

Categorie escluse o meno dal mercato delle vacanze: un'analisi di segmentazione del caso italiano

Giovanni Tonini¹

Dipartimento di Metodi Quantitativi per l'Economia, Università di Milano-Bicocca

Riassunto: Il presente lavoro si basa su una applicazione dell'analisi di segmentazione al mercato delle vacanze degli italiani, i cui obiettivi sono: selezionare i predittori più rilevanti del rischio o probabilità di non fare vacanza; classificare le modalità di tali predittori in due gruppi caratterizzati da una alta/bassa percentuale di "non vacanzieri"; individuare le categorie sociali più/meno a rischio di non andare in vacanza. I risultati di tale applicazione presentano un interesse non solo in sé, ma anche come base per ulteriori sviluppi degli studi sul turismo.

Parole chiave: "Non vacanzieri", analisi di segmentazione, procedura *CART*.

1. Introduzione

La propensione ad andare in vacanza varia in funzione diretta del grado di sviluppo socio-economico di una popolazione. Infatti, al crescere di tale sviluppo aumenta pure la frazione di vacanzieri, ovvero diminuisce la quota di "non vacanzieri", per cui quest'ultima può essere considerata come un indicatore di "malessere sociale".

Riveste quindi un indubbio interesse l'individuazione delle categorie sociali maggiormente a rischio di non fare vacanza, sia per caratterizzare le fasce sociali più "deboli" a questo proposito, sia per identificare gli utenti potenziali del mercato turistico (su cui intervenire per incrementare il numero dei vacanzieri).

A tal fine sembra appropriato ricorrere ad un'analisi di segmentazione (Fabbris, 1997, cap. 9), inizialmente di tipo binario, con la quale si cercherà di rispondere ai seguenti quesiti:

- 1) Quali sono i predittori più rilevanti del rischio o probabilità di non andare in vacanza?

¹ Il presente lavoro è stato finanziato nell'ambito del progetto "La ricerca di determinanti del rischio mediante analisi di segmentazione di campioni", cofinanziato dal MIUR. Coordinatore nazionale è Luigi Fabbris, coordinatore del gruppo di Milano-Bicocca è M. Civardi.

- 2) Come vanno classificate (dicotomizzate) le modalità di tali predittori per massimizzare l'omogeneità interna ai gruppi di individui, ovvero l'eterogeneità tra i gruppi stessi?
- 3) Quali sono le categorie sociali più/meno a rischio di non fare vacanza?

Prima di dare una risposta ai precedenti interrogativi con l'analisi di segmentazione, si cercherà di farlo con un approccio non multivariato, in modo da avere anche degli utili elementi di confronto con i successivi sviluppi. Pertanto, dopo una breve presentazione dei dati e delle variabili utilizzate, si procederà all'analisi univariata e bivariata delle quote di "non vacanzieri" (par. 2.1) e dei motivi della mancata vacanza (par. 2.2); quindi, si passerà all'applicazione della procedura di segmentazione, illustrandone brevemente sia alcuni aspetti metodologici (par. 3), sia i principali risultati ottenuti (par. 4).

2. Variabile dipendente, predittori e analisi iniziali dei dati

Per soddisfare le suddette esigenze conoscitive si sono utilizzati i dati elementari sulle vacanze degli italiani relativi all'Indagine multiscopo sulle famiglie – Aspetti della vita quotidiana, anno 1997 (ISTAT, 1999, cap. 10), dati disponibili sotto forma di *file standard*. In tale contesto, la variabile dipendente o "criterio" è costituita dalla frazione (percentuale) di "non vacanzieri"² sul totale della popolazione di riferimento, la quale percentuale rappresenta un indicatore *proxy* del rischio o probabilità di non andare in vacanza.

Per converso, le variabili esplicative utilizzabili a fini di segmentazione sono 11 e comprendono: 2 variabili di contesto (ripartizione geografica di residenza e tipo di comune o dominio), 4 di natura demografica (sesso, età, stato civile e numero di componenti della famiglia attuale), e 5 di tipo socio-economico (titolo di studio, fonte principale di reddito, condizione professionale unica o prevalente, posizione nella professione e ramo di attività economica)³.

