere l'unità. Anche per questi metodi è possibile una suddivisione tra metodi gerarchici e non gerarchici che sono spesso delle estensioni dei metodi classici al caso *fuzzy*.

L'obiettivo della classificazione sfumata è quello di assegnare le unità ad un certo numero di *cluster* con un diverso grado di appartenenza che assume valori nell'intervallo $[0,1]$ dove il valore 0 indica che l'elemento non appartiene all'insieme sfocato, il valore 1 indica che l'elemento appartiene totalmente all'insieme sfocato, i valori intermedi indicano il rispettivo grado di appartenenza.

L'interesse teorico di questi metodi è dovuto al fatto che essi trattano bene l'imprecisione: le unità statistiche non sono sempre classificabili con esattezza perché non è raro il caso di unità che possono essere assegnate indifferentemente a più gruppi. Questi metodi assegnano ogni singola unità in parte a ciascun gruppo in modo che la classificazione che risulta non solo mostri come si aggregano le unità, ma riesca anche a mostrare quanto una unità

appartiene ad un gruppo. In tal modo l'assegnazione di una unità ad un gruppo non è mai una forzatura mentre la non assegnazione di una unità ad un gruppo indica con certezza che quella unità non appartiene a quel gruppo.

I metodi di classificazione sfocata non hanno quindi la pretesa di dare risposte precise su come si aggregano i dati, cosa che si può fare più agevolmente con un metodo di analisi classica, ma, al contrario, tentano di rappresentare proprio l'imprecisione insita nei dati. I primi algoritmi sulla *cluster analysis fuzzy* per dati *crisp* sono stati realizzati da Bezdek (1974) e Dunn (1974).

I metodi gerarchici di *fuzzy clustering* hanno la caratteristica di prevedere due fasi: nella prima si calcola una misura di similarità tra coppie di unità, nella seconda si assegna ciascuna unità ai gruppi formati con un certo grado di appartenenza. La prima fase è analoga alla classificazione *crisp*, la differenza sta nel modo in cui vengono attribuite le funzioni di appartenenza ai cluster. Restano, come nel caso della classificazione classica, le differenze nel modo in cui attribuire le distanze tra una unità e un gruppo, o tra due gruppi, oltre al modo in cui attribuire il grado di appartenenza di una unità ad un cluster sfocato. Rispetto all'attribuzione delle distanze è possibile trovare delle similitudini con i metodi, ad esempio, del legame singolo e del legame completo.

I metodi di classificazione non gerarchici hanno la caratteristica di fornire direttamente un determinato numero di gruppi fissato a priori, attraverso procedure di tipo iterativo che cercano di ottimizzare una funzione obiettivo che, in genere, rappresenta una misura della dispersione dei punti dai centri dei cluster.

La differenza principale tra i diversi algoritmi consiste nella diversa funzione obiettivo adottata e, dunque, nel differente processo iterativo scelto per calcolare i gradi di appartenenza delle unità ai vari gruppi. La funzione obiettivo determina per ogni soluzione una misura dell'errore, in termini di efficienza o costo basandosi sulla distanza tra i dati e gli elementi

rappresentativi dei cluster. Ogni formulazione della funzione obiettivo incorpora dei vincoli, la soluzione ottima corrisponde al valore ottimo della funzione.

Il metodo *fuzzy k-means*, che sarà oggetto del successivo paragrafo rappresenta sicuramente il metodo non gerarchico sfocato più noto e utilizzato.

3.5. Fuzzy K-means

Questo metodo proposto da Bezdek nel 1981 è quello più utilizzato e più diffuso tra quelli di classificazione sfocata, si tratta di un'estensione del metodo *crisp* delle k-medie, e risulta particolarmente idoneo per trattare *dataset* di notevoli dimensioni, grazie alla velocità con la quale converge verso una classificazione ottimale. Nella fase iniziale della procedura, analogamente alla versione *crisp* dell'algoritmo si sceglie il numero di cluster c in cui si vogliono classificare le n unità con p caratteri ciò avviene o in modo ragionato sulla base delle conoscenze a priori del ricercatore o, come si vedrà in seguito sulla base delle indicazioni tratte da criteri di *cluster validation*. Si procede iterativamente e attraverso la minimizzazione di una funzione obiettivo si ottiene una classificazione sfocata nella quale per ogni unità viene determinato il grado di appartenenza ai c gruppi. Il grado di appartenenza u_{ik} , delle n unità ai c gruppi soddisfa i seguenti vincoli:

$$0 \leq u_{i,k} \leq 1$$

$$\sum_{k=1}^c u_{ik} = 1$$

per l' i -esima unità con $i=1, \dots, n$ ed il k -esimo gruppo con $k=1, \dots, c$. Il primo vincolo stabilisce l'insieme di definizione della funzione di appartenenza, il secondo vincola la somma dei gradi di appartenenza di ogni unità ad 1. La matrice U contenente i gradi di appartenenza u_{ik} , avrà dimensioni $n \times c$.

Indichiamo con J_m la funzione obiettivo da minimizzare utilizzata per calcolare i valori ottimi dei gradi di appartenenza; questa è funzione del quadrato della distanza d_{ik} , tra l'unità i -esima e il centroide del k -esimo gruppo e dipende dal parametro m che può assumere qualsiasi valore reale maggiore o uguale a 1 e che regola il livello di *fuzziness*:

$$J_m(U, v) = \sum_{k=1}^c \sum_{i=1}^n (u_{ik})^m (d_{ik})^2$$

Dove $(d_{ik})^2 = |x_i - v_k|^2$ cioè un'opportuna norma su R^p (ad esempio la norma euclidea); $v_k \in R^p$ è la componente k -esima del vettore dei centroidi, $x_i \in R^p$ è la componente i -esima del vettore delle unità; $U[u_{ik}]$ è la matrice dei gradi di appartenenza di dimensione $(n \times c)$. Le variabili rispetto alle quali effettuare la minimizzazione sono quindi i centri dei cluster e i gradi di appartenenza. Il significato della funzione obiettivo è che ogni centroide è la migliore rappresentazione delle unità che compongono il gruppo, in quanto rende minima la somma dei quadrati degli errori $x_i - v$. La funzione obiettivo J_m misura l'errore quadratico in cui si incorre quando si rappresentano le n unità con i c centroidi dei gruppi, essa dipende da come le unità sono disposte nei gruppi e misura la dispersione delle stesse intorno ai centri. La partizione ottima definita partizione di minima varianza è quella che minimizza J_m .

Il valore del parametro m che deve essere scelto all'inizio della procedura esprime il grado di sfocatura (*fuzziness*), cioè quanto sarà sfocata la partizione risultante. L'algoritmo delle *K-medie fuzzy* può essere sintetizzato nei seguenti passi:

1. Si fissano i valori di m (parametro di *fuzziness*) e di c (numero di gruppi) e la metrica da utilizzare e si sceglie una partizione iniziale delle unità in c gruppi rappresentata con la matrice

$$U^0 = \begin{bmatrix} u_{i,k} \end{bmatrix}$$

dove l'esponente indica il numero di iterazioni;

2. si calcolano i centri dei gruppi

$$V_k^0 = \frac{\sum_{i=1}^n (u_{i,k})^m * x_i}{\sum_{i=1}^n (u_{i,k})^m}$$

3. si calcola la matrice U^1 alla prima iterazione;

4. si calcola la differenza tra l'ultima e la penultima iterazione secondo una opportuna distanza, se $|U^1 - U^0| < \delta$ dove δ è un parametro stabilito a priori. A questo punto ci si ferma e si considera come classificazione finale quella dell'ultima iterazione, altrimenti si itera il passo 2 fino a quando la condizione non è soddisfatta.

Nel passo 3 possono verificarsi le seguenti condizioni:

1. Se per qualche gruppo (supponiamo r) si ha che $d_{ir}=0$ si pone $u_{ir}=1$ e $u_{ik}=0$ per tutti i $k \neq r$;

2. se la condizione precedente non è soddisfatta allora si applica la seguente formula:

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left[\frac{d_{ik}}{d_{jk}} \right]^{2/(m-1)}}$$

Come già accennato una delle motivazioni principali per cui il metodo delle *k-medie fuzzy* è il più utilizzato e la velocità con cui converge verso la classificazione finale ciò è dovuto anche al fatto che ad ogni successiva iterazione il valore della funzione obiettivo decresce rispetto a quello

dell'iterazione precedente cosa che invece non sempre si verifica negli altri metodi di questo tipo suscitando qualche dubbio sulla loro convergenza.

Inoltre tale metodo si è mostrato relativamente indipendente dalla partizione iniziale fornendo sempre lo stesso risultato. Tuttavia i tempi del processo di convergenza risultano tanto più rapidi quanto più la partizione di partenza si avvicina a quella finale, dunque risulta senz'altro conveniente scegliere sempre come $U_{(0)}$ quella ottenuta mediante un altro metodo di classificazione (classico o sfocato).

Poiché questo metodo rappresenta una generalizzazione del metodo delle k-medie classico, presenta in generale lo stesso tipo di problemi che consistono principalmente nella difficoltà di scelta delle condizioni iniziali.

Le maggiori questioni relative a tale algoritmo sulle quali si sono concentrati gli studiosi riguardano la scelta del valore del parametro m che regola il livello di *fuzziness* e la scelta del numero di gruppi, tali scelte sono essenziali per inizializzare l'algoritmo.

Le applicazioni empiriche effettuate con il metodo delle k-medie sfocato hanno messo in evidenza che non esiste un valore ottimale per il parametro m ma esso è variabile in base alle diverse applicazioni evidenziando così un limite di tale metodo poiché la scelta di m può essere fatta solamente osservando le classificazioni finali ottenute in corrispondenza dei diversi valori e, in ogni caso, se non si hanno conoscenze del fenomeno indagato risulta difficile scegliere il livello di sfocatura ideale delle classificazioni.

Una possibile soluzione proposta è il calcolo di un indice che misuri il grado di sfocatura delle diverse classificazioni considerando che una classificazione si dice totalmente sfocata se, detto c il numero di cluster, per ogni unità tutti i valori della funzione di appartenenza assumono valore $1/c$, oppure si dice classica se ogni unità appartiene ad un unico cluster.

Si noti che, secondo queste definizioni, il concetto di sfocatura è assimilabile a quello di eterogeneità, per cui una classificazione può essere detta più o meno sfocata a seconda se essa sia più o meno eterogenea. Quindi

si può considerare un qualunque indice di eterogeneità come un indice di sfocatura per cui per misurare il grado di sfocatura delle classificazioni si propone di utilizzare l'indice relativo di eterogeneità di Gini:

$$I = \frac{c}{c-1} \left[1 - \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^c u_{ik}^2 \right]$$

Tale indice varia nell'intervallo [0,1] assumendo valore 1 nel caso di massima sfocatura e valore 0 nel caso in cui la classificazione è di tipo classico. L'indice aiuta quindi nella scelta di m poiché fornisce una misura sintetica dell'effetto dei diversi valori di m sulle relative classificazioni.

Da diverse applicazioni effettuate il valore migliore di m sembra essere quello ottenuto in corrispondenza del valore dell'indice I compreso tra 0.4 e 0.5.

3.6. Criteri di cluster validity

Con il termine *cluster validity* si indica il calcolo di alcuni indici che siano in grado di misurare la qualità della partizione ottenuta con l'algoritmo di clustering e in particolare siano in grado di fornire informazioni utili per la scelta dei valori ottimali di c e di m . Gli indici di *cluster validity* possono classificarsi in

- Indici di *fuzziness*, che misurano il grado di sfocatura della partizione
- Indici di compattezza e separazione che inizialmente sono stati studiati per i metodi *crisp* ma poi sono stati estesi ai metodi *fuzzy*.
- Indici di compattezza e separazione specificatamente proposti per partizioni *fuzzy*

- Indici di *fuzziness* e compattezza/separazione che misurano sia il grado di sovrapposizione dei gruppi sia il loro grado di compattezza e eterogeneità tra i gruppi

3.6.1. Indici di *fuzziness*

Un primo criterio di *cluster validity* parte dall'osservazione dei valori della matrice dei gradi di appartenenza, infatti la qualità della classificazione può essere misurata attraverso l'associazione dei punti x_i ai centroidi v , se il valore di un grado di appartenenza a un gruppo per un unità è superiore a tutti gli altri gradi di appartenenza della stessa unità ad altri gruppi allora si può affermare che l'unità in questione appartiene al gruppo con il grado di appartenenza più elevato. Per valutare la *fuzziness* è necessario sintetizzare l'informazione contenuta nella matrice dei gradi di appartenenza in un unico numero che indichi il grado di accuratezza con cui avviene la classificazione.

Gli indici più diffusi sono il coefficiente di partizione *PC* e il coefficiente di entropia *PE* proposti da Bezdek:

$$PC = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^c u_{ik}^2$$

$$PE = \frac{1}{n} \sum_{i=1}^n \left(\sum_{k=1}^c u_{ik} \log u_{ik} \right)$$

L'indice *PC* vale $1/c$ in caso di *fuzziness* massima della partizione ovvero quando i gradi di appartenenza sono equidistribuiti tra i cluster, *PC* vale 1 se la partizione ottenuta è di tipo *crisp* ovvero ogni unità appartiene ad un solo gruppo, quindi i gradi di appartenenza sono tutti pari a 1 o a 0.

L'indice PE varia tra 0 e $\log c$ e aumenta all'aumentare della *fuzziness*, in caso di massima *fuzziness*, PE vale $\log c$, in caso di partizione *crisp*, PE vale 0 che equivale a dire che i gradi di appartenenza sono tutti o 1 o 0.

Gli indici PE e PC sono i due criteri più diffusi per misurare il grado di sovrapposizione tra i gruppi, tuttavia soffrono di una forte sensibilità al parametro m . Quando m è molto vicino a 1 o molto elevato i due indici perdono la loro capacità discriminante tra i vari valori di c . Bedzek ha infatti dimostrato che quando m è prossimo a 1 si ha che $PC=1$ e $PE=0$, mentre quando m è molto elevato $PC=1/c$ e $PE=\log c$. In tal caso, i due indici, preferiscono sempre $m=2$. Per superare tale tendenza dei due indici, Davè ha proposto nel 1996 un indice compreso tra 0 e 1 che vale 0 in caso di partizione massimamente *fuzzy* e 1 in caso di partizione massimamente *crisp*.

$$MPC(c) = 1 - \frac{c}{c-1}(1 - PC(c))$$

Gli indici di *fuzziness* proposti presentano alcuni limiti e sono stati soggetti ad alcune critiche. In particolare viene criticata l'esclusiva dipendenza di questi indici dalla matrice dei gradi di appartenenza senza considerare in alcun modo le informazioni contenute nei dati di partenza così come relativamente ai centroidi dei *cluster*.

Si ritiene opportuno quindi illustrare nel successivo paragrafo un altro tipo di indice di *fuzziness* che vada a integrare le informazioni della matrice partizionata *fuzzy* con le informazioni della matrice dei dati.

3.6.2. Indice fuzzy silhouette

Gli autori Campello e Hruschka nel 2006 hanno proposto un indice di *cluster validity* denominato *Fuzzy Silhouette* che prende spunto e comprende al suo interno l'indice *Crisp Silhouette*. Analizziamo prima brevemente il

coefficiente silhouette per classificazioni di tipo *crisp*. Indichiamo con p il cluster cui appartiene l'unità j .

Indichiamo con a_{pj} la distanza tra j e tutte le altre unità del cluster p .

Sia d_{qj} la distanza tra j e tutte le altre unità appartenenti ad un altro cluster q .

Sia b_{pj} la minima di queste distanze d_{qj} , che rappresenta la diversità tra l'unità j e l'unità più vicina di quelle appartenenti ad un altro cluster q . Possiamo allora definire con s_j il coefficiente *silhouette*:

$$s_j = \frac{(b_{pj} - a_{pj})}{\max(a_{pj}, b_{pj})}$$

Per tale coefficiente si auspica naturalmente un valore positivo che sta ad indicare che a_{pj} è minore di b_{pj} . Se il valore tende ad 1 possiamo considerare l'appartenenza dell'unità j al gruppo p molto soddisfacente perché vuol dire che a_{pj} sarà molto piccolo.

Per calcolare tale coefficiente per un singolo cluster bisogna considerare la media dei coefficienti silhouette di tutte le unità del cluster, definiamo quindi *Crisp Silhouette* l'indice:

$$CS = \frac{1}{N} \sum_{j=1}^N s_j$$

La migliore partizione si ha quando CS è massimizzata cioè quando la distanza intracluster (a_{pj}) è molto minore della distanza intercluster (b_{pj}).

Nel caso di una *fuzzy cluster* bisognerà considerare il grado di appartenenza dell'unità j al cluster p rispetto all'appartenenza rispetto agli altri cluster *fuzzy*.

Si definisce quindi *Fuzzy Silhouette* l'indice:

$$FS = \frac{\sum_{j=1}^N (u_{pj} - u_{qj})^\alpha s_j}{\sum_{j=1}^N (u_{pj} - u_{qj})^\alpha}$$

Dove u_{pj} e u_{qj} sono il primo e il secondo grado di appartenenza più elevati della matrice di partizione *fuzzy*. Il coefficiente s_j è l'indice silhouette dell'unità j . Il coefficiente α è un criterio di ponderazione.

L'indice FS è la media ponderata della silhouette individuale delle singole unità in cui il peso è dato dalla differenza tra il primo e il secondo miglior abbinamento delle unità al cluster *fuzzy* di appartenenza.

Per quanto riguarda le scelte a priori del *fuzzy k-means*, nei paragrafi precedenti sono state proposte alcune soluzioni, ma diversi problemi restano di difficile soluzione come per esempio la scelta iniziale del numero c di gruppi la quale può essere fatta solamente dopo aver compiuto un'accurata analisi dei dati a disposizione e dopo aver ripetuto la procedura di classificazione per diversi valori di c in modo tale da poter valutare quale sembra essere il valore migliore.

Nei paragrafi successivi verrà analizzata la metodologia dell'analisi in componenti principali utilizzata nella ricerca per ridurre il *dataset* iniziale e sintetizzarlo in un numero di componenti molto inferiore alle variabili di partenza con una limitata perdita di informazione.

3.7. L'analisi in componenti principali

L'analisi in componenti principali²¹ è una metodologia statistica multivariata per la semplificazione dei dati di partenza. Lo scopo primario di questa tecnica è la riduzione di un numero più o meno elevato di variabili

²¹ L'ACP nella formulazione attuale è stata proposta da Hotelling nel 1933 ma le basi sono già contenute in alcuni lavori di Bravais (1994) e Pearson (1901.)

(rappresentanti altrettante caratteristiche del fenomeno analizzato) in alcune variabili latenti.

L'informazione di partenza è costituita da una matrice di dati $n \times p$ dove n rappresenta le unità statistiche e p le variabili quantitative. Alla base di questa metodologia vi è la correlazione delle p variabili, poiché se le variabili sono correlate tra loro vuol dire che le dimensioni d'interesse spesso sono minori di p . L'obiettivo dell'ACP consiste nell'individuare opportune trasformazioni lineari delle variabili osservate facilmente interpretabili e capaci di evidenziare e sintetizzare l'informazione insita nei dati d'origine.

Tale strumento risulta utile soprattutto quando si ha un numero di variabili considerevole da cui si vogliono estrarre le maggiori informazioni possibili pur lavorando con un set più ristretto di variabili.

Nell'ambito della ricerca oggetto del presente lavoro, l'ACP è stata utilizzata per ridurre i molteplici indicatori individuati per la valutazione della competitività turistica delle regioni. È stato possibile data la correlazione tra gruppi di tali indicatori individuare un numero ridotto di dimensioni della competitività e utilizzare tali dimensioni per la successiva classificazione *fuzzy*.

L'ACP è particolarmente utile quando un certo aspetto non è direttamente quantificabile ma si dispone di più indicatori del medesimo, il ricercatore tende dunque a scoprire delle dimensioni sottese ("latenti") atte ad interpretare un fenomeno collettivo altrimenti difficilmente decifrabile.

L'analisi in componenti principali consente dunque di sostituire alle p variabili un nuovo insieme di variabili, definite componenti principali, che godono delle seguenti proprietà (Zani, 1999):

- sono incorrelate tra loro,
- sono elencate in ordine decrescente della loro varianza.

Una variabile con alta variabilità fornisce di norma più informazioni di una con bassa variabilità in quanto tende ad essere più dispersa e quindi a differenziare maggiormente le unità in termini della variabile in questione.

La determinazione della prima componente principale, y_1 , richiede l'individuazione del vettore p -dimensionale a_1 dei coefficienti della seguente combinazione lineare delle p variabili:

$$y_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1s}X_s + \dots + a_{1p}X_p$$

Ovvero in forma ridotta:

$$y_1 = Xa_1$$

dove X rappresenta la matrice dei dati mentre a_1 rappresenta il vettore dei coefficienti, infatti $a_1 = [a_{11}, \dots, a_{1s}, \dots, a_{1p}]'$.

Poiché, come già detto, la prima componente principale è la combinazione lineare di massima variabilità, e poiché la più utilizzata misura di variabilità è la varianza, si deve calcolare la varianza di Xa . La varianza totale di una trasformazione lineare di X è esprimibile in funzione della matrice di covarianza S ²²:

$$VAR(Xa_1) = a_1' S a_1$$

Si deve calcolare dunque il vettore dei coefficienti a_1 tale da massimizzare l'espressione precedente. Al fine di dare soluzione a tale problema si introduce il vincolo che $a_1' a_1 = 1$.

Il problema precedente diviene quindi un problema di massimo condizionato e richiede l'impiego di un moltiplicatore di Lagrange, λ , calcolando le derivate prime parziali dell'espressione da massimizzare diminuita del prodotto del moltiplicatore di Lagrange per il vincolo posto nella forma di uguaglianza a 0:

²² Nel caso in cui le variabili sono espresse in diverse unità di misura, si applica l'ACP partendo dalla matrice di correlazione.

$$\frac{\partial [a_1' S a_1 - \lambda (a_1' a_1 - 1)]}{\partial a_1}$$

Calcolando la derivata parziale si ottiene:

$$\begin{aligned} 2S a_1 - 2\lambda a_1 &= \\ &= 2(S - \lambda I) a_1 \end{aligned}$$

ove I indica la matrice di identità di dimensioni $p \times p$ in cui gli elementi della diagonale principale sono tutti uguali a 1, gli altri elementi sono tutti pari a 0.

L'uguaglianza a 0 delle derivate parziali conduce al sistema lineare di p equazioni con p incognite:

$$(S - \lambda_1 I) a_1 = 0$$

dove 0 è un vettore p -dimensionale i cui elementi sono tutti uguali a 0.

Il sistema ammette soluzioni non nulle se il suo determinante è uguale a 0, cioè se:

$$|S - \lambda_1 I| = 0$$

La suddetta uguaglianza definisce l'equazione caratteristica della matrice S , che è un polinomio di ordine p , con p soluzioni chiamate autovalori. Gli autovalori sono tutti non negativi perché la matrice di covarianza S è semidefinita positiva.

L'obiettivo è massimizzare la varianza della prima componente principale, per cui si sceglie come λ_1 il massimo di tali autovalori, in quanto vale la seguente uguaglianza:

$$\text{VAR}(Xa_1) = a_1' S a_1 = a_1' a_1 \lambda_1 = 1 \lambda_1 = \lambda_1$$

Il primo autovalore è quindi uguale alla varianza della prima CP.

Dopo aver trovato λ_1 lo andiamo a sostituire nell'espressione seguente:

$$(S - \lambda_1 I) a_1 = 0$$

trovando così i coefficienti del vettore a_1 , definito autovettore.

A questo punto: *“si definisce prima componente principale di p variabili, espresse in termini di scostamenti dalla loro media, la combinazione lineare $y_1 = Xa_1$, in cui a_1 è l'autovettore corrispondente all'autovalore più grande, λ_1 , della matrice di covarianza S ”* (Zani, 1999).

La seconda componente principale, ortogonale alla precedente, si calcola in maniera analoga, scegliendo come λ_2 il secondo autovalore in ordine decrescente della matrice di covarianza. A questo corrisponde l'autovettore a_2 .

In termini generali si può quindi definire la v -esima componente principale come la combinazione lineare $y_v = Xa_v$, in cui a_v è l'autovettore associato al v -esimo autovalore λ_v , in ordine decrescente, della matrice di covarianza.

Ogni autovalore λ_v è uguale alla varianza della corrispondente v -esima componente principale, per cui la somma di tutti gli autovalori è uguale alla varianza totale.

E' possibile inoltre calcolare il coefficiente di correlazione lineare tra la v -esima componente principale e la s -esima variabile:

$$r(Y_v, X_s) = \frac{a_{vs} \sqrt{\lambda_v}}{\sqrt{\text{var}(X_s)}}$$

Il segno di questo coefficiente segnala il tipo di relazione lineare, diretta o inversa, tra la componente e la variabile, mentre il valore numerico, in modulo, indica l'entità del legame.

L'obiettivo dell'analisi in componenti principali è quello di sostituire le variabili originarie con un numero di componenti che siano in grado di "spiegare" una percentuale elevata della varianza totale, in modo da avere una perdita limitata dell'informazione di partenza.

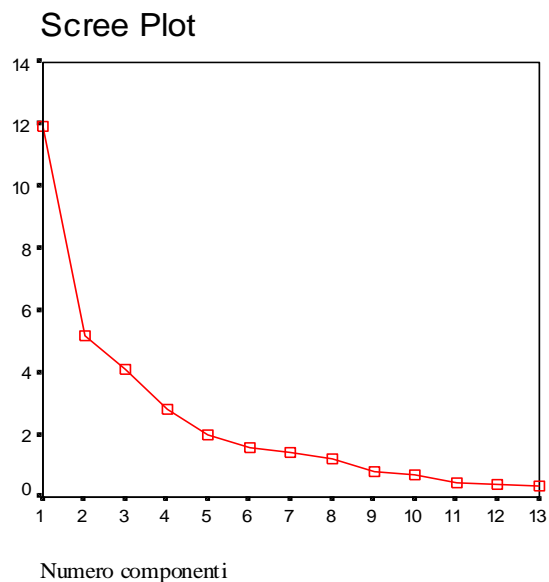
Il numero appropriato di componenti si individua pertanto considerando quelle componenti principali che spiegano una percentuale elevata della varianza totale (ad esempio l'80%). Naturalmente tale criterio va rivisto in funzione del numero di variabili che si cerca di sintetizzare, al crescere del numero delle p variabili la varianza totale cresce per cui ci si potrebbe accontentare di un numero di componenti che spieghi almeno il 95% della varianza di ognuna delle variabili di partenza e quindi abbassare la soglia di percentuale spiegata di varianza seguendo la formula:

$$0,95^p * 100.$$

La "Regola di Kaiser" è un altro criterio per la scelta del numero di componenti, si sceglie di prendere in considerazione solo quelle componenti che abbiano un autovalore maggiore o uguale ad uno. L'autovalore di una componente principale è uguale alla sua varianza. Quando le variabili sono standardizzate la varianza è unitaria. Pertanto la scelta di mantenere solo le componenti con autovalore >1 è dovuta al fatto che si sceglie di mantenere solo le componenti che spieghino una quota di varianza maggiore di quella di una singola variabile.

Un metodo alternativo consiste nella costruzione di una rappresentazione grafica degli autovalori λ_v , in funzione del numero v di componenti principali. Tale grafico chiamato *scree-plot* si presenta come una spezzata discendente. Il numero di componenti da prendere in considerazione si ottiene quando si manifesta una brusca variazione della pendenza, ad esempio se k componenti sono importanti e le restanti trascurabili, la variazione della pendenza si avrà tra k e $k+1$. Di seguito è riportato un esempio di *screeplot* (Fig.11).

Fig.11 Esempio di *screeplot*



Fonte: ns. elaborazione

Le componenti principali ottenute partendo dalla matrice di covarianza sono combinazioni lineari degli scostamenti dalla media delle variabili originarie. Il confronto fra tali variabili risulta possibile solo se le variabili sono espresse nella stessa unità di misura. L'ACP mira a riprodurre la varianza totale delle variabili di partenza sarà maggiormente o esclusivamente influenzata da quelle che presentano una variabilità molto maggiore delle altre.

