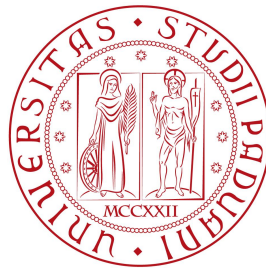


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze statistiche



L'INFLUENZA DELL'ENOGASTRONOMIA NEL TURISMO

Indagine su un campione di viaggiatori Italiani

Relatore Prof. Francesca Bassi
Dipartimento di Scienze Statistiche

Laureanda: Sofia Carpanese
N.matricola: 2123733

Anno Accademico 2025/2026

Indice

Introduzione	1
1 Turismo Enogastronomico	3
1.1 La gastronomia nella scelta della destinazione	3
1.2 Questionario sul turismo gastronomico	5
1.2.1 Analisi del questionario	8
1.3 Obiettivo di ricerca	9
2 Metodologia	11
2.1 Analisi fattoriale	11
2.1.1 Formalizzazione matematica del modello fattoriale	12
2.1.2 Scomposizione della varianza e matrice di covarianza	13
2.1.3 Le fasi procedurali dell'analisi fattoriale	14
2.2 Cluster Analysis Gerarchica	17
2.2.1 Single Linkage (Legame Singolo)	17
2.2.2 Complete Linkage (Legame Completo)	18
2.2.3 Average Linkage	18
2.2.4 Metodo di Ward	18
2.3 Latent Class Analysis	19
2.3.1 Struttura del Modello	19
2.3.2 Come decidere il numero di classi	21
2.3.3 Obiettivi e Segmentazione	22
2.3.4 Il Modello di Regressione LC	23
3 Analisi dei dati	25
3.1 Indagine Gastronomica in Italia	25
3.1.1 Pulizia del dataset	26
3.1.2 Analisi Esplorative	28
3.2 Analisi della Struttura Latente e Dimensionalità	35
3.2.1 Procedura Metodologica e Verifica della Fattoriz- zabilità	35

3.2.2	Analisi della Dimensionalità (Parallel Analysis e Omega)	36
3.3	Cluster Analysis e Segmentazione	39
3.3.1	Cluster Analysis Gerarchica	39
3.3.2	Latent Class Analysis	47
3.3.3	Confronto tra Clustering Gerarchico e LCA	52
3.3.4	Modello di Regressione LC	54
	Conclusioni	58

Introduzione

La gastronomia è un fattore che influisce sulla decisione della meta turistica?

In questo elaborato si intende analizzare se i viaggiatori selezionino la propria meta di viaggio considerando anche la cultura gastronomica locale, al fine di comprendere se un investimento nell'enogastronomia possa favorire un incremento del turismo all'interno del territorio italiano.

A tal fine è stato sviluppato un questionario composto da 51 domande, di cui 45 basate su scale Likert con punteggi da 1 a 7, dove 1 indicava “per nulla interessato” e 7 “pienamente interessato” all’aspetto considerato. Queste domande avevano l’obiettivo di valutare quali elementi legati alla gastronomia fossero di maggiore interesse per i turisti italiani. Le restanti domande erano invece finalizzate alla descrizione del profilo sociodemografico dei rispondenti. Il questionario è stato somministrato a un campione di cittadini italiani, ottenendo un totale di 1.208 osservazioni.

Per analizzare in modo appropriato i dati raccolti, si è inizialmente proceduto con una fase di pulizia del dataset, necessaria per garantire l’affidabilità delle analisi. Successivamente è stata condotta un’analisi esplorativa dei dati, al fine di ottenere una visione generale della distribuzione delle risposte. In seguito, per verificare se il questionario misurasse un unico costrutto latente e per valutare la possibilità di ridurre la dimensionalità delle variabili, è stata applicata un’Analisi Fattoriale, utilizzando eventualmente i punteggi fattoriali nelle analisi successive.

Una volta completate le fasi preliminari di pulizia e riduzione del dataset, sono state applicate tecniche di cluster analysis gerarchica, basate sulle distanze, e tecniche di latent class analysis, basate sulle probabilità, con l’obiettivo di individuare eventuali gruppi omogenei di rispondenti in base alle valutazioni fornite nelle domande a scala Likert. Dopo l’iden-

tificazione dei gruppi, si è proceduto all'analisi demografica dei profili dei turisti appartenenti a ciascun cluster. Tale analisi è stata condotta utilizzando le variabili derivanti dalle domande demografiche rivolte ai rispondenti.

La tesi è strutturata in tre macro-capitoli. Il primo capitolo introduce il concetto di turismo enogastronomico, analizza il questionario proposto e discute eventuali miglioramenti applicabili per ottimizzare le analisi. Il secondo capitolo è dedicato alla descrizione delle metodologie di analisi utilizzate. Il terzo capitolo presenta le analisi empiriche applicate ai dati, illustrando i risultati ottenuti e consentendo un confronto tra i metodi impiegati.

Capitolo 1

Turismo Enogastronomico

1.1 La gastronomia nella scelta della destinazione

In uno scenario sempre più globalizzato e caratterizzato da una rapida circolazione delle conoscenze, il patrimonio paesaggistico e storico-artistico non costituisce più l'unico fattore di attrazione verso una determinata destinazione. Attualmente, infatti, anche la scoperta delle eccellenze agroalimentari e delle tradizioni culinarie locali riveste un ruolo cruciale nel processo decisionale che porta a intraprendere un viaggio. In tale prospettiva, diviene fondamentale indagare quali siano le determinanti che orientano la scelta della meta turistica e come sia possibile segmentare efficacemente la domanda.

Numerosi studi confermano che l'enogastronomia non si limita a favorire la crescita del mercato turistico, ma agisce come leva strategica per lo sviluppo economico e culturale del territorio ospitante.

A tal proposito, il "*Rapporto sul Turismo Enogastronomico Italiano*" [Garibaldi, 2020] evidenzia come questo settore abbia subito un'evoluzione profonda: l'esperienza non si esaurisce più nel semplice acquisto di prodotti locali o nel consumo di piatti tipici, ma abbraccia una dimensione esperienziale a 360 gradi. Il turista contemporaneo cerca un contatto diretto con i luoghi di produzione (quali cantine, frantoi e caseifici) frequenta ristoranti storici o gourmet, sperimenta il cibo di strada e partecipa attivamente a cooking class o food tour. Si delinea così una pluralità di esperienze in cui il coinvolgimento multisensoriale diventa la chiave per fruire del patrimonio culturale locale in modo attivo e partecipativo.

Questa simbiosi tra territorio e prodotto è confermata da "*Il turismo enogastronomico*" [Perri, 2018], le quali sottolineano come il legame tra enogastronomia e turismo sia in grado di ampliare significativamente la soddisfazione del visitatore. Di conseguenza, le strategie di marketing non possono limitarsi alla promozione del solo pacchetto di viaggio, ma devono valorizzare l'intero sistema territoriale. Poiché il turista enogastronomico si muove all'interno del perimetro che unisce il "cru" (il prodotto d'eccellenza legato a un micro-territorio specifico) all'area geografica di appartenenza, l'intera esperienza turistica si trasforma in un percorso multisensoriale e identitario.

La letteratura di settore evidenzia come la gastronomia locale rappresenti una delle variabili più influenti nel processo decisionale del turista [Berbel-Pineda et al., 2019]. Da qui emerge la necessità di approfondire i fattori che condizionano tale scelta e di comprendere le sfumature comportamentali dei viaggiatori in relazione alle loro preferenze alimentari. Diversi contributi accademici si sono focalizzati su questo obiettivo, analizzando prodotti tipici, piatti della tradizione e offerta vinicola come componenti imprescindibili dell'esperienza turistica complessiva. Sulla base di queste premesse, la gastronomia locale assume una duplice valenza: motore di crescita per il comparto turistico e strumento di sviluppo territoriale in senso ampio. In quest'ottica, è possibile classificare i comportamenti dei turisti in quattro macro-categorie principali, come teorizzato da Chen e Huang [Chen and Huang, 2016]:

1. il cibo come espressione della cultura locale;
2. il cibo come attrazione turistica;
3. il cibo come prodotto turistico;
4. il cibo come esperienza turistica.

Muovendo da tali considerazioni, è stato sviluppato un questionario volto ad analizzare l'incidenza della componente gastronomica sulle dinamiche del turismo italiano. L'obiettivo è l'identificazione di distinti segmenti di visitatori, al fine di delineare scelte strategiche efficaci per la valorizzazione del territorio e dell'offerta turistica.

1.2 Questionario sul turismo gastronomico

Al fine di esaminare il peso esercitato dalla componente gastronomica sulle motivazioni di viaggio, è stato predisposto uno specifico questionario. Lo strumento di indagine comprende cinque gruppi di domande valutate tramite scala Likert a 7 punti (con estremi varianti da 1, “per nulla d’accordo”, a 7, “pienamente d’accordo”), volte a misurare il grado di rilevanza attribuito dai soggetti ai singoli aspetti, e sei quesiti a risposta multipla finalizzati a delineare il profilo socio-demografico del rispondente.

Nello specifico, il questionario è strutturato come segue:

GASTRONOMIA E TURISMO

L’Università di Padova, in collaborazione con l’Università di Granada, sta conducendo una ricerca per conoscere la percezione e l’opinione dei visitatori di destinazioni turistiche in relazione alla gastronomia.

Si garantisce che la partecipazione allo studio è volontaria e gratuita, che i partecipanti possono ritirarsi dallo studio in qualsiasi momento senza alcun danno e che i soggetti dispongono di informazioni sufficienti per dare il proprio consenso alla partecipazione.

1. Qual è il motivo principale per cui viaggia?

- Svago o riposo
- Convegni
- Gastronomia e vini
- Cultura
- Studio
- Visita a familiari o amici
- Religioso
- Lavoro
- Non viaggio – fine intervista

2. Per favore, valuti da 1 a 7 (1: molto poco; 7: molto) i seguenti aspetti relativi alla gastronomia nei suoi viaggi:

	1	2	3	4	5	6	7
Importanza gastronomia nella motivazione	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ricerca di esperienze nella scelta meta	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Esperienza gastronomica e soddisfazione	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cultura e patrimonio nella scelta	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Conoscere piatti o ricette famosi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Presenza di ristoranti famosi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. Per favore, valuti da 1 a 7 la Sua intenzione ad andare in determinati luoghi legati alla gastronomia:

	1	2	3	4	5	6	7
Ristoranti	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Bar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Produttori agricoli	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Agroindustrie (caseifici, cantine, ecc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Festival/Eventi gastronomici	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mercati gastronomici	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Musei gastronomici	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Negozi di alimentari tradizionali	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. Valuti l'importanza dei seguenti aspetti (1: molto basso; 7: eccellente):

	1	2	3	4	5	6	7
Qualità dei piatti	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Varietà dei piatti	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Prezzi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Strutture ristorative	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ambiente degli esercizi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Innovazione e nuovi sapori	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Servizio e ospitalità	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gastronomia tradizionale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5. Valuti le Sue percezioni (1: fortemente in disaccordo; 7: fortemente d'accordo):

	1	2	3	4	5	6	7
Accresce conoscenza culinaria	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Scoperta altri elementi culturali	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Conoscere il vero sapore tipico	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Scoprire cose nuove	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Esperienze gastronomiche uniche	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cucina nel luogo di origine	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Esperienze che rilassano	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Godersi l'ambiente con calma	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dimenticare la routine quotidiana	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ingredienti freschi locali	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
La gastronomia locale è sana	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
La gastronomia locale è nutriente	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
La gastronomia km0 è sostenibile	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Momenti piacevoli con famiglia/amici	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rafforzare relazioni sociali	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Parlare con altri dell'esperienza	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consigliare la meta ad altri	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

6. Quali emozioni prova? (1: totalmente in disaccordo; 7: totalmente d'accordo):

	1	2	3	4	5	6	7
Allegria	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sorpresa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ammirazione	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fiducia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ottimismo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delusione	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

7. Genere:

- Maschio
- Femmina
- Non binario

8. Età: _____ (in anni compiuti)

9. Livello di istruzione:

- Licenza elementare
- Licenza media
- Diploma
- Laurea (triennale e/o magistrale)
- Titolo post-laurea

10. Condizione professionale:

- Dipendente pubblico
- Dipendente azienda privata
- Lavoratore autonomo
- Studente
- Disoccupato
- Pensionato
- Casalinga

11. Reddito familiare disponibile (euro netti al mese):

- Meno di 1.000€
- Da 1.001 a 1.500€
- Da 1.501 a 2.500€
- Da 2.501 a 3.500€
- Più di 3.500€
- Preferisco non rispondere

1.2.1 Analisi del questionario

Dall'analisi critica dello strumento di indagine emergono alcuni limiti metodologici che meritano di essere discussi. In primo luogo, l'adozione di una scala Likert a 7 punti [Revilla et al., 2014], pur garantendo una maggiore sensibilità statistica, può risultare dispersiva: la difficoltà del rispondente nel distinguere tra gradi intermedi di accordo favorisce spesso una polarizzazione dei voti sui valori più alti, saturando la variabile e riducendo il realismo delle risposte.

In secondo luogo, l'elevato numero di item nella quinta batteria di domande può innescare il fenomeno della *survey fatigue*. La stanchezza cognitiva induce il soggetto al cosiddetto *straight-lining*, ovvero la tendenza a replicare meccanicamente lo stesso punteggio per velocizzare il completamento del compito, [Krosnick, 1991]. Inoltre, la formulazione degli item, orientata esclusivamente verso accezioni positive della

gastronomia (senza l’inserimento di *reversed items*), [Podsakoff et al., 2003], non permette di bilanciare il *social desirability bias*, rendendo difficile intercettare eventuali risposte casuali o poco attente.

In terzo luogo, analizzando l’ultima domanda presente nel questionario, relativa al reddito mensile percepito dai rispondenti, emerge un limite nella struttura delle classi di risposta. La variabile è infatti costruita attraverso una combinazione di fasce di reddito chiuse e categorie aperte (“Meno di 1.000€” e “Più di 3.500€”), che rende meno omogeneo il confronto tra i gruppi, [Moore et al., 2000]. Questa impostazione può aver ridotto la precisione informativa della variabile, limitando la possibilità di cogliere differenze reddituali più puntuali all’interno del campione e, di conseguenza, la sua efficacia in analisi di segmentazione.

Infine, sotto il profilo socio-demografico, lo strumento appare limitato per una segmentazione esaustiva della domanda, [Dolnicar, 2008]. Sarebbe stato opportuno integrare variabili cruciali quali la regione di residenza, per valutare l’influenza della provenienza geografica, e la tipologia di *travel party* (viaggio in famiglia, coppia o amici). Analogamente, l’inserimento della frequenza di viaggio e della composizione del nucleo familiare avrebbe permesso di distinguere i viaggiatori abituali e di comprendere come le necessità logistiche e familiari condizionino, di fatto, le scelte e le abitudini gastronomiche.

1.3 Obiettivo di ricerca

Il campione analizzato comprende 1.208 rispondenti maggiorenni di nazionalità italiana. Data l’ampiezza del dataset, l’obiettivo primario della segmentazione è mappare il diverso grado di coinvolgimento del turista nei confronti dell’offerta gastronomica territoriale. Lo studio mira a discriminare tra profili caratterizzati da un interesse marginale o meramente strumentale (viaggiatori meno interessati) e profili per i quali la gastronomia rappresenta il *driver* principale nella scelta della destinazione e nella fruizione dell’esperienza di viaggio (viaggiatori alto-coinvolti).

Per l’identificazione di tali gruppi sono stati adottati due differenti approcci metodologici, al fine di garantire la massima robustezza dei risultati:

1. **Cluster Analysis Gerarchica:** un metodo basato sulle distanze geometriche (distanza Euclidea) e sul criterio di aggregazione di Ward, utilizzato per una prima esplorazione della struttura dei dati e per la determinazione del numero potenziale di gruppi;
2. **Latent Class Analysis (LCA):** un approccio probabilistico basato su modelli (*model-based*) che permette di identificare classi latenti non osservabili direttamente, partendo dalle probabilità di risposta ai singoli item del questionario.

L'integrazione di questi due metodi consente di convalidare la stabilità dei segmenti individuati e di fornire una profilazione dettagliata delle tipologie di viaggiatore presenti nel mercato italiano.