² I "non vacanzieri" sono coloro che, nei 12 mesi precedenti l'indagine, non si sono mai recati in vacanza per un periodo di almeno 4 notti consecutive. Per informazioni di natura congiunturale sulle vacanze "brevi" (con al più 3 pernottamenti consecutivi) vedasi ISTAT (1998). In un prossimo lavoro si cercherà, per quanto possibile, di effettuare un'analisi comparata tra i dati qui utilizzati e quelli ricavabili dall'indagine trimestrale *CATI* sui viaggi degli italiani appena citata.

³ Il fatto che le variabili di cui sopra siano caratterizzate da un diverso numero di modalità elementari non sembra influire più di tanto sui risultati del processo di segmentazione. Infatti, tali risultati non variano apprezzabilmente se, ad esempio, anziché operare con tutte e 102 le età annuali, queste ultime vengono aggregate nelle 13 classi pluriennali previste dall'ISTAT nelle sue pubblicazioni, e riportate anche nella successiva Tabella 1.

2.1 *Analisi preliminari dell'incidenza relativa dei "non vacanzieri"*

Prima di applicare le tecniche di segmentazione ai dati di interesse sembra opportuno condurre un'analisi preliminare dei dati stessi, con l'obiettivo di cogliere due aspetti rilevanti ai fini dei successivi sviluppi:

- 1) – il grado di connessione esistente tra la variabile dipendente e ciascuna delle potenziali variabili esplicative;
- 2) – le variazioni delle quote di "non vacanzieri" al mutare delle modalità dei possibili predittori.

Nel primo caso, sulla base di un indice normalizzato di connessione⁴, si sono individuati i fattori che più/meno influiscono sul rischio di non andare in vacanza; nel secondo caso, invece, si è cercato di classificare euristicamente le modalità dei predittori, distinguendole a seconda che la corrispondente quota di "non vacanzieri" fosse alta o bassa.

- 1) – Dalla graduatoria che si ottiene ordinando in senso decrescente i valori dell'indice di connessione risulta che i predittori più connessi con la variabile criterio sono (regione a parte): la ripartizione geografica (0,285), il titolo di studio (0,282) e la posizione nella professione (0,246). All'estremo opposto, i predittori meno connessi sono: il sesso (0,008), il numero di componenti della famiglia attuale (0,115) e lo stato civile (0,125, ovvero 0,058 senza la modalità "vedovo/a").

Si tratterà in seguito di verificare se tale graduatoria trovi conferma anche con l'analisi multivariata di segmentazione la quale, a differenza degli indici di connessione, tiene congiuntamente conto di tutte le variabili esplicative via via considerate.

- 2) – Con riferimento al secondo obiettivo delle analisi preliminari, si possono individuare tre diversi gruppi di variabili esplicative a seconda che le quote di "non vacanzieri":

- a) – varino in modo monotono (o quasi) al mutare delle corrispondenti modalità;
- b) – non abbiano un andamento monotono, ma siano suddivisibili in almeno due gruppi, nettamente distinti tra loro;
- c) – non presentino differenze significative passando da una modalità all'altra.

2.a) – Al primo raggruppamento appartengono, in senso lato, le variabili di contesto (ripartizione geografica e tipo di comune o dominio), nonché il titolo di studio e, seppur parzialmente, la posizione nella professione, l'età e il numero di componenti della famiglia attuale (vedasi Tabella 1).

⁴ Come misura normalizzata di connessione si è utilizzato l'indice medio di contingenza di Cramér, il quale, nel caso in cui una delle due variabili considerate sia dicotomica, coincide con l'indice ϕ di Pearson (Leti, 1983, pp. 766-768).