Per cui l'applicazione dell'ACP partendo dalla matrice di covarianza è corretta solo nel caso di variabili espresse nella stessa unità di misura, con ordini di grandezza non molto differenti e variabilità non marcatamente diversa. Si possono quindi considerare le variabili standardizzate che è equivalente ad applicare l'ACP alla matrice di correlazione R.

Le fasi della procedura sono simili all'ACP partendo dalla matrice di covarianza e possiamo quindi definire la v -esima componente principale come la combinazione lineare :

$$y_v = Z a_v$$

in cui a_v è l'autovettore associato al v -esimo auto valore λ_v in ordine decrescente della matrice di correlazione R.

Un grafico denominato *biplot* risulta molto utile nell'interpretazione dei risultati dell'ACP, in cui vengono sintetizzate le informazioni ricavate dall'ACP rappresentando sul piano sia le unità sia le variabili.

Nel grafico i vettori rappresentano le variabili e sono sempre compresi nel cerchio di raggio unitario con il centro nell'origine. La lunghezza del vettore coincide con la quota di varianza della variabile spiegata dalle prime due componenti , per cui quando il vettore è pari al raggio vuol dire che la variabile è riprodotta quasi perfettamente nelle componenti estratte.

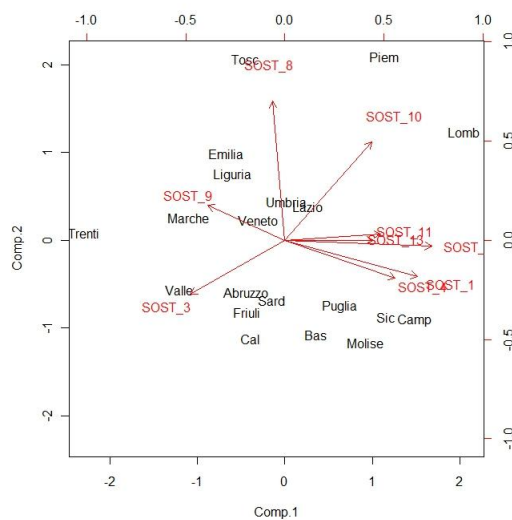
L'angolo tra il vettore e gli assi cartesiani segnala la correlazione tra la variabile e le due componenti. Se l'angolo è molto piccolo la variabile è fortemente correlata in maniera diretta o inversa secondo la direzione del vettore. Se l'angolo è prossimo a 90° la variabile e la corrispondente CP sono quasi incorrelate.

L'angolo tra due vettori, ovvero tra due variabili, segnala la correlazione tra esse, quando l'angolo è molto piccolo la correlazione è elevata e diretta, se l'angolo si avvicina ai 180° la correlazione è elevata ma inversa. L'angolo di 90° segnala una correlazione prossima a 0.

Nel grafico sono rappresentate anche le unità statistiche tenendo in considerazione che i punti vicini al centroide segnalano che le unità in questione ha valori prossimi alla media delle p variabili. Punti lontani dal centroide e nella direzione di quegli assi indicando che le relative unità sono caratterizzate da scores di quella variabile particolarmente elevati in modulo.

Il *biplot* quindi riesce a evidenziare le relazioni tra unità e variabili che sarebbe difficile ottenere dai singoli output dell'ACP. Naturalmente però tale grafico è utile solo nel caso in cui le prime due componenti siano in grado di spiegare una percentuale di varianza sufficientemente elevata (fig.12).

Fig.12 Esempio di biplot



Fonte: ns. elaborazione

3.8. Il software R per l'implementazione della metodologia

Nella ricerca oggetto della tesi si è scelto di utilizzare il software statistico R.

Da alcuni anni, soprattutto in ambito universitario (ma non solo), si sta sempre più diffondendo l'utilizzo di questo *package* che costituisce una valida alternativa ai *software* statistici più noti (SAS, SPSS, STATA, S PLUS) e merita l'approfondimento di coloro che si occupano di analisi dei dati.

Più che come *software*, R è definito come un ambiente statistico costituito da una varietà di strumenti, orientato alla gestione, all'analisi dei dati e alla produzione di grafici, in cui è possibile implementare una molteplicità di funzioni di calcolo. È proprio la flessibilità di questo *software* a renderlo sempre più utilizzato nella ricerca scientifica. Oltre alla flessibilità, R ha altre importanti caratteristiche che lo differenziano dagli altri *software* statistici. La prima rilevante differenza sta nel fatto che R è completamente gratuito e scaricabile da Internet sotto i vincoli della GPL (*General Public Licence*) sul sito di **The R Project for Statistical Computing** il cui indirizzo Internet è *www.r-project.org*.

R è un *software open source*, è quindi possibile per chiunque accedere al codice sorgente e modificarlo aumentando e integrando le possibilità di calcolo e utilizzo e mettendo a disposizione di tutti gli utilizzatori il proprio lavoro.

Inizialmente il codice sorgente di R fu sviluppato nel 1996 da Ross Ihaka e Robert Gentleman del dipartimento di Statistica dell'Università di Auckland in Nuova Zelanda. Da allora numerosi ricercatori statistici e informatici di fama mondiale hanno dato il loro contributo allo sviluppo e al miglioramento di questo *software*.

Nel 1997 è nato l'**R Development Core Team**, un gruppo formato da statistici di tutto il mondo che si occupano dello sviluppo e della distribuzione del programma.

Attualmente R è un prodotto molto affidabile, stabile e in continua evoluzione: nell'anno 2008 sono state rilasciate 6 diverse versioni (dalla 2.6.2 del mese di febbraio alla 2.8.1 del mese di dicembre). A partire dal 2000 sono apparsi nel mercato librario internazionale diversi testi di statistica basati su R, ed il loro numero è cresciuto di anno in anno. A questo proposito va anche segnalato che dal 2001 è iniziata la pubblicazione online di una rubrica

tecnico-scientifica, [R News](#) che poi, dal 2009, è diventata [The R Journal](#) ovvero una rivista scientifica *online*.

R si presenta come uno strumento completo e ben si adatta a molteplici esigenze: è il *software* maggiormente utilizzato a livello mondiale dai ricercatori in campo statistico soprattutto per la sua versatilità, ognuno ha la possibilità di crearsi strumenti personali di analisi statistica ad hoc per le proprie esigenze.

Esiste un'importante differenza a livello di filosofia tra **R** e la maggior parte degli altri *packages* statistici. Con gli altri *software* un'analisi statistica porta ad una grande quantità di output di informazioni e dati, mentre con **R**, l'analisi statistica è condotta come una serie di passi, con risultati intermedi, memorizzati in oggetti. Ad ogni passo dell'analisi gli output sono minimi e l'utente ha la possibilità di visualizzarli e manipolarli richiamando, quando è necessario, gli oggetti nei quali i risultati dell'analisi sono contenuti.

Nonostante le infinite potenzialità di questo valido strumento, spesso chi si avvicina a questo software lo considera “difficile”, soprattutto per l'interfaccia che è sostanzialmente molto differente dai principali software statistici. Infatti, in R i comandi vanno digitati e questo potrebbe risultare troppo “tecnico” per chi si avvicina per la prima volta a questo strumento.

Tuttavia, se si superano le perplessità iniziali, questo software offre delle potenzialità enormi e l'analisi statistica con questo software risulta davvero facilitata rispetto agli altri *software* commerciali.

CAPITOLO 4

L'analisi esplorativa della destination competitiveness delle regioni italiane

SOMMARIO: 4.1. Obiettivi e fasi della ricerca.- 4.2. La progettazione e creazione della matrice dei dati.- 4.3. Le determinanti della competitività e gli indicatori.- 4.3.1. Territorio.- 4.3.2. Clima. - 4.3.3. Offerta.- 4.3.4. Flussi.- 4.3.5. Accessibilità. - 4.3.6. Profilo economico-turistico.- 4.3.7. Sostenibilità e qualità.- 4.3.8. Performance imprese turistiche.- 4.4. Applicazione e risultati dell'analisi in componenti principali.- 4.4.1. ACP Territorio.- 4.4.2. ACP Clima.- 4.4.3. ACP Offerta.- 4.4.4. ACP Flussi.- 4.4.5. ACP Accessibilità.- 4.4.6. ACP Profilo economico-turistico.- 4.4.7. ACP Sostenibilità e qualità.- 4.4.8. ACP Performance imprese turistiche.- 4.4.8.1. ACP Performance settore alberghi.- 4.4.8.2. ACP Performance settore agenzie di viaggio e T.O. - 4.5. Applicazione e risultati della cluster analysis crisp. - 4.5.1. Applicazione e risultati cluster analysis gerarchica.- 4.5.2. Applicazione e risultati cluster analysis non gerarchica.- 4.6. Classificazione fuzzy della regioni italiane. - 4.7. I cluster regionali.- 4.7.1. Gruppo 1: Destinazioni di prossimità.- 4.7.2. Gruppo 2: Destinazioni leader.- 4.7.3. Gruppo 3: Destinazioni mature.- 4.7.3. Gruppo 4: Destinazioni da sviluppare.- 4.8 Conclusioni

4.1. Obiettivi e fasi della ricerca

Nei capitoli precedenti si è definito il contesto di riferimento (cap.1), i modelli teorici (cap.2), la metodologia statistica e il *software* (cap.3) utilizzati nell'ambito della ricerca.

Questo capitolo è dedicato alla descrizione dell'analisi empirica della *destination competitiveness* delle regioni italiane con l'esposizione delle varie fasi del lavoro e dei risultati conseguiti.

I modelli di competitività delle destinazioni analizzati nei capitoli precedenti mostrano l'elevata complessità che il tema della competitività assume e come siano numerosi i fattori da analizzare per valutare, misurare e gestire la competitività di una destinazione.

La competitività di un territorio è il risultato dell'interazione di numerosi fattori che concorrono insieme a definire quelle condizioni di unicità e inimitabilità che sono i presupposti del vantaggio competitivo.

Inserendosi nel dibattito sulla competitività turistica, lo scopo principale della ricerca è quello di individuare una metodologia d'analisi in chiave statistica-territoriale in grado di misurare quantitativamente la competitività dei sistemi turistici regionali e identificare cluster di regioni che esprimono analoghe potenzialità e performance turistiche competitive.

Lo studio della competitività turistica regionale attraverso l'analisi multidimensionale ha portato in una prima fase alla sintesi su poche dimensioni dei molteplici aspetti della competitività turistica considerati e successivamente si è potuta applicare sulle nuove variabili individuate la *cluster analysis* che ha permesso di classificare le regioni.

L'obiettivo del lavoro di ricerca è quello di rilevare se esistano sostanziali differenze in termini di competitività tra le varie regioni, cercando di indagarne le cause principali.

La prima fase del lavoro ha riguardato la progettazione di una base di dati che potesse raccogliere tutte le informazioni utili ai fini della ricerca in modo da poter applicare successivamente i metodi fattoriali al *dataset* definito.

Tale fase è risultata cruciale poiché, come evidenziato il grado di competitività turistica di una destinazione, è caratterizzato da una molteplicità di elementi da prendere in considerazione.

La selezione delle variabili è stata effettuata considerando vari e molteplici aspetti della competitività in termini turistici di una destinazione, sulla base dei principali *driver* di competitività studiati e analizzati nella

letteratura, considerando naturalmente gli aspetti quantitativi e i dati realmente disponibili per l'unità d'analisi considerata, ovvero le regioni italiane.

La scelta delle regioni italiane come destinazioni turistiche a confronto è dovuta principalmente a motivazioni di ordine pratico. In effetti dovendo valutare le potenzialità turistiche di un territorio attraverso l'utilizzo di una molteplicità di variabili, la dimensione regionale è quella che permette una disaggregazione sufficiente dal punto di vista territoriale, compatibile con una disponibilità di dati abbastanza ampia e riferibile ad anni sostanzialmente recenti. La scelta di porre come oggetto di osservazione le regioni nasce anche dall'attuale assetto normativo in tema di valorizzazione e promozione turistica, che attribuisce assoluta centralità ai sistemi regionali delegando loro poteri legislativi ed attribuendo agli stessi la stragrande maggioranza delle risorse finanziarie, è infatti a livello regionale che si definiscono le strategie e le azioni di sviluppo del turismo.

Nei successivi paragrafi è riportato l'insieme delle variabili prese in considerazione per ciascun aspetto della competitività considerato nella ricerca.

Successivamente alla fase determinante della predisposizione della base di dati, la seconda fase del lavoro ha riguardato la raccolta dei dati dalle fonti ufficiali e non del sistema statistico nazionale. Nell'attuale sistema statistico italiano, questa fase, ha rilevato elementi di criticità soprattutto relativamente alla tempestività, aggiornamento e completezza delle fonti ufficiali. Rispetto alla fase di progettazione, non è stato possibile reperire la totalità dei dati aggiornati all'anno di riferimento (2008). La ricerca di dati aggiornati si è spostata, quindi, su fonti di tipo secondario, quali enti ed istituti riconosciuti a livello nazionale (*ISSIRFA, ISPRA, TERNA, ACCREDIA, MINISTERI, ecc.*)

La terza fase, dopo il controllo e l'incrocio dei dati raccolti dalle varie fonti, ha interessato la creazione di indicatori che permettessero i confronti tra le regioni eliminando le influenze della dimensione territoriale o della popolazione.

I successivi passi riguardano l'elaborazione statistica sugli indicatori attraverso l'utilizzo del *software* R.

L'analisi in componenti principali ha permesso di ridurre il numero molto vasto degli indicatori considerati in un numero gestibile di fattori (componenti).

Sulle componenti estratte attraverso l'ACP sono state applicate varie tecniche di classificazione, in un primo momento si è applicata una cluster analysis di tipo gerarchico che attraverso il taglio del dendrogramma ha fornito le indicazioni utili sul numero di gruppi da considerare nella successiva analisi *cluster* di tipo non gerarchico, sia *crisp* sia *fuzzy*.

Con l'analisi di tipo *crisp* si sono definiti i gruppi di regioni simili per gli aspetti della competitività considerati mentre attraverso l'analisi *fuzzy* si sono potuti individuare cluster sfocati, evidenziando per ogni gruppo il grado di appartenenza di ciascuna regione a quel gruppo.

L'analisi, dunque, ha consentito di individuare gruppi regionali che mostrano un grado di competitività sostanzialmente omogeneo e allo stesso tempo evidenziare le cause delle diverse performance competitive tra i cluster individuati.

4.2 La progettazione e la creazione della matrice dei dati..

La matrice dei dati alla base delle elaborazioni statistiche effettuate nell'ambito dell'analisi sulla competitività delle regioni italiane, è costruita con dati provenienti da fonti ufficiali e non del sistema statistico italiano.

Il punto di riferimento principale è l'Istat con le rilevazioni sul movimento dei turisti e sulla capacità ricettiva degli esercizi alberghieri e complementari²³. Per quanto concerne l'andamento del turismo in Italia,

²³ L'indagine "Movimento dei clienti negli esercizi ricettivi" è un'indagine censuaria in cui sono rilevati con cadenza mensile gli arrivi, le presenze e la permanenza media negli esercizi ricettivi determinando quindi i flussi di turisti italiani e internazionali nel territorio italiano. L'altra indagine censuaria "Capacità degli esercizi ricettivi" rileva con periodicità annuale il numero di esercizi, di camere, di letti e di bagni delle strutture alberghiere e il numero di

l'ISTAT (www.istat.it) offre un quadro molto ricco di dati sull'offerta e sulla domanda dettagliati a livello delle singole circoscrizioni turistiche. Tali dati, pur sottostimando il 'fenomeno' turismo nel suo complesso in ragione dell'estrema difficoltà di 'misurare' il sommerso delle seconde case, costituiscono la principale fonte utilizzata da tutti i centri di ricerca nazionali per la stesura di rapporti sulla competitività della destinazione Italia. Nonostante la ricchezza dei dati Istat a livello di domanda e offerta, è da rilevare che il sistema statistico italiano presenta molteplici carenze per ciò che riguarda la tempestività e varietà delle ulteriori informazioni necessarie in un lavoro come quello oggetto della tesi.

Nella ricerca, infatti, i fattori considerati a livello regionale investono vari aspetti caratterizzanti un territorio dal punto di vista sia strettamente turistico che ad esso correlato. Non ci si limita solo ed esclusivamente alla valutazione della presenza di strutture ricettive o dei flussi turistici, ma si inserisce nel *dataset* una serie di elementi che sono indissolubilmente legati allo sviluppo turistico (sostenibilità, qualità, accessibilità, performance delle imprese), è proprio nella ricerca di questo tipo di informazioni, dettagliate a livello regionale, che si sono riscontrate le maggiori difficoltà. Data la carenza di informazioni Istat in questo senso, la ricerca dei dati si è orientata verso altre fonti.

4.3. Le determinanti della competitività e gli indicatori.

Nella fase di progettazione del *dataset* si sono presi in considerazione i vari aspetti della competitività misurabili attraverso variabili di tipo

esercizi e di posti letto per le strutture complementari. Anche l'indagine svolta dall'Ufficio italiano cambi, soggetto pubblico ente strumentale della Banca d'Italia è un'indagine campionaria che quantifica i flussi turistici monetari e fisici in entrata in Italia e in uscita dall'Italia.

quantitativo, necessarie per l'applicazione della strategia d'analisi statistica ipotizzata.

Nelle pagine successive vengono descritte le varie categorie ritenute rilevanti nello studio della competitività turistica delle regioni. Ogni categoria a sua volta raggruppa gli indicatori utilizzati per l'analisi statistica esplorativa con l'indicazione delle fonti dei dati.²⁴

4.3.1. Territorio

Fonti: ISTAT, Ministero del Lavoro della Salute e delle Politiche sociali, Centro Studi e Ricerche sul Mezzogiorno

L'analisi della competitività delle regioni come destinazioni turistiche non può prescindere dall'individuazione delle risorse turistiche ovvero dagli elementi di attrattività turistica del territorio regionale. Nel gruppo Territorio sono raggruppati gli indicatori relativi al patrimonio di risorse turistiche delle regioni. Le risorse considerate sono risorse naturali e culturali, ma anche quelle definite "immateriali" che sono però analizzate considerando gli aspetti tangibili dell' "heritage" territoriale. Alla luce di quanto detto nel primo capitolo sui cambiamenti nella domanda turistica, che è sempre più alla ricerca di esperienze e luoghi autentici in cui si respira la tradizione del territorio, si è ritenuto interessante l'inserimento nel *dataset* di variabili relative al numero delle città dell'olio, del vino, del pane, al numero dei prodotti agroalimentari tradizionali e al numero dei borghi con il riconoscimento del marchio Borghi più belli d'Italia, poiché, queste risorse hanno le potenzialità per essere sfruttate e valorizzate come veri e propri prodotti turistici. Tali variabili sono state affiancate agli indicatori sulle risorse più comunemente sfruttate dal punto di vista turistico, quali il mare, la montagna e le risorse culturali.

Di seguito sono elencate le variabili inserite per il gruppo Territorio:

²⁴ *Gli indicatori riportati sono quelli inizialmente ipotizzati, tuttavia in alcuni casi non sono stati tutti effettivamente utilizzati nell'applicazione della strategia di analisi statistica.*

TERR_1: Km costa % (*km costa/superficie regionale *100*)

TERR_2: Superficie montana %

TERR_3: Siti Unesco.

TERR_4: Beni culturali censiti.

TERR_5: Borghi più belli d'Italia.

TERR_6: Città del vino, dell'olio e del pane.

TERR_7: Prodotti agroalimentari tradizionali.

Tab.7 Territorio – Matrice dei dati.

REGIONI	TERR_1	TERR_2	TERR_3	TERR_4	TERR_5	TERR_6	TERR_7
Piemonte	0	51.08	2	4080	10	98	369
Valle D'Aosta	0.00	100.00	0	370	1	5	29
Lombardia	0.00	43.30	6	7579	12	42	199
Trentino Alto Adige	0.00	100.00	0	1806	6	20	200
Veneto	0.86	32.00	4	5212	4	49	373
Friuli-Venezia Giulia	1.42	56.9	1	1350	6	28	143
Liguria	6.44	81.5	2	2999	17	41	287
Emilia-Romagna	0.59	38.5	3	3881	8	44	218
Toscana	2.61	47.30	6	4436	16	105	455
Umbria	0.00	85.80	1	2097	21	51	70
Marche	1.78	59.40	1	2531	15	54	148
Lazio	2.10	44.20	4	5718	10	38	336
Abruzzo	1.17	76.60	0	1614	17	38	143
Molise	0.80	79.70	0	474	1	39	159
Campania	3.46	56.40	5	3204	6	81	329
Puglia	4.47	24.80	2	1584	9	52	190
Basilicata	0.62	71.30	1	913	4	18	42
Calabria	4.75	65.70	0	1528	8	18	273
Sicilia	5.77	36.70	5	3708	7	61	239
Sardegna	7.19	74.50	1	2093	2	40	170
Italia	2.45	51.86	43	57177	180	922	4372

4.3.2. *Clima*

Fonti: CRA-CMA Consiglio per la ricerca e la sperimentazione in Agricoltura- Unità di ricerca per la climatologia e la meteorologia applicate all'agricoltura.

In tale gruppo rientrano gli indicatori relativi al clima delle regioni. Il clima favorevole è sicuramente un elemento di attrattiva che si aggiunge alle risorse turistiche del territorio²⁵.

CLI_1: Temperatura media max annua

CLI_2: Temperatura minima media annua

CLI_3: Precipitazione cumulata mm

Tab.8 Clima – Matrice dei dati

REGIONE	CLI_1	CLI_2	CLI_3
Piemonte	13.70	4.1	952.7
Valle D'Aosta	8.50	-1.7	786.4
Lombardia	15.40	6.1	913.1
Trentino	10.40	1.3	1066.4
Veneto	16.50	6.9	1037
Friuli	15.60	6.4	1327.7
Liguria	17.40	8.8	987.4
Emilia	18.00	8.2	858.1
Toscana	18.40	9	923.5

²⁵ *Nell'applicazione dell'ACP si è preferito tenere separate le variabili che si riferiscono al clima da quelle concernenti il territorio e creare un indicatore che sintetizzasse le condizioni climatiche della regione.*

<i>Segue tab.8 Clima – Matrice dei dati</i>			
Umbria	17.70	8.5	943.4
Marche	18.00	8.9	843.1
Lazio	19.40	8.8	984.2
Abruzzo	18.30	6	936.4
Molise	19.40	8.5	884.6
Campania	19.60	10.1	841.2
Puglia	21.30	11.7	579
Basilicata	19.90	10.5	637.9
Calabria	21.20	12.6	608
Sicilia	22.80	14	518
Sardegna	21.80	12.1	556.5
Italia	18.8	9.4	789.8

4.3.3. Offerta

(Fonte: Istat)

Le variabili inserite nel gruppo *Offerta* sono relative a tutti i servizi e le *facilities* che rendono fruibile un territorio e le sue risorse da parte dei turisti.

Una destinazione per essere competitiva deve saper costruire il prodotto turistico, e gli elementi dell'offerta sono indiscutibilmente importanti per la competitività globale. Nell'analisi sono stati inseriti indicatori relativi alla capacità ricettiva sia di tipo alberghiero che extralberghiero in termini di posti letto regionali. Gli indicatori sono stati opportunamente costruiti per eliminare l'influenza della diversa dimensione regionale.

Gli indicatori del gruppo OFFERTA sono :

OFF_1: N° medio posti letto ogni 1000 abitanti (*Totale posti letto/Popolazione *1000*)

OFF_2: Dimensione media esercizi ricettivi (*Posti letto/Strutture ricettive*)

OFF_3: N° medio posti letto alberghieri ogni 1000 abitanti (*Totale posti letto alberghi/Popolazione*1000*)

OFF_4: Dimensione media strutture alberghiere (*Numero letti / Numero totale esercizi*)

OFF_5: N° medio posti letto extralberghieri ogni 1000 abitanti (*Totale posti letto extralberghieri/Popolazione*1000*)

OFF_6: Dimensione media strutture extra alberghiere (*Numero letti / Numero totale esercizi*)

OFF_7: Indice di ricettività alberghiera % (*n° alberghi e residenze turistica/totale esercizi ricettivi*)

OFF_8: Indice di qualità delle strutture alberghiere (*Totale letti alberghi 4/5 stelle / Totale letti alberghieri*)

Tab. 9 Offerta –Matrice dei dati

REGIONE	OFF_1	OFF_2	OFF_3	OFF_4	OFF_5	OFF_6	OFF_7	OFF_8
Piemonte	41,20	38,01	19,03	53,82	22,17	30,36	32,61	28,15
Valle d'Aosta	419,37	54,54	183,63	47,33	235,74	61,89	50,46	19,13
Lombardia	33,49	57,54	19,77	65,13	13,71	49,26	52,17	43,29
Trentino	374,48	29,29	240,18	41,74	134,29	19,10	45,01	15,91
Veneto	141,59	14,49	45,15	67,91	96,44	10,59	6,80	32,43
Friuli	125,08	16,08	32,25	53,71	92,83	12,94	7,72	23,03
Liguria	100,82	40,46	46,00	46,32	54,82	36,58	39,86	19,75
Emilia	99,69	51,50	68,31	64,17	31,38	36,03	55,00	17,96
Toscana	136,35	44,47	51,97	65,35	84,38	37,16	25,94	32,07
Umbria	94,83	23,87	32,44	51,35	62,38	18,67	15,90	25,60
Marche	148,02	75,09	39,87	62,65	108,15	81,02	32,29	14,85
Lazio	51,29	36,95	27,92	82,08	23,37	22,30	24,51	46,99
Abruzzo	78,82	51,70	37,79	61,20	41,04	45,23	40,49	21,78
Molise	41,77	42,27	20,03	58,95	21,74	33,52	34,38	32,88
Campania	31,96	48,09	18,47	66,04	13,48	35,03	42,09	47,13
Puglia	54,53	61,59	19,98	88,23	34,55	52,43	25,58	40,89
Basilicata	64,84	62,78	38,00	96,31	26,85	42,06	38,20	36,51
Calabria	98,39	90,74	49,05	120,00	49,34	73,04	37,70	39,16
Sicilia	37,08	45,18	23,22	96,83	13,86	23,86	29,22	40,06
Sardegna	121,83	58,56	60,35	112,80	61,48	39,79	25,72	49,59
ITALIA	77,43	33,15	36,67	64,47	40,76	23,06	24,35	31,43

4.3.4. Flussi

Fonte: ISTAT

Nell'analisi del livello di competitività di un territorio risulta sicuramente essenziale la valutazione della performance in termini di flussi turistici attratti. Per questa ragione sono stati introdotti gli indicatori relativi alla domanda turistica attraverso i quali è possibile conoscere la rilevanza complessiva del turismo per una determinata regione.

Particolare attenzione è stata attribuita all'attrazione nella regione dei flussi turistici stranieri, poiché spesso rappresentano la componente della domanda con maggiore capacità di spesa e con impatto positivo sulla bilancia turistica regionale e nazionale. Gli indicatori sono:

FLS_1: Arrivi Italiani ogni 100 abitanti (*Arrivi italiani/popolazione*100*)

FLS_2: Permanenza media italiani (*Presenze italiani/Arrivi italiani*)

FLS_3: Arrivi Stranieri ogni 100 abitanti (*Arrivi stranieri /popolazione *100*)

FLS_4: Permanenza media stranieri (*Presenze stranieri/Arrivi stranieri*)

FLS_5: Variazione % arrivi totali(2003-2008)

FLS_6: Variazione % presenze totali(2003-2008)

FLS_7: Indice attrazione stranieri

(*totale presenze stranieri nella regione/totale presenze stranieri in Italia*)

FLS_8: % presenze stranieri/tot presenze nella regione

FLS_9: Indice di utilizzazione netta²⁶ italiani

FLS_10: Indice di utilizzazione netta stranieri

²⁶ E' dato dal rapporto tra le presenze (P) registrate e il numero di giornate letto di effettiva apertura (G o altre chiusure temporanee), dove il numero di giornate letto di effettiva apertura è dato dal prodotto tra il numero di **giorni di apertura dell'esercizio** (ge) per il **numero di posti letto** (L.)