Capitolo 2

Metodologia

2.1 Analisi fattoriale

L'Analisi Fattoriale Esplorativa (EFA) è una tecnica statistica multivariata impiegata per indagare la struttura latente di un set di dati. L'obiettivo principale di questo metodo risiede nella **riduzione della dimensionalità**, [Molino, 2011]: attraverso l'estrazione dei fattori, è possibile sintetizzare l'informazione derivante da numerose variabili osservate in un numero limitato di costrutti ipotetici. Ciò permette di sostituire le variabili originali con i fattori individuati, ottenendo un modello più parsimonioso ed efficiente, pur preservando la quota più significativa della varianza totale.

L'analisi fattoriale risulta particolarmente efficace nello studio di fenomeni **non direttamente osservabili** (variabili latenti), quali, a titolo esemplificativo, il grado di soddisfazione o le attitudini psicometriche, [Molino, 2011]. A differenza dei modelli di natura predittiva, l'EFA si configura come un metodo volto all'individuazione delle dimensioni sottostanti che spiegano le interrelazioni tra le variabili manifeste.

Il processo analitico ha inizio con l'esame della *matrice di correlazione*, [Alaimo, 2020], che quantifica il grado di associazione tra i vari item del test. Successivamente, mediante l'estrazione e la rotazione dei fattori, si ottiene la matrice delle saturazioni (*factor loadings*). Tale output rivela la forza del legame tra ogni singola variabile osservata e il relativo fattore latente, indicando in che misura ogni quesito sia effettivamente in grado di rappresentare il costrutto teorico oggetto di indagine.

2.1.1 Formalizzazione matematica del modello fattoriale

L'analisi fattoriale si basa sull'assunto che un vettore di p variabili osservate \mathbf{X} sia il risultato di una combinazione lineare di un numero ridotto m di fattori latenti comuni \mathbf{F} (con $m \ll p$) e una componente di errore specifica \mathbf{U} , [Alaimo, 2020]. A differenza della regressione lineare classica, i predittori f_m non sono direttamente osservabili. Pertanto, il modello viene formalizzato definendo una struttura di dipendenza latente che lega le variabili osservate ai costrutti non osservati.

L'equazione del modello

Il legame tra le variabili manifeste e i fattori comuni può essere espresso, per ogni variabile j [Alaimo, 2020], come:

$$X_j = \mu_j + \sum_{k=1}^m \lambda_{jk} f_k + u_j \quad \text{con } j = 1, \dots, p \quad (2.1)$$

dove:

- μ_j : l'intercetta del modello;
- λ_j : influenza delle variabili esplicative;
- f_m : variabili latenti;
- u_j : errore residuo.

In forma matriciale contratta, il modello si definisce come:

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F} + \mathbf{U} \quad (2.2)$$

dove $\boldsymbol{\Lambda}$ rappresenta la matrice dei *factor loadings*, la quale misura l'intensità della relazione tra i fattori e le variabili osservate. Elevati valori di covarianza tra due variabili X_j e X_l suggeriscono che esse condividano una forte influenza da parte del medesimo fattore latente.

Assunzioni e ipotesi statistiche

Affinché il modello sia identificabile e stimabile, si introducono le seguenti assunzioni fondamentali sulla natura dei fattori e dei residui, [Alaimo, 2020]:

- **Centralità e Normalità:** I fattori comuni \mathbf{F} e i fattori specifici \mathbf{U} sono variabili casuali multinormali con media nulla: $E[\mathbf{F}] = \mathbf{0}$ e $E[\mathbf{U}] = \mathbf{0}$. Di conseguenza, $E[\mathbf{X}] = \boldsymbol{\mu}$.
- **Standardizzazione dei fattori:** Si assume che i fattori latenti siano standardizzati e indipendenti, ovvero la loro matrice di varianza-covarianza è l'identità: $Var(\mathbf{F}) = \mathbf{I}_m$.
- **Indipendenza dei residui:** I fattori specifici u_j sono tra loro incorrelati e indipendenti dai fattori comuni: $Cov(\mathbf{F}, \mathbf{U}) = \mathbf{0}$ e $Var(\mathbf{U}) = \boldsymbol{\Phi}$, dove $\boldsymbol{\Phi}$ è una matrice diagonale.

Queste ipotesi implicano che tutta la struttura di correlazione tra le variabili osservate venga spiegata solo dai fattori comuni. Poiché sia \mathbf{F} che \mathbf{U} sono ignoti, il modello non è stimabile tramite il metodo dei minimi quadrati, pertanto si procede con l'utilizzo di tecniche specifiche quali la fattorizzazione all'asse principale o la Massima Verosimiglianza.

2.1.2 Scomposizione della varianza e matrice di covarianza

Dalle assunzioni del modello derivano importanti implicazioni sulla struttura dei dati. In particolare, la varianza totale di ogni variabile osservata X_j può essere scomposta in due componenti distinte, [Alaimo, 2020]:

$$Var(X_j) = \sigma_j^2 = \underbrace{\sum_{m=1}^M \lambda_{jm}^2}_{c_j} + \phi_j^2 \quad (2.3)$$

dove si distinguono:

- **Comunalità** (c_j): rappresenta la quota di varianza della variabile X_j spiegata dai fattori comuni f_m .
- **Varianza specifica** (ϕ_j^2): costituisce la parte residua della varianza, ovvero la parte di variabilità non spiegata dai fattori comuni f_m .

Un aspetto cruciale dell'analisi fattoriale riguarda la covarianza tra due variabili distinte X_j e X_s . Sotto l'ipotesi di indipendenza dei residui, la

relazione si riduce a:

$$\begin{aligned}
\text{Cov}(X_j, X_s) &= \text{Cov}\left(\sum_{m=1}^M \lambda_{jm} f_m + u_j, \sum_{m=1}^M \lambda_{sm} f_m + u_s\right) \\
&= \sum_{m_1=1}^M \sum_{m_2=1}^M \lambda_{jm_1} \lambda_{sm_2} \text{Cov}(f_{m_1}, f_{m_2}) + \sum_{m=1}^M \lambda_{jm} \text{Cov}(f_m, u_s) \\
&\quad + \sum_{m=1}^M \lambda_{sm} \text{Cov}(f_m, u_j) + \text{Cov}(u_j, u_s) \\
&= \sum_{m=1}^M \lambda_{jm} \lambda_{sm} \text{Var}(f_m) = \sum_{m=1}^M \lambda_{jm} \lambda_{sm}
\end{aligned} \tag{2.4}$$

Questo dimostra che la correlazione tra le variabili osservate è interamente generata dalla loro dipendenza dai medesimi fattori comuni. In termini matriciali, la matrice di varianza-covarianza delle variabili osservate, indicata con Σ , può essere espressa come:

$$\Sigma = \Lambda \Lambda' + \Phi \tag{2.5}$$

Tuttavia, la matrice Λ non è univocamente determinata. Data una matrice ortogonale Q tale che $QQ' = I$, è possibile definire una nuova matrice $\Lambda^* = \Lambda Q$ per cui la relazione fondamentale rimane invariata:

$$\Lambda^* \Lambda^{*'} = (\Lambda Q)(Q' \Lambda') = \Lambda \Lambda' \tag{2.6}$$

e quindi:

$$\Sigma = \Lambda^* \Lambda^{*'} + \phi \tag{2.7}$$

Ne segue che, fissate Σ e ϕ , la matrice Λ non è univocamente identificata, ma è determinata solo a meno di una moltiplicazione a destra per una matrice ortogonale.

2.1.3 Le fasi procedurali dell'analisi fattoriale

L'applicazione dell'analisi fattoriale richiede lo svolgimento di alcuni passaggi metodologici essenziali. Viene sviluppato in fasi che vanno dalla valutazione dei presupposti iniziali alla validazione dei risultati finali.

Prima di procedere all'estrazione del numero di fattori, è necessario valutare le caratteristiche delle variabili osservate, [Alaimo, 2020]:

- **Distribuzione e linearità:** Si analizza la forma delle distribuzioni tramite gli indici di asimmetria e curtosi, procedendo a trasformare le variabili con distribuzione asimmetrica. Eventuali *outliers* multivariati possono essere identificati attraverso il calcolo della distanza di Mahalanobis, punto di partenza per l'individuazione di valori anomali e per la rappresentazione dei quantili.
- **Fattorializzabilità della matrice:** L'idoneità della matrice di correlazione viene verificata attraverso test diagnostici:
 - **Determinante della matrice:** Deve essere diverso da zero per garantire l'invertibilità della matrice. Se pari a zero, indica la presenza di variabili linearmente dipendenti, di conseguenza l'analisi fattoriale non può essere applicata.
 - **Test di sfericità di Bartlett:** Verifica l'ipotesi nulla che la matrice di correlazione sia una matrice identità; valori di significatività inferiori a 0,05 indicano che la matrice è idonea alla fattorizzazione.
 - **Test di Kaiser-Meyer-Olkin (KMO) [Kaiser, 1974]:** Confronta la grandezza delle correlazioni osservate rispetto a quelle parziali. I criteri di interpretazione sono riassunti nella Tabella 2.1.

Tabella 2.1: Interpretazione dei valori del test KMO

Valori soglia	Interpretazione
$> 0,90$	Eccellenti
$0,80 - 0,90$	Buoni
$0,70 - 0,80$	Accettabili
$0,60 - 0,70$	Mediocri
$< 0,60$	Scarsi

Successivamente si passa alla scelta del metodo di estrazione dei fattori, che dipende dalle proprietà distributive dei dati. In particolare, il metodo della Massima Verosimiglianza e quello dei Minimi Quadrati Generalizzati richiedono l'assunzione di normalità multivariata, mentre l'Analisi delle Componenti Principali (ACP) e il metodo dei Minimi Quadrati Ordinari non richiedono specifiche assunzioni distributive.

Per determinare il numero ottimale di fattori, oltre all'ispezione dello *scree-plot* e della varianza spiegata, si raccomanda l'utilizzo della **Parallel Analysis (PA)** [Horn, 1965]. Tale procedura confronta gli autovalori estratti dai dati osservati con quelli ottenuti da matrici di dati casuali generate mediante simulazione Monte Carlo. Vengono mantenuti esclusivamente i fattori i cui autovalori empirici superano quelli casuali, riducendo significativamente il rischio di sovrastima del numero di costrutti latenti.

Al fine di rendere la soluzione fattoriale maggiormente interpretabile, si procede alla rotazione dei fattori. Si distinguono:

- **Soluzioni ortogonali (Varimax, Quartimax, Equamax):** assumono l'indipendenza tra i fattori.
- **Soluzioni oblique (Promax, Oblimin):** consentono la correlazione tra i fattori.

Infine, la soluzione fattoriale viene valutata attraverso l'analisi delle communalità, dei punteggi fattoriali e degli indici di affidabilità. In particolare, la coerenza interna dei fattori può essere stimata tramite il coefficiente Ω di McDonald [McDonald, 1999], che rappresenta una misura di affidabilità basata sui carichi fattoriali e risulta meno restrittiva rispetto al coefficiente α di Cronbach [Dunn et al., 2013]. Valori di Ω superiori a 0,70 indicano una buona affidabilità del fattore.

2.2 Cluster Analysis Gerarchica

Il **clustering gerarchico** rappresenta un approccio fondamentale nell'analisi dei cluster, il cui obiettivo è costruire una gerarchia di gruppi basata sulla similarità. Tale struttura viene rappresentata graficamente tramite un diagramma ad albero multilivello, noto come *dendrogramma*. Attraverso questa rappresentazione è possibile esaminare la struttura dei dati a diversi livelli di aggregazione, permettendo così l'individuazione di sottogruppi naturali e annidati.

Nell'algoritmo gerarchico, il processo di raggruppamento è governato dalla scelta della **regola di legame** (*linkage*). Tale regola definisce il modo in cui viene calcolata la distanza tra due cluster distinti, determinando quali gruppi verranno fusi ad ogni iterazione. I metodi di *linkage* più utilizzati in letteratura sono, [Sneath et al., 1973]:

- *Single Linkage* (legame singolo);
- *Complete Linkage* (legame completo);
- *Average Linkage* (legame medio);
- Metodo di Ward.

2.2.1 Single Linkage (Legame Singolo)

Il metodo del *Single Linkage*, ovvero il criterio del vicino più prossimo (*nearest neighbor*), si basa sulla distanza minima per effettuare l'aggregazione. In questo approccio, la similarità tra due cluster è determinata dalla coppia di elementi più vicini appartenenti ai due gruppi distinti, [Johnson and Wichern, 2007].

Dato un insieme di distanze $D = \{d_{ik}\}$, l'algoritmo individua la distanza minima per unire i due oggetti, siano essi U e V . Una volta formato il nuovo cluster ($U \cup V$), è necessario aggiornare la matrice delle distanze per determinare la prossimità tra questa nuova entità e un qualsiasi altro cluster W . La regola di aggiornamento è definita formalmente come:

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \quad (2.8)$$

dove d_{UW} e d_{VW} rappresentano rispettivamente le distanze tra i cluster costituenti (U e V) e il terzo cluster W . In sostanza, la distanza tra il

nuovo gruppo e il resto dei dati è pari alla distanza del suo componente più vicino.

2.2.2 Complete Linkage (Legame Completo)

Al contrario, il *Complete Linkage*, è il criterio del vicino più lontano (*furthest neighbor*), definisce la distanza tra cluster in base alla coppia di elementi più distanti. La regola di aggiornamento in questo caso è:

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\} \quad (2.9)$$

Questo metodo garantisce che tutti gli elementi all'interno di un cluster risultante si trovino entro una determinata distanza massima l'uno dall'altro, favorendo la formazione di gruppi più compatti.

2.2.3 Average Linkage

L'*Average Linkage* rappresenta un compromesso tra i due precedenti, considerando la distanza media tra tutte le possibili coppie di elementi appartenenti ai due cluster, [Sokal and Michener, 1958]. La distanza tra il cluster (UV) e il cluster W è definita come:

$$d_{(UV)W} = \frac{1}{N_{(UV)} \cdot N_W} \sum_{i \in (UV)} \sum_{k \in W} d_{ik} \quad (2.10)$$

dove $N_{(UV)}$ e N_W indicano il numero di elementi nei rispettivi cluster, mentre d_{ik} è la distanza tra l'oggetto i del cluster (UV) e l'oggetto k del cluster W . Questo metodo è ampiamente utilizzato per la sua capacità di gestire sia oggetti che variabili.

2.2.4 Metodo di Ward

A differenza dei criteri basati esclusivamente sulla distanza, il *Metodo di Ward*, [Jr., 1963], mira a minimizzare la perdita di informazione derivante dall'unione di due gruppi. Tale perdita è definita come l'incremento della somma dei quadrati degli errori (*Error Sum of Squares*, ESS). L'ESS per un generico cluster k è dato dalla somma dei quadrati degli scarti tra ogni elemento e il centroide del cluster stesso:

$$ESS = \sum_{j=1}^N (x_j - \bar{x})'(x_j - \bar{x}) \quad (2.11)$$

Ad ogni iterazione, l'algoritmo fonde i due cluster la cui combinazione comporta il minore aumento possibile di ESS. Questo metodo è particolarmente efficace per identificare cluster con forma approssimativamente ellittica.

2.3 Latent Class Analysis

L'approccio delle **Classi Latenti (LC)** al clustering si configura come un metodo *model-based*, fondato su un modello probabilistico che descrive l'eterogeneità non osservata della popolazione mediante una variabile latente. L'assunzione fondamentale, su cui si basa LC, è quella di **indipendenza locale**, secondo cui le variabili osservate non sono necessariamente indipendenti marginalmente, ma risultano indipendenti una volta condizionate all'appartenenza alla stessa classe latente. In questo modo, la variabile latente funge da fattore esplicativo delle associazioni osservate tra gli indicatori [Vermunt, 2003].

La Latent Class Analysis può essere applicata a variabili risposta di diversa natura, quali continue, categoriali, di conteggio oppure misurate su scala likert [Jay Magidson, 2020]. Più in generale, i modelli LC costituiscono una famiglia di modelli statistici di classificazione che rappresentano la popolazione come una miscela di un numero finito L di sottogruppi latenti, ciascuno caratterizzato da una specifica distribuzione delle risposte osservate. Ogni unità statistica appartiene a una sola classe latente, non direttamente osservabile, e l'obiettivo dell'analisi consiste nello stimare sia il numero di classi sia la proporzione di unità appartenenti a ciascuna di esse, in quanto esse sono non note a priori.