Tabella 1 – Indice del grado di connessione (valori in neretto) e percentuale di “non vacanzieri” per ciascuna modalità dei potenziali predittori. Italia, anno 1997

Regione	0,309	Sesso	0,008
Piemonte e Valle d’Aosta	40,0%	Maschio	51,7%
Lombardia	30,3%	Femmina	52,5%
Trentino-Alto Adige	41,4%	Classe d’età	0,212
Veneto	45,5%	00-05 anni	48,6%
Friuli-Venezia Giulia	53,6%	06-10 anni	43,3%
Liguria	53,1%	11-14 anni	43,2%
Emilia-Romagna	40,7%	15-17 anni	48,3%
Toscana	48,0%	18-19 anni	45,4%
Umbria	49,8%	20-24 anni	46,3%
Marche	67,1%	25-34 anni	42,7%
Lazio	44,5%	35-44 anni	44,9%
Abruzzo	71,8%	45-54 anni	53,6%
Molise	68,1%	55-59 anni	56,4%
Campania	60,7%	60-64 anni	60,5%
Puglia	67,8%	65-74 anni	66,9%
Basilicata	73,7%	75 e più anni	80,0%
Calabria	75,5%	Numero componenti famiglia	0,115
Sicilia	77,2%	1 componente	61,1%
Sardegna	71,3%	2 componenti	55,5%
		3 componenti	48,1%
Ripartizione geografica	0,285	4 componenti	47,4%
Italia Nord-Occidentale	35,6%	5 componenti	57,7%
Italia Nord-Orientale	44,2%	6 componenti	64,3%
Italia Centrale	49,0%	7 componenti	71,4%
Italia Meridionale	66,6%	8 componenti	80,1%
Italia Insulare	75,8%	9 componenti	100,0%
		10 componenti	100,0%
Tipo di comune o Dominio	0,152	Stato civile	0,125
Comuni metropolitani	38,8%	Celibe/Nubile	47,3%
Comuni limitrofi ai metropolitani	43,0%	Coniugato/a	53,2%
Comuni con oltre 50.000 ab.	54,7%	Separato/a di fatto	48,0%
Comuni con 10.001-50.000 ab.	56,5%	Separato/a legalmente	49,7%
Comuni con 2001-10.000 ab.	57,0%	Divorziato/a	52,0%
Comuni con al più 2000 ab.	62,9%	Vedovo/a	71,6%

Fonte: Nostre elaborazioni del file standard di tipo B (e di tipo A per il carattere “regione”) relativo a ISTAT (1999).

Tabella 1 (continua)

Titolo di studio	0,282	Posizione nella professione	0,246
Titoli post-laurea	24,2%	<i>Lavoratore dipendente in qualità di:</i>	
Laurea	24,5%	Dirigente	18,8%
Diploma universitario	26,9%	Direttivo, quadro	22,0%
Diploma scuola superiore (4-5 a.)	36,9%	Impiegato, intermedio	32,8%
Diploma sc. professionale (2-3 a.)	42,2%	Capo operaio, operaio, assimilati	54,1%
Licenza scuola media inferiore	53,4%	Apprendista	55,7%
Licenza elementare	67,9%	Lavorante a domicilio	54,0%
Nessun titolo (sa leggere-scrivere)	65,3%	<i>Lavoratore autonomo in qualità di:</i>	
Analfabeta (parziale o totale)	79,5%	Imprenditore	40,9%
Fonte principale di reddito	0,198	Libero professionista	24,0%
Da lavoro dipendente	40,6%	Lavoratore in proprio	54,4%
Da lavoro autonomo	46,4%	Socio di cooperativa	41,9%
Pensione	67,3%	Coadiuvante	46,4%
Indennità e provvidenze varie	70,9%	Ramo di attività economica	0,215
Patrimoniale	52,6%	Agricoltura, caccia e pesca	77,4%
Mantenimento familiare	50,8%	Industria, estrazione	40,6%
Condizione professionale	0,242	Costruzioni	56,1%
Occupato	42,5%	Commercio, alberghi, ristoranti	44,4%
In cerca di nuova occupazione	66,4%	Trasporti, magaz., comunicazioni	39,7%
In cerca di prima occupazione	66,8%	Intermediazioni, attività profes.	25,3%
In servizio di leva o civile	67,9%	Pubblica amministrazione, difesa	39,1%
Casalinga	63,0%	Istruzione, sanità, servizi sociali	33,0%
Studente	39,7%	Altri servizi	40,3%
Inabile	80,5%		
Ritirato dal lavoro	65,5%		
Altra condizione	76,0%		