Tab.10 Flussi – Matrice dei dati

REGIONI	FLS_1	FLS_2	FLS_3	FLS_4	FLS_5	FLS_6	FLS_7	FLS_8	FLS_9	FLS_10
Piemonte	52.23	3.23	26.18	3.52	23.70	29.20	0.03	35.36	18.2	9.9
Valle d'Aosta	450.43	3.57	224.90	3.74	-1.30	-11.00	0.01	34.36	23.6	15.0
Lombardia	59.26	2.33	51.37	2.96	21.30	9.00	0.09	52.39	17.8	18.9
Trentino	394.30	4.85	435.60	5.20	13.60	7.60	0.14	54.23	26.6	31.0
Veneto	114.46	4.46	174.75	4.18	20.10	10.00	0.22	58.87	18.1	27.5
Friuli	91.48	4.53	67.28	4.56	14.60	0.20	0.02	42.49	17.9	13.9
Liguria	147.56	4.19	76.31	3.36	5.20	-4.30	0.03	29.34	30.1	13.3
Emilia	154.20	4.38	47.74	4.36	13.90	4.80	0.06	23.56	35.9	10.9
Toscana	146.14	3.97	146.96	3.62	13.20	12.00	0.12	47.82	16.8	16.4
Umbria	173.00	2.54	66.74	3.49	8.60	3.70	0.01	34.68	22.8	9.2
Marche	110.75	5.64	21.12	5.04	-1.20	-14.70	0.01	14.56	28.9	5.8
Lazio	67.04	2.80	122.25	3.07	28.70	31.70	0.13	66.67	14.4	34.9
Abruzzo	107.40	4.56	14.45	5.29	14.00	6.30	0.01	13.50	28.0	4.0
Molise	56.32	3.37	4.59	3.37	-0.10	-14.30	0.00	7.53	14.6	1.5
Campania	48.55	3.94	28.53	4.59	-1.10	-5.00	0.05	40.64	22.1	15.6
Puglia	61.26	4.19	10.37	4.05	27.20	13.80	0.01	14.07	25.5	4.7
Basilicata	70.19	4.05	8.76	3.51	9.20	5.70	0.00	9.74	20.0	2.6
Calabria	64.60	5.41	11.45	6.38	17.50	15.80	0.01	17.29	23.8	5.5
Sicilia	51.48	3.23	31.99	3.45	2.90	6.00	0.03	39.87	19.5	14.1
Sardegna	93.61	5.38	47.88	4.85	23.50	18.40	0.02	31.57	24.1	12.2
Italia	89.52	3.94	69.61	3.87	15.50	8.50	1.00	43.30	22.4	17.6

4.3.5. Accessibilità

(Fonti: Ministero delle Infrastrutture e dei Trasporti, Ferrovie regionali, ENAC)

In tale gruppo sono inseriti gli indicatori relativi all'accessibilità della regione in termini di collegamenti stradali, ferroviari e aeroportuali. E' evidente infatti che più una regione risulta facilmente accessibile da parte dei turisti più potrà competere nel mercato nazionale e internazionale abbattendo i costi di trasporto e i tempi per gli spostamenti. Gli indicatori inseriti in tale gruppo sono:

ACS_1: km strade regionali e provinciali ogni 100 km di superficie regionale

ACS_2: km altre strade d'interesse nazionale ogni 100 km di superficie regionale

ACS_3: Km autostrade ogni 100 km di superficie regionale

ACS_4: Km totali ferrovie (ferrovie dello stato e regionali) ogni 100 km di superficie regionale

ACS_5: % rete elettrificata su totale rete ferroviaria

ACS_6: n° aeroporti attivi

Tab.11 Accessibilità - Matrice dei dati

REGIONE	ACS_1	ACS_2	ACS_3	ACS_4	ACS_5	ACS_6
Piemonte	51.16	2.913	3.216	7.8	70.0	8
Valle d'Aosta	15.32	4.689	3.494	2.5	0.0	1
Lombardia	59.54	4.140	2.460	8.8	79.5	14
Trentino	37.94	0.000	1.551	3.1	84.3	2
Veneto	51.73	4.392	2.679	6.8	62.1	12
Friuli	40.62	2.431	2.672	5.9	80.9	3
Liguria	87.09	2.490	6.916	9.7	97.0	4
Emilia	50.90	5.539	2.568	7.1	78.0	12
Toscana	54.34	4.205	1.896	6.6	66.8	8
Umbria	54.07	6.623	0.698	6.1	67.8	1
Marche	60.44	5.271	1.733	4.0	62.2	2
Lazio	59.88	3.174	2.727	8.1	92.1	9
Abruzzo	70.70	9.635	3.270	6.0	65.4	2
Molise	61.00	12.866	0.811	6.1	24.4	0
Campania	71.23	10.000	3.252	10.5	80.1	3
Puglia	52.58	8.110	1.617	8.5	45.6	5
Basilicata	48.67	10.395	0.290	3.6	59.9	0
Calabria	53.31	9.442	1.956	7.1	45.6	3
Sicilia	56.47	9.319	2.544	5.8	54.0	6
Sardegna	22.34	12.744	0.000	4.3	0.0	5
Italia	52.36	6.401	2.200	1.1	64.7	98

4.3.6. Profilo economico-turistico.

Fonti : ISTAT, ISSIRFA, UIC, Banca d'Italia

Attraverso gli indicatori inseriti in questo gruppo si vuole misurare la capacità delle regioni di trasformare in impatti positivi lo sfruttamento turistico

del territorio in termini di occupazione nel settore e di entrate misurate dalla bilancia dei pagamenti turistici. Sono inseriti anche indicatori sull'impegno economico della regione in materia di turismo. Gli indicatori del gruppo sono:

PET_1: % occupati del settore turistico sul totale occupati

PET_2: Bilancia dei pagamento turistica (*Spesa stranieri in Italia – Spesa italiani all'estero*)

PET_3: Spesa per pernottamento stranieri

PET_4: Spesa per viaggiatore straniero

PET_5: Incidenza spesa turistica su spesa complessiva della regione

PET_6: Spesa turistica regionale per arrivo

PET_7: Spesa turistica regionale per presenza

PET_8: Pil pro-capite ai prezzi di mercato

Tab.12 Profilo economico-turistico: matrice dei dati

REGIONE	PET_1	PET_2	PET_3	PET_4	PET_5	PET_6	PET_7	PET_8
Piemonte	4.5	-456	82.35	411.17	0.64	22.44	6.75	28721
Valle d'Aosta	8.7	236	98.57	266.49	3.82	72.25	19.91	33834
Lombardia	3.9	-664	128.02	287.85	0.11	2.13	0.81	33648
Trentino-Alto Adige	8.3	819	73.49	274.09	1.42	15.61	3.10	32696
Veneto	5.5	2958	85.95	405.66	0.42	3.75	0.87	30456
Friuli-Venezia Giulia	4.4	299	90.31	126.55	1.25	34.80	7.66	29394
Liguria	6.8	652	103.74	211.69	0.65	10.23	2.62	27143
Emilia-Romagna	4.7	-148	77.84	434.28	0.24	4.57	1.04	32397
Toscana	6.8	2664	91.26	508.67	0.34	3.04	0.80	28727
Umbria	5.3	30	70.75	495.81	0.60	9.33	3.33	24455
Marche	5.2	-141	72.40	319.65	0.14	2.90	0.52	26652
Lazio	5.2	2400	112.99	554.60	0.33	8.26	2.78	30623
Abruzzo	5.2	-115	61.05	611.97	0.15	4.30	0.93	21949
Molise	4.6	-39	47.75	538.46	0.27	20.47	6.07	20259
Campania	5.8	163	108.06	619.42	0.21	7.59	1.82	16867
Puglia	4.3	-63	59.36	468.38	0.10	3.08	0.74	17520
Basilicata	3.9	-68	72.12	493.33	2.32	186.58	46.71	18954
Calabria	5.0	-15	57.54	713.71	0.75	46.48	8.36	17008
Sicilia	4.4	674	59.91	563.93	0.13	7.85	2.37	17443
Sardegna	6.4	249	80.58	770.97	0.33	11.00	2.11	20402

4.3.7. Sostenibilità e qualità

Fonti: ISTAT, Ministero del lavoro, della salute e delle politiche sociali- Rapporto acque di balneazione 2009, Touring Club, ISNART, APAT, TERNA, ACCREDIA

La sostenibilità è sicuramente una delle determinanti più importanti per la competitività di un territorio. La sostenibilità e la responsabilità determinano sempre di più le scelte del turista concernenti le modalità di viaggio e di mobilità nel territorio, le destinazioni turistiche, le strutture ove soggiornare o i ristoranti da frequentare. La progettazione della base di dati ha cercato di identificare le variabili quantitative che esprimessero al meglio l'impatto dell'attività turistica sui territori regionali e l'attenzione delle regioni e delle attività turistiche al tema ambientale.

Le prime due variabili misurano l'impatto del fenomeno turistico rispetto alla popolazione sia in termini di presenze che di posti letto, misurando allo stesso tempo anche la maggior propensione turistica di una regione.

La variabile relativa alla permanenza media è stata inserita in relazione al fatto che, presumibilmente, laddove i giorni di permanenza siano maggiori, si riducono gli impatti negativi del turismo di breve durata e si aumentino gli impatti positivi in termini economici.

L'indice di qualità delle strutture oltre ad indicare la presenza sul territorio di strutture di alta qualità, fornisce indirettamente una misura del grado di attrattività di un territorio e del suo sfruttamento ambientale, si presuppone infatti che le località servite da strutture di fascia più alta siano meno votate al turismo di massa quindi indirizzate ad un certo tipo di utenza più sensibile alle tematiche ambientali.

Nella selezione delle variabili, si è voluto dare ampio spazio alle certificazioni ambientali e marchi di qualità come strumenti capaci di elevare il livello qualitativo dei prodotti e dei processi e indirizzare il mercato verso la sostenibilità attraverso la loro riconoscibilità da parte dei turisti. Si ritiene,

infatti che il numero delle certificazioni ottenute sia uno strumento in grado di sintetizzare al meglio la propensione alla sostenibilità di una determinata regione sia a livello di imprese che di territori. Le certificazioni si ottengono grazie ad una serie di criteri molto restrittivi per cui possono fungere sicuramente da metro di giudizio per gli aspetti che si stanno indagando. Un marchio ecologico indica ai consumatori i prodotti/servizi migliori dal punto di vista ambientale, che rispettano determinati criteri ecologici stabiliti da un ente indipendente.

Analizziamo brevemente i vari tipi di “certificazione” inseriti nell’analisi.

Le Bandiere Arancioni del *Touring Club Italiano* rappresentano il marchio di qualità turistico ambientale rivolto alle piccole località dell'entroterra che si distinguono per un’offerta di eccellenza e un’accoglienza di qualità.

L’ottenimento del marchio avviene in base a diversi criteri tra i quali: la valorizzazione del patrimonio culturale, la tutela dell'ambiente, la cultura dell'ospitalità, l'accesso e la fruibilità delle risorse, la qualità della ricettività, della ristorazione e dei prodotti tipici, il rafforzamento dell’identità locale. Il marchio ha validità biennale e richiede il mantenimento dei requisiti nel tempo.

La Bandiera arancione intende stimolare una crescita sociale ed economica attraverso lo sviluppo sostenibile del turismo.

Il progetto Marchio Ospitalità Italiana viene promosso dall’Istituto Nazionale Ricerche Turistiche (IS.NA.R.T.) in collaborazione con le Camere di Commercio d’Italia fin dal 1997.

La variabile relativa al Marchio di Ospitalità è inserita nel lavoro per l’approccio ecocompatibile e di tutela ambientale richiesto alle strutture che intendono aderire al progetto. La struttura che intende seguire un approccio eco-compatibile deve dimostrare particolare attenzione alle problematiche relative agli aspetti ambientali:

1. limitazione al consumo delle risorse naturali (consumi energetici ed idrici)

2. riduzione della produzione di rifiuti (raccolta differenziata)
3. promozione di comportamenti e scelte per acquisti eco compatibili
4. formazione al personale e sensibilizzazione sui temi dell'ambiente anche verso gli ospiti.

L'*Ecolabel* europeo è il marchio europeo di certificazione ambientale per i prodotti e i servizi. Il servizio di ricettività turistica è il primo servizio a cui è stato concesso il marchio *Ecolabel*. I criteri ambientali per ottenere l'assegnazione del marchio hanno l'obiettivo di limitare i principali impatti ambientali relativi alle tre fasi caratterizzanti il ciclo di vita del servizio: acquisto, erogazione del servizio e produzione dei rifiuti.

In particolare tali criteri dovrebbero essere rispettati al fine di:

- limitare il consumo energetico
- limitare il consumo idrico
- limitare la produzione di rifiuti
- favorire l'utilizzo di fonti rinnovabili e di sostanze che risultano meno pericolose per l'ambiente
- promuovere la comunicazione e l'educazione ambientale.

Si elencano di seguito tutti gli indicatori considerati per il gruppo Sostenibilità e qualità:

SOST_1: Tasso di turisticità (*presenze/abitanti*)

SOST_2: Tasso di ricettività (*posti letto/abitanti*)

SOST_3: Permanenza media

SOST_4: Letti 4/5 stelle su totale letti alberghieri %

SOST_5: Verde per abitante (*Metri quadri di verde urbano, gestito da comuni, province, regioni e stato, nei comuni capoluogo di provincia per abitante*)

SOST_6: % costa balneabile su costa controllata

SOST_7: N° bandiere blu ogni 10 km di costa balneabile

SOST_8: N° bandiere arancioni

SOST_9: N° strutture turistiche certificate Ecolabel

SOST_10: N° strutture con Marchio Ospitalità Italiana

SOST_11: % Consumi elettrici settore Alberghi,Bar e ristoranti su totale consumi elettrici della regione

SOST_12: % rifiuti urbani oggetto di raccolta differenziata

SOST_13: N° totale organizzazioni certificate settore Alberghi e ristoranti

Tab.13 Sostenibilità e qualità -Matrice dei dati

REGIONE	SOST_1	SOST_2	SOST_3	SOST_4	SOST_5	SOST_6	SOST_7	SOST_8	SOST_9	SOST_10	SOST_11	SOST_12	SOST_13
Piemonte	3	0,041	3.33	28.15	42.5	0	0	19	16	932	2,72	48,5	6
Valle D'Aosta	25	0,423	3.63	19.13	26.2	0	0	0	0	0	5,82	38,6	0
Lombardia	3	0,034	2.63	43.29	28.6	0	0	9	1	956	2,77	46,2	37
Trentino	42	0,377	5.04	15.91	70.3	0	0	4	67	0	9,50	56,8	1
Veneto	12	0,142	4.29	32.43	62.3	89.8	0,55	9	2	315	3,80	52,9	29
Friuli	7	0,126	4.54	23.03	22.1	97.7	0,34	1	1	23	2,55	42,6	2
Liguria	9	0,101	3.91	19.75	35.4	97.6	0,50	10	2	451	5,81	21,8	8
Emilia	9	0,100	4.38	17.96	157.7	100	0,80	18	6	268	4,10	42,7	15
Toscana	11	0,137	3.8	32.07	152.1	99.8	0,38	32	11	278	4,64	33,6	29
Umbria	7	0,095	2.8	25.6	187.6	0	0,00	9	1	182	3,00	28,9	9
Marche	7	0,149	5.55	14.85	186.1	99.2	0,99	16	0	144	4,26	26,3	2
Lazio	6	0,052	2.97	46.99	121	92.8	0,11	17	1	91	5,39	12,9	6
Abruzzo	6	0,079	4.65	21.78	710	95.8	1,16	2	0	197	4,06	21,9	5
Molise	2	0,042	3.37	32.88	18.5	100	0,58	1	0	26	3,30	6,5	7
Campania	3	0,032	4.18	47.13	25.9	80.8	0,32	2	0	298	4,52	19,0	87
Puglia	3	0,055	4.17	40.89	8.1	98	0,07	2	3	286	3,22	10,6	12
Basilicata	3	0,065	3.99	36.51	545.6	100	0,17	1	0	70	3,05	9,1	3
Calabria	4	0,098	5.56	39.16	20.8	93.3	0,05	1	1	286	5,59	12,7	6
Sicilia	3	0,037	3.31	40.06	73.3	98.8	0,04	1	1	165	3,66	6,7	38
Sardegna	7	0,122	5.2	49.59	85.9	99.9	0,02	6	4	240	3,46	34,7	13
Italia	6	0,078	3.94	31.43	93,6	96,2	0,21	160	117	5208	3,82	30,6	315

4.3.8. Performance imprese turistiche

Fonti: Banca dati Aida BVD

L'introduzione nell'analisi di tale gruppo di variabili è dovuta alla considerazione che la competitività di un territorio passa inevitabilmente anche dalla competitività delle imprese che in esso operano. Nell'analisi si sono voluti introdurre per questo motivo indicatori che misurino lo *status* di salute delle imprese turistiche italiane mediante l'analisi della redditività delle stesse.

L'analisi ha preso in considerazione due tipiche attività connesse al turismo: il settore degli alberghi e il settore delle agenzie di viaggio e tour operator. La fonte dei dati è la Banca dati Aida.

Alcune precisazioni su questi dati risultano necessarie, i dati sono riferiti esclusivamente alle società di capitali obbligate alla presentazione del bilancio, naturalmente il numero di società differisce da regione a regione, e inoltre nei dati sono incluse anche società con fatturato a zero.

Appare interessante ricordare che il tessuto imprenditoriale nel settore della ricettività e dell'intermediazione nel nostro paese è composto in maggioranza da piccoli gruppi familiari e a gestione individuale, quindi probabilmente il dato non è esaustivo dell'intero comparto che si sta indagando ma comunque rappresenta ai fini dell'analisi un dato interessante da non sottovalutare.

Gli indicatori di redditività inseriti sono i più comuni indici di redditività ROA, ROS, ROE. Il ROA è dato dal rapporto tra reddito operativo e totale attivo e misura l'efficienza nella gestione del capitale, ovvero la capacità di trarre profitto dagli investimenti. Il ROE misura l'efficienza nella gestione del capitale proprio investito nell'azienda, quindi la capacità di trarre profitto dai propri mezzi. Il ROS è costituito dal rapporto tra utile operativo e fatturato. Indica la redditività operativa derivante dalla gestione caratteristica cioè dalle vendite realizzate.

Performance Alberghi (ATECO 55.1)

- **PA1** : N° medio dipendenti per società (*Dipendenti / N°società*)
- **PA2** : Fatturato medio (in migliaia di euro)(*Ricavi delle vendite/N°società*)
- **PA3** : Utile o perdita media (in migliaia di euro) (*Utile o perdita/N° società*)
- **PA4** : ROA (*Redditività del totale attivo*)
- **PA5** : ROS (*Redditività delle vendite*)
- **PA6** : ROE (*Redditività del capitale proprio*)

Performance Agenzie di viaggio e tour operator (ATECO 79.1)

- **PE_ADV1** : N° medio dipendenti per società (*Dipendenti / N°società*)
- **PE_ADV2** : Fatturato medio (in migliaia di euro) (*Ricavi delle vendite/N°società*)
- **PE_ADV3** : Utile o perdita media (in migliaia di euro) (*Utile o perdita/N° società*)
- **PE_ADV4** : ROA (*Redditività del totale attivo*)
- **PE_ADV5** : ROS (*Redditività delle vendite*)
- **PE_ADV6** : ROE (*Redditività del capitale proprio*)

Tab. 14 Performance Alberghi - Matrice dati

REGIONE	PA_1	PA_2	PA_3	PA_4	PA_5	PA_6
Piemonte	6.96	833	-63.48	-2	-0.48	-7.09
Valle D'Aosta	7.94	587	-104.17	-17.18	-3.01	-6.53
Lombardia	37.00	3161	-107.53	2.22	0.93	-3.94
Trentino	8.31	1112	-22.01	7.53	2.32	-2.15
Veneto	11.61	1404	-161.45	-3.37	-1.09	-9.95
Friuli	6.17	602	-39.17	2.23	0.56	-4.1
Liguria	8.58	888	-19.89	5.36	2.17	-2.72
Emilia	10.16	831	-42.49	-0.47	-0.17	-6.63
Toscana	10.13	927	-89.98	-1.77	-0.53	-8.16
Umbria	5.84	622	-51.84	-4.18	-1.20	-5.64
Marche	5.83	422	-31.45	-0.14	-0.04	-4.82
Lazio	12.47	1404	-90.55	1.77	0.46	-4.39
Abruzzo	11.42	825	-35.25	-1.12	-0.42	-4.75
Molise	7.41	722	-37.33	5.37	1.54	-4.45
Campania	7.64	824	-31.10	4.72	10.68	-2.01
Puglia	6.01	702	-55.16	-0.16	-0.03	-4.52
Basilicata	2.29	424	-21.77	1.67	0.37	-3.17
Calabria	6.98	613	-63.31	-1.47	-0.31	-6.82
Sicilia	7.95	795	-117.90	-4.08	-0.79	-7.15
Sardegna	10.03	1222	-71	3.98	0.81	-3.61

Tab.15 Performance Agenzie di viaggio e Tour operator - Matrice dei dati

Regioni	PE_ADV1	PE_ADV2	PE_ADV3	PE_ADV4	PE_ADV5	PE_ADV6
Piemonte	13.73	6,961	-121.49	3.17	1.37	-22.5
Valle D'Aosta	4.75	766	0.00	4.02	2.33	-0.05
Lombardia	7.33	2,990	-6.80	0.90	0.64	-1.91
Trentino	3.60	1,855	12.92	5.28	2.42	10.21
Veneto	3.41	1,327	-4.44	1.86	0.79	-5.8
Friuli	4.71	937	1.43	3.10	2.06	1.57
Liguria	6.44	3,214	18.21	7.71	2.11	14.89
Emilia	7.21	2,371	7.03	3.61	1.46	3.11
Toscana	2.99	1,285	-2.39	3.17	1.27	-2.65
Umbria	2.25	2,350	-4.06	0.77	0.13	-4.08
Marche	2.42	980	-9.33	0.61	0.25	-19.98
Lazio	5.18	2,445	0.76	2.01	0.84	0.41
Abruzzo	2.28	672	2.20	4.38	2.21	5.52
Molise	1.90	948	-5.50	2.92	0.86	-13.02
Campania	3.47	1,642	8.00	3.67	1.64	5.34
Puglia	1.44	446	0.04	3.81	2.06	0.27
Basilicata	6.00	412	2.86	3.48	2.44	7.51
Calabria	2.14	1,346	-29.26	3.68	1.64	-36.61
Sicilia	3.48	1,286	-5.54	2.77	1.38	-3.92
Sardegna	3.43	1,121	-5.15	1.25	0.57	-9.66

4.4 Applicazione e risultati dell'analisi in componenti principali.

L'analisi in componenti principali è stata applicata ai gruppi di indicatori definiti nei paragrafi precedenti. Tale scelta è stata motivata dal vasto *set* di indicatori da dover sintetizzare per un numero di unità abbastanza contenuto (le 20 regioni italiane). Inoltre l'applicazione dell'analisi in componenti principali sulle singole matrici di dati consente di ottenere degli indicatori sintetici per ciascuno dei driver di competitività considerati nella ricerca.

L'analisi in componenti principali è stata applicata sulle matrici di correlazione degli indicatori, tale scelta risulta appropriata quando le variabili sono espresse in diverse unità di misura come nel caso in esame. I risultati dell'analisi statistica saranno riportati seguendo la suddivisione in gruppi²⁷.

La scelta del numero di componenti da estrarre è stata fatta sulla base dei criteri descritti nel capitolo dedicato alla metodologia, in particolare analizzando lo *screeplot*, la quota di varianza spiegata e la regola di *Kaiser*.

La definizione e interpretazione delle componenti estratte è stata fatta sulla base dei coefficienti di correlazione tra le componenti e le variabili e attraverso l'osservazione del *biplot*.

4.4.1. ACP Territorio

Per tale gruppo di indicatori i criteri descritti hanno portato alla scelta di considerare le prime due componenti principali.

La prima componente estratta considera il 51% della varianza su un totale di 7 variabili, considerando anche la seconda componente raggiungiamo il 67% della varianza originaria spiegata dalle componenti estratte (tab.16).

²⁷ In Appendice sarà riportata la lista dei comandi per l'implementazione in R della procedura statistica dell'analisi in componenti principali.

Nella tabella denominata “*Matrice dei loadings*” sono riportati i pesi di tutte le componenti principali, ossia gli autovettori delle combinazioni lineari.

Per interpretare il significato delle CP si utilizzano i coefficienti di correlazione tra ognuna di esse e ciascuna delle variabili (tab.16) e l’osservazione del *biplot* (fig.13).

Dall’analisi si può evidenziare che la prima componente è correlata positivamente in maniera elevata con il numero di siti Unesco (TERR_3), con il numero di beni culturali censiti (TERR_4), con la variabile relativa al numero delle città del pane, dell’olio e del vino (TERR_6), con il numero dei prodotti agroalimentari (TERR_7) e negativamente con la percentuale di superficie montana (TERR_2).

Le regioni che presentano punteggi elevati di tale componente sono le regioni con un numero elevato di risorse culturali e maggiormente legate alle tradizioni agroalimentari ed enogastronomiche. Tali regioni sono Toscana, Lombardia, Veneto, Lazio, Campania, Piemonte e Sicilia.

La seconda componente è correlata positivamente con la variabile relativa ai km di costa (TERR_1) e negativamente con la variabile del numero di Borghi più belli d’Italia (TERR_5) presenti nella regione.

Data la maggiore incidenza della variabile relativa ai km di costa nel definire la componente possiamo considerarla in via generale come un indicatore della risorsa “mare” nelle regioni e infatti sono le regioni del sud e le isole a presentare i valori più elevati per tale fattore.

Si definiscono, quindi, le due componenti estratte:

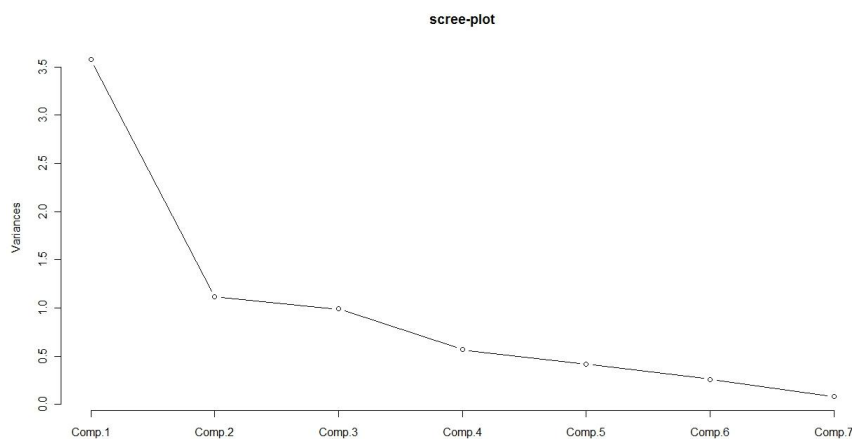
TERR_C1: Indicatore delle risorse culturali e del turismo legato alla tradizione.

TERR_C2: Indicatore della risorsa “mare”.