2.3.1 Struttura del Modello

Nei modelli LC si distinguono due tipologie di variabili:

- **Variabili manifeste (indicatori):** misurate direttamente, spesso categoriali (nominali oppure ordinali);
- **Variabili latenti:** non osservabili direttamente, le cui categorie definiscono le *classi latenti*.

L'idea alla base del modello consiste nell'introdurre una variabile latente X con L classi [Lazarsfeld and Henry, 1968], tale da spiegare le associazioni osservate tra gli indicatori. Con questa prospettiva, la popolazione

viene interpretata come un insieme di sottopopolazioni non osservabili, ciascuna caratterizzata da una distribuzione specifica delle risposte sugli item.

La Latent Class Analysis può essere interpretato come un modello di misura, poichè i parametri stimati definiscono la variabile latente attraverso gli indicatori osservati. In particolare, esso include:

- le proporzioni di classe $P(x = l)$. Esse rappresentano la probabilità di appartenenza alla classe l ;
- le probabilità di risposta condizionate $P(y_{ik} = s_k | x = l)$, dove viene descritta la distribuzione delle risposte per ciascun indicatore all'interno della classe l .

Attraverso le differenze di tali probabilità nelle diverse classi è possibile interpretare e caratterizzare le classi latenti.

Come detto in precedenza, l'assunzione centrale del modello a classi latenti è l'*indipendenza locale* [Collins and Lanza, 2010], secondo la quale gli indicatori risultano mutuamente indipendenti una volta condizionati alla classe latente. Di conseguenza, dato $x = l$, la probabilità congiunta di osservare uno specifico vettore di risposte \mathbf{y}_i può essere espressa come prodotto delle probabilità sui singoli indicatori.

Siano definiti i seguenti termini:

- Y_{ik} (con $i = 1, \dots, I$ e $k = 1, \dots, K$): la risposta dell'individuo i sull'indicatore k ;
- s_k : un particolare livello dell'item k ;
- X_i : la variabile latente composta da L classi ($l = 1, \dots, L$);
- \mathbf{Y}_i : il vettore completo delle risposte del caso i .

La struttura probabilistica che definisce un modello LC semplice può essere espressa come segue, [Goodman, 1974]:

$$P(\mathbf{Y}_i = s) = \sum_{l=1}^L P(X_i = l) \prod_{k=1}^K P(Y_{ik} = s_k | X_i = l) \quad (2.12)$$

Come evidenziato dall'equazione, la probabilità di osservare un determinato pattern di risposta è una media ponderata delle probabilità specifiche di classe $P(Y_{ik} = s_k | X_i = l)$, dove il peso è rappresentato dalla probabilità che l'unità i appartenga alla classe latente l . Per l'assunto di indipendenza locale, si ipotizza che gli indicatori siano indipendenti data l'appartenenza alla classe latente.

2.3.2 Come decidere il numero di classi

Un problema cruciale nella specificazione di un modello a classi latenti consiste nella determinazione del numero appropriato di classi. Questa scelta può essere effettuata tramite l'utilizzo di test di bontà dell'adattamento, criteri informativi oppure statistiche locali di bontà dell'adattamento, come i residui bivariati (Bivariate Residuals, BVR), [Jay Magidson, 2020].

Il test di bontà dell'adattamento tradizionale confronta le frequenze attese del modello, ottenute tramite stima di massima verosimiglianza, con le frequenze osservate. Questo confronto viene effettuato utilizzando la statistica del rapporto di verosimiglianza L^2 oppure la statistica di Pearson X^2 . Il corrispondente valore-p valuta se le differenze tra frequenze osservate e stimate possano essere attribuite alla sola variabilità campionaria.

Nel caso ideale in cui le frequenze attese riproducano perfettamente quelle osservate, l'adattamento del modello risulta perfetto e sia L^2 sia X^2 assumono valore pari a zero. All'aumentare di L^2 , la statistica quantifica il grado di mancato adattamento del modello, ossia la quota di associazione residua tra gli indicatori non spiegata dal modello. Poiché il valore di L^2 viene calcolato per il modello nullo di indipendenza e misura l'associazione totale presente nei dati, è possibile costruire una misura analoga a un coefficiente R^2 che esprima la percentuale di associazione spiegata dal modello a classi latenti.

Le frequenze attese possono essere ricavate moltiplicando le probabilità stimate dal modello per la dimensione campionaria N , ottenendo così le frequenze previste per ciascun pattern di risposta.

Inoltre, quando il numero di indicatori è elevato, la tabella a doppia entrata tende a diventare sparsa (con molte celle vuote o con frequenze

molto piccole). In queste situazioni le statistiche L^2 e X^2 non seguono più una distribuzione χ^2 , per cui non rendendo più affidabili i valori-p associati. Per risolvere tale problema è possibile andare a utilizzare valori-p bootstrap, che non dipendono dall'approssimazione asintotica χ^2 . Poiché il bootstrap può risultare computazionalmente oneroso nei modelli complessi, nella pratica si ricorre frequentemente a criteri informativi basati sulla log-verosimiglianza, come l'AIC o il BIC.

Nelle applicazioni esplorative, la procedura usuale consiste nello stimare modelli successivi con un numero crescente di classi e selezionare la soluzione più parsimoniosa che fornisca un adattamento adeguato ai dati. In generale, si preferisce il modello con il valore minimo del BIC, oppure quello per cui l'ipotesi di indipendenza locale non viene più respinta a un livello di significatività prefissato.

Per individuare eventuali associazioni residue tra coppie di indicatori, possono inoltre essere esaminate statistiche locali di bontà dell'adattamento, come i residui bivariati, utili per identificare possibili violazioni dell'assunzione di indipendenza locale.

2.3.3 Obiettivi e Segmentazione

L'obiettivo dell'analisi LC è determinare il numero minimo di classi latenti L sufficiente per spiegare le associazioni osservate tra le variabili manifeste. Per cui l'ultimo passaggio dell'analisi consiste nell'utilizzare i risultati del modello per classificare le unità [Jay Magidson, 2020].

Per ogni pattern di risposta osservato (y_1, \dots, y_J) , le **probabilità di appartenenza a posteriori** sono calcolate applicando il **teorema di Bayes**:

$$P(X = l \mid y_1, \dots, y_J) = \frac{P(X = l) \prod_{j=1}^J P(y_j \mid X = l)}{\sum_{l'=1}^L P(X = l') \prod_{j=1}^J P(y_j \mid X = l')} \quad (2.13)$$

dove $P(x = l)$ rappresenta la probabilità incondizionata della classe latente l e $P(y_j \mid x = l)$ le probabilità condizionate associate agli indicatori per i soggetti appartenenti a quella classe. L'assegnazione delle unità avviene secondo la **regola modale**, cioè ogni soggetto viene assegnato alla classe per la quale la probabilità a posteriori è massima [Jay Magidson, 2020].

Magidson e Vermunt [Magidson and Vermunt, 2001] definiscono questo processo come *LC cluster model*, poiché l'obiettivo di identificare L gruppi omogenei è concettualmente simile a quello della cluster analysis tradizionale.

L'utilizzo della Latent Class Analysis, in un contesto di marketing, vuol dire individuare andare ad individuare le classi latenti che corrisponderanno a specifici segmenti di mercato. Questo aspetto dunque, permetterà la profilazione di gruppi omogenei di consumatori con caratteristiche e comportamenti simili tra coloro appartenenti allo stesso gruppo e differenti dagli altri gruppi.

2.3.4 Il Modello di Regressione LC

Il modello di *regressione a classi latenti* (LC) differisce dal modello di classificazione LC: in questo contesto, i parametri che variano tra i sottogruppi non osservati non sono semplici probabilità condizionate, ma coefficienti di regressione. A differenza dei modelli lineari classici, in cui i coefficienti sono stimati come effetti fissi sull'intera popolazione, la regressione LC permette ai parametri di variare tra le classi. Questo approccio offre una soluzione all'eterogeneità non osservata analoga ai **modelli misti** [Jay Magidson, 2020].

L'applicazione di questi modelli riguarda i casi in cui si dispone di risposte multiple o misure ripetute per ogni unità campionaria. Un esempio tipico è la *conjoint analysis*, dove i segmenti di mercato vengono identificati in base agli attributi che guidano le preferenze e le scelte individuali. In tale scenario, la variabile latente interagisce con i predittori osservati, consentendo di catturare diverse strutture di risposta e di individuare eventuali *outlier* [Wedel and Kamakura, 2000].

La struttura probabilistica del modello è definita dalla seguente funzione di densità miste:

$$P(Y_i | \mathbf{z}_i^{cov}, \mathbf{z}_i^{pred}) = \sum_{l=1}^L P(X_i = l | \mathbf{z}_i^{cov}) P(Y_i | X_i = l, \mathbf{z}_i^{pred}) \quad (2.14)$$

dove \mathbf{z}_i^{cov} è il vettore delle covariate che spiegano l'appartenenza alle L classi latenti (modello di appartenenza), mentre \mathbf{z}_i^{pred} rappresenta i predittori che influenzano la variabile risposta y_i specificamente all'interno

di ciascuna classe (modello di misura). Questo doppio livello di analisi permette di identificare simultaneamente i segmenti della popolazione e i relativi *driver* comportamentali.

Capitolo 3

Analisi dei dati

3.1 Indagine Gastronomica in Italia

Dopo la fase di sviluppo del questionario, composto da 51 domande, si è proceduto alla sua somministrazione con l'obiettivo di individuare gruppi di rispondenti caratterizzati da comportamenti simili. Come primo step è stata applicata l'analisi fattoriale, al fine di ridurre la numerosità delle variabili e verificare se il questionario proposto ai viaggiatori italiani misurasse un unico costrutto, in questo caso l'interesse per la gastronomia nella scelta della destinazione di viaggio.

A tal fine è stata effettuata una segmentazione dei dati: inizialmente sono state utilizzate le distanze euclidee sia sui valori medi delle batterie di item sia sui punteggi fattoriali delle batterie, per determinare il numero ottimale di gruppi da individuare. Successivamente è stato applicato il metodo LCA, un metodo probabilistico, per verificare la presenza di classi latenti non direttamente osservabili in grado di spiegare il comportamento dei rispondenti. Per analizzare quali variabili demografiche distinguessero in modo significativo i diversi gruppi, sono stati applicati test ANOVA per verificare le differenze tra le medie e le distribuzioni delle variabili demografiche nei diversi cluster nel modello gerarchico, mentre per LCA sono stati usati test di Wald.

3.1.1 Pulizia del dataset

Inizialmente il dataset si componeva di 1208 osservazioni e 67 variabili:

Tabella 3.1: Tipologia delle variabili del dataset `dati`

Variabile/i	Tipo R	Descrizione
CollectorNm	factor	Nome del collector
respondent_id	numeric	Identificativo rispondente
collector_id	integer	Identificativo collector
date_created	factor	Data creazione risposta
date_modified	factor	Data modifica risposta
ip_address	factor	Indirizzo IP
email_address	logical	Email (valori mancanti)
first_name	logical	Nome (valori mancanti)
last_name	logical	Cognome (valori mancanti)
custom_1	factor	Identificativo personalizzato
q0001	integer	Risposta domanda singola
q0002_0001 – q0002_0006	integer	Batterie domanda q0002
q0003_0001 – q0003_0008	integer	Batterie domanda q0003
q0004_0001 – q0004_0008	integer	Batterie domanda q0004
q0005_0001 – q0005_0017	integer	Batterie domanda q0005
q0006_0001 – q0006_0006	integer	Batterie domanda q0006
q0007 – q0011	integer	Domande singole finali (demografiche)
p0001 – p0006	integer	Variabili aggiuntive (tutte NA)

Come si può osservare, per ciascuna domanda del questionario strutturata su scala Likert, **ogni modalità di risposta è stata considerata come una variabile distinta**. Analizzando il dataset, si è notato che le **ultime sei variabili (p000...)** risultavano essere **variabili nulle e non presenti nel questionario originale**. Tali variabili, generate automaticamente durante l'esportazione del dataset da **SPSS a formato .csv**, sono state pertanto eliminate.

Oltre a questa prima pulizia, sono state rimosse anche le **variabili identificative** dei rispondenti e degli intervistatori (dalla variabile *CollectorNm* fino a *custom_1*), in modo da mantenere esclusivamente le variabili riferite ai contenuti del questionario.

Effettuata questa prima selezione, si è passati alla **verifica dei dati mancanti**. È emerso che **67 rispondenti** avevano selezionato, alla prima domanda filtro, la modalità “**9 = non viaggio**”. Poiché per questi soggetti l’intervista si interrompeva, si è deciso di escluderli dal campione, riducendo così le osservazioni da **1.208 a 1.141**, con un totale di **51 variabili**.

Dopo tale esclusione, è stato eseguito un nuovo controllo sui valori mancanti, riscontrando che alcune variabili presentavano al massimo un 2% di dati mancanti.

Si è quindi proceduto come segue:

- per le variabili quantitative, i valori mancanti sono stati imputati con la **media**;
- per le variabili qualitative, è stata creata una nuova categoria denominata “**missing**”.

Infine, si è osservato che tutte le variabili presenti nel dataset erano di **tipo intero**. Per evitare problemi nelle successive procedure di imputazione e analisi, si è quindi deciso di trasformare in fattori la variabile q0001 (motivazioni nel viaggiare) e le variabili da q0007 a q0011 (relative alle caratteristiche demografiche dei rispondenti).

Tabella 3.2: Informazioni generali del dataset

Numero osservazioni	1141
Numero variabili	51

Tabella 3.3: Variabili categoriche

Variabile	Tipo	Descrizione
q0001	Factor (8 livelli)	Motivazione principale (es. svago/riposo, ...)
q0007	Factor (3 livelli)	Genere (femmina, maschio, altro)
q0008	Factor (6 livelli)	Fasce di età
q0009	Factor (6 livelli)	Titolo di studio
q0010	Factor (8 livelli)	Condizione lavorativa
q0011	Factor (7 livelli)	Fascia di reddito

Tabella 3.4: Variabili Likert (scale numeriche intere)

Batteria	Variabili	Numero item
q0002	q0002_0001 – q0002_0006	6
q0003	q0003_0001 – q0003_0008	8
q0004	q0004_0001 – q0004_0008	8
q0005	q0005_0001 – q0005_0017	17
q0006	q0006_0001 – q0006_0006	6

3.1.2 Analisi Esplorative

Successivamente alla fase di pulizia del dataset, si è passati ad analizzare la composizione del campione di intervistati attraverso **analisi esplorative supportate da rappresentazioni grafiche**, al fine di comprendere come ciascun aspetto fosse valutato dai rispondenti.

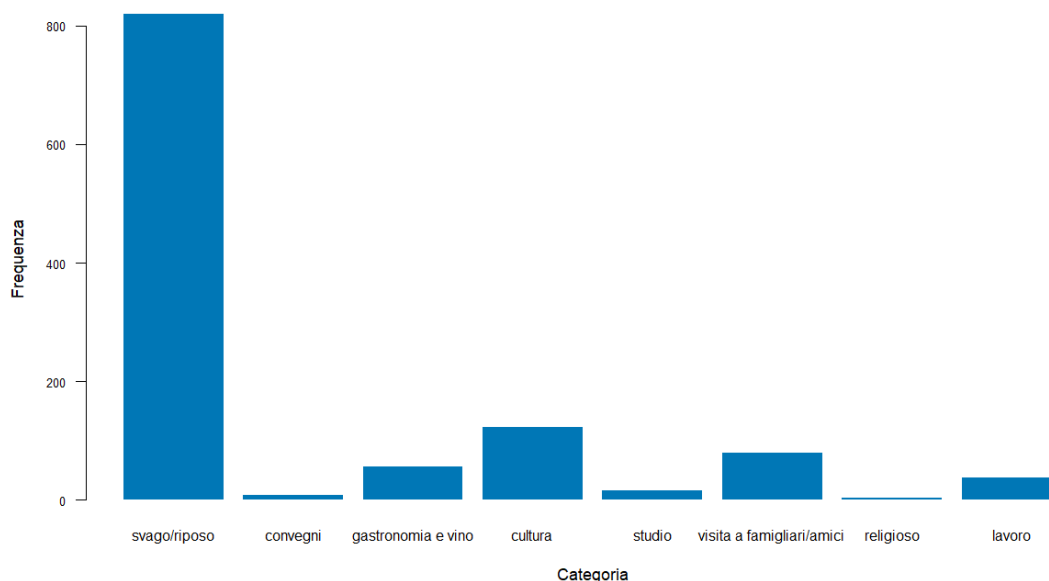


Figura 3.1: Motivazione del viaggio

Come prima analisi, si è osservato quali fossero le **principali motivazioni di viaggio**. Dalla figura si può notare che la motivazione più frequentemente indicata è quella legata allo **svago e/o al riposo**, seguita da **cultura, visita ad amici e familiari, gastronomia e vino, lavoro, studio, convegni e infine religione**. È interessante osservare

come la gastronomia ricopra un ruolo significativo tra le motivazioni di viaggio, pur non rappresentando la principale spinta alla partenza.