In particolare, l'analisi per ripartizione evidenzia la netta scissione tra l'Italia Centro-Settentrionale e l'Italia Meridionale-Insulare, confermando il forte dualismo di cui soffre il nostro Paese (quantunque vi siano differenze notevoli anche tra le due ripartizioni del Nord, come pure tra le due del Mezzogiorno⁵). Tale dualismo dipende,

⁵ Passando al dettaglio regionale, si può notare che la regione meglio posizionata è la Lombardia (con appena il 30,3% di "non vacanzieri"), mentre la più svantaggiata è la Sicilia (77,2%), assieme alla Calabria (75,5%). Tra le regioni del Centro, quella che più si avvicina alle basse percentuali del Nord è il Lazio, mentre quella che più se ne allontana è rappresentata dalle Marche (assimilabili, sotto questo aspetto, alle regioni del Sud-Isole). La variabile "regione" non verrà più utilizzata nel seguito in quanto la struttura dei *file standard* ISTAT ne impedisce l'elaborazione congiunta, a differenza del carattere "ripartizione", con l'altro predittore di contesto considerato (tipo di comune).

in gran parte, dalle note disparità di livello e di struttura dei redditi e dei consumi che caratterizzano le due macroaree del Paese.

Nel caso della tipologia di comune e del titolo di studio, la percentuale di “non vacanzieri” varia in senso opposto all’ordinamento delle modalità dei due caratteri; infatti, tale percentuale cresce sia al diminuire della dimensione demografica dei comuni, sia al decrescere del titolo di studio posseduto. In particolare, nel primo caso le differenze più marcate si hanno passando dai due gruppi iniziali dei comuni metropolitani e limitrofi agli altri raggruppamenti di comuni (compresi quelli extrametropolitani con oltre 50.000 abitanti). Nel secondo caso, invece, si possono distinguere nettamente i titoli medio-alti (dalla licenza media in poi) da quelli più bassi (fino alla licenza elementare), benché si notino differenze non trascurabili anche all’interno di ciascuna delle due classi individuate.

Per quanto riguarda infine la posizione nella professione, la quota di “non vacanzieri” cresce quando si passa dalle posizioni funzionalmente più elevate a quelle più modeste, se non altro nel caso dei dipendenti. La disparità più rilevante si ha tra chi è almeno impiegato-intermedio e le altre tipologie di dipendenti; a queste ultime sono associabili anche i lavoratori in proprio, mentre gli altri lavoratori autonomi, e soprattutto i liberi professionisti e gli imprenditori, sono assimilabili al gruppo dei “colletti bianchi”.

Quanto precede trova una sua spiegazione nel fatto che la vacanza risponde a esigenze di evasione da condizioni stressanti di vita e di lavoro, le quali risultano particolarmente pesanti nei centri metropolitani e limitrofi. Inoltre, la natura di bene superiore o di lusso della vacanza ne condiziona il godimento, e ciò giustifica la maggior presenza relativa di vacanzieri tra le posizioni professionali più qualificate, alle quali sono solitamente associati alti livelli di istruzione e di reddito.

All’interno di questo primo gruppo di variabili si possono far rientrare anche i due unici predittori quantitativi considerati, cioè l’età⁶ e il numero di componenti della famiglia attuale. Quanto al primo, si può notare una crescita monotona della quota di “non vacanzieri” solo da una certa classe di età in poi (dopo i 25-34 anni, dove si registra il minimo assoluto); per le classi precedenti si ha invece un andamento altalenante, con un minimo relativo nella classe di età preattiva 6-14 anni.