Tab.16 Territorio - Matrice dei loadings²⁸

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
TERR_1		0.713	0.585	-0.311		-0.112	-0.157
TERR_2	-0.411	-0.212	0.258	0.223	0.712	-0.403	
TERR_3	0.478		-0.196	-0.201	0.108	-0.593	0.574
TERR_4	0.447	-0.182	-0.266	-0.369	0.407		-0.627
TERR_5	0.168	-0.614	0.631	-0.351		0.173	0.182
TERR_6	0.415	-0.118	0.272	0.606	-0.313	-0.377	-0.364
TERR_7	0.44	0.146		0.432	0.449	0.549	0.294

Fig.13 Territorio- Screeplot



Tab. 17 Territorio - Quota di varianza spiegata dalle componenti e radice degli autovalori²⁹

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Standard deviation	1.891	1.056	0.995	0.753	0.644	0.510	0.281
Proportion of variance	0.511	0.159	0.141	0.081	0.059	0.037	0.011
Cumulative proportion	0.511	0.670	0.811	0.892	0.952	0.989	1.000

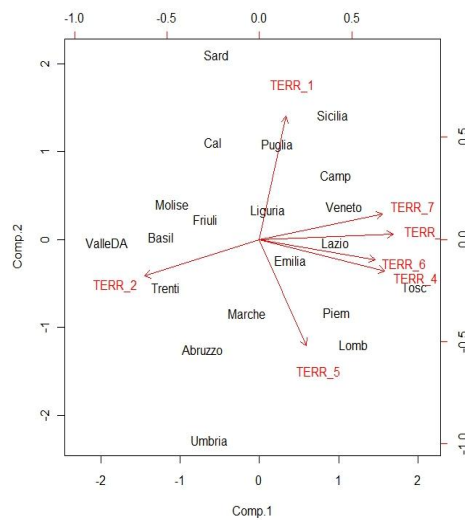
²⁸ La tabella Loadings rappresenta la matrice dei pesi o dei coefficienti delle componenti principali ossia gli autovettori.

²⁹ Nella tabella, la standard deviation rappresenta la radice quadrata degli autovalori, si ricorda che per la regola di Kaiser si considerano le componenti con autovalore maggiore di 1.

Tab.18 Territorio - Matrice di Correlazione tra le componenti estratte e le variabili

	Comp1	Comp2
TERR_1	0.180	0.753
TERR_2	-0.777	-0.224
TERR_3	0.903	0.032
TERR_4	0.846	-0.192
TERR_5	0.318	-0.649
TERR_6	0.784	-0.124
TERR_7	0.832	0.154

Fig.14 Territorio - Biplot



4.4.2. ACP Clima

Dall'analisi dello *scree-plot*, degli autovalori e della percentuale di varianza spiegata si è scelto di considerare una sola componente che spiega un'alta percentuale (80%) di varianza originaria. Nei dati si è trasformata la variabile precipitazione cumulata in valori negativi per avere dei dati che esprimano tutti una condizione di clima favorevole (alte temperature e poche precipitazioni). La componente estratta è correlata negativamente con tutte le

variabili del gruppo quindi si può interpretare come un indicatore di condizioni climatiche sfavorevoli. Questo comporta che a valori elevati del punteggio fattoriale corrisponderanno le regioni con clima "peggiore" in un'ottica di competitività turistica.

Dall'osservazione dei punteggi fattoriali³⁰ si può notare che le regioni che presentano condizioni climatiche migliori sono Sicilia, Sardegna, Calabria, Puglia e Basilicata. Mentre le regioni più svantaggiate dal punto di vista climatico risultano essere Valle d'Aosta, Trentino, Friuli e Piemonte.³¹

Si definisce dunque la componente estratta come:

CLI_C1: Indicatore di condizioni climatiche sfavorevoli

Tab.19 Clima - Matrice dei Loadings

	Comp.1	Comp.2	Comp.3
CLI_1	-0.617	0.351	0.705
CLI_2	-0.618	0.339	-0.709
CLI_3	-0.488	-0.873	

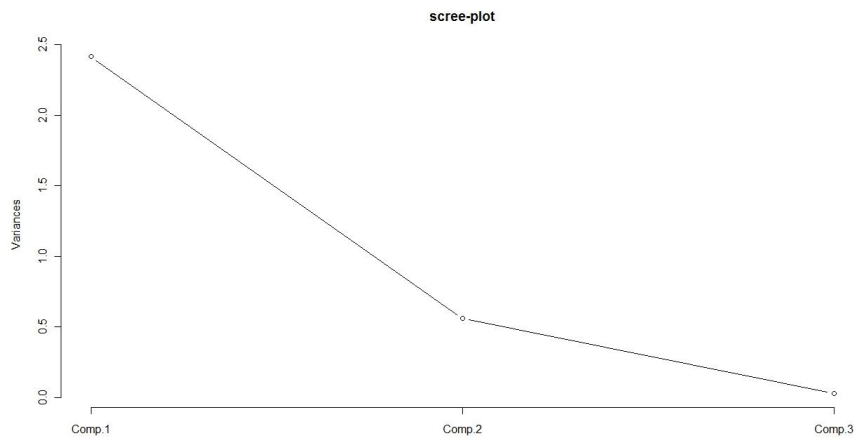
Tab.20. Clima - Quota di varianza spiegata dalle componenti e radice degli autovalori

	Comp.1	Comp.2	Comp.3
<i>Standard deviation</i>	1.553	0.747	0.162
<i>Proportion of Variance</i>	0.804	0.186	0.008
<i>Cumulative Proportion</i>	0.804	0.991	1

³⁰ I punteggi fattoriali sono riportati nella tabella denominata Matrice delle componenti pag.

³¹ L'analisi del biplot non risulta utile nel caso di una sola componente.

Fig.15 Clima - Screeplot



Tab. 21 Clima - Matrice di Correlazione tra le componenti estratte e le variabili

	Comp_1
CLI_1	-0.958
CLI_2	-0.960
CLI_3	-0.757

4.4.3. ACP Offerta

Per il gruppo Offerta si scelgono le prime due componenti che spiegano il 70% della varianza dei dati di partenza. L'analisi del *biplot* e delle correlazioni ci offre informazioni utili alla definizione delle componenti estratte. La prima componente è correlata positivamente con le variabili relative alla capacità ricettiva regionale in termini di posti letto alberghieri (OFF_3) ed extralberghieri. (OFF_5). La stessa componente presenta correlazione negativa con le variabili che rappresentano la dimensione media delle strutture alberghiere (OFF_4) e la percentuale di letti di strutture di lusso (4-5 stelle) sul totale dei letti alberghieri (OFF_8).

La seconda componente è correlata negativamente con la variabile relativa alla dimensione media delle strutture extralberghiere (OFF_6) e con l'indice di ricettività alberghiera (OFF_7) quindi si può interpretare come indicatore della ricettività extralberghiera di medio-piccole dimensioni

Il *biplot* ci fornisce informazioni sia sulle correlazioni tra componenti e variabili sia sulle singole unità rispetto ai fattori estratti. Si può evidenziare infatti che il Trentino e Valle d'Aosta sono le regioni con una capacità ricettiva proporzionata alla popolazione maggiore rispetto alle altre regioni. Per la seconda componente abbiamo valori elevati per le regioni Friuli, Veneto e Umbria che quindi sono le regioni che una maggiore percentuale di ricettività di tipo extralberghiero.

Le due componenti estratte vengono definite:

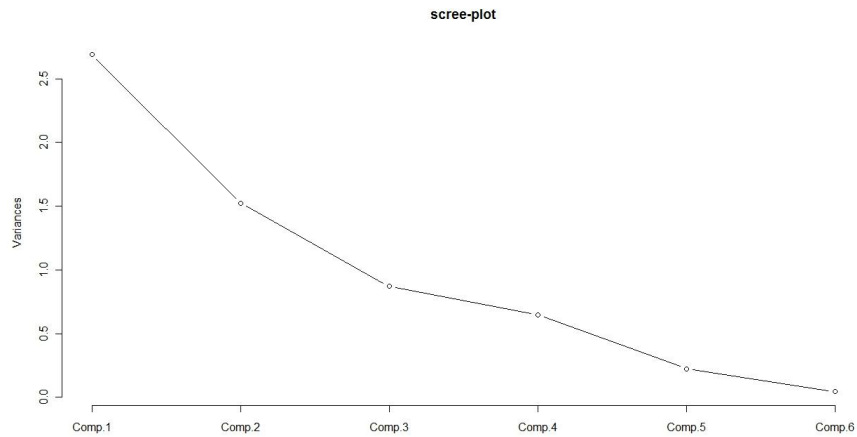
OFF_C1: Indicatore della capacità ricettiva (alberghiera ed extralberghiera) regionale rispetto alla popolazione

OFF_C2: Indicatore dell'offerta ricettiva extralberghiera

Tab.22 Offerta - Matrice dei loadings

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
OFF_3	0.503	0.323	0.548	-0.196	0.544	
OFF_4	-0.432	-0.354	0.503	0.111	-0.585	-0.284
OFF_5	0.502	0.559	-0.124	0.35	-0.545	
OFF_6	-0.711	-0.554	0.169	0.384		
OFF_7	0.21	-0.588	-0.512	0.426	-0.406	
OFF_8	-0.511	-0.122	0.244	0.429	0.683	0.12

Fig. 16 Offerta - Screeplot



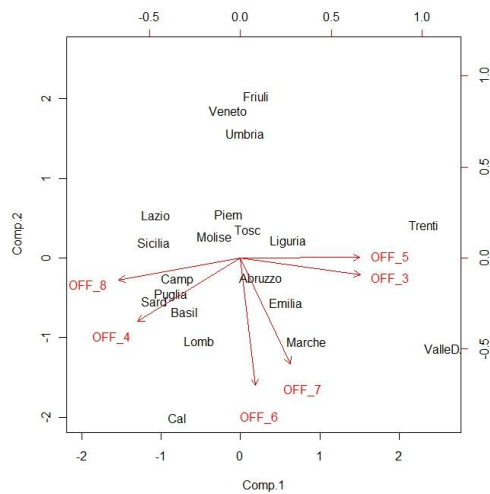
Tab. 23 Offerta - quota di varianza spiegata dalle componenti e radice degli autovalori

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	1.640	1.235	0.933	0.805	0.473	0.208
Proportion of variance	0.448	0.254	0.145	0.108	0.037	0.007
Cumulative Proportion	0.448	0.702	0.847	0.956	0.993	1.000

Tab.24 Offerta - Matrice di correlazione tra le componenti estratte e le variabili

	Comp_1	Comp_2
OFF_3	0.824	-0.115
OFF_4	-0.708	-0.437
OFF_5	0.823	0.002
OFF_6	0.105	-0.877
OFF_7	0.344	-0.727
OFF_8	-0.837	-0.151

Fig.17 Offerta - Biplot



4.4.4. ACP Flussi

Le prime due componenti principali esprimono il 64% della varianza originaria, percentuale abbastanza soddisfacente considerando che si vogliono sintetizzare 10 variabili.

I dati di partenza esprimono la propensione delle regioni ad attrarre flussi turistici, considerando sia la componente nazionale che internazionale.

L'analisi delle correlazioni evidenzia un legame diretto forte tra la prima componente e le variabili FLS_8 (presenze stranieri) , FLS_10 (indice di utilizzazione netta strutture da parte degli stranieri), FLS_7 (indice di attrazione degli stranieri nella regione) e in misura più lieve con FLS_3 (arrivi stranieri) che sono tutte variabili relative alla componente internazionale della domanda turistica sul territorio regionale. Quindi le regioni con il punteggio fattoriale maggiore sono quelle che hanno la maggior capacità di attrarre e trattenere flussi stranieri.

La seconda componente principale è correlata inversamente con FLS_1 e FLS_3 quindi con il numero di arrivi sia degli italiani che degli stranieri, e in misura più lieve con FLS_2 (permanenza media italiani) e FLS_9 (indice di

utilizzazione netta delle strutture da parte degli italiani) cioè con la capacità di trattenere gli italiani in regione. Quindi le regioni che hanno valori alti per questa componente sono quelle che hanno il minor numero di arrivi rispetto alla popolazione e che riescono a trattenere meno la domanda nazionale. Le variabili che sono rappresentate meglio dalle due componenti sono quelle i cui vettori si avvicinano al raggio unitario quindi FLS7, FLS8, FLS10, FLS1 E FLS3.

Dall'osservazione del *biplot* possiamo evidenziare che la regione Toscana e la regione Veneto si collocano distanti dal centroide lungo l'asse della prima componente, il Lazio si colloca distante sia rispetto alla prima che alla seconda componente principale evidenziando valori elevati per entrambi i fattori, mentre la Valle D'Aosta si colloca lungo l'asse negativo della seconda componente principale.

Definiamo, quindi le due componenti estratte:

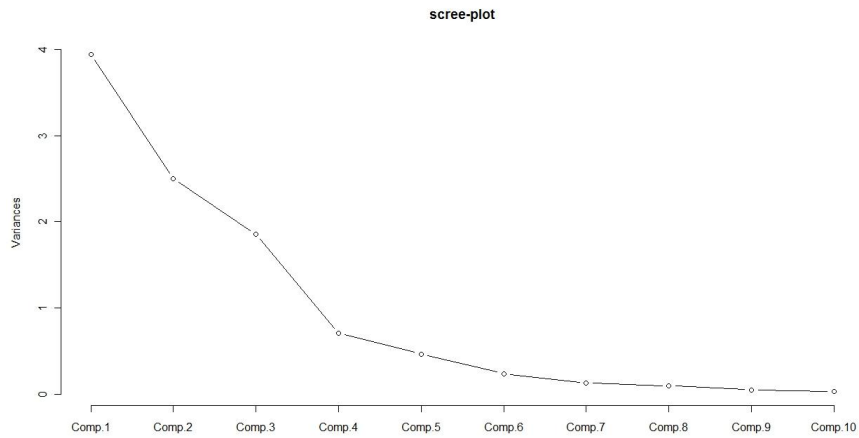
FLS_C1: Indicatore della capacità di attrarre flussi turistici stranieri

FLS_C2: Indicatore (in negativo) della capacità di attrarre e trattenere i flussi turistici

Tab.25 Flussi - Matrice dei loadings

	Comp.1	Comp.2
FLS_1	0.105	-0.509
FLS_2	-0.198	-0.362
FLS_3	0.319	-0.449
FLS_4	-0.174	-0.343
FLS_5	0.249	0.223
FLS_6	0.275	0.261
FLS_7	0.429	-0.124
FLS_8	0.474	-0.060
FLS_9	-0.231	-0.361
FLS_10	0.468	-0.158

Fig.18 Flussi - Screeplot



Tab.26 Flussi - Quota di varianza spiegata dalle componenti e radice degli autovalori

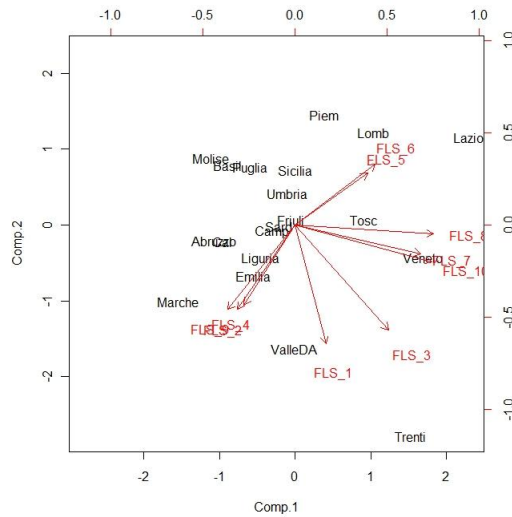
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	1.984	1.581	1.362	0.841	0.683
Proportion of Variance	0.394	0.250	0.185	0.071	0.047
Cumulative Proportion	0.394	0.644	0.829	0.900	0.946

	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
Standard deviation	0.486	0.358	0.310	0.218	0.166
Proportion of Variance	0.024	0.013	0.010	0.005	0.003
Cumulative Proportion	0.970	0.983	0.993	0.997	1.000

Tab.27 Flussi-Matrice di Correlazione tra le componenti estratte e le variabili

	Comp_1	Comp_2
FLS_1	0.209	-0.805
FLS_2	-0.392	-0.572
FLS_3	0.632	-0.710
FLS_4	-0.346	-0.542
FLS_5	0.493	0.353
FLS_6	0.545	0.412
FLS_7	0.850	-0.196
FLS_8	0.941	-0.061
FLS_9	-0.459	-0.571
FLS_10	0.928	-0.250

Fig.19 Flussi- Biplot



4.4.5. ACP Accessibilita'

L'analisi in componenti principali sul gruppo di variabili relative all'accessibilità delle regioni evidenzia che la prima componente sintetizza il 50% della varianza originaria mentre le prime due componenti esprimono il 71%.

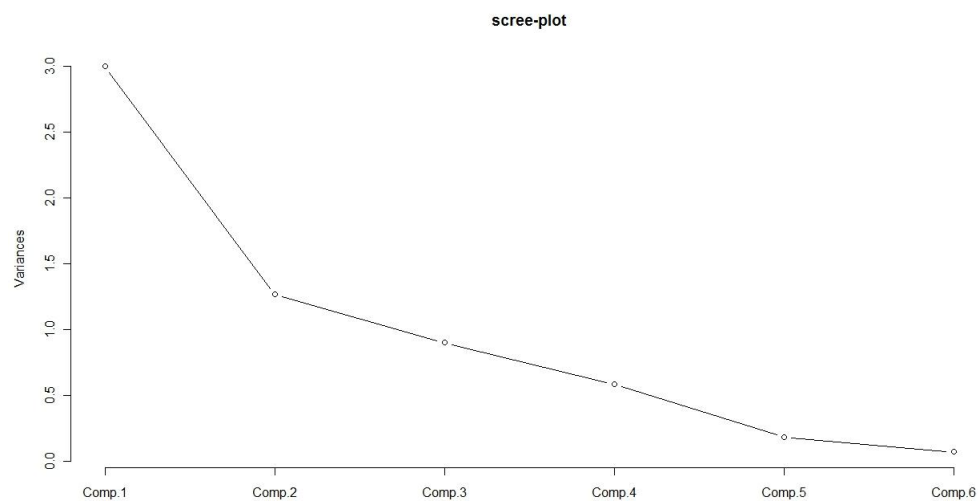
Nonostante la percentuale di varianza spiegata non sia pienamente soddisfacente si è ritenuto giusto considerare una sola componente che sintetizzasse le poche variabili di partenza. L'osservazione delle correlazioni tra variabili e componenti segnala una relazione inversa abbastanza forte tra la prima componente e quasi tutte le variabili esclusa la variabile relativa alla dotazione di altre strade di interesse nazionale (ACS_2). Quindi le regioni con dotazione di infrastrutture inferiore rispetto alle altre regioni avranno valori alti di tale fattore. Tali regioni sono Basilicata, Sardegna, Valle d'Aosta e Molise. Definiamo quindi la componente come:

ACS_C1: Indicatore (in negativo) della dotazione di infrastrutture per la mobilità e accessibilità nella Regione

Tab.28 Accessibilità- Matrice dei loadings

	Comp.1
ACS_1	-0.433
ACS_2	0.305
ACS_3	-0.426
ACS_4	-0.458
ACS_5	-0.489
ACS_6	-0.298

Fig.20 Accessibilità - Screeplot



Tab.29 Accessibilità-Quota di varianza spiegata dalle componenti e radice degli autovalori

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	1.732	1.126	0.949	0.763	0.424	0.263
Proportion of Variance	0.500	0.211	0.150	0.097	0.030	0.012
Cumulative Proportion	0.500	0.711	0.861	0.958	0.988	1.000

Tab. 30 Accessibilità' - Matrice di Correlazione tra le componenti estratte e le variabili

	Comp_1
ACS_1	-0.751
ACS_2	0.529
ACS_3	-0.737
ACS_4	-0.793
ACS_5	-0.847
ACS_6	-0.516

4.4.6. ACP Sostenibilità e Qualità

Per il gruppo Sostenibilità e Qualità si è ritenuto opportuno estrarre 2 componenti che spiegano complessivamente il 56% di varianza dei dati di partenza.

L'analisi delle correlazioni e del *biplot* evidenzia i legami più evidenti relativi alla prima componente con i tassi di ricettività e turisticità³², con il numero di letti 4/5 stelle sul totale e con la percentuale di consumi elettrici del settore Alberghi, Bar e Ristoranti. Potremmo quindi definire la componente come un Indicatore di impatto sul territorio dell'attività turistica.

La seconda componente è correlata positivamente con il numero di Bandiere Arancioni, il n° di Strutture con il Marchio Ospitalità Italiana e il n° di strutture turistiche certificate *Ecolabel* (in misura minore) e con il numero di siti certificati di qualità quindi potremmo definirlo come Indicatore di Certificazione Ambientale e di Qualità.

Suddividendo il *biplot* in quadranti le regioni appartenenti al primo quadrante (Piemonte, Lombardia, Lazio, Umbria) sono quelle che presentano i valori più elevati dei due indicatori individuati, per cui tali Regioni presentano

³² Le variabili analizzate sono state espresse in positivo, nel senso che a valori elevati corrispondono impatti limitati e maggior attenzione al tema della sostenibilità. Per questo motivo i classici indicatori d'impatto (tasso di turisticità e di ricettività) sono stati trasformati nel rispettivo reciproco così come la variabile relativa ai consumi energetici di alberghi e ristoranti sul totale dei consumi energetici della regione.

impatti dell'attività turistica abbastanza limitati e maggiore attenzione alle tematiche ambientali attraverso gli strumenti delle certificazioni e attraverso un'offerta di qualità.

Infatti tra le Regioni appartenenti a tale quadrante i tassi di ricettività e turisticità presentano valori al di sotto della media, segnalando un'incidenza del turismo sulla popolazione ancora sotto controllo.

Inoltre Lazio e Lombardia hanno un numero elevato di strutture di categoria alta, il Piemonte e la Lombardia si distinguono per un vasto numero di strutture con Marchio Ospitalità Italiana.

Le regioni nel secondo quadrante sono quelle a maggior propensione turistica e nelle quali la proporzione del turismo rispetto alla popolazione assume un'importanza rilevante. Allo stesso tempo però hanno valori elevati anche per quanto riguarda l'Indicatore di certificazione, quindi si tratta delle regioni per così dire "virtuose" poiché pur avendo un maggior flusso turistico e di conseguenza anche impatti sui territori abbastanza importanti hanno saputo cogliere la sfida della sostenibilità dotandosi di alcuni strumenti idonei a limitare il più possibile gli impatti negativi (Toscana, Emilia-Romagna, Liguria, Trentino, Marche, Veneto).

Tali regioni si distinguono soprattutto per il numero di Bandiere Arancioni e per le strutture con Marchio Ospitalità Italiana.

Il Trentino si distingue nettamente dal resto d'Italia per le strutture certificate Ecolabel (67).

Il terzo quadrante ospita le regioni con bassi valori per entrambi gli indicatori quindi le regioni che vi sono collocate sono quelle che dovrebbero mostrare maggiore attenzione al tema ambientale poiché presentano impatti dell'attività turistica significativi non mitigati da azioni a sostegno della sostenibilità. (Valle d'Aosta, Friuli, Sardegna, Abruzzo, Calabria)

Nel quarto quadrante si collocano le regioni per le quali l'impatto turistico non è ancora rilevante rispetto alla popolazione e che non si sono

ancora impegnate nella certificazione ambientale e di qualità (Sicilia, Puglia, Campania, Basilicata, Molise).

L'analisi del *biplot* fornisce ulteriori informazioni sulle regioni. Il Piemonte si trova lungo la traiettoria del vettore della variabile SOST_10 per cui tale regione è quella con il maggior numero di strutture certificate con il Marchio Ospitalità Italiana. La Toscana è lungo la traiettoria del vettore Sost_8, quindi, è la regione con valore più elevato della variabile relativa alle Bandiere Arancioni, mentre il Trentino si distingue per il maggior numero di strutture certificate *Ecolabel*.

Si definiscono le componenti estratte:

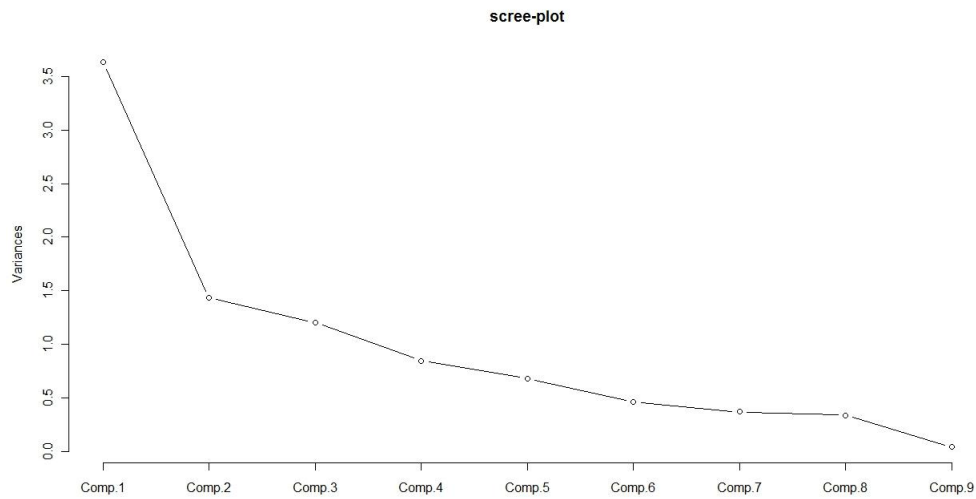
SOST_C1 : Indicatore (in negativo) degli impatti dell'attività turistica

SOST_C2: Indicatore delle certificazioni ambientali e di qualità

Tab.31 Sostenibilità – Matrice dei loadings

	Comp.1	Comp.2
SOST_1	0.440	-0.191
SOST_2	0.486	
SOST_3	-0.313	-0.283
SOST_4	0.366	-0.200
SOST_8		0.734
SOST_9	-0.255	0.184
SOST_10	0.289	0.518
SOST_11	0.318	
SOST_13	0.294	

Fig.21 Sostenibilità – Screeplot



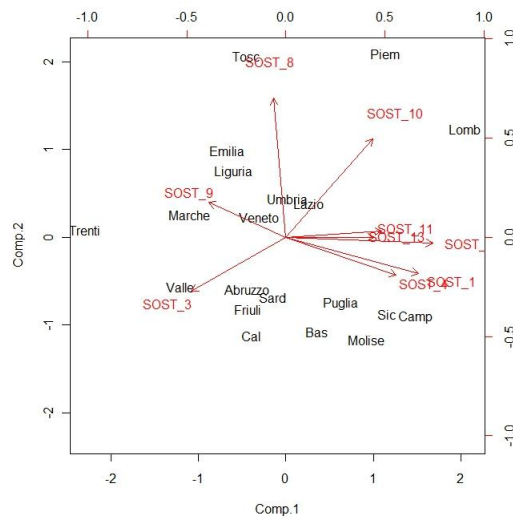
Tab.32 Sostenibilità- Quota di varianza spiegata dalle componenti e radice degli autovalori

	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6	Comp. 7	Comp. 8	Comp. 9
Standard deviation	1.906	1.197	1.096	0.920	0.826	0.678	0.608	0.581	0.202
Proportion of Variance	0.403	0.159	0.133	0.094	0.076	0.051	0.041	0.037	0.005
Cumulative Proportion	0.403	0.563	0.696	0.790	0.866	0.917	0.958	0.995	1.000

Tab. 33 Sostenibilità-- Matrice di correlazione

	Comp_1	Comp_2
SOST_1	0.838	-0.229
SOST_2	0.927	-0.037
SOST_3	-0.596	-0.339
SOST_4	0.697	-0.239
SOST_8	-0.078	0.878
SOST_9	-0.487	0.220
SOST_10	0.551	0.621
SOST_11	0.607	0.039
SOST_13	0.561	0.001

Fig. 22 Sostenibilità' - Biplot



4.4.7. ACP Profilo economico-turistico

La categoria sul profilo economico turistico delle regioni è rappresentata nell'analisi attraverso otto variabili che trovano la loro sintesi nelle due componenti estratte che spiegano il 69% della varianza di partenza.

La prima componente è definita come un Indicatore dell'Impegno economico della Regione nel settore turistico infatti essa è correlata negativamente con le variabili relative alla spesa della regione per il turismo e alla spesa rapportata al numero di arrivi e presenze. Quindi le regioni con alto valore di tale componente sono quelle che spendono meno per il settore turistico.