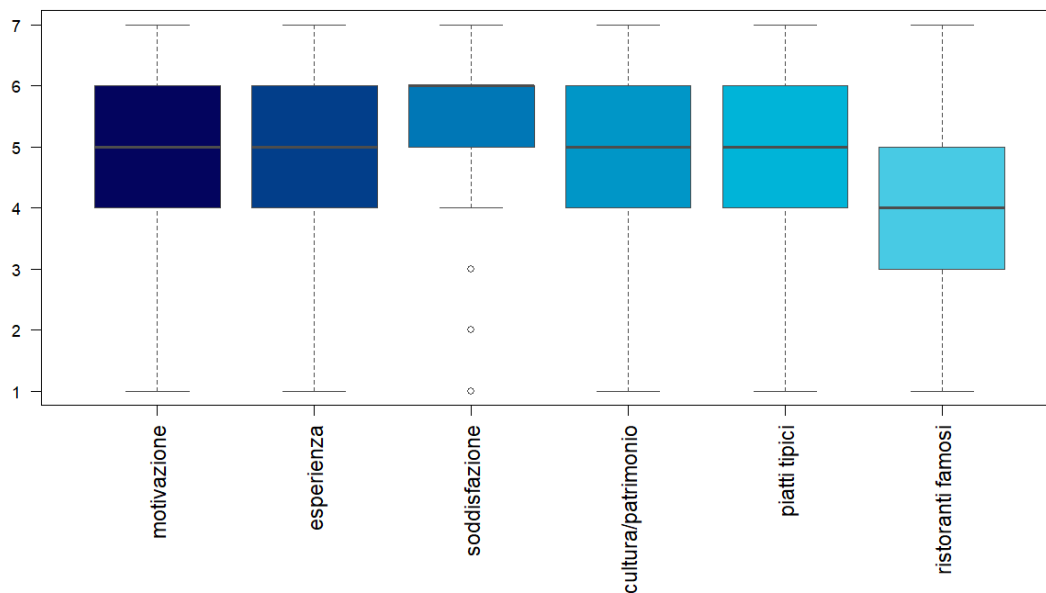


Figura 3.2: Gastronomia nel viaggio

Analizzando la batteria q0002, relativa alla **valutazione degli aspetti gastronomici nei viaggi**, i boxplot mostrano in generale una **media pari a 5 su 7** (valutazione “buona”) per le prime cinque categorie considerate. Solo l’ultima categoria presenta un valore medio leggermente inferiore, pari a 4 (“sufficiente”). Ciò suggerisce che i turisti non sono particolarmente interessati a visitare ristoranti rinomati, ma attribuiscono maggiore importanza alla soddisfazione complessiva, all’esperienza e alla cultura culinaria del luogo visitato.

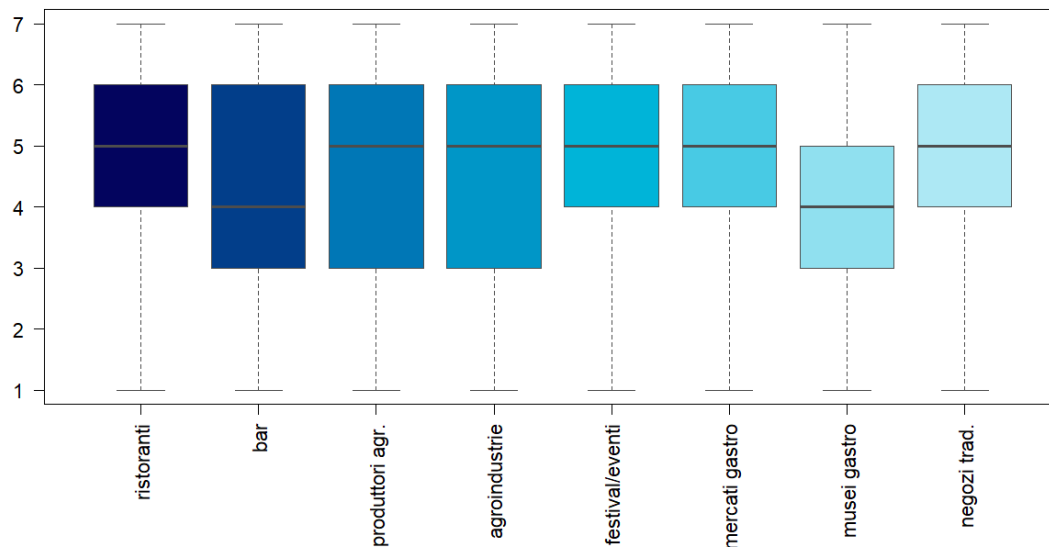


Figura 3.3: Intenzione della destinazione turistica

Proseguendo con la batteria q0003, dedicata alla **valutazione degli aspetti che influenzano la scelta di destinazioni legate alla gastronomia**, emerge che i rispondenti considerano rilevante l'offerta gastronomica complessiva della destinazione. La maggior parte delle modalità ottiene infatti punteggi elevati, indicando che le persone ricercano luoghi in grado di offrire esperienze complete, autentiche e di qualità. Le uniche modalità con valutazioni inferiori sono quelle relative a bar e musei gastronomici, percepite come meno attrattive.

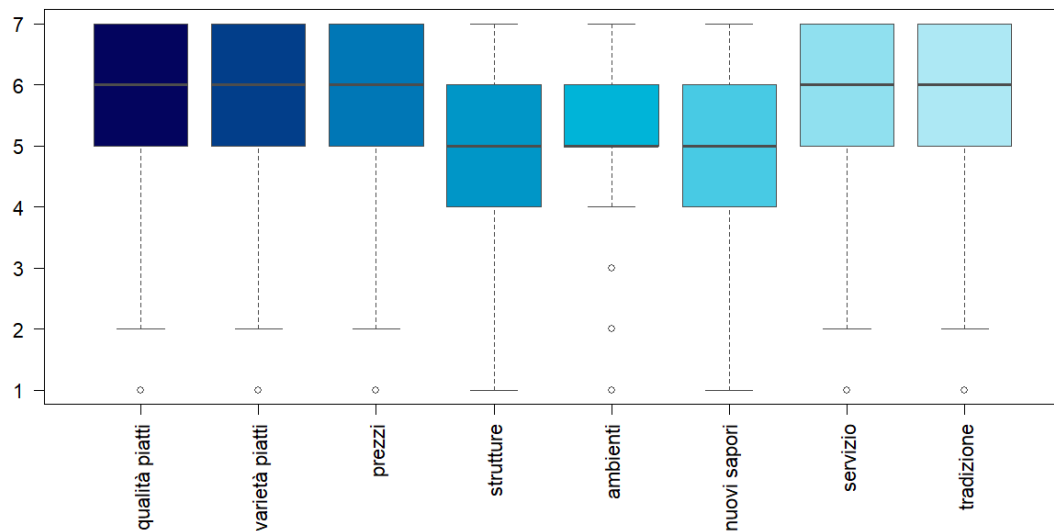


Figura 3.4: Importanza della destinazione turistica

Con la batteria q0004, volta ad indagare **cosa ricercano le persone nella gastronomia**, si osserva che le dimensioni qualità, varietà, prezzo, ospitalità e tradizione ottengono valori medi intorno al 6, risultando quindi molto importanti per i rispondenti. Ciò indica che i turisti italiani ricercano non solo la buona cucina, ma anche un'esperienza autentica, legata al contatto diretto con i ristoratori e alla scoperta delle tradizioni locali. Al contrario, aspetti come la struttura del locale, gli interni e la novità dei sapori risultano meno rilevanti.

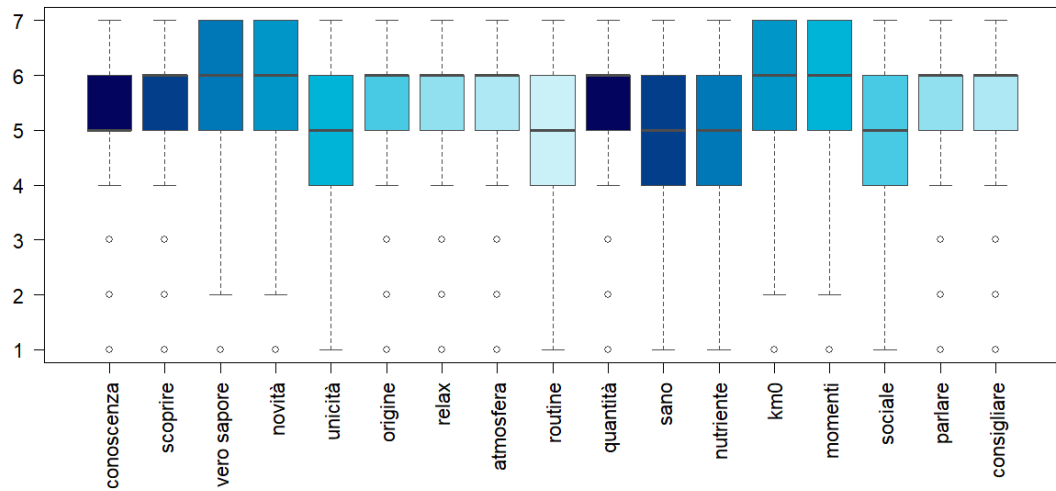


Figura 3.5: Percezioni sull'esperienza gastronomica

Analizzando la batteria q0005, che indagava **l'esperienza gastronomica vissuta**, emerge che gli italiani attribuiscono grande importanza al vero sapore dei piatti tipici e alla tradizione culinaria del territorio. Inoltre, la provenienza dei prodotti risulta un elemento di forte interesse: i rispondenti mostrano una preferenza per i prodotti a km 0, segno di una crescente attenzione verso la sostenibilità.

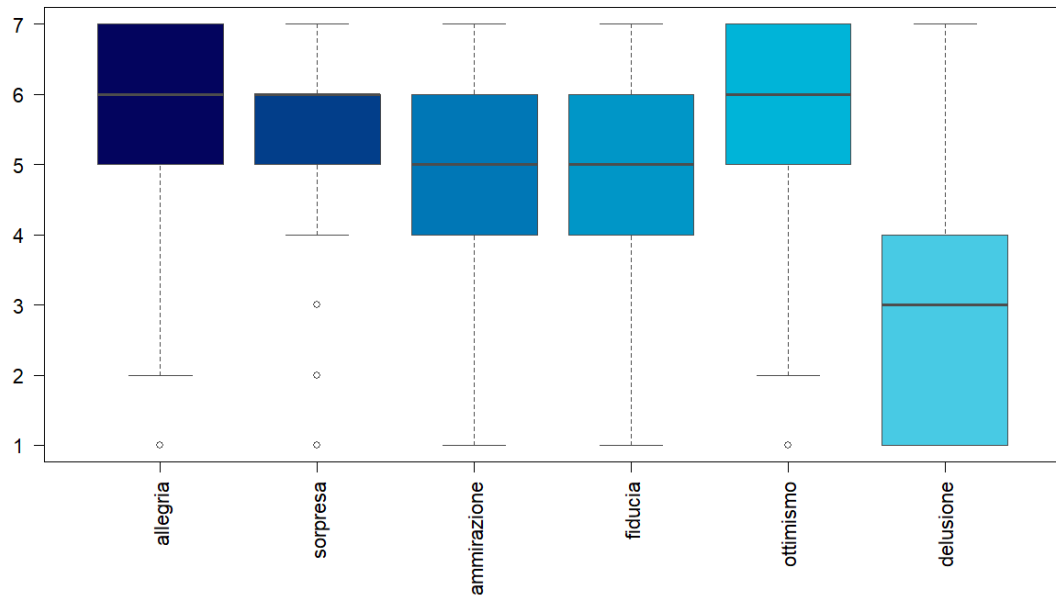


Figura 3.6: Emozioni

Infine, con la batteria q0006, dedicata alle **emozioni legate all'esperienza gastronomica**, si osserva che la maggior parte dei rispondenti dichiara di provare soddisfazione e sensazioni positive legate al consumo di prodotti locali.

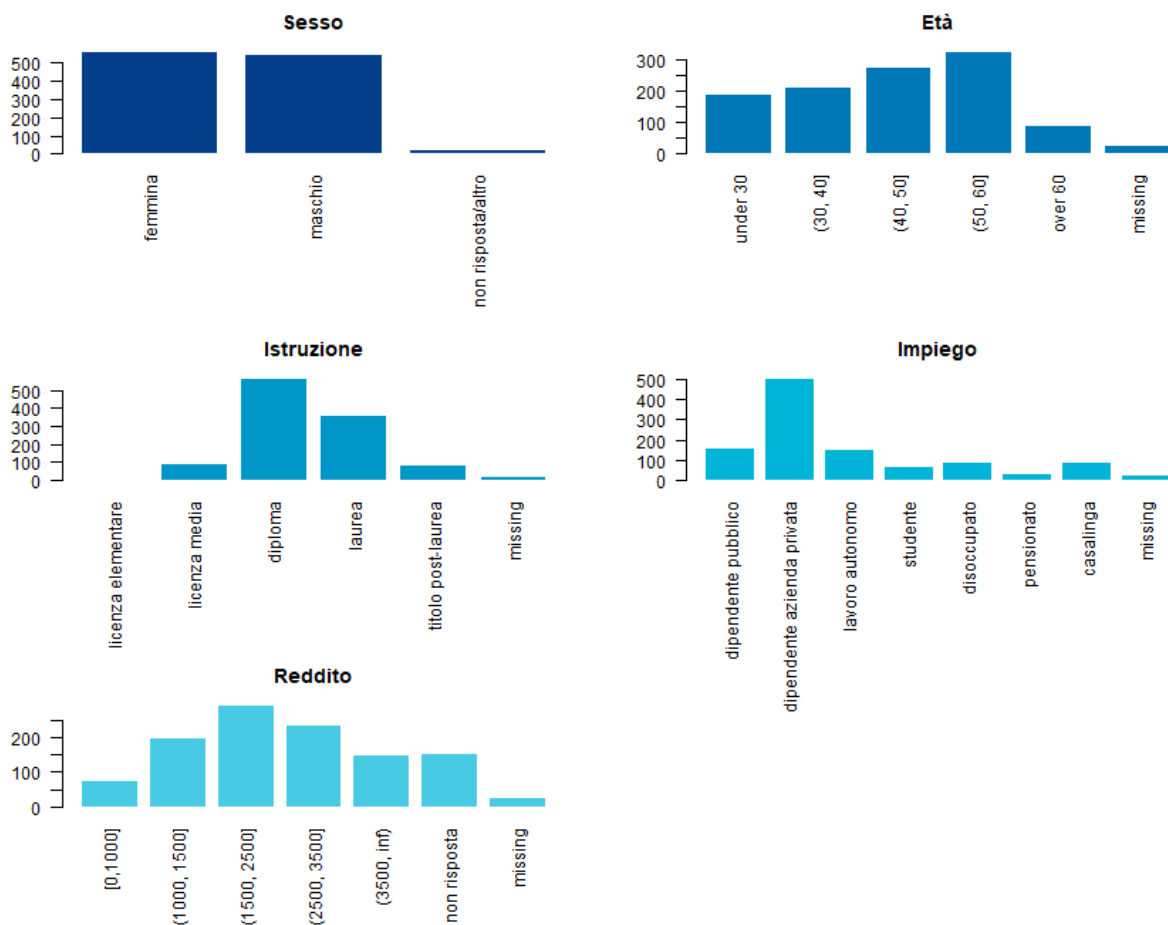


Figura 3.7: Socio-demografiche

Per quanto riguarda le **variabili demografiche**, il campione risulta bilanciato tra uomini e donne. La maggior parte dei partecipanti ha un'età compresa tra i **40 e i 60 anni**, quindi si tratta prevalentemente di **lavoratori** (sia pubblici che privati). In termini di istruzione, la quota maggiore possiede un **diploma di scuola secondaria**, seguita da individui con **laurea**. Il reddito medio dichiarato si colloca tra **1.500 e 3.500 euro mensili**, corrispondente a una fascia medio-alta.