La seconda componente è correlata negativamente al valore della bilancia turistica e alla percentuale dell'occupazione turistica nella Regione, alla spesa degli stranieri per turismo e al PIL procapite. Ne deriva, quindi, che tale componente si può identificare come un Indicatore dell'impatto economico del turismo sulla regione.

Le regioni con più alto valore di tale componente saranno quelle che non riescono a beneficiare pienamente degli impatti economici positivi del turismo.

Osservando il *biplot* potremmo dire che le regioni più competitive sono quelle che pur investendo minori risorse nel settore riescono a trarre maggiori benefici e sono le regioni situate nel quarto quadrante che hanno bassi valori delle due componenti.

PET_C1: Indicatore (in negativo) dell'impegno economico della Regione nel settore turistico

PET_C2: Indicatore (in negativo) dei benefici economici del turismo sulla regione

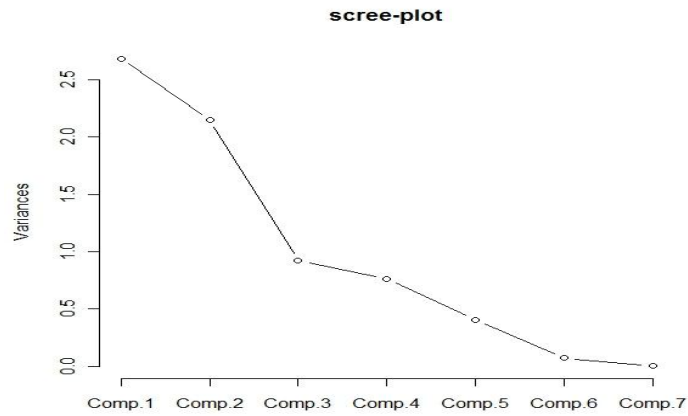
Tab.34 Profilo economico-turistico-Matrice dei loadings

	Comp.1	Comp.2
PET_1	-0.499	0.564
PET_2	0.18	-0.344
PET_3	-0.462	-0.613
PET_5	-0.495	-0.338
PET_6	-0.593	-0.229
PET_7	-0.594	-0.222
PET_8	-0.551	-0.333

Tab.35 Profilo economico-turistico:Quota di varianza spiegata dalle componenti e radice degli autovalori

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
<i>Standard deviation</i>	1.639	1.465	0.960	0.873	0.637	0.272	0.064
<i>Proportion of variance</i>	0.384	0.307	0.132	0.109	0.058	0.011	0.001
<i>Cumulative proportion</i>	0.384	0.690	0.822	0.931	0.989	0.999	1.000

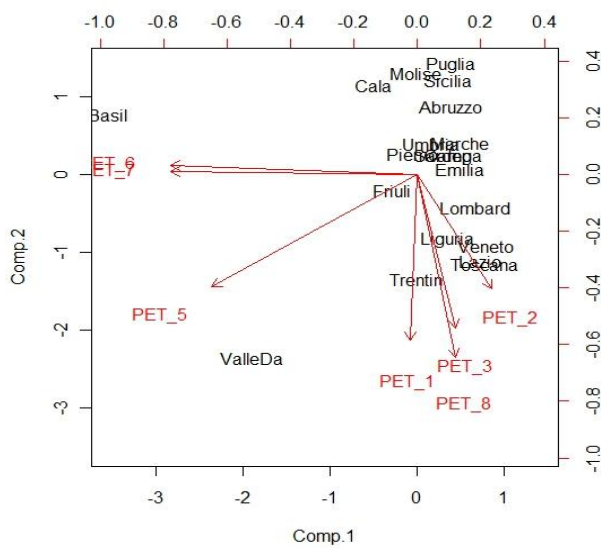
Fig.23 Profilo economico-turistico: Screeplot



Tab.36 Profilo economico –turistico: Matrice di correlazione tra le componenti estratte e le variabili

	Comp_1	Comp_2
PET_1	-0.030	-0.731
PET_2	0.295	-0.505
PET_3	0.152	-0.677
PET_5	-0.811	-0.495
PET_6	-0.972	0.040
PET_7	-0.974	0.012
PET_8	0.150	-0.807

Fig.24 Profilo economico –turistico: Biplot



4.4.8. ACP Performance imprese turistiche

4.4.8.1.ACP Performance Settore Alberghi

Le prime due componenti estratte spiegano l'82% della varianza di partenza.

La prima, denominata *Indicatore di redditività delle imprese alberghiere*, rileva la performance economica del settore essendo correlata negativamente con i più noti indici di redditività (ROA, ROE, ROS). La seconda, definita come indicatore delle dimensioni aziendali, evidenzia la correlazione negativa con le variabili relative alla dimensione delle imprese in termini di numero di dipendenti e fatturato.

Si ritiene opportuno precisare che quando si analizzano questi dati bisogna tenere in considerazione le modifiche che nel frattempo sono intervenute nel mercato turistico. In tal senso, per esempio, occorre ricordare come il turista moderno tenda sempre più a preferire l'esperienza al posto della standardizzazione. Ne deriva che non è detto che la presenza di grandi strutture sia direttamente sinonimo di competitività regionale proprio perché viceversa la presenza di un sistema diffuso di micro e media ricettività può rappresentare un'interessante driver di attrattività.

Le componenti estratte vengono definite:

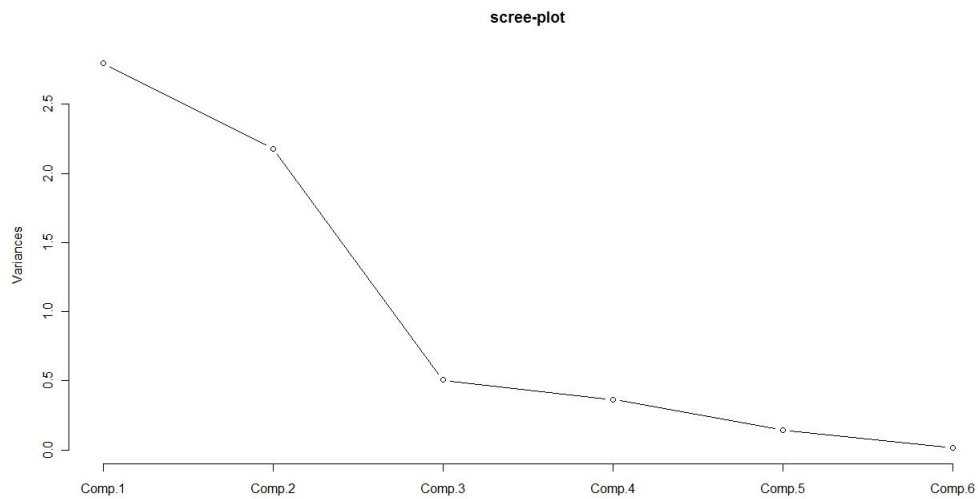
PA_C1: Indicatore (in negativo) della redditività degli alberghi

PA_C2: Indicatore (in negativo) delle dimensioni delle società alberghiere nella regione

Tab.37 Performance Alberghi – Matrice dei loadings

	Comp.1	Comp.2
PA_1	0.155	0.632
PA_2	0.129	0.657
PA_3	-0.515	-0.204
PA_4	-0.468	0.257
PA_5	-0.452	0.19
PA_6	-0.521	0.157

Fig.25 Performance alberghi- Screeplot



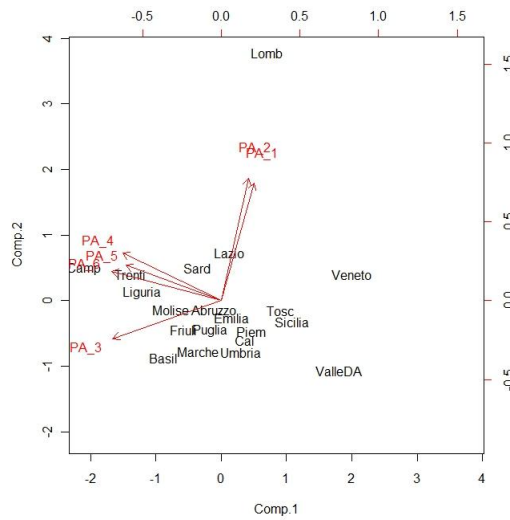
Tab.38 Performance Alberghi – Quota di varianza spiegata e radice degli autovalori

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	1.672	1.476	0.710	0.603	0.380	0.111
Proportion of variance	0.466	0.363	0.084	0.061	0.024	0.002
Cumulative proportion	0.466	0.829	0.913	0.974	0.998	1.000

Tab.39 Performance Alberghi- Matrice delle correlazioni tra componenti e variabili

	Comp_1	Comp_2
PA_1	0.260	0.933
PA_2	0.216	0.971
PA_3	-0.861	-0.301
PA_4	-0.782	0.379
PA_5	-0.755	0.281
PA_6	-0.871	0.232

Fig.26 Performance Alberghi - Biplot



4.4.8.2 ACP Performance Settore Agenzie di viaggio e tour operator

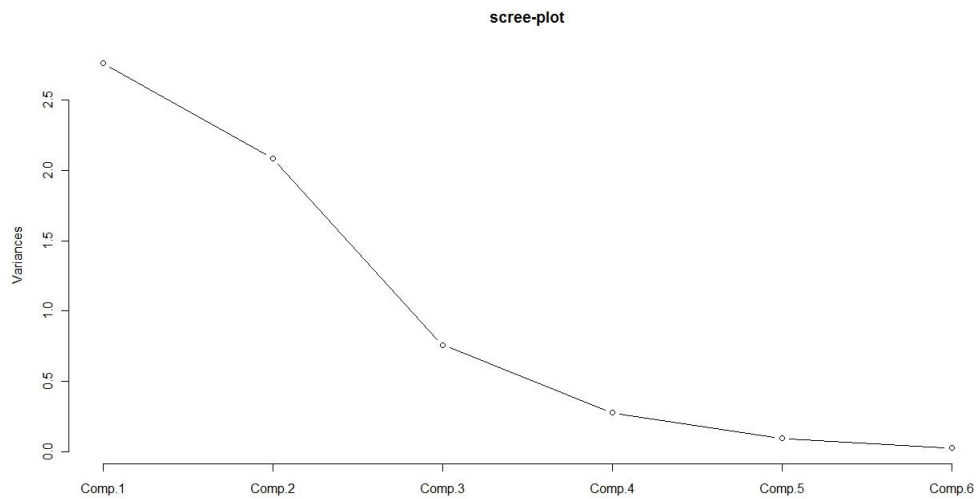
Nell'analisi dei dati relativi alle performance delle imprese del settore dell'intermediazione turistica (agenzie di viaggio e *tour operator*) si è scelto di considerare una sola componente che da sola spiega il 46% della varianza di base.

La scelta è stata fatta sulla base dei risultati ottenuti in termini di correlazione della componente con le variabili. Infatti tale componente è correlata con tutte le variabili di partenza, in maniera positiva con le variabili relative alle dimensioni aziendali e negativamente con gli indicatori di redditività. Quindi per sintetizzare al massimo l'informazione di partenza si è deciso di tenere in considerazione solo questa componente anche perché la seconda componente in termini di correlazioni con le variabili di partenza non aggiungeva alcuna informazione rilevante rispetto alla prima

Tab.40 Performance Agenzie di viaggio e TO – Matrice dei loadings

	Comp.1	Comp.2
PE_ADV1	0.401	-0.446
PE_ADV2	0.483	-0.335
PE_ADV3	-0.545	0.14
PE_ADV4	-0.255	-0.557
PE_ADV5	-0.299	-0.512
PE_ADV6	-0.393	-0.311

Fig.27 Performance Agenzie di viaggio e T.O. - Screeplot



Tab.41 Performance Agenzie di viaggio e T. O. - Quota di varianza spiegata e radice degli autovalori

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
<i>Standard deviation</i>	1.661	1.444	0.871	0.525	0.308	0.159
<i>Proportion of Variance</i>	0.460	0.348	0.126	0.046	0.016	0.004
<i>Cumulative Proportion</i>	0.460	0.807	0.934	0.980	0.996	1.000

Tab.42 Performance Agenzie di viaggio e T. O. – Matrice di correlazione tra le componenti e le variabili

	Comp_1
PE_ADV1	0.666
PE_ADV2	0.802
PE_ADV3	-0.906
PE_ADV4	-0.423
PE_ADV5	-0.496
PE_ADV6	-0.653

Nella tabella successiva (tab. 43) sono riportati tutti i punteggi fattoriali delle componenti estratte e definite precedentemente per le 20 regioni italiane. Tali componenti saranno utilizzate come nuove variabili sulle quali saranno applicati le tecniche di *cluster analysis* di tipo *crisp* e *fuzzy*.

I successivi paragrafi saranno dedicati alla descrizione dell'applicazione di tali tecniche e ai risultati conseguiti.

Tab.43 Matrice delle componenti

Regione	TERR_C1	TERR_C2	CLI_C1	OFF_C1	OFF_C2	FLS_C1	FLS_C2
Piemonte	1.833	-0.878	1.593	-0.237	0.676	0.764	2.289
Valle D'Aosta	-3.636	-0.035	3.068	4.252	-1.402	-0.030	-2.587
Lombardia	2.241	-1.258	0.859	-0.848	-1.281	2.051	1.935
Trentino	-2.211	-0.573	2.928	3.796	0.519	3.048	-4.402
Veneto	2.010	0.401	0.841	-0.252	2.287	3.356	-0.685
Friuli	-1.289	0.247	1.810	0.324	2.515	-0.103	0.106
Liguria	0.199	0.344	0.238	0.995	0.260	-0.897	-0.700
Emilia	0.740	-0.249	-0.088	0.941	-0.696	-1.102	-1.067
Toscana	3.688	-0.572	-0.130	0.158	0.444	1.809	0.105
Umbria	-1.174	-2.408	0.127	0.100	1.931	-0.201	0.665
Marche	-0.301	-0.882	-0.244	1.364	-1.299	-3.070	-1.604
Lazio	1.801	-0.037	-0.120	-1.750	0.666	4.543	1.833
Abruzzo	-1.351	-1.322	0.427	0.424	-0.307	-2.122	-0.324
Molise	-2.086	0.425	-0.318	-0.531	0.335	-2.198	1.391
Campania	1.815	0.757	-0.732	-1.294	-0.337	-0.596	-0.141
Puglia	0.425	1.130	-1.955	-1.426	-0.565	-1.176	1.177
Basilicata	-2.332	0.032	-1.360	-1.152	-0.840	-1.762	1.245
Calabria	-1.106	1.167	-2.017	-1.306	-2.480	-1.910	-0.347
Sicilia	1.754	1.493	-2.759	-1.782	0.247	0.018	1.136
Sardegna	-1.019	2.220	-2.167	-1.776	-0.674	-0.421	-0.027

Segue Tab.43- Matrice delle componenti

Regione	ACS_C1	PET_C1	PET_C2	SOST_C1	SOST_C2	PA_C1	PA_C2	PE_ADV_C1
Piemonte	-1.204	-0.088	0.393	2.174	2.510	0.791	-0.696	6.004
Valle D'Aosta	2.837	-3.107	-3.449	-2.282	-0.679	3.010	-1.569	-1.063
Lombardia	-1.941	1.095	-0.615	3.889	1.484	1.174	5.587	1.452
Trentino	0.597	-0.012	-1.968	-4.365	0.102	-2.317	0.591	-1.692
Veneto	-0.861	1.291	-1.350	-0.573	0.273	3.346	0.593	0.188
Friuli	-0.218	-0.472	-0.296	-0.821	-0.984	-0.951	-0.646	-0.837
Liguria	-3.961	0.563	-1.224	-1.133	0.897	-2.019	0.185	-1.329
Emilia	-1.069	0.809	0.107	-1.276	1.190	0.262	-0.377	0.022
Toscana	-0.473	1.276	-1.675	-0.853	2.479	1.514	-0.221	-0.438
Umbria	0.693	0.279	0.580	0.041	0.532	0.510	-1.156	0.744
Marche	0.582	0.800	0.605	-2.095	0.315	-0.599	-1.145	0.914
Lazio	-1.832	1.194	-1.633	0.508	0.468	0.218	1.090	0.472
Abruzzo	-0.279	0.632	1.285	-0.838	-0.713	-0.189	-0.210	-1.691
Molise	1.883	-0.029	1.915	1.761	-1.402	-1.289	-0.199	-0.095
Campania	-1.582	0.678	0.343	2.835	-1.082	-3.504	0.730	-0.949
Puglia	0.194	0.634	2.076	1.190	-0.899	-0.282	-0.655	-1.522
Basilicata	2.049	-5.807	1.127	0.676	-1.286	-1.475	-1.279	-1.269
Calabria	0.637	-0.808	1.670	-0.757	-1.343	0.593	-0.886	0.882
Sicilia	0.272	0.589	1.770	2.204	-1.046	1.818	-0.481	-0.245
Sardegna	3.676	0.482	0.339	-0.285	-0.818	-0.613	0.744	0.450

4.5 Applicazione e risultati cluster analysis crisp

4.5.1. Applicazione e risultati cluster analysis gerarchica

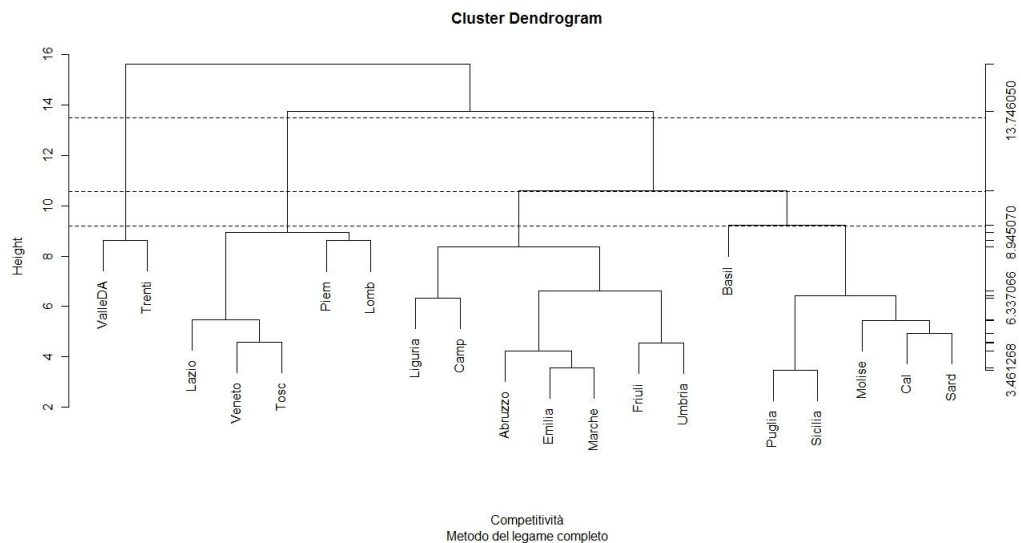
L'obiettivo dell'individuazione di cluster regionali omogenei per il loro grado di competitività ha reso utile l'applicazione della tecnica statistica della *cluster analysis*. Come visto nel Cap. 3 dedicato alla metodologia tali metodi di classificazione possono distinguersi in metodi di tipo *crisp* in cui le unità appartengono ad un solo *cluster* e metodi *fuzzy* in cui le unità possono appartenere a più cluster con un diverso grado di appartenenza. Nel lavoro di ricerca, agli indicatori individuati mediante l'applicazione dell'analisi in componenti principali, sono state applicate dapprima tecniche di *clustering* di tipo *crisp* e in una seconda fase tecniche di *clustering* di tipo *fuzzy*. La sostanziale convergenza di risultati per tutti i vari metodi di classificazione evidenzia la bontà e validità delle conclusioni raggiunte.

La *cluster analysis* gerarchica si è basata sull'applicazione dei più noti metodi gerarchici aggregativi variando la misura di distanza utilizzata.

Di seguito si riportano i dendrogrammi ottenuti che sono risultati maggiormente significativi e utili per fornire una prima informazione sulle strutture di gruppo presenti nei dati. Il dendrogramma rappresenta graficamente la mappa delle successive aggregazioni delle unità statistiche riproducendo sull'asse delle ordinate i livelli di distanza che caratterizzano le aggregazioni delle diverse partizioni.

Un criterio per definire i gruppi consiste nell'ispezione diretta del dendrogramma effettuando un "taglio" in corrispondenza di un salto nei livelli di distanza in cui è avvenuta l'aggregazione. In questo modo si individuano in corrispondenza del taglio i gruppi che corrispondono alla partizione ottimale.

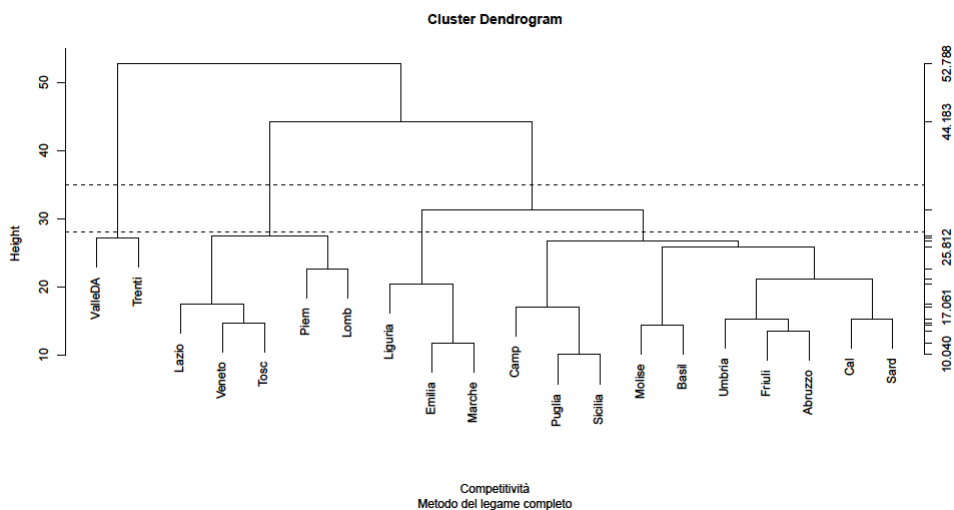
Fig.28 Metodo del legame completo³³-Dendrogramma



Come si può osservare dalla figura (fig.28) è evidente una struttura di gruppo nei dati. In particolare ci si è soffermati sulla suddivisione in 3, 4 e 5 cluster in base alla posizione dell' α -taglio. Per la suddivisione in 3 gruppi abbiamo un cluster formato dalle regioni Valle D'Aosta e Trentino, il secondo gruppo formato dalle regioni Piemonte, Lombardia, Veneto, Toscana, Lazio e il terzo macro-gruppo formato dalle restanti regioni. Questo macro-gruppo viene scisso in due cluster scendendo nella classificazione a quattro gruppi, per cui ai due gruppi già definiti si aggiungono il gruppo che si può definire delle regioni del Sud (Basilicata, Puglia, Sicilia, Molise, Calabria e Sardegna) e un altro gruppo che comprende Liguria, Campania, Abruzzo, Emilia- Romagna, Marche, Friuli e Umbria. Scendendo ancora nei livelli di distanza e tagliando a livello di distanza 9,21 si può evidenziare che la classificazione a 5 gruppi è composta dai gruppi precedenti più un gruppo formato dalla sola regione Basilicata.

³³ Il metodo del legame completo individua gruppi compatti al loro interno ma di forma circolare, sferica o ipersferica, assegnando eventuali punti intermedi ai gruppi principali individuati (si veda par.3.2)

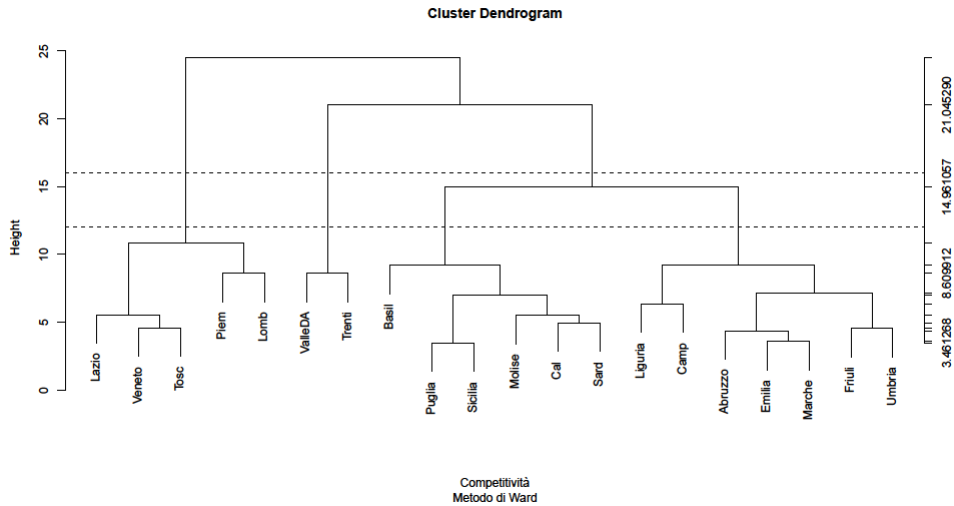
Fig.29 Dendrogramma - Metodo del legame completo – Distanza City-block



Applicando lo stesso metodo aggregativo, ma variando la misura di distanza otteniamo pressoché gli stessi risultati esposti in precedenza. L'unica differenza è che con tale metodo si considera solo la partizione a 3 e 4 gruppi poiché scendendo nell'aggregazione la classificazione diviene troppo frammentata e quindi meno utile.

Dalla figura 29 si può evidenziare la similitudine delle partizioni individuate. Anche in questo caso abbiamo per la suddivisione in 3 gruppi ($h=35$) un *cluster* formato dalle regioni Valle D'Aosta e Trentino, il secondo gruppo formato dalle regioni Piemonte, Lombardia, Veneto, Toscana, Lazio e il terzo macro-gruppo formato dalle restanti regioni. Scendendo nei livelli di distanza e posizionando l' α -taglio a livello di distanza 28, otteniamo la partizione in 4 cluster con la divisione del macro-gruppo in due gruppi formati rispettivamente dalle regioni Liguria, Emilia, Marche il primo e Campania, Puglia, Sicilia, Molise, Basilicata, Umbria, Friuli, Abruzzo, Calabria e Sardegna il secondo.

Fig.30 Metodo di Ward - Dendrogramma



L'applicazione del metodo gerarchico aggregativo di *Ward* porta ad una partizione molto simile a quella dei metodi precedentemente visti. Come si può osservare nella figura 30 sia per il livello di distanza (16) in cui si individuano tre gruppi che per il livello (12) che corrisponde alla partizione in quattro gruppi i risultati restano sostanzialmente omogenei. Con la partizione a tre gruppi Lazio, Veneto, Toscana, Piemonte e Lombardia formano il primo *cluster*, Valle d'Aosta e Trentino si aggregano nel secondo e infine abbiamo il macro-gruppo delle restanti regioni che con la partizione a quattro gruppi si suddivide nei due gruppi già visti: uno formato da Basilicata, Puglia, Sicilia, Molise, Calabria e Sardegna e un altro gruppo che comprende Liguria, Abruzzo, Emilia- Romagna, Marche, Friuli, Umbria e la Campania. Proprio l'aggregazione della Campania a questo gruppo rappresenta l'eccezione rispetto alle classificazioni viste in precedenza in cui si aggregava al gruppo delle regioni del Sud Italia e Isole.

4.5.2 Applicazione e risultati cluster analysis non gerarchica

Nei metodi gerarchici l'algoritmo cerca, ad ogni passo, la migliore scissione o aggregazione tra *cluster*, nel caso dei metodi non gerarchici l'algoritmo mira a ripartire le n unità in un numero predefinito di gruppi fornendo come risultato finale un'unica partizione in c gruppi basandosi sulla ottimizzazione di un criterio predeterminato.

Ciò consente di suddividere le unità statistiche in gruppi non sovrapposti, minimizzando la devianza interna ai gruppi stessi, e quindi massimizzando la loro omogeneità.