3.2 Analisi della Struttura Latente e Dimensionalità

Al fine di ridurre la complessità delle variabili osservate e individuare le dimensioni latenti che guidano il comportamento turistico, è stata applicata un'analisi di riduzione delle dimensioni (Analisi Fattoriale). L'obiettivo di tale procedura è triplice: sintetizzare l'informazione contenuta negli item del questionario, verificare se lo strumento misuri un unico costrutto coerente e ottenere punteggi sintetici da utilizzare successivamente come input per la *Cluster Analysis*.

3.2.1 Procedura Metodologica e Verifica della Fattorizzabilità

La procedura ha previsto inizialmente la selezione di tutti gli item appartenenti alle batterie di domande relative a motivazioni, importanza, percezioni, intenzioni ed emozioni. Data la natura ordinale delle variabili, misurate su scala Likert a 7 punti, si è proceduto al calcolo della **matrice di correlazione policorica** [Olsson, 1979]. Tale scelta metodologica è necessaria per evitare la sottostima delle relazioni tra le variabili, tipica della classica correlazione di Pearson quando applicata a dati categoriali.

Per confermare l'idoneità dei dati alla fattorizzazione, sono stati calcolati i due indici statistici di riferimento, i quali hanno restituito i seguenti risultati:

- Il test di sfericità di **Bartlett** [Bartlett, 1951] è risultato altamente significativo ($p < 0.001$), permettendo di rigettare l'ipotesi nulla che la matrice di correlazione sia una matrice identità. Ciò conferma la presenza di legami lineari strutturati e significativi tra le variabili.
- L'indice di adeguatezza campionaria **KMO** (Kaiser-Meyer-Olkin), [Kaiser, 1974], ha restituito un valore pari a **0.977**. Secondo la letteratura metodologica, un punteggio superiore a 0.90 definisce il campione come "eccellente" per l'analisi fattoriale, garantendo che la struttura delle correlazioni sia estremamente solida e non dovuta a errori casuali.

3.2.2 Analisi della Dimensionalità (Parallel Analysis e Omega)

Per determinare il numero ottimale di fattori, è stata condotta una **Parallel Analysis**.

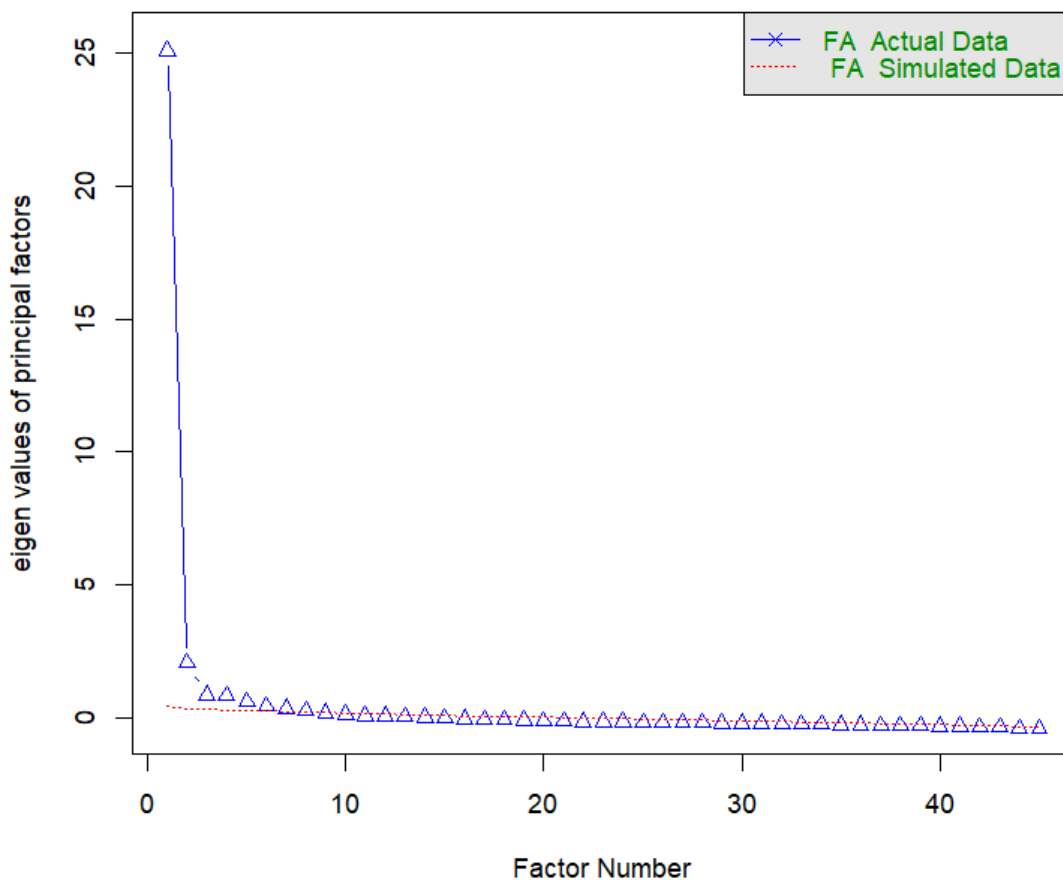


Figura 3.8: Parallel Analysis

Come si evince dal grafico (Figura 3.8), il metodo suggerisce l'estrazione di **8 fattori**, corrispondenti ai punti della curva dei dati reali (*Actual Data*) che si posizionano al di sopra della linea soglia dei dati simulati. Tuttavia, si osserva che il primo fattore è dominante, raggiungendo un autovalore pari a circa 25, tale sproporzione rispetto ai fattori successivi suggerisce la presenza di un robusto fattore generale sottostante.

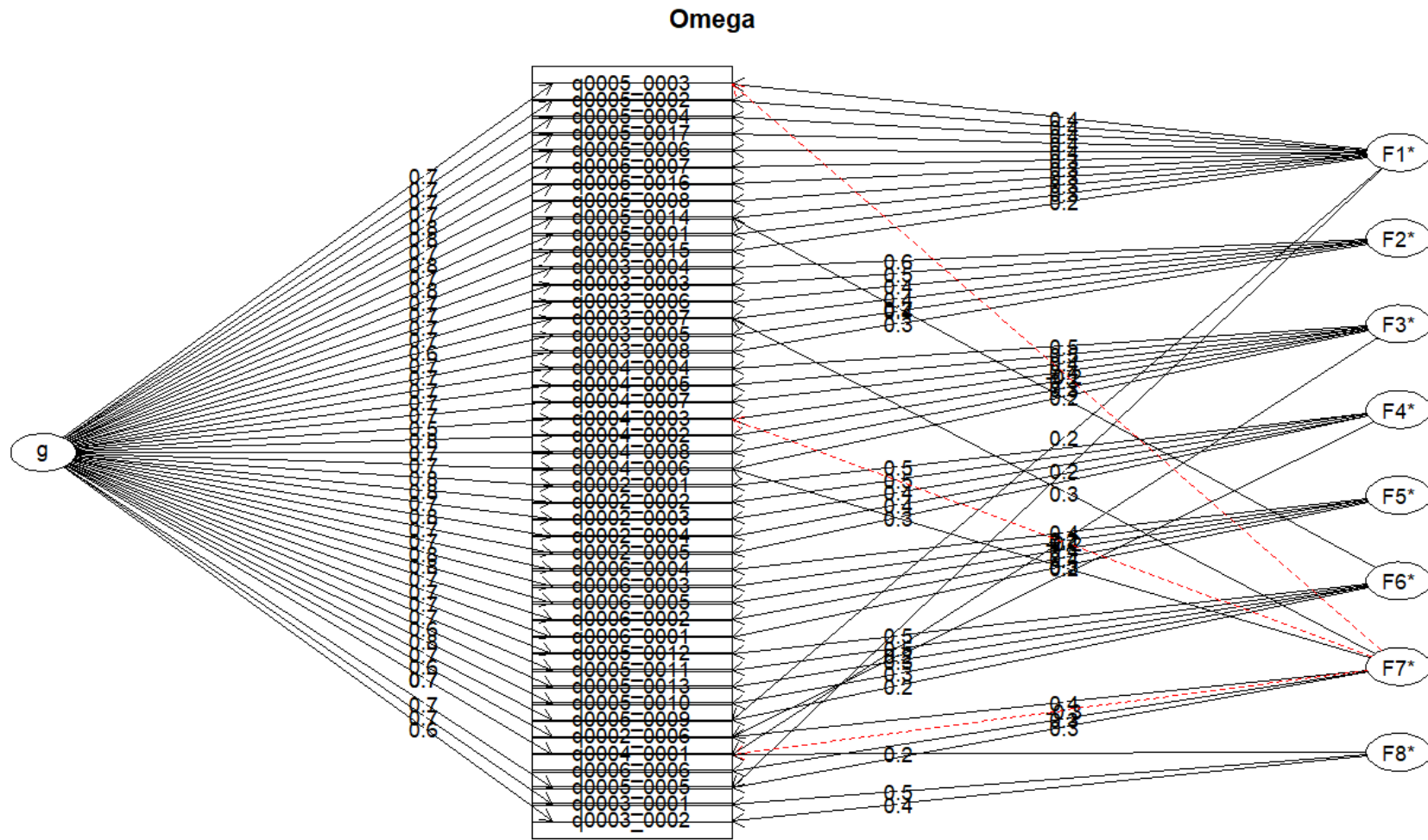


Figura 3.9: Omega Hierarchical

Per approfondire la natura di questi 8 fattori, è stato stimato un modello **Omega Hierarchical**, [McDonald, 1999]. Come si può osservare dalla Figura 3.9, l'ispezione della struttura bifattoriale ha confermato una **forte unidimensionalità essenziale**:

- **Dominanza del Fattore Generale (g):** La quasi totalità degli item presenta saturazioni (*loadings*) estremamente elevate sul fattore generale, con valori compresi tra 0.70 e 0.90. Questo indica che la maggior parte della varianza comune è spiegata da un unico nucleo tematico.
- **Ruolo dei Fattori Specifici ($F1 - F8$):** Sebbene la Parallel Analysis identifichi 8 dimensioni, l'analisi Omega mostra che, una volta rimosso l'effetto del fattore generale, le saturazioni specifiche residue (*group factors*) sono sensibilmente inferiori, attestandosi prevalentemente in un range compreso tra 0.20 e 0.40.

In conclusione, l'evidenza empirica suggerisce che il questionario possiede una solida coerenza interna e "parla della stessa cosa", [Reise, 2012], sebbene esistano sfumature tematiche identificabili negli 8 fattori estratti.

Tuttavia, poiché l'analisi Omega indica che il legame latente tra tutti gli item è talmente forte da rendere il fattore generale il principale motore delle risposte, l'utilizzo di punteggi fattoriali puri potrebbe appiattare la diversità tra i rispondenti. Per tali motivi, si è deciso di procedere con una riduzione dimensionale forzata: anziché utilizzare i punteggi degli 8 fattori statistici, verranno utilizzati gli indici medi calcolati per ogni batteria originale. Questa scelta permette di ridurre il numero di parametri mantenendo intatta la distinzione semantica tra le diverse aree del questionario, facilitando una *Cluster Analysis* più interpretabile ai fini della ricerca.

3.3 Cluster Analysis e Segmentazione

Al fine di individuare profili turistici omogenei e verificare la stabilità della segmentazione, sono stati adottati due approcci metodologici complementari. In prima istanza, si è proceduto con una **Cluster Analysis Gerarchica** applicata agli indici sintetici (punteggi fattoriali delle batterie), ottenendo una partizione basata su metriche di distanza geometrica. Successivamente, è stata applicata una **Latent Class Analysis (LCA)** direttamente sui singoli item, con l'obiettivo di far emergere eventuali strutture latenti non rilevabili attraverso l'uso dei soli punteggi aggregati.

3.3.1 Cluster Analysis Gerarchica

Con lo scopo di individuare gruppi omogenei di viaggiatori, si è proceduto in primo luogo all'applicazione del metodo di clustering **Gerarchico**. Tale metodo si fonda sull'analisi delle distanze: tramite queste, infatti, vengono raggruppate iterativamente le unità statistiche che presentano maggiore similarità.

Per i dati a disposizione, è stata utilizzata la **Distanza Euclidea**, misura standard per quantificare la separazione geometrica tra le osservazioni nello spazio multidimensionale delle variabili. Congiuntamente a tale metrica, è stato adottato il metodo di aggregazione di **Ward** (Ward's Linkage), scelto per la sua capacità di minimizzare la varianza interna ai gruppi e generare cluster compatti e di dimensioni equilibrate.

Al fine di testare la robustezza della segmentazione, l'analisi è stata condotta seguendo due diversi approcci di sintesi dei dati: l'utilizzo delle **medie aritmetiche** calcolate per ogni batteria di item e l'utilizzo dei **punteggi fattoriali**, anch'essi estratti per ciascuna batteria. Il confronto tra i due dendrogrammi (Fig. 3.10 e 3.11) ha confermato in entrambi i casi la scelta di una soluzione a quattro cluster come la più adeguata per interpretare la struttura dei dati.

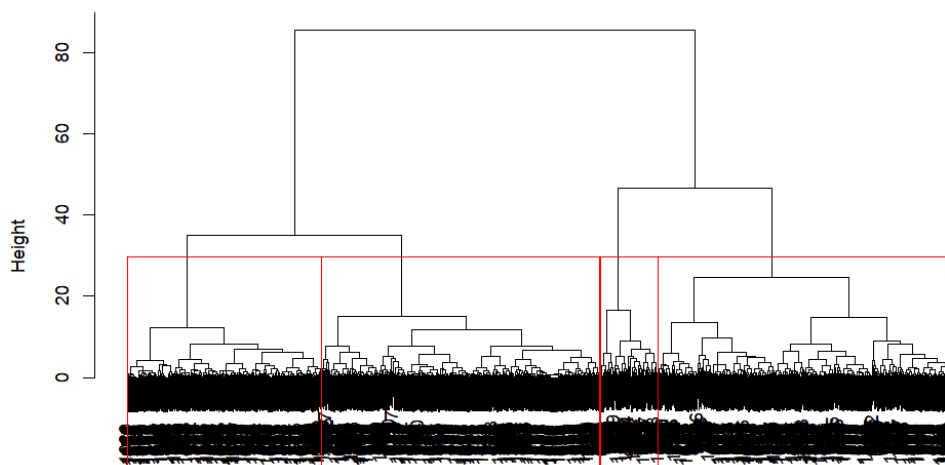


Figura 3.10: Clustering Gerarchico - Medie

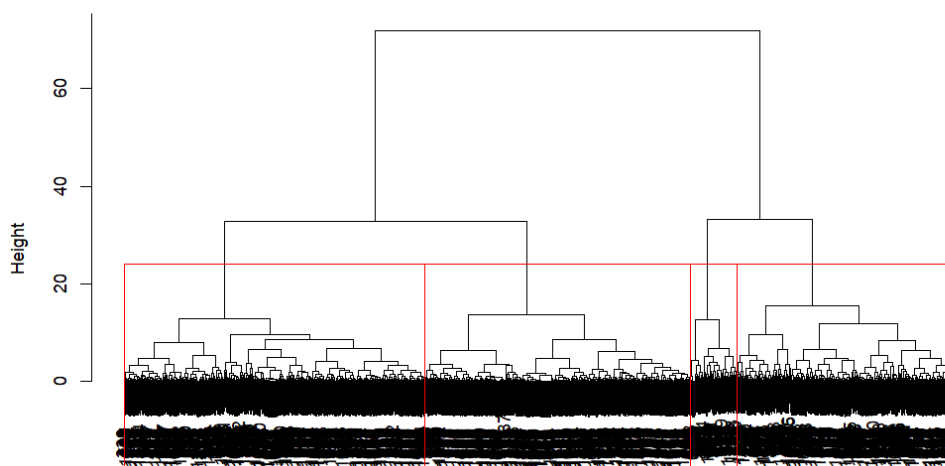


Figura 3.11: Clustering Gerarchico - Fattori

Per valutare il grado di concordanza tra le due procedure, è stata costruita una matrice di contingenza (Tab. 3.5). L'analisi evidenzia una buona

stabilità della soluzione a quattro cluster, in particolare per i gruppi caratterizzati da profili più distintivi. Il **cluster 4** risulta sostanzialmente coincidente nelle due procedure (64 unità), mentre il **cluster 2** mostra una forte coerenza, con 267 unità assegnate stabilmente allo stesso gruppo, si osserva tuttavia che 99 osservazioni vengono associate al cluster 1 nell'approccio basato sulle medie, indicando una limitata area di confine tra tali gruppi. Si osserva una parziale sovrapposizione tra i **cluster 1 e 3**, tale fenomeno è atteso per unità con caratteristiche intermedie, prossime alla media generale, e riflette una minore separazione nello spazio delle variabili.