Quanto al tipo di distanza, la scelta è ricaduta su quella euclidea, che, oltre ad essere la più utilizzata in quanto garantisce la convergenza dell'algoritmo in pochi passi (<10), è anche facilmente interpretabile: infatti, essa non è altro che l'estensione nello spazio p -dimensionale della distanza tra due punti del piano cartesiano.

Il metodo delle k -medie è il metodo di classificazione non gerarchica più semplice computazionalmente e altrettanto semplice da implementare nei principali *software* statistici, fornisce buoni risultati a patto di fornire una ragionevole soluzione di partenza e un numero adeguato di cluster.

L'applicazione dei metodi gerarchici e l'individuazione dell' α -taglio nei dendrogrammi ci fornisce le informazioni utili sul numero adeguato di *clusters*.

Sulla base delle indicazioni della cluster gerarchica si è scelto di applicare il metodo delle k -medie per la classificazione a 3, 4 e 5 gruppi. La tabella 44 riporta le relative classificazioni ottenute. Tale metodo è stato utilizzato in maniera da poter avere una classificazione non gerarchica di tipo *crisp* prima di effettuare la classificazione *fuzzy*. Si sono potute così confrontare le partizioni ottenute. Per tali motivi il commento dei risultati ottenuti e la valutazione sulla migliore partizione sono rimandati al paragrafo relativo alla *cluster fuzzy k-means*.

Tab. 44 Partizioni ottenute con il metodo delle *K-medie*

	<i>K=3</i>	<i>K=4</i>	<i>K=5</i>
Piemonte	1	3	3
Valle D'Aosta	2	2	1
Lombardia	1	3	3
Trentino	2	2	1
Veneto	1	3	3
Friuli	2	1	5
Liguria	2	1	5
Emilia	2	1	5
Toscana	1	3	3
Umbria	2	1	5
Marche	2	1	5
Lazio	1	3	3
Abruzzo	2	1	5
Molise	3	4	2
Campania	3	4	2
Puglia	3	4	2
Basilicata	3	4	4
Calabria	3	4	2
Sicilia	3	4	2
Sardegna	3	4	2

4.6. *Classificazioni fuzzy delle regioni italiane*

Dopo l'applicazione delle tecniche di *clustering* di tipo *crisp*, sulla base delle prime indicazioni ottenute si è applicata una tecnica di classificazione di tipo sfocato. In particolare si è applicato all'insieme dei dati l'algoritmo *fuzzy k-means*.

Questo metodo è quello più utilizzato e più diffuso tra quelli di classificazione sfocata, si tratta di un'estensione del metodo *crisp* delle *k-medie*, e risulta particolarmente idoneo per trattare *dataset* di notevoli dimensioni, grazie alla velocità con la quale converge verso una classificazione ottimale.

Nella fase iniziale della procedura, analogamente alla versione *crisp* dell'algoritmo si sceglie il numero di cluster *c* in cui si vogliono classificare le

n unità e il valore del parametro m che esprime il livello di *fuzziness*, cioè il grado di *overlap* (sovrapposizione) tra i cluster della partizione. Nella ricerca si è scelto di classificare le regioni con un valore di c pari a 3, 4 e 5 gruppi con vari gradi di sfocatura. In particolare le partizioni si sono effettuate per $m=1.5$, $m= 1.7$, e $m=2$.

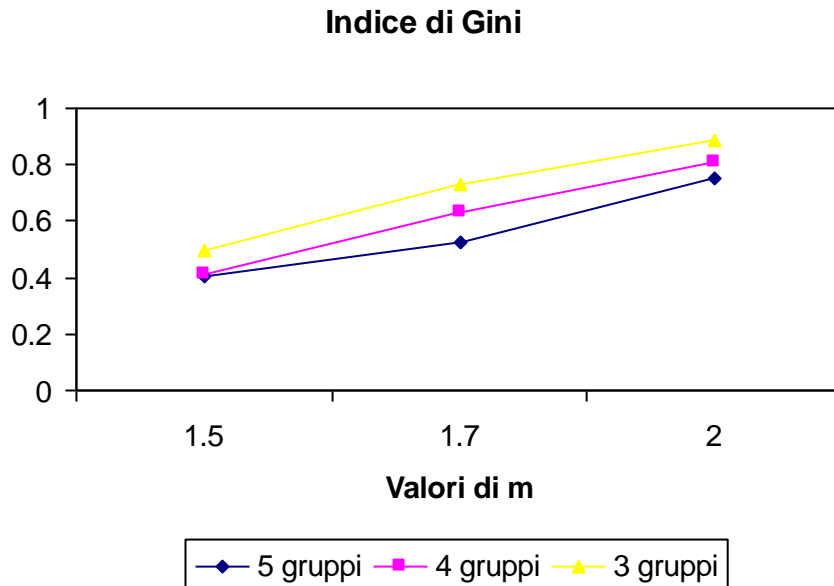
Tali scelte sono essenziali per inizializzare l'algoritmo e su tale problematica si sono concentrati gli studiosi. In particolare sono stati proposti vari indici che forniscono informazioni utili per valutare sia il numero di gruppi ottimale (c) che il miglior grado di sfocatura (m).

Per misurare il grado di sfocatura delle classificazioni si propone di utilizzare l'indice relativo di eterogeneità di Gini:

$$I = \frac{c}{c-1} \left[1 - \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^c u_{ik}^2 \right]$$

Tale indice varia nell'intervallo $[0,1]$ assumendo valore 1 nel caso di massima sfocatura e valore 0 nel caso in cui la classificazione è di tipo classico. L'indice aiuta quindi nella scelta di m poiché fornisce una misura sintetica dell'effetto dei diversi valori di m sulle relative classificazioni. Dalle diverse applicazioni effettuate in letteratura il valore migliore di m sembra essere quello ottenuto in corrispondenza del valore dell'indice I compreso tra 0.4 e 0.5. Nella ricerca tale indice è stato calcolato per i diversi valori di m per le differenti partizioni.(fig.31, tab.45).

Fig.31 Indice di Gini



Tab. 45 Indice di Gini

M	$c=5$	$c=4$	$c=3$
1.5	0.401	0.41	0.5
1.7	0.522	0.63	0.73
2	0.75	0.81	0.89

Da quanto appena illustrato il livello migliore di *fuzziness* è in corrispondenza del valore di $m=1.5$.

Un altro modo per valutare la *fuzziness* della partizione è quello di cercare di sintetizzare l'informazione contenuta nella matrice dei gradi di appartenenza in un unico numero che indichi il grado di accuratezza con cui avviene la classificazione.

Gli indici più diffusi che si basano su tale criterio sono il coefficiente di partizione PC e il coefficiente di entropia PE proposti da Bezdek:

$$PC = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^c u_{ik}^2$$

$$PE = \frac{1}{n} \sum_{i=1}^n \left(\sum_{k=1}^c u_{ik} \log u_{ik} \right)$$

L'indice PC vale $1/c$ in caso di *fuzziness* massima della partizione, se la partizione ottenuta è di tipo *crisp* PC vale 1 .

L'indice PE varia tra 0 e $\log c$ e aumenta all'aumentare della *fuzziness*, in caso di massima *fuzziness*, PE vale $\log c$, in caso di partizione *crisp*, vale 0 .

Gli indici PE e PC sono i due criteri più diffusi per misurare il grado di sovrapposizione tra i gruppi, tuttavia, soffrono di una forte sensibilità al parametro m . Quando m è molto vicino a 1 , o molto elevato, i due indici perdono la loro capacità discriminante tra i vari valori di c . Bedzek ha infatti dimostrato che quando m è prossimo a 1 , $PC=1$ e $PE=0$, mentre quando m è molto elevato, $PC=1/c$ e $PE=\log c$, e in tal caso i due indici preferiscono sempre $m=2$.

Per superare tale tendenza dei due indici, Davè ha proposto nel 1996 un indice compreso tra 0 e 1 , che vale 0 in caso di partizione massimamente *fuzzy*, e vale 1 in caso di partizione massimamente *crisp*.

$$MPC(c) = 1 - \frac{c}{c-1} (1 - PC(c))$$

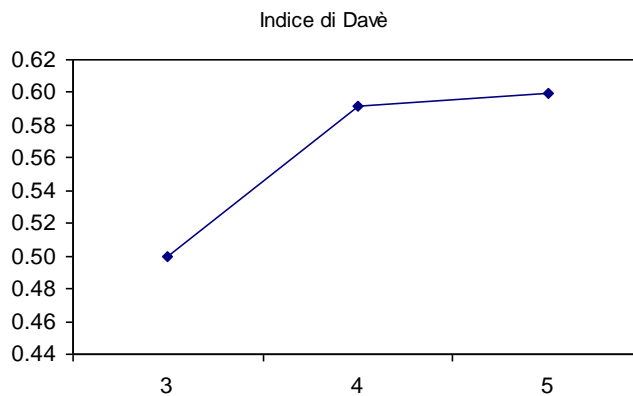
Nella successiva tabella 46 sono riportati i valori degli indici di *fuzziness* per le partizioni delle regioni italiane a 3, 4 e 5 gruppi.

Sulla scorta di quanto appena illustrato, l'indice di Davè (fig.32) porterebbe a scegliere la partizione a 3 gruppi come classificazione maggiormente sfocata.

Tab.46 Indici di cluster validity

Cluster	Gini	PC	PE	Davè
3	0,5	0,667	0,26	0,5
4	0,498	0,694	0,39	0,592
5	0,401	0,679	0,29	0,599

Fig. 32 Indice di Davè



Gli indici descritti si basano esclusivamente sulla matrice partizionata *fuzzy* è questo rappresenta il più grande limite di tali indici.

Nel lavoro si è ritenuto opportuno calcolare un altro tipo di indice di *fuzziness*, che vada a integrare le informazioni della matrice partizionata *fuzzy* con le informazioni della matrice dei dati.

Tale indice, denominato *Fuzzy Silhouette*, è stato proposto dagli autori Campello e Hruschka³⁴ ed è basato sull'indice *Crisp Silhouette*.

Data una partizione a k gruppi, si indica con p il cluster cui appartiene l'unità j , e con a_{pj} la distanza tra l'unità j e tutte le altre unità del cluster p cui essa appartiene.

Sia d_{qj} la distanza tra j e tutte le altre unità appartenenti ad un altro cluster q diverso da p .

³⁴ Campello, R.J.G.B., Hruschka, (2006). E.R. A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy sets and systems*, 157, 21, 2858-2875.

Sia b_{pj} la minima di queste distanze d_{qj} , che rappresenta la diversità tra l'unità j e l'unità più vicina a j di quelle appartenenti ad un altro cluster q . Si definisce con s_j il coefficiente silhouette (*riferito ad una sola unità*):

$$s_j = \frac{(b_{pj} - a_{pj})}{\max(a_{pj}, b_{pj})}$$

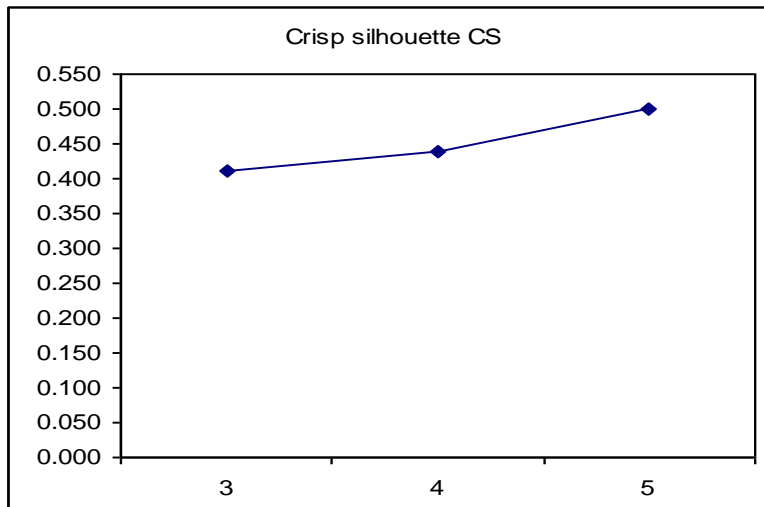
Naturalmente un valore positivo di questo indice sta ad indicare che a_{pj} è minore di b_{pj} , quindi, che l'unità j è più vicina alle unità appartenenti al suo cluster che alle unità appartenenti ad un cluster differente. Se il valore tende ad 1 possiamo considerare l'appartenenza dell'unità j al gruppo p molto soddisfacente perché vuol dire che a_{pj} sarà molto piccolo. Si definisce *Crisp Silhouette* l'indice:

$$CS = \frac{1}{N} \sum_{j=1}^N s_j \text{ (riferito ad un cluster di tipo crisp)}$$

La migliore partizione si ha quando CS è massimizzato cioè quando la distanza intracluster (a_{pj}) è molto minore della distanza intercluster (b_{pj}).

La figura 33 riporta il valore del *crisp silhouette* calcolato nella ricerca per le differenti partizioni delle regioni italiane. Si evidenzia che il punto di massimo corrisponde alla classificazione con 5 gruppi (*per $m=1.5$*).

Fig.33 Crisp Silhouette



L'estensione del *Crisp Silhouette* al caso di una classificazione di tipo *fuzzy* si definisce *Fuzzy Silhouette*:

$$FS = \frac{\sum_{j=1}^N (u_{pj} - u_{qj})^\alpha s_j}{\sum_{j=1}^N (u_{pj} - u_{qj})^\alpha}$$

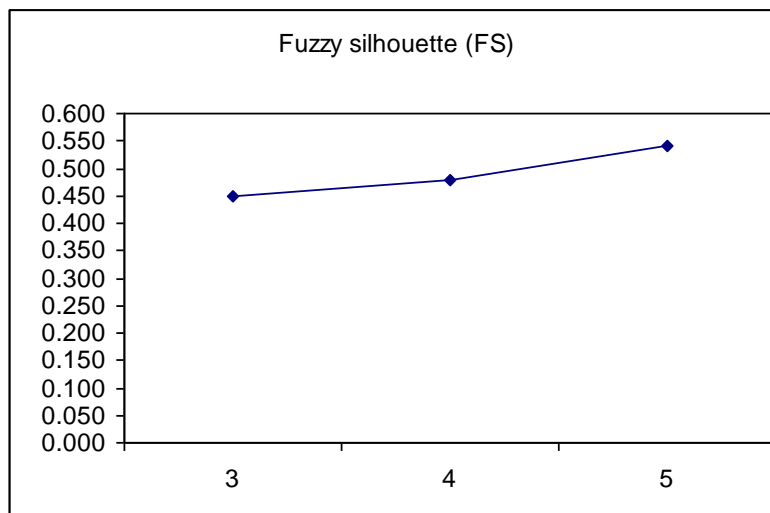
Nella formula dell'indice *FS*, u_{pj} e u_{qj} sono rispettivamente il primo e il secondo grado di appartenenza più elevati della matrice di partizione *fuzzy*. Il coefficiente s_j è l'indice silhouette dell'unità j , il coefficiente α è un criterio di ponderazione.

L'indice *FS* è la media ponderata della *silhouette* individuale delle singole unità, il peso è dato dalla differenza tra il primo grado di appartenenza e il secondo grado di appartenenza delle unità al *cluster fuzzy*.

La miglior partizione corrisponde, anche nel caso *fuzzy*, alla massimizzazione dell'indice *FS*.

Nell'analisi il valore massimo dell'indice corrisponde alla partizione con 5 gruppi. (fig.34)

Fig.34 Fuzzy Silhouette



Per confrontare le partizioni ottenute ed valutare la somiglianza tra di esse si è utilizzato l'indice di Rand, già descritto nel terzo capitolo, nella versione per partizioni *crisp*.

Il *Fuzzy Rand Index*³⁵ è un'estensione del *Rand index*, definito con $\omega=(a+d)/(a+b+c+d)$, basato sul confronto tra concordanze e discordanze tra due partizioni (a e d indicano il numero di coppie di oggetti che appartengono rispettivamente allo stesso cluster o a cluster diversi nelle due partizioni; b e c il numero di coppie di oggetti che appartengono rispettivamente allo stesso cluster nella prima partizione e a cluster diversi nella seconda e a cluster diversi nella seconda partizione e allo stesso cluster nella prima). Le due partizioni possono anche avere diverso numero di cluster.

³⁵ Per un approfondimento sul Fuzzy Rand Index si vedano:

Anderson, D.T., Bezdek, J.C., Popescu, M., Keller, J.M. (2010) Comparing Fuzzy, Probabilistic, and Possibilistic Partitions, *IEEE Transaction on Fuzzy Systems*, 18, 906-918.

Campello, R.J.G.B. (2007) A Fuzzy Extension of the Rand Index and Other Related Indexes for Clustering and Classification Assessment, *Pattern Recognition Letters*, 28, 833 – 841.

Nella tabella 47 sono elencati i valori di tale indice ottenuti per i confronti a coppie tra le partizioni *fuzzy* a 3,4,5 gruppi.

Tab.47 Fuzzy Rand Index

$k=3,k=4$	0,660
$k=3,k=5$	0,660
$k=4,k=5$	0,720

Dal valore del *fuzzy rand index* possiamo evidenziare che tra le partizioni a 4 e 5 gruppi c'è il miglior accordo, infatti l'indice di Rand assume valore 1 se le partizioni sono identiche, mentre un valore dell'indice pari a 0 indica che le due partizioni sono massimamente divergenti.

Nelle tabelle successive si riportano i gradi di appartenenza delle regioni ai cluster individuati attraverso l'applicazione della *cluster fuzzy*. Sono indicati i valori ottenuti per le partizioni a 3,4 e 5 gruppi con l'indicazione del cluster a cui viene assegnata l'unità.

Tab. 48 Gradi di appartenenza delle Regioni alla partizione con 3 cluster sfocati

Regione	Cluster	Grado di appartenenza		
		Cluster 1	Cluster 2	Cluster 3
Piemonte	3	0.148	0.158	0.694
ValleD'Aosta	1	0.668	0.195	0.137
Lombardia	3	0.087	0.107	0.805
Trentino	1	0.737	0.121	0.143
Veneto	3	0.080	0.040	0.880
Friuli	1	0.816	0.129	0.056
Liguria	1	0.707	0.133	0.160
Emilia	1	0.845	0.087	0.068
Toscana	3	0.085	0.041	0.874
Umbria	1	0.557	0.282	0.161
Marche	1	0.782	0.180	0.038
Lazio	3	0.023	0.023	0.953
Abruzzo	1	0.753	0.222	0.025
Molise	2	0.060	0.924	0.016
Campania	2	0.198	0.589	0.213
Puglia	2	0.023	0.963	0.014
Basilicata	2	0.212	0.723	0.065
Calabria	2	0.085	0.892	0.022
Sicilia	2	0.072	0.767	0.161
Sardegna	2	0.092	0.863	0.045

Tab. 49 Gradi di appartenenza delle Regioni alla partizione con 4 cluster sfocati

Regione	Cluster	Gradi di appartenenza			
		Cluster 1	Cluster 2	Cluster 3	Cluster 4
Piemonte	3	0.200	0.033	0.626	0.142
ValleD'Aosta	2	0.032	0.938	0.012	0.017
Lombardia	3	0.103	0.023	0.782	0.092
Trentino	2	0.051	0.911	0.021	0.017
Veneto	3	0.079	0.022	0.861	0.037
Friuli	1	0.768	0.044	0.058	0.130
Liguria	1	0.764	0.037	0.109	0.090
Emilia	1	0.944	0.005	0.022	0.029
Toscana	3	0.110	0.015	0.836	0.039
Umbria	1	0.716	0.022	0.095	0.167
Marche	1	0.820	0.031	0.027	0.122
Lazio	3	0.024	0.004	0.954	0.018
Abruzzo	1	0.896	0.008	0.011	0.085
Molise	4	0.112	0.006	0.014	0.868
Campania	4-1	0.338	0.020	0.165	0.478
Puglia	4	0.046	0.002	0.010	0.943
Basilicata	4	0.224	0.060	0.056	0.660
Calabria	4	0.135	0.009	0.019	0.837
Sicilia	4	0.118	0.008	0.126	0.748
Sardegna	4	0.108	0.015	0.037	0.840

Tab.50 Gradi di appartenenza delle Regioni alla partizione con 5 cluster sfocati

Regione	Cluster	Grado di appartenenza				
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Piemonte	3	0.031	0.193	0.562	0.154	0.060
ValleD'Aosta	1	0.916	0.034	0.013	0.015	0.022
Lombardia	3	0.022	0.101	0.730	0.111	0.035
Trentino	1	0.904	0.049	0.019	0.015	0.012
Veneto	3	0.019	0.069	0.861	0.041	0.010
Friuli	2	0.044	0.720	0.060	0.121	0.055
Liguria	2	0.034	0.740	0.101	0.094	0.031
Emilia	2	0.004	0.948	0.017	0.027	0.004
Toscana	3	0.013	0.101	0.830	0.045	0.010
Umbria	2	0.021	0.680	0.091	0.160	0.048
Marche	2	0.026	0.820	0.023	0.099	0.031
Lazio	3	0.004	0.022	0.949	0.021	0.005
Abruzzo	2	0.008	0.889	0.010	0.075	0.018
Molise	4	0.008	0.138	0.018	0.690	0.146
Campania	4	0.016	0.258	0.129	0.525	0.073
Puglia	4	0.001	0.018	0.004	0.970	0.007
Basilicata	5	0.000	0.000	0.000	0.001	0.999
Calabria	4	0.011	0.158	0.022	0.688	0.121
Sicilia	4	0.005	0.061	0.067	0.838	0.029
Sardegna	4	0.015	0.105	0.037	0.759	0.085

Come si può evidenziare dalle tabelle precedenti, le partizioni a 4 e 5 gruppi differiscono solo per il fatto che la Basilicata, nella partizione più disaggregata a 5 gruppi, si stacca dal cluster 4 e va a formare un gruppo a se stante.

Gli indici *CS* e *FS*, calcolati in precedenza, indicavano la classificazione con 5 gruppi come partizione ottimale. L'indice di *Rand* segnala, tuttavia, un buon raccordo tra le partizioni a 4 e 5 gruppi. In virtù di queste considerazioni e valutando che la partizione a 5 gruppi differisce solo per il cluster in più formato dalla sola regione Basilicata, si è scelto di considerare come risultato finale dell'analisi la partizione sfocata con 4 cluster. Nei paragrafi seguenti si commentano i risultati ottenuti cercando di evidenziare le caratteristiche comuni ai gruppi.

4.7. I cluster regionali

La *cluster analysis fuzzy* ha individuato 4 cluster sfocati che raggruppano le regioni italiane nel modo seguente (le regioni sono ordinate sulla base del grado di appartenenza al gruppo):

- **Cluster 1:** Emilia (0.944), Abruzzo (0.896), Marche (0.820), Friuli (0.768), Liguria (0.764), Umbria (0.716)
- **Cluster 2:** Valle d'Aosta (0.938) e Trentino (0.911)
- **Cluster 3:** Lazio (0.954), Toscana (0.834), Veneto (0.861), Lombardia (0.782), Piemonte (0.626)

- **Cluster 4:** Puglia (0.943), Molise (0.868), Sardegna (0.840), Calabria (0.837), Sicilia (0.748), Basilicata (0.660), Campania (0.478)

Dall'osservazione dei valori dei gradi di appartenenza è possibile notare quanto la maggioranza delle unità siano assegnate al relativo gruppo con un grado di appartenenza abbastanza elevato. Le unità che presentano un grado di appartenenza non sufficientemente elevato sono la Campania e la Basilicata per il gruppo 4, e, il Piemonte per il cluster 3.

Di seguito sono elencati tutti gli indicatori individuati attraverso l'analisi in componenti principali:

- ***TERR_C1: Indicatore delle risorse culturali e del turismo legato alla tradizione.***
- ***TERR_C2: Indicatore della risorsa “ mare”.***
- ***CLI_C1: Indicatore di condizioni climatiche sfavorevoli***
- ***OFF_C1: Indicatore della capacità ricettiva (alberghiera ed extralberghiera) regionale rispetto alla popolazione***
- ***OFF_C2: Indicatore dell'offerta ricettiva extralberghiera***
- ***FLS_C1: Indicatore della capacità di attrarre flussi turistici stranieri***
- ***FLS_C2: Indicatore (in negativo) della capacità di attrarre e trattenere i flussi turistici***
- ***ACS_C1: Indicatore (in negativo) della dotazione di infrastrutture per la mobilità e accessibilità nella Regione***
- ***SOST_C1 : Indicatore (in negativo) degli impatti dell'attività turistica***
- ***SOST_C2: Indicatore delle certificazioni ambientali e di qualità***
- ***PET_C1: Indicatore (in negativo) dell'impegno economico della Regione nel settore turistico***
- ***PET_C2: Indicatore (in negativo) dei benefici economici del turismo sulla regione***

- **PA_C1: Indicatore (in negativo) della redditività degli alberghi**
- **PA_C2: Indicatore (in negativo) delle dimensioni delle aziende alberghiere nella regione**
- **PERF_ADV_C1: Indicatore delle dimensioni aziendali e di scarsa redditività**

L'interpretazione e il commento sui gruppi individuati è basato sui valori dei punteggi fattoriali ottenuti dall'analisi in componenti principali, sui valori medi calcolati per gruppo e , naturalmente, con un riferimento costante ai dati di partenza.

La tabella successiva riporta i centroidi finali per gruppo dei punteggi fattoriali, ottenuti con la *cluster analysis fuzzy*.

Tab.51 Centroidi finali dei cluster sfocati

Cluster	TERR_C1	TERR_C2	CLI_C1	OFF_C1	OFF_C2	FLS_C1
1	-0.339	-0.551	0.230	0.503	0.189	-1.180
2	-2.849	-0.293	2.900	3.886	-0.439	1.408
3	2.214	-0.353	0.430	-0.651	0.587	2.549
4	-0.478	0.965	-1.530	-1.218	-0.561	-1.149
	FLS_C2	ACS_C1	PET_C1	PET_C2	ALB_C1	ALB_C2
1	-0.399	-0.624	0.334	0.251	-0.512	-0.448
2	-3.342	1.678	-1.577	-2.628	0.374	-0.504
3	0.956	-1.240	0.985	-1.027	1.275	1.230
4	0.664	1.098	-0.411	1.349	-0.407	-0.326
	ADV_C1	SOST1	SOST2			
1	-0.288	-0.732	0.156			
2	-1.319	-3.181	-0.287			
3	1.049	0.856	1.203			
4	-0.250	0.875	-1.007			

4.7.1. Gruppo 1: Destinazioni di prossimità

Il gruppo 1 è stato definito “*Destinazioni di prossimità*” poiché l’analisi statistica evidenzia che le regioni appartenenti al cluster sono interessate principalmente da flussi nazionali che riescono a trattener con una buona

permanenza media. Il cluster ha, infatti, il valore più basso per l'indicatore sui flussi stranieri segnalando la difficoltà di queste regioni di "vendersi" all'estero.

Il gruppo si presenta omogeneo al suo interno, infatti i valori dell'indice fuzzy silhouette (tab.53) e, in maniera visiva più immediata, la figura 35, mostrano la compattezza del cluster, evidenziando come la regione Emilia Romagna, che è anche la regione con il grado di appartenenza più alto si distacchi, presenti il valore maggiore dell'indice Fuzzy segnalando il maggior peso che la regione ha nel determinare il cluster.

Le regioni del gruppo presentano in generale un sufficiente livello competitivo del turismo, ma con ampi margini di miglioramento. Tali regioni presentano, infatti, buone potenzialità in termini di risorse, in misura maggiore per le risorse legate alle peculiarità dei territori e alla unicità dei piccoli borghi, infatti le regioni del gruppo presentano il numero più elevato di borghi riconosciuti come Borghi più belli d'Italia.

Sebbene l'indicatore sulla risorsa mare presenti i valori più bassi, molte regioni appartenenti al cluster possono sicuramente puntare anche sulla risorsa turistica del mare (Emilia, Liguria, Abruzzo e Marche) esprimendo buone potenzialità per il turismo balneare.