Tabella 3.5: Matrice di coincidenza tra Cluster su Fattori e Cluster su Medie

Cluster su Fattori	Cluster su Medie			
	1	2	3	4
1	274	2	138	0
2	99	267	0	0
3	13	0	268	16
4	0	0	0	64

Nota: I valori sulla diagonale indicano le unità classificate in modo coerente.

Nel complesso, la predominanza della diagonale principale conferma la **robustezza della segmentazione** a quattro gruppi quale struttura di base dei dati. Tuttavia, successive analisi condotte tramite modelli di classificazione probabilistica hanno suggerito la possibile presenza di una ulteriore articolazione interna dei gruppi, con soluzioni a cinque o sei classi che tendono a suddividere alcuni cluster centrali senza modificare sostanzialmente i gruppi estremi.

Per le analisi descrittive dei profili si è scelto di proseguire con la soluzione basata sui **punteggi fattoriali**, in quanto maggiormente capace di isolare le dimensioni latenti del fenomeno. Inoltre, essendo variabili incorrelate, esse garantiscono una maggiore stabilità numerica del processo di clustering.

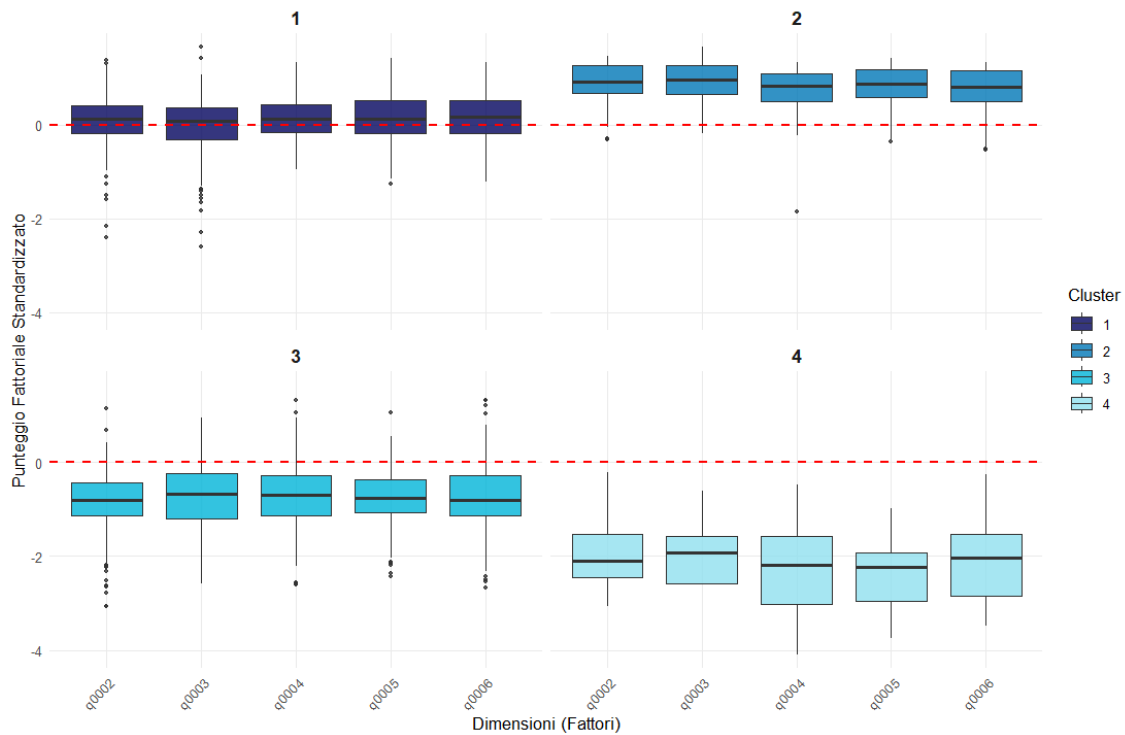


Figura 3.12: Profili cluster

cluster	q0002	0001	mean	q0002	0002	mean	q0002	0003	mean	q0002	0004	mean
1		5.26			5.13			5.49			5.32	
2		6.33			6.24			6.40			6.42	
3		4.07			3.77			4.43			4.08	
4		2.66			2.17			2.61			2.56	

Tabella 3.6: Medie per cluster

Dall'analisi delle Figure 3.11 e 3.12 si osserva la presenza di quattro cluster distinti, che riflettono differenti livelli di interesse verso la gastronomia come componente dell'esperienza di viaggio. Per verificare se le differenze tra i cluster risultassero statisticamente significative, sono state confrontate le medie dei punteggi fattoriali nei diversi gruppi mediante analisi della varianza (*ANOVA*). I risultati mostrano, per tutte le dimensioni considerate, valori di *p*-value inferiori a 0.001, indicando che i gruppi individuati differiscono in modo statisticamente significativo rispetto ai fattori analizzati.

Osservando la Tabella 3.6 e la Figura 3.12, si può affermare che i gruppi si compongono come segue:

- **Cluster 1:** rappresenta il gruppo più numeroso con il 36,3% del campione, è caratterizzato da valori prossimi alla media. Gli individui appartenenti a questo cluster mostrano un interesse moderato per la gastronomia: essa è considerata un aspetto rilevante del viaggio, ma non costituisce il fattore principale nella scelta della destinazione;
- **Cluster 2:** pari al 32,1% del campione, presenta valori nettamente superiori alla media. Questo gruppo è composto da soggetti per i quali la gastronomia assume un ruolo centrale nella pianificazione e nella scelta del viaggio, configurandosi come un vero e proprio driver decisionale. Tali viaggiatori attribuiscono grande importanza alle esperienze enogastronomiche e tendono a selezionare le destinazioni in funzione dell'offerta culinaria;
- **Cluster 3:** include il 26% degli intervistati, è caratterizzato da valori leggermente inferiori alla media. Gli individui appartenenti a questo gruppo mostrano un interesse piuttosto limitato per la gastronomia, che viene percepita come un elemento accessorio dell'esperienza turistica, secondario rispetto ad altri fattori motivazionali;
- **Cluster 4:** risulta il meno popolato con il 5,6% del campione, presenta valori molto bassi. I soggetti appartenenti a questo cluster manifestano uno scarso o nullo interesse per l'aspetto gastronomico, che non influisce in maniera significativa sulle loro scelte di viaggio né sull'esperienza complessiva della destinazione.

Avendo a disposizione le variabili demografiche fornite dal campione, risulta interessante analizzare se e in che misura i diversi gruppi individuati presentino una composizione demografica differente.

Tabella 3.7: Identikit descrittivo dei Cluster

Variabile	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Gastronomia	3.4%	9%	2%	4.07%
Motivazione	svago/riposo (72.7%)	svago/riposo (71.3%)	svago/riposo (70.7%)	svago/riposo (75%)
Genere	femmina (48.1%)	femmina (51.1%)	maschio (51.2%)	femmina (57.8%)
Età	(50,60] (30.4%)	(50,60] (30.3%)	(50,60] (24.6%)	(40,50] (29.7%)
Istruzione	diploma (51.9%)	diploma (49.5%)	diploma (48.5%)	diploma (43.8%)
Occupazione	dipendente azienda priva- ta (43%)	dipendente azienda priva- ta (47.5%)	dipendente azienda priva- ta (43.1%)	dipendente azienda priva- ta (43.8%)
Reddito	(1500,2500] (23.2%)	(1500,2500] (27%)	(1500,2500] (27.3%)	(1500,2500] (29.7%)

Come si evince dalla Tabella 3.7, i quattro cluster presentano una composizione socio-demografica apparentemente quasi omogenea, senza differenze immediatamente evidenti tra i gruppi. L'unica parziale eccezione riguarda il **cluster 3**, che si distingue dagli altri per una lieve prevalenza maschile, mentre negli altri cluster si osserva una maggiore presenza femminile, seppur con una distribuzione complessivamente equilibrata tra i generi.

Dall'analisi della tabella emerge inoltre un profilo complessivamente comune ai diversi cluster: la maggior parte dei soggetti viaggia principalmente per motivi di svago o riposo, appartiene a una fascia d'età compresa tra i 50 e i 60 anni, possiede prevalentemente un diploma di scuola secondaria superiore, è occupata come dipendente di aziende private e dichiara un reddito mensile compreso tra i 1.500 e i 2.500 euro.

Per verificare formalmente l'omogeneità delle distribuzioni delle caratteristiche socio-demografiche tra i gruppi, è stato applicato il test di indipendenza χ^2 sulle principali variabili categoriali. I risultati indicano

che, pur in presenza di differenze non immediatamente evidenti a livello descrittivo, emergono differenze statisticamente significative tra i cluster ($p < 0.05$). Ciò suggerisce che la distribuzione delle caratteristiche socio-demografiche non è completamente uniforme tra i gruppi individuati.

Al fine di valutare il ruolo delle variabili socio-demografiche nella caratterizzazione dei cluster, è stato successivamente stimato un modello di *Regressione Logistica Multinomiale*, considerando l'appartenenza al cluster come variabile dipendente.

I risultati del test del rapporto di verosimiglianza (*Likelihood Ratio Test*), riportati in Tabella 3.8, indicano che l'insieme delle variabili considerate contribuiscono in modo significativo alla distinzione tra i gruppi (p -value < 0.05), evidenziando quindi la presenza di associazioni sistematiche tra caratteristiche socio-demografiche e appartenenza ai cluster.

Tabella 3.8: Significatività delle variabili

Variabile	Chi-quadrato (χ^2)	G.d.L.	Sig. (p -value)
Motivazione viaggio (q0001)	44.85	21	0.0018
Genere (q0007)	15.47	6	0.0169
Età (q0008)	31.32	15	0.0080
Istruzione (q0009)	33.61	15	0.0039
Occupazione (q0010)	31.90	21	0.0599
Reddito (q0011)	46.44	18	0.0003

Analizzando i parametri del modello, emergono specificità che definiscono l'identità dei cluster al di là del profilo medio. Poiché le variabili indipendenti sono state inserite come fattori (scala nominale), i coefficienti stimati esprimono il contributo di ogni specifica modalità rispetto alla categoria di riferimento.

In particolare, la variabile relativa alla **motivazione del viaggio** (*q0001*) mostra associazioni differenziate con l'appartenenza ai cluster. Alcune motivazioni, come la partecipazione a *convegni* e i viaggi a carattere *religioso*, risultano fortemente associate a specifici gruppi, soprattutto ai cluster 2 e 3, indicando che la finalità del viaggio rappresenta una dimen-

sione rilevante nella segmentazione osservata. Per quanto riguarda la modalità **gastronomia e vino**, si osserva un'associazione significativa nel cluster 2 ($p = 0.0037$), suggerendo che questo tipo di motivazione è più presente in questo gruppo rispetto agli altri. Nei cluster 3 e 4, invece, l'effetto non risulta significativo, indicando una rilevanza più contenuta o assente per la segmentazione in quei gruppi. Altre motivazioni, quali la *visita a familiari e amici*, la *cultura* e il *lavoro*, non mostrano effetti statisticamente significativi nel distinguere sistematicamente i cluster.

L'**età** ($q0008$) e il **reddito** ($q0011$) mostrano un effetto significativo nel test globale, ma contribuiscono in modo limitato alla distinzione tra cluster, con la maggior parte delle fasce anagrafiche e di reddito che non presentano effetti statisticamente rilevanti rispetto alla categoria di riferimento. Analogamente, il **genere** ($q0007$) ha un impatto contenuto: pur risultando significativo nel test globale, i coefficienti non mostrano differenze marcate tra uomini e donne, fatta eccezione per una leggera predominanza osservata nel cluster 3.

Per quanto riguarda il **livello di istruzione** ($q0009$) i risultati evidenziano un'associazione positiva e significativa ($p < 0.001$) per tutte le modalità (licenza media, diploma, laurea e titolo post-laurea) in relazione al cluster 2. Al contrario, per i cluster 3 e 4 le medesime categorie presentano coefficienti fortemente negativi.

Diversamente, la variabile **occupazione** ($q0010$) non risulta statisticamente significativa al livello globale del 5%, suggerendo che la condizione occupazionale non costituisca un fattore determinante nella distinzione tra i gruppi individuati, una volta controllato per le altre caratteristiche socio-demografiche.

Nel complesso, i risultati della regressione logistica multinomiale confermano che esistono **differenze statisticamente significative** nella composizione socio-demografica dei gruppi. Tali differenze contribuiscono in modo sostanziale alla definizione dell'identità dei cluster e ne rafforzano l'interpretazione sostantiva.

3.3.2 Latent Class Analysis

In una fase successiva all'analisi di clustering gerarchico, si è proceduto all'applicazione della *Latent Class Analysis* (LCA - LC). A differenza dei metodi di clustering basati su metriche di distanza, la LCA è un approccio *model-based* fondato su modelli di probabilità finita, che consente di stimare, per ciascuna osservazione, la probabilità a posteriori di appartenenza a una specifica classe. Il modello assume che l'eterogeneità osservata nel campione possa essere spiegata dall'esistenza di un numero finito di classi latenti non direttamente osservabili, all'interno delle quali le risposte alle variabili considerate risultano condizionalmente indipendenti. In tal modo, la LCA permette di identificare strutture latenti nei dati e di interpretare le associazioni tra le variabili osservate in termini di appartenenza ai diversi cluster.

Per la costruzione dei cluster sono stati utilizzati tutti gli item appartenenti alle batterie a scala Likert. Tale approccio ha consentito di cogliere le dimensioni latenti sottostanti alle risposte fornite dai rispondenti, permettendo una segmentazione basata sui pattern di risposta. A differenza dei metodi di clustering gerarchico, che spesso trattano le scale Likert come variabili continue basandosi su distanze, la LCA permette di trattare tali variabili coerentemente con la loro natura ordinale.

In questo modello, gli item Likert sono stati considerati come indicatori categoriali ordinati: la LCA stima infatti la probabilità che un individuo scelga una specifica categoria di risposta condizionatamente all'appartenenza a una classe latente. Tali item rappresentano le variabili manifeste che definiscono il profilo dei cluster. Attraverso l'utilizzo dell'algoritmo *Expectation-Maximization* (EM), il modello è stato stimato ripetendo l'analisi cinquanta volte, ciascuna con condizioni iniziali differenti, al fine di ridurre il rischio di convergenza verso massimi locali. La soluzione riportata corrisponde al valore massimo della log-verosimiglianza osservato, suggerendo la convergenza a un massimo globale.

Per la definizione del numero ottimale di classi latenti, sono stati stimati diversi modelli di **Latent Class Analysis** con numero di classi compreso tra tre e sette, con l'inclusione e l'esclusione delle covariate (variabili descrittive).

Tabella 3.9: Indici di valutazione del numero di classi

Model	LL	BIC	AIC	Npar	df
3-cluster	-72561.89	147672.14	145847.78	362	779
3-cluster+cov	-71487.57	147974.04	145827.14	426	715
4-cluster	-70845.44	144563.06	142506.88	408	733
4-cluster+cov	-70762.22	145072.43	142532.44	504	637
5-cluster	-69932.45	143060.90	140772.90	454	687
5-cluster+cov	-69821.30	143739.69	140806.61	582	559
6-cluster	-69390.34	142300.51	139780.68	500	641
6-cluster+cov	-69246.51	143139.19	139813.01	660	481
7-cluster	-69035.02	141913.70	139162.05	546	595
7-cluster+cov	-68857.63	142910.53	139191.26	738	403

Tabella 3.10: Indici di qualità della classificazione

Model	L2	ClassErr	Entropy
3-cluster	129172.45	0.0141	0.9675
3-cluster+cov	143227.73	0.0141	0.9682
4-cluster	125739.55	0.0169	0.9661
4-cluster+cov	139777.03	0.0172	0.9665
5-cluster	123913.56	0.0253	0.9560
5-cluster+cov	137895.20	0.0239	0.9578
6-cluster	122829.35	0.0384	0.9424
6-cluster+cov	136745.60	0.0360	0.9455
7-cluster	122118.71	0.0444	0.9390
7-cluster+cov	135967.86	0.0393	0.9390

Come si evince dalle Tabelle 3.9 e 3.10, la selezione del numero ottimale di classi latenti è stata condotta valutando congiuntamente i criteri di informazione (BIC, AIC), la log-verosimiglianza e gli indici di qualità della classificazione (entropia e tasso di errata classificazione).