Anche il turismo di montagna potrebbe essere intensificato per alcune regioni del cluster (Abruzzo, Umbria) come un'opportunità per destagionalizzare i flussi turistici. Il cluster raggruppa regioni con una discreta dotazione in termini di posti letto (rispetto alla popolazione) sia alberghieri che della ricettività complementare.

Le strutture presenti sono principalmente di tipo extralberghiero e di medio-piccole dimensioni, con valori elevati rispetto a tale ricettività per le Marche e il Friuli Venezia Giulia.

La regione Emilia-Romagna mostra, al contrario, una maggiore disponibilità di posti letto nel settore alberghiero, con una percentuale

dell'indice di ricettività alberghiera più alto a livello nazionale e non solo di gruppo.

Le peculiarità del territorio sono comunque favorevoli a questo tipo di ricettività, si pensi agli alberghi diffusi che trovano la loro naturale integrazione nei piccoli centri storici e borghi in modo da offrire quell'atmosfera unica che il turista moderno ricerca nella sua esperienza di viaggio senza dover rinunciare ai tradizionali comfort dei servizi alberghieri.

La dotazione di infrastrutture per l'accessibilità delle regioni è buona, migliore rispetto ad altri cluster (2 e 4), considerando anche che il gruppo è composto da regioni di medie dimensioni e con un percentuale di superficie montana abbastanza elevata. Si distinguono nel gruppo per la dotazione stradale, autostradale e ferroviaria la Liguria, l'Abruzzo e l'Emilia Romagna. Rilevante anche il fatto che in ogni regione di questo gruppo, seppur di medie dimensioni, ci sia almeno un aeroporto fino al massimo di 12 aeroporti della regione Emilia.

Per quanto concerne l'indicatore sull'impegno economico delle regioni a livello di gruppo non si evidenziano grandi risorse investite per il turismo. A livello di singole regioni però le percentuali rispetto alla spesa totale regionale presentano alcune distinzioni all'interno del gruppo. Friuli, Liguria, Emilia investono sicuramente in misura maggiore rispetto alle altre regioni del gruppo che presentano percentuali di spesa inferiore alla media nazionale. Gli effetti positivi del turismo sul benessere economico e sull'occupazione misurati dall'indicatore PET_C2, restituiscono una situazione sostanzialmente omogenea tra le regioni collocandosi come gruppo in fondo alla graduatoria. Infatti i valori della variabile relativa alla bilancia turistica sono negativi per molte regioni del gruppo poiché come già detto le regioni mostrano difficoltà nel richiamare flussi stranieri. La percentuale di occupati nel settore (ad eccezione della Liguria) e i valori del PIL pro-capite sono al di sotto della media nazionale (ad eccezione dell'Emilia-Romagna).

I dati relativi alla sostenibilità del gruppo 1 segnalano un livello di intensità turistica abbastanza elevato, in maniera più significativa per le regioni Emilia, Liguria, Friuli, Marche.

Il cluster si differenzia dagli altri, anche per l'attenzione agli strumenti della certificazione ambientale e di qualità, con un buon numero di comuni con il riconoscimento delle Bandiere Arancioni (56) e delle strutture certificate con il Marchio Ospitalità Italiana (1265).

Analizzando la competitività delle imprese turistiche il gruppo comprende regioni con una dimensione media in termini di dipendenti e fatturato. Emilia e Abruzzo sono le regioni del gruppo con una media più elevata di dipendenti per struttura. Per ciò che riguarda la redditività delle imprese del settore alberghiero, a livello nazionale, si rileva un'inefficienza nella gestione del capitale proprio soprattutto nel settore alberghiero (ROE), mentre per la redditività della gestione caratteristica (ROS) gli indici presentano valori sicuramente più positivi. Il settore dell'intermediazione, invece, a livello nazionale, produce una maggiore redditività anche dei mezzi propri oltre che della gestione caratteristica.

La redditività del settore alberghiero per il cluster 1 è maggiore rispetto al livello nazionale (valore più basso dell'indicatore in negativo) e degli altri gruppi, soprattutto per ciò che riguarda l'efficienza della gestione caratteristica. Situazione simile anche per il settore dell'intermediazione che, però, presenta valori leggermente inferiori di redditività della gestione caratteristica.

Tab.52 Cluster 1- Punteggi fattoriali

Regioni	TERR_C1	TERR_C2	CLI_C1	OFF_C1	OFF_C2	FLS_C1	FLS_C2
Friuli	-1.289	0.247	1.810	0.324	2.515	-0.103	0.106
Liguria	0.199	0.344	0.238	0.995	0.260	-0.897	-0.700
Emilia	0.740	-0.249	-0.088	0.941	-0.696	-1.102	-1.067
Umbria	-1.174	-2.408	0.127	0.100	1.931	-0.201	0.665
Marche	-0.301	-0.882	-0.244	1.364	-1.299	-3.070	-1.604
Abruzzo	-1.351	-1.322	0.427	0.424	-0.307	-2.122	-0.324
Media	-0.339	-0.551	0.230	0.503	0.189	-1.180	-0.399

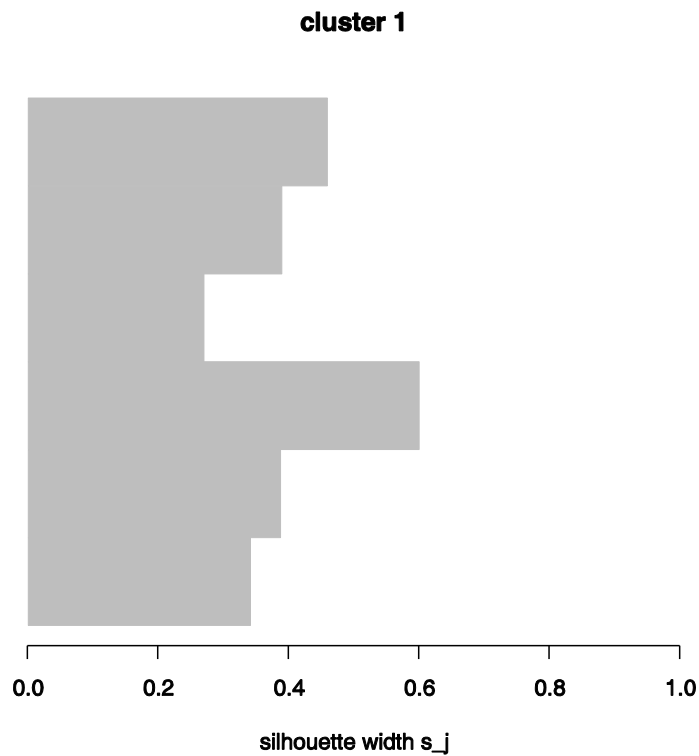
Tab.52 Cluster 1- Punteggi fattoriali

Regioni	ACS_C1	PET_C1	PET_C2	SOST_C1	SOST_C2	PA_C1	PA_C2	PE_ADV_C1
Friuli	-0.218	-0.472	-0.296	-0.821	-0.984	-0.951	-0.646	-0.837
Liguria	-3.961	0.563	-1.224	-1.133	0.897	-2.019	0.185	-1.329
Emilia	-1.069	0.809	0.107	-1.276	1.190	0.262	-0.377	0.022
Umbria	0.693	0.279	0.580	0.041	0.532	0.510	-1.156	0.744
Marche	0.582	0.800	0.605	-2.095	0.315	-0.599	-1.145	0.914
Abruzzo	-0.279	0.632	1.285	-0.838	-0.713	-0.189	-0.210	-1.691
Media	-0.624	0.334	0.251	-0.732	0.156	-0.512	-0.448	-0.288

Tab. 53 Cluster 1 – Fuzzy Silhouette

Regione	Silhouette
Friuli V.Giulia	0,53
Liguria	0,59
Emilia-Romagna	0,65
Umbria	0,49
Marche	0,56
Abruzzo	0,56

Fig.35 Cluster 1- Grafico fuzzy silhouette



4.7.2. Gruppo 2: Destinazioni “leader”

Il cluster 2 è formato dalle regioni Trentino e Valle d’Aosta, che vengono definite destinazioni leader poiché presentano alcune aspetti di eccellenza rispetto alle altre regioni e, quindi, agli altri cluster.

Il gruppo risulta il più omogeneo anche in virtù del fatto che è composto da due sole regioni, l’omogeneità del gruppo è evidente osservando la figura 36 e la tabella 55, le due regioni, infatti hanno un valore poco differente del fuzzy silhouette.

La dotazione di risorse è sicuramente inferiore rispetto agli altri cluster, ciò è evidenziato dai valore massimamente negativi degli indicatori TERR_C1 e TERR_C2 . La situazione si presenta molto svantaggiosa per la Valle d’Aosta, mentre il Trentino dispone di qualche elemento di attrattiva in più da poter sfruttare turisticamente ma, comunque, rispetto ad altre regioni, le risorse restano limitate. Le regioni sono caratterizzate da un territorio montuoso che è la loro principale risorsa e da un clima abbastanza sfavorevole rispetto agli altri cluster.

Le caratteristiche geologiche sono sicuramente una delle cause della scarsa dotazione di infrastrutture stradali e ferroviarie delle regioni che presentano il valore più basso per l’indicatore sull’accessibilità.

Le due regioni del gruppo, tuttavia, sono state definite come destinazioni leader per i risultati raggiunti nonostante il territorio non sia ricchissimo di fattori d’attrattiva.

Il sistema di offerta evidenzia la forte propensione turistica delle due regioni con un numero di posti letto sia alberghieri che extralberghieri di gran lunga superiori alla media, infatti il valore dell’indicatore OFF_C1 è il più elevato. Mentre il Trentino offre in misura maggiore posti letto in strutture alberghiere di piccole dimensione la Valle D’Aosta si distingue per la prevalenza dei posti letto in strutture complementari di grandi dimensioni.

Si segnala che le due regioni presentano in assoluto i più bassi valori per gli indici di ricettività di alta categoria, quindi le strutture presenti sono maggiormente di livello medio-basso.

Le *performances* molto positive delle regioni in termini di flussi attratti evidenziano quanto questi territori abbiano saputo costruire un vantaggio competitivo che va al di là degli elementi di attrattiva e delle risorse disponibili. (il valore dell'indicatore FLS_1 è il secondo più elevato dopo il cluster 3, mentre l'indicatore FLS_2 presenta il valore più alto, considerando che l'indicatore è in negativo). Il confronto con il numero di arrivi di regioni quali Campania, Sicilia, Puglia deve sicuramente far riflettere. Ogni 100 abitanti il Trentino ha un numero di arrivi di turisti pari a 394 italiani e 436 stranieri, la Valle d'Aosta ha arrivi per 450 italiani e 325 stranieri, la Puglia si ferma a 61 italiani e 10 stranieri, la Campania 49 italiani e 28 stranieri, la Sicilia 51 italiani e 32 stranieri. Il gap è senz'altro evidente e non giustificabile solo con la localizzazione sicuramente più favorevole ai flussi stranieri per Trentino e Valle D'Aosta. Il Trentino Alto- Adige è la seconda regione in Italia dopo il Veneto ad attrarre flussi stranieri e i dati rivelano anche la capacità di trattenere e destagionalizzare i flussi con indici di utilizzazione e permanenza media al di sopra della media nazionale.

La maggiore competitività di tali regioni si rispecchia anche negli impatti positivi che il turismo genera in termini di occupazione e bilancia dei pagamenti turistica. Le due regioni presentano le più alte percentuali in Italia di addetti al turismo e valori in attivo della bilancia dei pagamenti e risulta elevato anche il Pil procapite . Anche l'investimento economico delle regioni in termini di spese per il turismo è sicuramente superiore agli altri cluster individuati.(il cluster occupa il primo posto nella graduatoria dei cluster per gli indicatori *PET_C1* e *PET_C2*)

Gli indicatori sulla sostenibilità indicano per le regioni un impatto sicuramente elevato dell'attività turistica in termini di tasso di turisticità e ricettività confermando la forte propensione turistica delle due regioni.

Relativamente agli strumenti di certificazione ambientale e di qualità, sebbene il Trentino si distingua per il maggior numero in assoluto di strutture turistiche certificate Ecolabel (67), per ciò che concerne gli altri strumenti di certificazione considerati nell'analisi il cluster registra il numero più esiguo di strutture certificate, ciò si evidenzia, infatti, dal valore più basso del cluster per l'indicatore SOST_C2.

La competitività maggiore mostrata investe anche le società alberghiere e le agenzie di viaggio con sede nelle regione Trentino con tutti gli indici di redditività positivi tranne il ROE del settore alberghiero che comunque resta il valore più alto a livello nazionale. La Valle d'Aosta, al contrario, segnala una redditività dei due settori in linea con la situazione abbastanza negativa a livello nazionale.

Tab.54 Cluster 2- Punteggi fattoriali

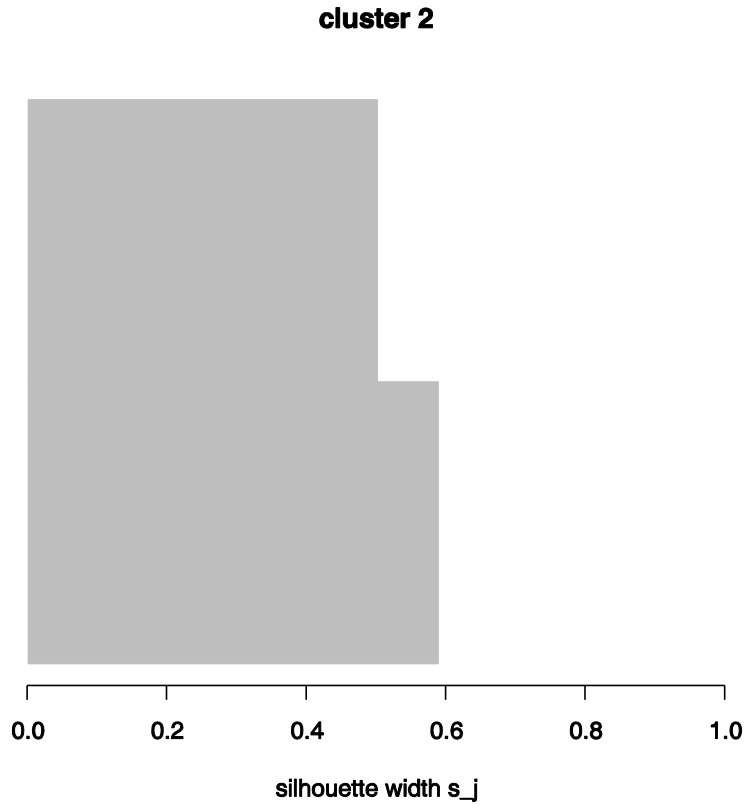
Regione	TERR_C1	TERR_C2	CLI_C1	OFF_C1	OFF_C2	FLS_C1	FLS_C2	ACS.C1
Valle D'Aosta	-3.636	-0.035	3.068	4.252	-1.402	-0.030	-2.587	2.837
Trentino	-2.211	-0.573	2.928	3.796	0.5 19	3.048	-4.402	0.597
Media	-2.849	-0.293	2.9	3.886	-0.439	1.408	-3.342	1.678

Regione	PET_C1	PET_C2	SOST_C1	SOST_C2	PA_C1	PA_C2	PE_ADV_C1
Valle D'Aosta	-3.107	-3.449	-2.282	-0.679	3.010	-1.569	-1.063
Trentino	-0.012	-1.968	-4.365	0.102	-2.317	0.591	-1.692
Media	-1.577	-2.628	-3.181	-0.287	0.374	-0.504	-1.319

Tab. 55 Cluster 2-Fuzzy Silhouette

Regione	silhouette
Valle d'Aosta	0,648928
Trentino A.A.	0,582878

Fig.36 Cluster 2 – Grafico Fuzzy Silhouette



4.7.3. Gruppo 3: Destinazioni mature

Il gruppo comprende le regioni Lazio, Toscana, Veneto, Piemonte e Lombardia che rappresentano sicuramente le regioni con le destinazioni più note del nostro paese sia per il turista nazionale che, soprattutto, per i turisti internazionali. Tali regioni rappresentano sicuramente, anche se per motivazioni differenti, le regioni più turistiche del nostro paese. Mentre Lazio, Veneto e Toscana si distinguono per essere maggiormente votate al turismo culturale, Piemonte e Lombardia sono sicuramente indirizzate maggiormente verso un turismo di tipo *business*.

La tabella 57 e il grafico 37 mostrano una discreta omogeneità del gruppo, anche se minore rispetto al primo e al secondo gruppo, con le regioni Lazio e Lombardia che assumono un ruolo maggiormente caratterizzante per il cluster in esame.

Tutte le regioni del gruppo rappresentano le regioni più dotate di risorse culturali, ma anche le potenzialità delle risorse legate al patrimonio agroalimentare sono sicuramente superiori agli altri cluster.

Tale set di risorse è supportato dalla migliore dotazione di infrastrutture (valore più elevato in negativo di ACS_C1) che rende sicuramente più agevole l'accessibilità alle destinazioni regionali del cluster.

Per ciò che riguarda l'offerta ricettiva il gruppo presenta il valori più elevato dell'indicatore OFF_C2, riguardante l'offerta extralberghiera. Infatti, Veneto e Toscana presentano un elevato numero di posti letto per questa tipologia di ricettività. A livello generale di cluster, la capacità ricettiva rispetto alla popolazione non risulta elevata, anche se, nel cluster sono presenti un maggior numero di strutture di lusso (4-5 stelle) rispetto agli altri due gruppi già commentati e strutture alberghiere di dimensione più grande in termini di posti letto per struttura.

Il cluster mostra una maggiore capacità di attrarre principalmente flussi stranieri (maggior valore per l'indicatore FLS_C1), anche perché a livello nazionale queste destinazioni hanno ormai raggiunto la fase di maturità quindi è naturale registrare una fase di stagnazione degli arrivi italiani.

Le regioni infatti, ad eccezione del Piemonte (che ha un minor grado di appartenenza), rappresentano insieme al Trentino le regioni preferite dal turismo internazionale con indici di utilizzazione delle strutture da parte degli stranieri superiori alla media nazionale e in alcuni casi superiori agli indici di utilizzazione degli italiani. Il Lazio e il Veneto sono le destinazioni maggiormente interessate da flussi stranieri soprattutto per l'incidenza di destinazioni come Roma e Venezia. La capacità di attrarre e trattenere flussi turistici stranieri si manifesta anche nei valori della bilancia turistica che

presenta per Veneto, Toscana e Lazio valori estremamente positivi rispetto al resto d'Italia, mentre Piemonte e Lombardia presentano un valore passivo della bilancia turistica.

Il cluster si colloca in fondo alla classifica dei cluster per l'indicatore PET_C1 relativo all'impegno economico delle regioni per il turismo, gli investimenti di queste regioni, infatti, sono sostanzialmente al di sotto della media nazionale. Ciò può essere giustificato anche dal fatto che queste regioni godono ormai di una rendita di posizione, dovuta alla loro immagine e notorietà nel mercato nazionale e internazionale quindi non necessitano di grandi investimenti nella promozione e comunicazione delle loro destinazioni.

. L'indicatore sull'impatto economico del turismo colloca il cluster al secondo posto come capacità di trarre i massimi benefici economici dal turismo, la percentuale di occupati nelle regioni presenta valori nella media ad eccezione della Toscana che registra una percentuale di 6,8% di occupati nel settore, il Pil pro-capite è superiore alla media nazionale.

Il gruppo si differenzia dagli altri per l'attenzione alla sostenibilità e alla qualità con un vasto numero di siti certificati, con numerose strutture riconosciute con il marchio dell'I.S.N.A.R.T. (Lombardia, Piemonte, Veneto e Toscana sono in testa alla classifica per numero di strutture certificate), con vari comuni delle regioni riconosciuti con il riconoscimento del Touring club (Toscana, Piemonte e Veneto soprattutto) e, ugualmente, il marchio Ecolabel è abbastanza diffuso tra le imprese delle regioni appartenenti al cluster.

Gli impatti dell'attività turistica in termini di tasso di turisticità e ricettività sono sostanzialmente sotto controllo come testimonia il valore dell'indicatore SOST_C1 che colloca il gruppo al terzo posto.

Infine, la valutazione sulle performance aziendali ci porta ad osservare la mediocre redditività dei settori analizzati anche se nel gruppo sono concentrate le società di maggiori dimensioni in termini di fatturato e dipendenti.

Le regioni del gruppo che manifestano una maggiore redditività delle imprese sono Lombardia e Lazio.

Tab.56 Cluster 3 – Punteggi fattoriali

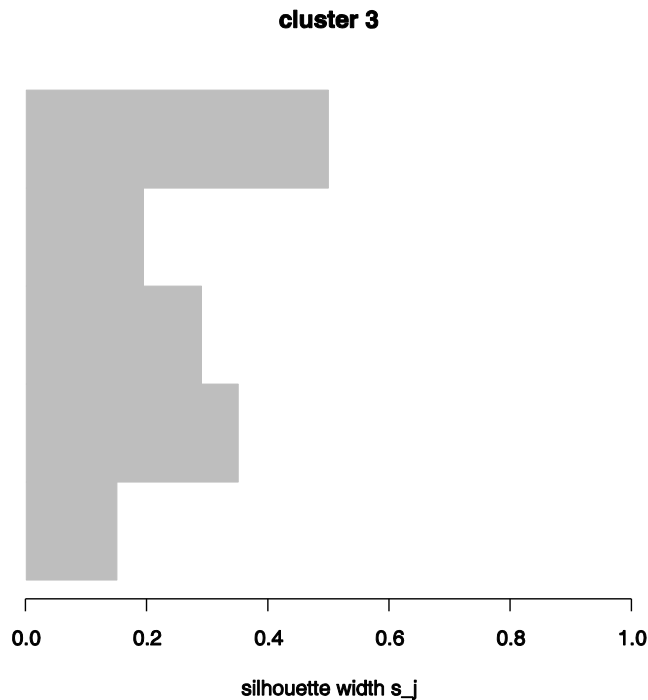
Regione	TERR_C1	TERR_C2	CLI_C1	OFF_C1	OFF_C2	FLS_C1	FLS_C2	ACS_C1
Piemonte	1.833	-0.878	1.593	-0.237	0.676	0.764	2.289	-1.204
Lombardia	2.241	-1.258	0.859	-0.848	-1.281	2.051	1.935	-1.941
Veneto	2.010	0.401	0.841	-0.252	2.287	3.356	-0.685	-0.861
Toscana	3.688	-0.572	-0.130	0.158	0.444	1.809	0.105	-0.473
Lazio	1.801	-0.037	-0.120	-1.750	0.666	4.543	1.833	-1.832
Media	2.214	-0.353	0.43	-0.651	0.587	2.549	0.956	-1.24

Regione	PET_C1	PET_C2	SOST_C1	SOST_C2	PA_C1	PA_C2	PE_ADV_C1
Piemonte	-0.088	0.393	2.174	2.510	0.791	-0.696	6.004
Lombardia	1.095	-0.615	3.889	1.484	1.174	5.587	1.452
Veneto	1.291	-1.350	-0.573	0.273	3.346	0.593	0.188
Toscana	1.276	-1.675	-0.853	2.479	1.514	-0.221	-0.438
Lazio	1.194	-1.633	0.508	0.468	0.218	1.090	0.472
Media	0.985	-1.027	0.856	1.203	1.275	1.23	1.049

Tab.57 Cluster 3 – Fuzzy Silhouette

regione	silhouette
Piemonte	0,350944
Lombardia	0,515219
Veneto	0,369637
Toscana	0,266627
Lazio	0,536497

Fig. 37 Cluster 3 – Grafico Fuzzy Silhouette



4.7.4. Gruppo 4 : Destinazioni “da sviluppare”.

Il gruppo comprende, in ordine di grado di appartenenza, le regioni Puglia, Molise, Calabria, Sardegna, Sicilia, Basilicata e Campania. Il gruppo è stato definito come destinazioni da sviluppare perché comprende le regioni che sicuramente hanno un altissimo potenziale da sfruttare turisticamente ma non riescono a supportare tali risorse con un offerta integrata e politiche di promozione e sviluppo turistico.

La figura 38 e la relativa tabella 59 evidenziano la maggiore eterogeneità di questo cluster rispetto agli altri, con valori più bassi in generale per l'indice Fuzzy Silhouette. Ciò sta a significare che le regioni del cluster presentano maggiore variabilità e quindi sono più “distanti” tra loro. In

particolare i valori del fuzzy silhouette rivelano un distacco dal gruppo della Campania, che, infatti, presenta anche il più basso grado di appartenenza.

La principale risorsa del gruppo è sicuramente il mare (massimo valore per TERR_C2) che però dovrebbe essere tutelato in misura maggiore. Guardando ai dati, è evidente infatti che considerando il dato del numero di bandiere blu ogni 10 Km di costa balneabile si evidenzia che le regioni di questo gruppo (a parte per la Campania e il Molise) presentano la percentuale minore.

Il gruppo dispone di un vasto set di beni culturali e siti UNESCO, con Campania e Sicilia a guidare la classifica, ma altrettanto importanti sono le potenzialità del cluster in termini di risorse culturali legate al patrimonio agroalimentare e alle tradizioni. Tale risultato assume una valenza sicuramente significativa se si pensa all'opportunità che si ha, con questo immenso patrimonio, per destagionalizzare i flussi turistici orientati principalmente al turismo balneare, e attrarre la domanda internazionale maggiormente interessata a tali elementi di attrattiva.

Le regioni appartenenti al cluster sono inoltre favorite anche da un clima favorevole con temperature miti sempre al di sopra della media nazionale.

Il sistema ricettivo del cluster è caratterizzato da strutture di grandi dimensioni soprattutto per il comparto alberghiero rispetto alla media nazionale anche se in termini di proporzione dei posti letto sulla popolazione solo la Calabria e la Sardegna registrano un valore elevato mentre le altre regioni manifestano una capacità ricettiva esigua rispetto alla popolazione. Ciò si evidenzia nei valori delle componenti estratte per l'Offerta, infatti sia per OFF_C1 che per OFF_C2 il gruppo occupa l'ultima posizione,

Rilevante il dato sulle strutture di qualità, i letti 4/5 stelle sul totale dei letti alberghieri presentano i valori più alti insieme a Lombardia e Lazio per tutte le regioni del cluster individuato. Anche l'indice di ricettività alberghiera presenta valori superiori alla media nazionale.

La situazione relativa ai flussi registrati, in termini di arrivi sia italiani che stranieri, è sicuramente negativa rispetto agli altri cluster. Particolarmente esiguo il numero degli stranieri, con il Molise in fondo alla classifica con 5 arrivi di turisti stranieri ogni 100 abitanti. Per comprendere il *gap* competitivo con le regioni leader, basti pensare che, una regione come la Valle d'Aosta dotata di risorse sicuramente inferiori anche a quelle della nostra regione Molise, riesca ad attrarre 225 turisti stranieri ogni 100 abitanti. I numeri bastano per capire quanto sia scarsa la cultura pro-turismo nelle regioni che in realtà avrebbero tutte le carte in regola per competere nei mercati nazionali ed internazionali. Le differenze risultano elevate anche per gli arrivi dei turisti nazionali, per il Molise il dato si ferma a 56 turisti, la Valle d'Aosta attira 450 turisti domestici ogni 100 abitanti.

Il contesto è simile per tutte le regioni del cluster soprattutto in termini di arrivi stranieri, infatti le regioni mostrano una maggiore propensione ad attirare i flussi nazionali.

Ulteriore conferma a quanto detto ci è data dagli indici di utilizzazione che per gli stranieri presentano la media più bassa, con una situazione leggermente migliore per Campania e Sicilia che mostrano una capacità maggiore di attrarre flussi stranieri, dovuta principalmente all'immenso patrimonio culturale di queste regioni. Gli indici di utilizzazione degli italiani sono sostanzialmente nella media nazionale.

Le infrastrutture per il trasporto e la mobilità del gruppo è sicuramente di scarso livello collocandosi per quasi tutte le variabili considerate al terzo posto prima del cluster formato da Trentino e Valle d'Aosta, che come già detto mostrano elementi di criticità su questo aspetto.

Preoccupante il dato sulla rete autostradale che presenta una media di 1.50 Km ogni 100 km, tale proporzione è spinta in alto da regioni come Campania e Sicilia che presentano una maggiore dotazione.

Le considerazioni esposte si riflettono anche sulla scarsa capacità del turismo di apportare benefici economici nelle regioni del cluster con una

bilancia dei pagamenti negativa per alcune regioni del gruppo ad eccezione di Sicilia, Campania e Sardegna che presentano una spesa per viaggiatore straniero abbastanza elevata. In queste regioni gli stranieri arrivano meno numerosi rispetto ad altri cluster, tuttavia coloro che arrivano tendenzialmente restano di più e spendono di più che in altre regioni.

L'indicatore PET_C1 segnala che gli investimenti nel settore sono abbastanza elevati nel *cluster*, ciò comunque rappresenta un dato negativo poiché tali investimenti non riescono a tradursi in benefici per tali regioni evidenziando un probabile spreco di risorse non finalizzate in maniera efficace.

Il gruppo manifesta una maggiore attenzione alla qualità rispetto alla sostenibilità. Infatti il cluster si distingue per una buona media di imprese certificate di qualità, mentre per quanto riguarda i riconoscimenti relativi alla sostenibilità sia per l'*Ecolabel* che per il numero di Bandiere Arancioni sia ben al di sotto della media nazionale. Unica nota positiva il riconoscimento I.S.N.A.R.T. che in regioni quali Campania, Puglia, Sardegna e Calabria presenta un numero di strutture certificate abbastanza soddisfacente.

Da rilevare che gli impatti del turismo sul territorio a livello di intensità turistica rispetto alla popolazione non sono sicuramente preoccupanti anzi evidenziano la bassa propensione al turismo come attività primaria e lo scarso sfruttamento dal punto di vista turistico di tali territori.

A livello di redditività per il settore alberghiero si evidenzia una maggiore redditività della gestione caratteristica per questo gruppo rispetto agli altri *cluster* e anche l'efficienza della gestione del capitale proprio misurata attraverso il ROE risulta leggermente migliore rispetto agli altri gruppi.

Per il settore dell'intermediazione i risultati sono meno positivi e in linea con la tendenza generale del settore, con una maggiore redditività delle vendite per il cluster in esame.

Tab. 58 Cluster 4- Punteggi fattoriali

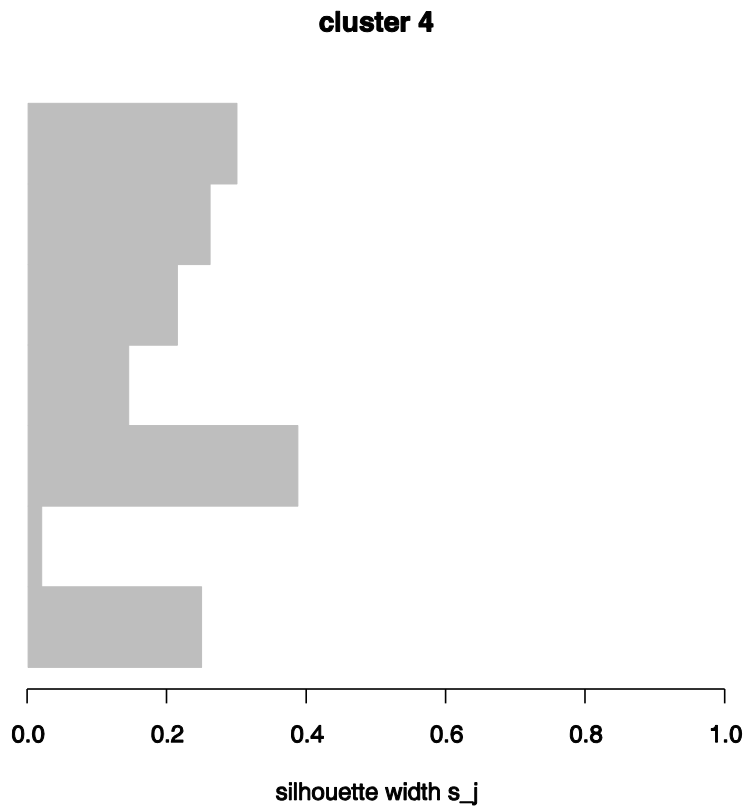
Regioni	TERR_C1	TERR_C2	CLI_C1	OFF_C1	OFF_C2	FLS_C1	FLS_C2	ACS_C1
Molise	-2,086	0,425	-0,318	-0,531	0,335	-2,198	1,391	1,883
Campania	1,815	0,757	-0,732	-1,294	-0,337	-0,596	-0,141	-1,582
Puglia	0,425	1,130	-1,955	-1,426	-0,565	-1,176	1,177	0,194
Basilicata	-2,332	0,032	-1,360	-1,152	-0,840	-1,762	1,245	2,049
Calabria	-1,106	1,167	-2,017	-1,306	-2,480	-1,910	-0,347	0,637
Sicilia	1,754	1,493	-2,759	-1,782	0,247	0,018	1,136	0,272
Sardegna	-1,019	2,220	-2,167	-1,776	-0,674	-0,421	-0,027	3,676
Media	-0.478	0.965	-1.53	-1.218	-0.561	-1.149	0.664	1.098

Regioni	PET_C1	PET_C2	SOST_C1	SOST_C2	PA_C1	PA_C2	PERF_ADV_C1
Molise	-0,029	1,915	1,761	-1,402	-1,289	-0,199	-0,095
Camp	0,678	0,343	2,835	-1,082	-3,504	0,730	-0,949
Puglia	0,634	2,076	1,190	-0,899	-0,282	-0,655	-1,522
Basilicata	-5,807	1,127	0,676	-1,286	-1,475	-1,279	-1,269
Calabria	-0,808	1,670	-0,757	-1,343	0,593	-0,886	0,882
Sicilia	0,589	1,770	2,204	-1,046	1,818	-0,481	-0,245
Sardegna	0,482	0,339	-0,285	-0,818	-0,613	0,744	0,450
Media	-0,411	1,349	0,875	-1,007	-0,407	-0,326	-0,25

Tab.59 Cluster 4- Fuzzy Silhouette

Regione	Silhouette
Molise	0,329
Campania	0,141
Puglia	0,431
Basilicata	0,331
Calabria	0,305
Sicilia	0,419
Sardegna	0,409

Fig.40 Cluster 4-Grafico fuzzy silhouette



4.8. Conclusioni

La crescente competizione tra destinazioni turistiche richiede una conoscenza approfondita delle determinanti che ne influenzano la competitività, in modo da poter agire rafforzando i punti di forza e riducendo i punti di debolezza, migliorando così l'attrattività del territorio.

Da Paese *leader* nel *ranking* del turismo, l'Italia oggi fatica a preservare il proprio livello competitivo sia verso i suoi concorrenti storici che rispetto alla competizione sempre più aggressiva di nuove destinazioni.

Analizzare i fattori che minano la capacità competitiva diventa quindi un esercizio fondamentale per il ricercatore ma anche per il policy-maker. Infatti, in questo modo esso dispone di un set di indicatori su cui poter sviluppare una strategia più coerente ed efficace.

L'evidente *gap* competitivo dell'Italia rispetto ai suoi concorrenti, tradizionali ed emergenti, ha motivato un'attenta analisi di come la minore competitività a livello nazionale si sia realmente manifestata a livello regionale.

Le potenzialità competitive e l'effettiva "turisticità" di ciascuna regione sono state valutate sulla base della dotazione di risorse territoriali e "antropiche" che generano valore per il territorio e sulla capacità di supportare tali risorse con un sistema ricettivo adeguato, un livello di infrastrutture che permetta l'accessibilità alle destinazioni, l'attenzione alla qualità e sostenibilità, la capacità di generare benefici per l'economia regionale, la capacità di attrazione dei flussi turistici domestici e internazionali, nonché la competitività delle imprese turistiche presenti sul territorio.

L'analisi si è basata sull'utilizzo di una combinazione di tecniche statistiche multivariate che hanno portato all'individuazione di 4 cluster in base ai quali sono state segmentate le venti regioni italiane.

I risultati ottenuti hanno evidenziato sostanziali differenze nel livello di competitività turistica delle regioni italiane, mostrando alcune eccellenze (Trentino Alto Adige) ma anche moltissime criticità proprio in quelle regioni che dovrebbero riuscire ad esprimere una maggiore competitività in virtù del patrimonio di risorse del territorio disponibili che potrebbero divenire prodotti turistici e quindi tradursi in domanda effettiva (ad esempio la Sardegna e la Puglia).

E' naturale che analisi di questo tipo conciliano il vantaggio della semplificazione delle molteplici informazioni da valutare con una inevitabile perdita del dettaglio informativo sulle singole regioni.

Dalle considerazioni di carattere generale è emerso che molte delle regioni più "attraenti", in termini di risorse, non riescono a valorizzare adeguatamente il loro potenziale turistico evidenziando performance negative con flussi turistici piuttosto esigui, che appaiono ancora più modesti se paragonati alle performance di regioni che hanno saputo valorizzare al massimo i loro limitati fattori di attrattiva presenti sul territorio.

Si è mostrato inoltre che oltre ai prodotti turistici tradizionali (mare, montagna, cultura) esiste un patrimonio antropico da valorizzare e sfruttare turisticamente che risponde alle nuove tendenze del turismo, offre l'opportunità per destagionalizzare i flussi e permette di conciliare sviluppo turistico e sostenibilità.

Si rileva indispensabile quindi agire sui punti di debolezza dei sistemi turistici regionali, primo fra tutti l'accessibilità e il sistema dei trasporti all'interno dei territori regionali. Molte regioni inoltre devono migliorare la qualità del sistema di offerta e promuovere maggiormente la sostenibilità.

Risulta necessario inoltre incrementare la comunicazione a livello di mercato d'origine dei flussi stranieri.

L'analisi ha mostrato come la competitività sia un concetto poliedrico influenzato da una molteplicità di fattori e non misurabile solo attraverso la quantificazione dei flussi turistici attratti e della capacità ricettiva. Moltissime

variabili concorrono a definire la competitività, alcune inserite nell'analisi, per altre invece, seppur considerate come elementi fondamentali come la formazione, non è stato possibile reperire i dati aggiornati dettagliati a livello regionale.

Il lavoro ha condotto un'analisi comparativa a livello regionale, poiché è in tale contesto che si determina e si sviluppa la programmazione strategica del turismo. E' importante però sottolineare come il sistema turistico italiano sia complesso e variegato per cui coesistono all'interno della stessa regione una moltitudine di prodotti e di esperienze turistiche e differenti sistemi di offerta. Potrebbe quindi essere interessante provare ad applicare la stessa strategia di analisi a livello di sub-sistemi regionali.

Tuttavia bisognerebbe poter disporre di informazioni esaustive e tempestive, ciò rappresenta il limite del sistema informativo turistico nazionale e regionale che dovrebbe essere ripensato e riorganizzato in modo da poter disporre di una visione integrata e aggiornata di tutte le fonti informative a tutti i livelli amministrativi.

Appendice

Implementazioni in R

Analisi in componenti principali in R

Analisi in componenti principali sulla matrice dei dati Territorio.

Territorio<-read.table("territorio.txt", head=T) :per leggere la matrice dei dati territorio.txt (in formato testo) e assegnare il nome Territorio all'oggetto matrice dei dati.

La funzione *read.table* permette di importare i file di testo con estensione .txt, *head=T* sta ad indicare che la prima riga del file contiene il nome delle variabili (altrimenti da porre uguale a False),

summary (Territorio) : per calcolare le statistiche descrittive della matrice Territorio

cor(Territorio): per calcolare la matrice di correlazione

ACP_TERR<-princomp(Territorio, cor=T, scores=T)

La funzione *princomp* per effettuare l'analisi in componenti principali della matrice Territorio, *cor=T* (true) sta ad indicare che l'analisi in componenti principali viene effettuata sulla matrice di correlazione, *scores=T* serve per calcolare la matrice dei punteggi delle componenti principali

L'output della funzione *princomp* comprende:

ACP_TERR\$scores : è la matrice dei punteggi

ACP_TERR\$loadings : è la matrice dei pesi o dei coefficienti delle componenti principali ossia gli autovettori

ACP_TERR\$sdev: è l'errore standard delle component principali, ovvero la radice quadrata degli autovalori

screeplot(ACP_TERR, type="lines",main="scree-plot") : comando per disegnare lo *screeplot* dove *main* è il nome del grafico

ACP_TERR\$sdev[1]*ACP_TERR\$loadings[,1] :per calcolare la correlazione tra la prima componente e le variabili

ACP_TERR\$sdev[2]*ACP_TERR\$loadings[,1] :per calcolare la correlazione tra la componente 2 e le variabili

biplot(ACP_TERR, pc.biplot=T) : comando per disegnare il *biplot* in cui viene visualizzata la correlazione di ciascuna delle variabili originali con le prime due componenti principali sotto forma di vettori geometrici

Cluster analysis gerarchica in R

dist(x, method = "euclidean", diag = FALSE, upper = FALSE) : la funzione *dist* si impiega per il calcolo della matrice delle distanze, *x* rappresenta la matrice dei dati, *method* indica la misura di distanza (*euclidean, manhattan, canberra, binary*), *diag* e *upper* si pongono come *FALSE* se non si vuole che la matrice di distanze contenga rispettivamente i valori nulli sulla diagonale e il triangolo superiore.

hclust(d, method = "complete")

La funzione *hclust* si impiega per effettuare la cluster gerarchica, *d* è una matrice di distanze e *method* indica il metodo gerarchico aggregativo che si vuole utilizzare (*complete, single, average, centroid, ward*).

La funzione *hclust* produce come output gli elementi:

\$merge: sequenza del processo di agglomerazione;

\$height: vettore che indica la distanza alla quale é avvenuta l'unione tra due cluster.

\$order: permutazione delle unita analizzata alla costruzione del dendrogramma;

\$labels: vettore delle etichette delle unita.

plot(x, hang = 0.1, main = "Cluster Dendrogram", sub = "Metodo del legame completo", xlab = null) : è il comando per costruire il dendrogramma, *x* è il risultato della funzione *hclust*, *hang* determina l'altezza a cui vengono poste l'etichette al di sotto del dendrogramma, *main*, *sub* e *xlab* sono istruzioni per la finestra grafica.

abline(h=, lty=2): dopo il comando plot per il taglio del dendrogramma in corrispondenza di un salto nelle distanze (*h* indica il livello di distanza dove inserire il taglio, *lty* definisce lo stile della retta del taglio).

cutree (x, k=, h=): per ricavare i cluster in corrispondenza di un taglio del dendrogramma o un numero prefissato di cluster, *x* indica l'output della funzione *hclust*, *k* è il numero di cluster prefissato, *h* è l'altezza alla quale viene tagliato il dendrogramma).

Cluster analysis non gerarchica con il metodo delle k-medie in R

kmeans(x, centers, iter.max=10) è la funzione per la cluster analysis con il metodo delle *k-medie*, *x* è la matrice dei dati, *center* è il numero prefissato di

cluster o un insieme di centroidi iniziali, *iter.max* è il massimo numero di iterazioni).

I risultati della funzione comprendono

\$cluster: vettore di allocazione delle unità

\$center: matrice dei centroidi

\$whitnss: devianza complessiva entro i gruppi

\$size: dimensione dei gruppi

Cluster fuzzy K-means in R

cmeans(x,center,iter.max=100, dist="euclidean", method="cmeans", m=2): è la funzione per effettuare la *cluster fuzzy k-means* in R, *x* è la matrice dei dati, *center* è il numero di cluster da fissare a priori, *iter.max* è il massimo numero di iterazioni, *dist* è il tipo di distanza utilizzata, *method = cmeans* per applicare l'algoritmo *fuzzy k-means*, *m* è il parametro per la scelta del livello di *fuzziness* della partizione.

L'output della funzione *cmeans* comprende:

\$centers: I centroidi finali dei cluster.

\$Size: Il numero di unità per ogni cluster.

\$Cluster: L'assegnazione delle unità ai cluster sfocati attraverso il massimo valore del grado di appartenenza

\$Iter: Numero di iterazioni

\$membership: Matrice dei gradi di appartenenza

BIBLIOGRAFIA

Ambrosetti - The European House (2008), *Sistema Turismo Italia, Proposte per essere vincenti, 2°Edizione.*

Anderberg M. R. (1973), *Cluster analysis for applications*, New York, Academic Press.

Anderson, D.T., Bezdek, J.C., Popescu, M., Keller, J.M., (2010), “*Comparing Fuzzy, Probabilistic, and Possibilistic Partitions*”, IEEE Transaction on Fuzzy Systems, 18, 906-918.

Andriola L., Manente M., (2005), “*Turismo durevole e sviluppo sostenibile: il quadro di riferimento italiano*”, ENEA, RT/AMB/2000/5.

Arabie P., Hubert L.J., De Soete G., (1998), *Classification and clustering*, Journal of classification, 15, 151-153.

Bezdek J. C. ,(1981), *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York.

Boggiani R., *Introduzione a R*, da www.statistica.unipd.it

Buhalis D. (2000). “*Marketing the competitive destination of the future*”, Tourism Management, 21(1), 97-116.

Campello, R.J.G.B. ,(2007), “*A Fuzzy Extension of the Rand Index and Other Related Indexes for Clustering and Classification Assessment*”, Pattern Recognition Letters, 28, 833 – 841.

Cantone L., Risitano M., et al., (2007) “*Strategie di sviluppo delle destinazioni turistiche e ruolo della marca territoriale*”, Mercati e competitività, 1(1), 21-47.

Cerbara L., Iacovacci G., (1998), *Tecniche sfocate per la classificazione di dati di popolazione*, Working paper, 2, Irp Cnr, Roma.

Ciccarelli A., (2003), “*Una metodologia statistica per l’analisi di competitività delle province*”, da www.tagliacarne.it/download

Coccorese P., Pellecchia A., (2005), “*Verso i sistemi turistici locali: un’analisi cluster*”, Studi e Note di Economia 3/2005.

Commissione Europea, “*Agenda for a sustainable and competitive European tourism*”, Comunicazione, COM(2007) 621 def.

Commissione Europea, “*L’Europa, prima destinazione turistica mondiale – un nuovo quadro politico per il turismo europeo*”, Comunicazione, COM(2010) 352 def.

Confturismo, (2007), “*La spesa delle regioni per il turismo*”.

Costa N., (2008) “*Il sistema Italia nel contesto internazionale*”, in E.B.I.T. (a cura di), Turismo. Prospettive & Governance.

Cracolici M., F., Nijkamp P., (2008) “*The attractiveness and competitiveness of tourist destinations: A study of Southern Italian regions*”, Tourism Management, 30, 336-344.

Crouch, G.I. e Ritchie, J.R.B.(1999) “*Tourism, competitiveness and social prosperity*”. Journal of Business Research, 44(3),137-152.

De Caprariis G., Rapacciuolo C., Terzulli A., (2006) *Un’analisi dei problemi strutturali del settore turistico italiano*, da www.econ-pol.unisi.it/opts/Quaderni/N6-decapraris-rapacciuolo-terzulli.pdf.

Dwyer L., Kim C. (2003), *Destination Competitiveness: Determinants and Indicators*, Current Issues in Tourism, 6(5), 369-414.

Dwyer L., Livaic Z.,Mellor R., (2003), “*Competitiveness of Australia as a Tourist Destination*”, Journal of Hospitality and Tourism Management, 10 (1), 60-78.

Ejarque J.,(2003) *La destinazione turistica di successo*, Milano, Hoepli

Enright M.J., Newton J., (2004), *“Tourism destination competitiveness: a quantitative approach”*. Tourism Management, 25, 2004, 777-788.

Everitt B. S. (1980), *Cluster analysis*, London, Heineman Educational Books.

Fabbris L. (1983), *Analisi esplorativa di dati multidimensionali*, Cleup Editore

Gismondi R., (2001), *“Le performances del turismo nelle Regioni e nelle province”*, in X Rapporto sul turismo Italiano, Mercury, 101-145.

Gismondi R., Russo M.A., (2004), *“Definizione e calcolo di un indice territoriale di turisticità: un approccio statistico multivariato”*, STATISTICA, anno LXIV, n. 3, 545-571.

Heat E., (2003) *“Towards a Model to enhance destination competitiveness: a southern African perspective”*, Journal of Hospitality and Tourism Management, 10(2), 124-141

IRPET (a cura di), (2009), *Turismo e Toscana - La congiuntura 2008*, REGIONE TOSCANA, Settore Sistema Statistico Regionale.

ISAE, Istituto di Studi e Analisi Economica, Quaderni di discussione, (2010), *Modelli comportamentali delle imprese italiane: l'applicazione della cluster analysis ai dati delle indagini ISAE*.

ISNART - UNIONCAMERE (2010), *“Competitività del sistema turistico Italiano”*.

ISNART-UNIONCAMERE (2009), *“Impresa Turismo 2009”*

ISPRA, *Annuario dei dati Ambientali*, Cap. 4 *“TURISMO”*, 277-306

ISPRA, *Guida Ecolabel 2008, Prodotti e servizi con il marchio Ecolabel europeo in Italia*.

ISTAT, *“Capacità e movimento turistico”*, anni 2003, 2007, 2008.

Kozak M., Rimmington M., (1999), “*Measuring tourist destination competitiveness: conceptual considerations and empirical findings*”. *Hospitality Management*, 18(3), 273–283.

Kozak, M. (2002). “*Destination benchmarking*”, *Annals of Tourism Research*, 29(2), 497-520

La Rocca A., *Fuzzy clustering: la logica, i metodi*. ISTAT – Direzione centrale per le indagini su condizioni e qualità della vita.

Leiper N. (1979), The framework of tourism, *Annals of Tourism Research*, 6,

Liguori M., (2008), “*La competitività dell’Italia e delle Regioni Italiane*”, in E.B.I.T. (a cura di), *Turismo. Prospettive & Governance*.

Massari R., *Analisi dei dati con R (dispense di Riccardo Massari)*.

Melian-Gonzalez A., Garcia-Falcon J.M., (2003), “*Competitive potential of tourism in destinations*”, *Annals of Tourism Research*, 30 (3), 720–740.

Mercury, (2000), *IX Rapporto sul turismo Italiano*.

Mercury, *Rapporto sul turismo italiano. Pensare turisticamente*, XVI edizione, 2008/2009, Franco Angeli.

Mihalic, T., (2000), “*Environmental management of a tourist destination. A factor of tourism competitiveness*”, *Tourism Management*, 21, 65-78

Milioli M.A. (1994), “*Confronto fra partizioni sfocate nell’analisi di dati territoriali*”

Minguzzi A., Presenza A., (2010), *Destination Building, Teorie e pratiche per il management della destinazione turistica*, Pearson.

Minguzzi A., Presenza A., Leonelli D., (2011), *Lo studio della competitività turistica delle regioni italiane attraverso l’analisi multivariata*, in *Rapporto sul Turismo Italiano*, XVII edizione, 2010-2011, Franco Angeli.

Murphy P., Pritchard M.P., Smith B., (2000), *“The destination product and its impacts on traveller perceptions”*, Tourism Management 21, 43-52.

Osservatorio del Turismo IULM (2005), *Competitività della città di Milano nel settore turistico*, Rapporto di ricerca.

Osservatorio del Turismo IULM (2006), *Competitività della destinazione Milano: indicatori per un confronto internazionale*, Rapporto di ricerca, Fondazione IULM.

Osservatorio del Turismo IULM (2008), *Osservatorio sulla competitività di Milano nel settore turistico: le performance del comparto alberghiero*, Rapporto di ricerca, Fondazione IULM.

Presenza A., (2007), *“Destination Management Organisation. Ruolo, organizzazione e indicatori di performance”*, Franco Angeli.

Ritchie J.R.B., Crouch G.I., (2000), *“The competitive destination: a sustainable perspective”*, Tourism Management, 21 (1), 1-7.

Rosignoli S., (2009) *“Incidenza del turismo sull'economia regionale: stima del contributo effettivo ed esercizio di valutazione del potenziale”*, IRPET.

Sainaghi R., Canali S. (2008), *Posizionamento competitivo delle urban destination e performance delle imprese alberghiere: il caso Milano*, working paper, ricerca MIUR.

Sainaghi R., (2008) *“Attrattori e prodotti turistici: il binomio della competitività”*, *Economia e diritto del Terziario*, 1, 118-139

Scoltock J., (1982), *“A survey of the literature of Cluster analysis”*, The computer journal, 25 (1), 130-134.

SRM, Associazione Studi e Ricerche per il Mezzogiorno, (2009) *Turismo & Mezzogiorno. Caratteri strutturali, potenzialità e dinamiche competitive dei contesti turistici meridionali*, Giannini Editore.

Touring Club Italiano, *“L'Annuario del turismo e della cultura”*, 2006

Touring Club Italiano, *“Tourism Monitor 2011”*.

Touring Club Italiano, *I Libri Bianchi del Touring Club Italiano*, (2005), *Sviluppo Sostenibile e competitività del settore turistico*.

UNWTO, (2008), *“Tourism Highlights”*.

UNWTO, (2009), *“Tourism Highlights”*.

UNWTO, *“World Tourism Barometer”*, 7(1), January 2009.

UNWTO, *“World Tourism Barometer”*, 7(1), January 2010.

UNWTO, *“World Tourism Barometer”*, 7(2), June 2009.

UNWTO, *“World Tourism Barometer”*, 8(2), June 2010.

World Economic Forum, (2008), *“The Travel & Tourism Competitiveness Report 2008”*.

Zadeh L. A. ,(1965), *“Fuzzy sets”*, Inf. Control, 8, 338-353.

Zadeh L.A., (1977), *“Fuzzy set and their application to pattern classification and clustering”*, in J. Van Ryzin, *Classification and Clustering*, , Accademic Press, New York, 251-299.

Zani S., (2000), *Analisi dei dati statistici II, Osservazioni multidimensionali*, Giuffrè Editore.

Siti Internet Consultati

- ✓ www.abruzzoturismo.it
- ✓ www.accredia.it
- ✓ www.ao.camcom.it/studi.aspx
- ✓ www.apat.it
- ✓ www.aptbasilicata.it
- ✓ www.assaeroporti.it
- ✓ www.cnel.it (*cnelstats*)
- ✓ www.ebtl.it
- ✓ www.enit.it
- ✓ www.eurostat.it
- ✓ www.federturismo.it
- ✓ www.incomingcampania.it
- ✓ www.infocamere.it
- ✓ www.irpet.it
- ✓ www.isnart.it
- ✓ www.ispra.it
- ✓ www.issirfa.it
- ✓ www.istat.it
- ✓ www.miur.it
- ✓ www.noi-italia.istat.it
- ✓ www.osservatorioturisticocalabria.it
- ✓ www.politicheagricole.it/ucea
- ✓ www.portal.euromonitor.com
- ✓ www.r-project.org
- ✓ www.regione.marche.it
- ✓ www.regione.molise.it/turismo/

- ✓ www.regione.sicilia.it
- ✓ www.regioneliguria.it
- ✓ www.regioneapiemonte.it/turismo/osservatorio/
- ✓ www.regioneumbria.eu
- ✓ www.salute.gov.it
- ✓ www.sardegna statistiche.it
- ✓ www.sardegna turismo.it
- ✓ www.sincert.it
- ✓ www.sis-statistica.it/magazine/
- ✓ www.sistan.beniculturali.it
- ✓ www.statistica.regione.lombardia.it
- ✓ www.statistica.regione.veneto.it
- ✓ www.tagliacarne.it
- ✓ www.terna.it
- ✓ www.touringclub.it
- ✓ www.turiscalabria.it
- ✓ www.turismofvg.it
- ✓ www.turismo.intoscana.it
- ✓ www.turismo.marche.it
- ✓ www.uif.bancaditalia.it
- ✓ www.unioncamere.it
- ✓ www.unwto
- ✓ www.venus.unive.it/ciset/
- ✓ www.viaggiareinpuglia.it
- ✓ www.weforum.org
- ✓ www.wttc.org