Analizzando i valori di log-verosimiglianza, si osserva un continuo miglioramento all'aumentare del numero di classi, raggiungendo il valore massimo nel modello a 7 classi con covariate (-68857.63). Parallelamente, i criteri di informazione tendono a diminuire in modo progressivo. È interessante notare come i modelli senza covariate presentino sistematicamente valori di BIC inferiori rispetto ai rispettivi modelli con covariate (il minimo assoluto si registra infatti per il modello a 7 classi senza covariate, pari a 141913.70), a causa della penalizzazione imposta dall'indice per il maggior numero di parametri stimati. Poiché il BIC e l'AIC continuano a decrescere all'aumentare dei gruppi senza mostrare un evidente punto di arresto (il cosiddetto "effetto gomito"), si è ritenuto inopportuno basare la scelta del numero di classi unicamente su tali

criteri, che tenderebbero a favorire soluzioni eccessivamente complesse e frammentate.

Valutando non solo gli indici citati prima ma andando a considerare gli indici di qualità della classificazione, i risultati offrono un quadro più chiaro. Il modello a 4 classi con covariate emerge come una soluzione con un buon compromesso: presenta un valore di entropia molto elevato (0.9665) e un tasso di errata classificazione estremamente contenuto (0.0172). Questi valori indicano un'alta stabilità nell'assegnazione delle osservazioni e una netta separazione tra i gruppi latenti, qualità che tendono a peggiorare nei modelli con un numero di classi superiore (dal modello a 5 classi in poi si nota un progressivo calo dell'entropia e un aumento dell'errore).

Pertanto, volendo ottenere un buon adattamento statistico ai dati con l'alta qualità della classificazione, la parsimonia parametrica e la chiarezza interpretativa, il modello a quattro classi con covariate è stato selezionato come soluzione finale.

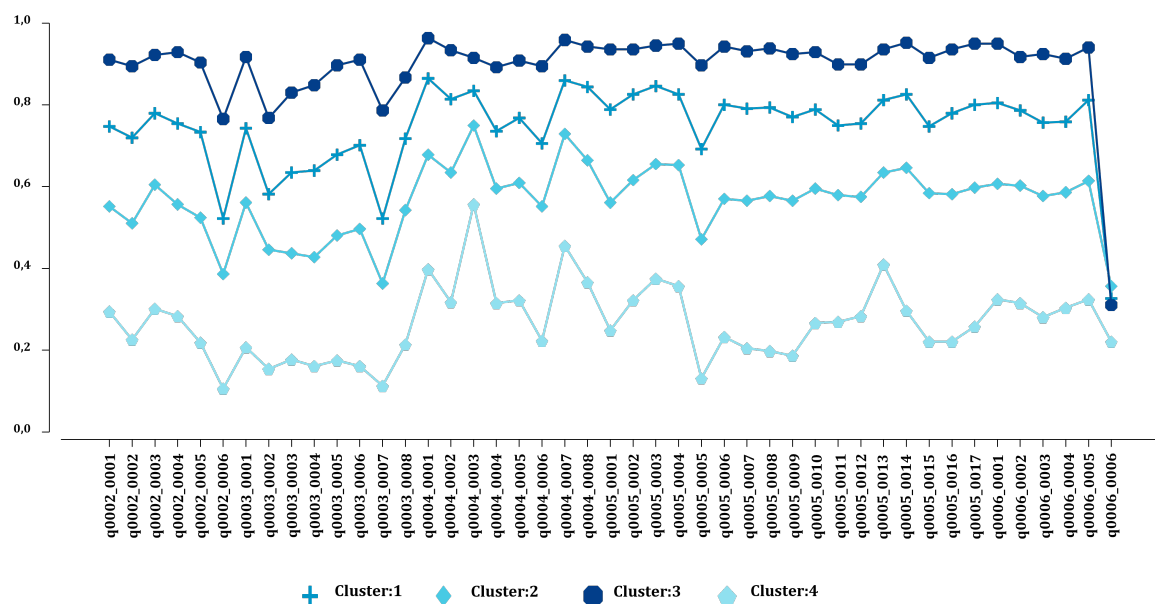


Figura 3.13: Clustering con metodo LCA

La soluzione a quattro classi latenti con covariate suddivide il campione come segue:

- **Cluster 1:** risulta essere il raggruppamento più popolato, rappresentando il 41% del campione osservato (468 soggetti). I viaggiatori di questo gruppo tendono ad assegnare punteggi medi, evidenziando un discreto interesse per la gastronomia nelle loro mete di viaggio, sebbene questa non ne costituisca la motivazione principale;
- **Cluster 2:** comprende il 31% dei rispondenti (355 soggetti). I membri di questa classe presentano punteggi leggermente al di sotto della media generale, suggerendo un interesse marginale o secondario per l'aspetto culinario;
- **Cluster 3:** costituito dal 22% del campione (246 soggetti), raggruppa gli individui maggiormente interessati alla gastronomia. I punteggi sistematicamente superiori alla media per quasi tutti gli item dimostrano che questi viaggiatori scelgono la propria meta turistica attribuendo forte rilevanza all'aspetto culinario e ai prodotti tipici del territorio;
- **Cluster 4:** rappresenta solo il 6% del totale (72 soggetti) ed è il gruppo meno numeroso. Include i turisti per i quali l'aspetto gastronomico è del tutto irrilevante nella scelta della destinazione.

Osservando la numerosità e la composizione dei singoli profili, emerge chiaramente come la ricerca della cultura del cibo rivesta un interesse generale e rilevante per la maggior parte del campione (in particolare per i Cluster 1 e 3) nella scelta delle destinazioni di viaggio.

Tabella 3.11: Profilo descrittivo delle Classi Latenti (LCA)

Variabile	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Gastronomia	3.47%	2.32%	11.62%	4.09%
Motivazione	svago/riposo (72.89%)	svago/riposo (70.68%)	svago/riposo (70.35%)	svago/riposo (76.69%)
Genere	femmina (48.34%)	maschio (52.03%)	femmina (51.57%)	femmina (59.15%)
Età	(50,60] (27.86%)	(50,60] (25.13%)	(50,60] (35.65%)	(40,50] (30.99%)
Istruzione	diploma (50.28%)	diploma (49.44%)	diploma (51.24%)	diploma (43.07%)
Occupazione	dipendente azienda priva- ta (42.20%)	dipendente azienda priva- ta (43.32%)	dipendente azienda priva- ta (51.44%)	dipendente azienda priva- ta (42.05%)
Reddito	(1500,2500] (26.46%)	(1500,2500] (25.73%)	(2500,3500] (25.50%)	(1500,2500] (30.55%)

Osservando la Tabella 3.11, si rileva come la composizione demografica dei quattro cluster non presenti differenze marcate. Da un primo sguardo, emergono tuttavia alcune peculiari distinzioni: il **Cluster 2** si caratterizza per una prevalenza di rispondenti di sesso maschile, il **Cluster 3** presenta un livello di reddito mediamente superiore rispetto agli altri gruppi, mentre il **Cluster 4** si differenzia per una classe di età leggermente più giovane (40-50 anni).

Un'analisi della prima riga della Tabella 3.11, relativa alla variabile motivazione con focus sulla modalità "*Gastronomia*", offre spunti particolarmente interessanti. In linea con le aspettative, il **Cluster 3** (precedentemente identificato per i punteggi superiori alla media in tutti gli item) registra la frequenza nettamente più elevata di selezione per questa modalità (11.62%). Un dato ancor più degno di nota, e in apparente contraddizione con quanto detto precedentemente, riguarda il **Cluster 4**: sebbene fosse emerso come il gruppo con i punteggi di interesse gastronomico più bassi, esso presenta una percentuale di selezione per la voce "*Gastronomia*" (4.09%) superiore a quella riscontrata nei Cluster 1 (3.47%) e 2 (2.32%).

Per approfondire se, nel modello a quattro classi latenti, le variabili demografiche e comportamentali influenzino significativamente la probabilità di appartenenza ai cluster, è stato analizzato l'output relativo alle covariate. La significatività statistica è stata valutata mediante il **test di Wald** verificando l'ipotesi nulla che i parametri della covariata siano uguali a zero tra le diverse classi latenti.

Analizzando i p -value, emerge che solo due variabili risultano statisticamente significative nel determinare la composizione dei quattro gruppi:

- **Titolo di Studio (q0009):** Risulta essere il predittore più robusto del modello, con un valore di $p < 0.001$ ($Wald = 50.80$). Ciò indica che il livello di istruzione non è solo un descrittore socio-demografico, ma un fattore discriminante fondamentale nella segmentazione del campione.
- **Motivazione (q0001):** Presenta un valore di $p = 0.026$ ($Wald = 35.32$), confermando che la motivazione di scelta differenzia in modo significativo i profili dei cluster.

Al contrario, le variabili relative al **genere (q0007)**, all'**età (q0008)**, al **lavoro (q0010)** e al **reddito (q0011)** non mostrano significatività statistica ($p > 0.05$). Questo suggerisce che un individuo appartiene a una determinata classe latente più in base alle proprie motivazioni e alla propria istruzione.

Infine, un aspetto rilevante emerge dall'analisi dei coefficienti individuali. Nonostante la forte significatività globale delle variabili $q0009$ e $q0001$, l'osservazione dei test per le singole modalità (valori z) rivela che nessuna categoria presa singolarmente risulta significativa. Questo indica che la capacità discriminante non è trainata da una singola modalità, ma dall'intera variabile.

3.3.3 Confronto tra Clustering Gerarchico e LCA

Confrontando le analisi precedenti, si nota come i due metodi concordino nell'identificare i soggetti "mediamente interessati" e quelli "per nulla interessati", assegnandoli rispettivamente alle classi 1 e 4. Diversamente, per quanto riguarda i cluster 2 e 3, gli algoritmi invertono le etichette numeriche assegnate ai profili: nel metodo gerarchico il cluster 2 rappresenta i "molto interessati" e il 3 i "poco interessati", mentre l'LCA effettua l'assegnazione opposta. Poiché l'ordine numerico nei modelli

di clustering è del tutto arbitrario, per rendere le due partizioni direttamente confrontabili è stato necessario allinearne la nomenclatura. Di conseguenza, si è proceduto a ricodificare l'output dell'LCA, scambiando le etichette delle classi 2 e 3 in modo da farle corrispondere ai medesimi profili concettuali individuati dal clustering gerarchico.

Dal confronto tra il *clustering Gerarchico* e la *Latent Class Analysis* emerge una sostanziale coerenza nella segmentazione dei soggetti, sebbene siano presenti alcune discrepanze nell'assegnazione di specifiche unità ai cluster. Tali differenze suggeriscono l'esistenza di una struttura latente non completamente intercettata dall'approccio basato esclusivamente su metriche di distanza, ma colta dal modello probabilistico della LCA, che consente un'assegnazione meno rigida delle osservazioni alle classi.

Al fine di quantificare il grado di concordanza tra le partizioni ottenute con i due metodi, è stato calcolato l'indice di Rand [Rand, 1971]. Tale metrica valuta la somiglianza tra due soluzioni di clustering calcolando il rapporto tra il numero di coppie di osservazioni classificate nel medesimo cluster (o in cluster differenti) da entrambi gli algoritmi e il numero totale di coppie possibili. Poiché questo indice assume valori compresi tra 0 (totale discordanza) e 1 (perfetta sovrapposizione), il valore di 0,496 ottenuto indica un livello di accordo moderato. I due metodi condividono l'assegnazione di una parte consistente delle osservazioni, pur differendo per alcune unità che vengono collocate in classi differenti. Questo esito supporta l'utilizzo della LCA come approccio complementare al clustering gerarchico, in quanto in grado di cogliere componenti latenti sottostanti e di fornire una rappresentazione più flessibile e probabilistica dell'eterogeneità del campione.

Tabella 3.12: Matrice di confronto tra Clustering Gerarchico e LCA

Gerarchico	LCA			
	Classe 1	Classe 2	Classe 3	Classe 4
Cluster 1	324	3	87	0
Cluster 2	122	243	0	1
Cluster 3	22	0	266	9
Cluster 4	0	0	2	62

La Tabella 3.12 riporta la matrice di confronto tra le due soluzioni di

clustering, evidenziando una prevalente corrispondenza lungo la diagonale principale, a conferma della sostanziale sovrapposizione tra i cluster individuati dai due metodi.

3.3.4 Modello di Regressione LC

Al fine di indagare quali fattori determinino la motivazione gastronomica dei turisti, è stato stimato un modello di *Regressione a Classi Latenti* (Latent Class Regression). Come variabile dipendente è stato selezionato l'item **q0002_0001**, che rileva tramite scala Likert il grado di importanza della "*gastronomia nella motivazione di viaggio*". Per la stima di tale modello, sono state utilizzate come predittori le restanti variabili su scala Likert, mentre le caratteristiche socio-demografiche sono state inserite come covariate.

Poiché questa analisi risponde a un differente quesito di ricerca, le classi latenti individuate in questa fase non coincidono con i quattro cluster estratti in precedenza. Nel primo modello, infatti, l'obiettivo esplorativo era comprendere come si segmentassero i turisti in base al loro comportamento generale e al profilo socio-demografico. In questo secondo approccio, invece, l'attenzione si sposta sull'eterogeneità dei processi decisionali: l'obiettivo è identificare l'esistenza di gruppi latenti all'interno dei quali pesi e driver differenti guidano l'importanza attribuita alla gastronomia nella scelta della meta turistica.

Tabella 3.13: Predittori principali (Wald > 25) e Covariate del Modello a Classi Latenti

Variabile	Classe 1	Classe 2	Classe 3	Classe 4	p-value	Wald
Predittori Principali (Scale Likert)						
Ricerca di esperienze nella meta scelta	1.5964*	1.0579*	2.8669*	0.1418	< 0.001	126.4119
Esperienza gastronomica e soddisfazione	0.7226*	1.3774*	1.9443*	1.5532*	< 0.001	82.8472
Cultura e patrimonio nella scelta	0.0510	2.1423*	-0.4848	7.0858*	< 0.001	61.5802
Presenza di ristoranti famosi	-0.1555	-0.4770*	0.4763	1.8022*	< 0.001	26.2495
Prodotti agricoli	0.3679*	0.3779	-1.1846*	0.1082	< 0.001	27.6380
Agroindustrie	-0.2100	0.7652*	1.0032*	-0.2947	< 0.001	25.8893
Negozi di alimentari tradizionali	-0.0746	0.6893*	-0.8471*	1.2010*	< 0.001	30.3805
Qualità dei piatti	0.2206	0.9167*	0.6742	1.8675*	< 0.001	25.7314
Varietà dei piatti	-0.3193*	2.2892*	1.2809*	-0.6849	< 0.001	53.2366
Strutture ristorative	0.4005*	1.0003*	-0.9905*	1.7805*	< 0.001	38.1146
Gastronomia tradizionale	0.1349	-1.2855*	-0.4672	-2.5212*	< 0.001	31.0954
Accresce conoscenza culinaria	0.7371*	-1.0925*	1.6782*	0.2877	< 0.001	59.4366
Scoperta altri elementi culturali	-0.3720*	-2.1728*	0.5380	2.3252*	< 0.001	43.9753
Scoprire cose nuove	0.1769	0.9168*	-0.8663*	-2.1825*	< 0.001	26.6444
Esperienze gastronomiche uniche	0.0794	1.3538*	0.8092*	-0.9367*	< 0.001	34.5214
Momenti piacevoli con famiglia/amici	0.4181*	-1.4258*	-1.3042*	0.1151	< 0.001	38.9144
Consigliare la meta ad altri	0.2479*	0.3486	-0.9186*	-2.5961*	< 0.001	28.3391
Covariate Socio-Demografiche						
Motivazione					0.0260	35.3319
convegni	-8.2547	0.0327	5.2497	2.9723		
cultura	-1.5126	-3.0376	4.1357	0.4144		
gastronomia e vino	-2.7763	12.2683*	-4.3592	-5.1328		
lavoro	3.1078	10.3075*	-7.9369	-5.4785		
religioso	-0.9345	-12.9521	8.3243	5.5624		
studio	2.6201	-4.0693	-0.5283	1.9775		
svago/riposo	2.7534	-2.8719	1.3903	-1.2718		
visita a famigliari/amici	4.9968	0.3223	-6.2756	0.9565		
Genere					0.0110	16.4581
femmina	-3.0969	2.8006	0.0324	0.2639		
maschio	-2.9867	0.3123	1.8105	0.8639		
Età					<0.001	45.4696
under 30	2.0589	4.3795	5.3516	-11.7900		
(30, 40]	-0.1833	-12.6283	8.2346	4.5770		
(40, 50]	1.9386	1.2272	-4.5189	1.3531		
(50, 60]	-0.7441	-1.9555	2.5117	0.1879		
over 60	3.2647	7.3054	-10.9066	0.3366		
Istruzione					<0.001	48.8551
licenza media	2.3419	5.8344	-1.0987	-7.0775		
diploma	-0.1502	1.8104	-0.0140	-1.6462		
laurea	-0.5359	2.1578	-0.8511	-0.7708		
titolo post-laurea	-6.4772	-14.2773	10.4421	10.3124		
Reddito					0.0059	36.6165
[0,1000]	2.7452*	8.5547*	-2.2222	-9.0777*		
(1000, 1500]	-4.6854*	2.2983	3.1307*	-0.7436		
(1500, 2500]	-1.9080*	-3.1437*	3.7525*	1.2991		
(2500, 3500]	-2.6615*	2.3179	1.5032	-1.1597		
(3500, inf)	0.6088	-1.2147	-3.6077*	4.2137*		

(*) p-value < 0.05 (ovvero statistica di Wald $z \geq |1.96|$ per il singolo coefficiente)

Al fine di garantire una maggiore chiarezza e focalizzare l'interpretazione

sui fattori maggiormente discriminanti, la Tabella 3.13 riporta esclusivamente i predittori che presentano un impatto più marcato sul modello, ovvero sono stati selezionati i predittori che hanno ottenuto un valore della statistica Wald superiore a 25. Differentemente, le covariate socio-demografiche sono state mantenute integralmente per definire i profili di appartenenza.

L'analisi dei coefficienti di regressione (β) rivela profondi contrasti tra le quattro classi latenti in merito a ciò che guida la motivazione gastronomica durante la scelta della meta turistica. Incrociando questi risultati con le covariate socio-demografiche, è possibile delineare i seguenti profili decisionali:

- **Classe 1:** questa classe si caratterizza per coefficienti più vicini allo 0 (non presenta nè picchi negativi nè positivi) rispetto agli altri raggruppamenti, rappresentando un approccio più generalista. La motivazione gastronomica è sostenuta in modo significativo dalla *Ricerca di esperienze nella meta scelta* ($\beta = 1,60$), dall'*Esperienza gastronomica e soddisfazione* ($\beta = 0,72$) e dal desiderio di *Accrescere la conoscenza culinaria* ($\beta = 0,74$). Non emergono covariate socio-demografiche particolarmente discriminanti per questo gruppo, suggerendo che si tratti di un turista trasversale, per il quale il cibo è un importante veicolo di soddisfazione generale del viaggio.
- **Classe 2:** questa classe presenta delle caratteristiche interessanti. Le covariate indicano in modo significativo che questi soggetti dichiarano esplicitamente "gastronomia e vino" come principale motivazione di viaggio ($\beta = 12,27$) oppure viaggiano per lavoro ($\beta = 10,31$). Dal punto di vista dei predittori, la loro motivazione gastronomica è spinta dalla *Cultura e patrimonio nella scelta* ($\beta = 2,14$) e dalla *Varietà dei piatti* ($\beta = 2,29$). Tuttavia, se si osservano i coefficienti di *Accrescere la conoscenza culinaria* ($\beta = -1,09$) o la *Scoperta di altri elementi culturali* ($\beta = -2,17$), si può notare che queste hanno un impatto fortemente negativo.
- **Classe 3:** questa classe basa la propria motivazione gastronomica prevalentemente sulla *Ricerca di esperienze nella meta scelta* ($\beta = 2,87$), sulla *Esperienza gastronomica e soddisfazione* ($\beta = 1,94$) e sulla volontà di *Accrescere la conoscenza culinaria* ($\beta = 1,68$). Al contrario, si evidenzia che questo gruppo non è tanto interessato a vivere il cibo come un momento di aggregazione sociale, infatti, si

osserva un impatto fortemente negativo e significativo della variabile relativa ai *Momenti piacevoli con famiglia/amici* ($\beta = -1,30$). Si tratta di turisti che vivono la gastronomia come un'esperienza formativa, intellettuale e individuale.

- **Classe 4:** per questo gruppo, la motivazione gastronomica è trainata in modo considerevole dalla valutazione della *Cultura e patrimonio nella scelta* ($\beta = 7,09$) e dal desiderio di *Scoperta di altri elementi culturali* ($\beta = 2,33$). Curiosamente, in questa classe elementi come la *Gastronomia tradizionale* ($\beta = -2,52$) o *Scoprire cose nuove* in ambito culinario ($\beta = -2,18$) presentano coefficienti negativi. Questo indica che, per la Classe 4, il cibo non è un'attrazione a sé stante, ma piuttosto un aspetto fondamentale dell'esplorazione storico-culturale della destinazione.

In sintesi, questa analisi ha permesso di dimostrare che la motivazione gastronomica non è uguale per tutti. Dietro un generico "interesse per il cibo" si nascondono infatti motivazioni molto diverse: si va dal desiderio di imparare e studiare la cucina locale (Classe 3), all'uso del cibo come mezzo per esplorare la cultura del territorio (Classe 4), fino alla semplice ricerca della buona tavola (Classi 1 e 2).

Conclusioni

Il presente studio si è posto l'obiettivo di indagare se la popolazione italiana consideri la cultura gastronomica come un fattore determinante nella definizione della propria destinazione di viaggio. L'indagine, condotta su un campione di cittadini maggiorenni attraverso un questionario strutturato, ha permesso di mappare le motivazioni legate al turismo-enogastronomia e definire i profili dei viaggiatori.

Le analisi applicate, attraverso l'integrazione del *clustering gerarchico* (basato sia sulle medie delle batterie sia sui punteggi fattoriali) e della *Latent Class Analysis* (LCA), hanno evidenziato la presenza di quattro profili distinti di viaggiatori. Nello specifico:

- **Cluster 1 (Interesse Moderato):** Risulta essere il raggruppamento più popolato in entrambi i metodi. Vi appartengono soggetti che esprimono valutazioni prossime alla media. Per questi individui esiste un interesse per la gastronomia, che tuttavia non rappresenta la motivazione primaria nella scelta della meta.
- **Cluster 2 (Gastronomi Puri):** Comprende i soggetti che ricercano attivamente la cucina locale nella scelta della destinazione. Questo gruppo è caratterizzato da punteggi superiori alla media in entrambi i metodi di clustering.
- **Cluster 3 (Interesse Marginale):** Raggruppa individui poco attratti dall'aspetto gastronomico, con punteggi che si attestano leggermente al di sotto della media generale.
- **Cluster 4 (Indifferenti):** Definisce il segmento minoritario per il quale l'offerta gastronomica risulta del tutto irrilevante nel processo decisionale turistico.

Un contributo metodologico rilevante del lavoro risiede nel confronto critico tra i due approcci. Sebbene entrambi gli algoritmi concordino sulla struttura a quattro gruppi, emerge una mancata corrispondenza

nella classificazione delle osservazioni. Tale divergenza evidenzia come il clustering gerarchico, basato su metriche di distanza, possa non cogliere appieno le particolarità dei soggetti. Al contrario, l'approccio probabilistico della *Latent Class Analysis* si è dimostrato superiore nel catturare i pattern di risposta latenti. La LCA, trattando le variabili coerentemente con la loro natura ordinale e operando attraverso una logica *model-based*, fornisce una segmentazione più raffinata e una stabilità maggiore nell'assegnazione ai cluster.

L'analisi delle covariate in questo primo modello ha rivelato un dato cruciale: le tradizionali variabili socio-demografiche (genere, età, occupazione, reddito) non esercitano un'influenza statisticamente significativa nel differenziare i turisti. Al contrario, il **livello di istruzione** e la **motivazione primaria di viaggio** sono risultati gli unici predittori capaci di discriminare l'appartenenza ai cluster, dimostrando come la propensione verso il turismo enogastronomico sia legata al background culturale dell'individuo piuttosto che aspetti anagrafici e economici.

Per comprendere maggiormente come si formula il processo decisionale del turista, lo studio ha compiuto un ulteriore passo analitico stimando un modello di **Regressione a Classi Latenti** (*Latent Class Regression*). L'obiettivo, in questa seconda fase, non era più segmentare i turisti per profilo generale, ma capire quali driver definiscono l'importanza attribuita alla gastronomia. Questo approccio ha permesso di identificare quattro ulteriori classi latenti che descrivono approcci motivazionali diversi:

- Un **approccio generalista** (Classe 1), tipico di un turista per il quale il cibo è solo un veicolo di soddisfazione complessiva e un'occasione per accumulare esperienze positive.
- Un **approccio trainato da motivazioni specifiche** (Classe 2), riconoscibile soprattutto in chi viaggia per lavoro o esplicitamente per vino/gastronomia, focalizzato sulla varietà e sull'impatto culturale del cibo, ma altrettanto disinteressato all'apprendimento di nuove conoscenze culinarie.
- Un **approccio intellettuale e formativo** (Classe 3), che vive la gastronomia come un'esplorazione individuale finalizzata ad accrescere la propria conoscenza personale. Infatti, si dissocia quasi del tutto dall'idea che il pasto sia un momento di convivialità sociale.

- Un **approccio storico-culturale** (Classe 4), dove l'interesse per la cucina viene usato con la finalità di esplorare e comprendere a fondo il patrimonio e le tradizioni della destinazione scelta.

In sintesi, questa seconda analisi ha permesso di dimostrare che dietro la facciata di un generico "interesse per il cibo" si nascondono processi decisionali eterogenei e sfaccettati.

In conclusione, l'evidenza empirica suggerisce che il cibo rappresenta un elemento identitario di inestimabile valore per la maggioranza dei viaggiatori italiani. Dunque, per intercettare e fidelizzare il turista non è sufficiente segmentare il mercato basandosi su età o reddito, né proporre un'offerta gastronomica standardizzata. Occorre, piuttosto, sviluppare linee di prodotto differenziate (introduzione di percorsi formativi per i turisti più "intellettuali", esperienze integrate col patrimonio per i turisti "storico-culturali") puntando su una comunicazione capace di dialogare con lo spessore culturale e le reali leve decisionali dei viaggiatori.

Bibliografia

- Roberta Garibaldi. Rapporto sul turismo enogastronomico italiano 2020. In *Associazione Italiana TURISMO ENOGASTRONOMICO*, 2020.
- Erica Croce Giovanni Perri. *Il turismo enogastronomico: progettare, gestire, vivere l'integrazione tra cibo, viaggio e territorio*. FrancoAngeli, 2018.
- Juan M. Berbel-Pineda, Beatriz Palacios-Florencio, José M. Ramírez-Hurtado, and Luna Santos-Roldán. Gastronomic experience as a factor of motivation in the tourist movements. *International Journal of Gastronomy and Food Science*, 18:100171, 2019. ISSN 1878-450X. doi: <https://doi.org/10.1016/j.ijgfs.2019.100171>. URL <https://www.sciencedirect.com/science/article/pii/S1878450X19300460>.
- Qian Chen and Rong Huang. Understanding the importance of food tourism to chongqing, china. *Journal of Vacation Marketing*, 22(1):42–54, 2016. doi: 10.1177/1356766715589427. URL <https://doi.org/10.1177/1356766715589427>.
- Melanie Revilla, Willem E. Saris, and Jon A. Krosnick. Scelta del numero di categorie nelle scale di accordo-disaccordo. *Sociological Methods & Research*, 43:73 – 97, 2014. URL <https://api.semanticscholar.org/CorpusID:33350432>.
- Jon A. Krosnick. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5:213–236, 1991. URL <https://doi.org/10.1002/acp.2350050305>.
- Philip Podsakoff, Scott MacKenzie, Jeong-Yeon Lee, and Nathan Podsakoff. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88:879–903, 10 2003. doi: 10.1037/0021-9010.88.5.879.

- Jeffrey Moore, Linda Stinson, and Edward Welniak. Income measurement error in surveys: A review. *Journal of Official Statistics*, 16:331, 01 2000.
- Sara Dolnicar. *Market Segmentation in Tourism*, 2008. doi: 10.1079/9781845933234.0129.
- R. Albano; D. Molino. *Quaderni di Ricerca*. Aracne editrice S.r.l., 2011. URL <http://hdl.handle.net/2318/806>.
- Maria Alaimo, LEONARDO SALVATORE; Barbato. L'analisi fattoriale. *Ragionando di sviluppo locale: una lettura nuova di tematiche antiche*, 2020.
- Henry F. Kaiser. An index of factorial simplicity. *Psychometrika*, 39(1): 31–36, 1974. doi: 10.1007/BF02291575.
- J.L. Horn. A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30:179–185, 1965. URL <https://doi.org/10.1007/BF02289447>.
- R.P. McDonald. Test theory: A unified treatment. *Psychology Pres*, 1999. URL <https://doi.org/10.4324/9781410601087>.
- Thomas Dunn, Thom Baguley, and Vivienne Brunnsden. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105, 08 2013. doi: 10.1111/bjop.12046.
- P.H.A. Sneath, P.H.A. Sneath, R.R. Sokal, and U.R.R. Sokal. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. A Series of books in biology. W. H. Freeman, 1973. ISBN 9780716706977. URL <https://books.google.it/books?id=iWwCQgAACAAJ>.
- Richard A. Johnson and Dean W. Wichern. *Applied multivariate statistical analysis*. Prentice Hall, Upper Saddle River N.J, sixth edition, pearson international edition. edition, 2007. ISBN 9780135143506.
- Robert R Sokal and Charles D Michener. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 38:1409–1438, 1958.
- Joe H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58 (301):236–244, 1963. doi: 10.1080/01621459.1963.10500845. URL

<https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500845>.

Jeroen K. Vermunt. 7. multilevel latent class models. *Sociological Methodology*, 33(1):213–239, 2003. doi: 10.1111/j.0081-1750.2003.t01-1-00131.x. URL <https://doi.org/10.1111/j.0081-1750.2003.t01-1-00131.x>.

Jeroen K. Vermunt John P. Madura Jay Magidson. *Latent Class Analysis*. SAGE Publications Ltd, 2020. doi: <http://dx.doi.org/10.4135/9781526421036>.

P. F. Lazarsfeld and N. W. Henry. *Latent Structure Analysis*. Houghton Mifflin, Boston, 1968.

{Linda M.} Collins and {Stephanie T.} Lanza. *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. John Wiley and Sons Inc., United States, January 2010. ISBN 9780470228395. doi: 10.1002/9780470567333. Publisher Copyright: © 2010 John Wiley & Sons, Inc. All rights reserved.

L. A. Goodman. The analysis of systems of qualitative variables when some of the variables are unobservable: Part i. a modified latent structure approach. *American Journal of Sociology*, 79:1179–1259, 1974. doi: 10.1086/225676. URL <https://doi.org/10.1086/225676>.

Jay Magidson and Jeroen Vermunt. Latent class models for clustering: A comparison with k-means. *Can J Marketing Research*, 20, 11 2001.

Michel Wedel and Wagner Kamakura. *Market Segmentation: Conceptual and Methodological Foundations*, volume 8. 01 2000. ISBN 978-1-4613-7104-5. doi: 10.1007/978-1-4615-4651-1.

Ulf Olsson. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4):443–460, 1979. doi: 10.1007/BF02296207.

M. S. Bartlett. The effect of standardization on a 2 approximation in factor analysis. *Biometrika*, 38(3/4):337–344, 1951. ISSN 00063444, 14643510. URL <http://www.jstor.org/stable/2332580>.

Steven P. Reise. The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5): 667–696, 2012. doi: 10.1080/00273171.2012.715555. URL <https://doi.org/10.1080/00273171.2012.715555>. PMID: 24049214.

William M Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336): 846–850, 1971.