Viceversa, la serie delle percentuali di “non vacanzieri” secondo la dimensione della famiglia segue un arco di parabola con concavità rivolta verso l’alto; infatti, tale percentuale decresce passando dalle famiglie monocomponente a quelle composte da quattro membri (dove tocca il minimo assoluto), mentre in seguito aumenta sistematicamente, fino a raggiungere il 100% in corrispondenza delle famiglie – invero assai poco diffuse – con almeno nove componenti.

⁶ Il carattere “età” presenta indubbie relazioni con altre variabili esplicative (in particolare con quelle economiche e col titolo di studio, a loro volta strettamente legate). A tale carattere risulta inoltre associato un effetto “generazione”, effetto che interagisce con quello proprio dell’età (Bernini, 1998).

2.b) – Le variabili con modalità che si possono utilmente classificare, ma che non presentano andamenti monotoni delle corrispondenti quote di “non vacanzieri”, sono la condizione professionale, la fonte principale di reddito e il ramo di attività economica. Relativamente alla condizione professionale vi è una distinzione netta tra occupati e non occupati (studenti esclusi, avendo questi una percentuale di “non vacanzieri” più simile a quella dei primi che dei secondi). Tale distinzione si ritrova anche nel caso della fonte di reddito⁷, variabile che fornisce informazioni analoghe, ma meno analitiche, rispetto agli altri predittori economici considerati finora (condizione e posizione professionale).

Forti differenze nelle quote di “non vacanzieri” sono altresì riscontrabili tra i vari rami di attività economica, in conseguenza anche della diversa quantità, qualità e distribuzione infrannuale del tempo libero che li caratterizza. In particolare, il ramo “agricoltura, caccia e pesca” registra il massimo assoluto di detta quota, mentre quello delle “costruzioni” presenta un massimo relativo. Nel caso del “primario” si tratta di un settore produttivo in cui certe attività non possono essere affatto sospese nel corso dell’anno, mentre altre risultano concentrate durante l’estate (la stagione altrimenti più propizia per le vacanze). Viceversa, le percentuali minime di “non vacanzieri” si rilevano in due settori dei servizi all’interno dei quali l’interruzione dell’attività lavorativa può essere relativamente più prolungata e, cioè, nei rami “intermediazioni e attività professionali ed imprenditoriali”, e “istruzione, sanità e servizi sociali” (i quali presentano, rispettivamente, il minimo assoluto e relativo).

2.c) – Infine, l’unica variabile priva di variazioni apprezzabili nelle quote di “non vacanzieri” è il sesso, essendo tali quote quasi le stesse per maschi e femmine (a motivo, tra l’altro, della diffusa abitudine a trascorrere le vacanze in coppia). E’ comunque opportuno verificare se questo risultato trovi conferma anche in seguito, quando si opererà in termini più disaggregati con l’analisi multivariata.

Questa scarsa variabilità si ritrova anche con riferimento alle prime cinque modalità del carattere “stato civile”, in corrispondenza delle quali la frazione di “non vacanzieri” oscilla leggermente attorno al 50%. Al contrario, detta frazione sale notevolmente nel caso dei vedovi, per ragioni legate ovviamente all’età più avanzata, nonché alle condizioni economiche, di salute e/o di altro tipo connesse con tale stato.

Da quanto precede si può dedurre l’esistenza di indubbie relazioni che legano la variabile dipendente “quota di non vacanzieri” a fattori quali le condizioni socio-economiche generali di un territorio, i livelli di reddito familiare e il grado di istruzione acquisito (Costa, Manente, 2000, pp. 62-63). In questo contesto i predittori

⁷ Secondo le definizioni ISTAT, la fonte principale di reddito dovrebbe riguardare solo la popolazione con almeno 14 anni d’età. In realtà, il totale dei rispondenti al relativo quesito supera quello degli ultratredicenni di oltre il 6%, per cui le informazioni attinenti a questo carattere non sembrano così affidabili come quelle ottenute per gli altri.

considerati e, in particolare, la ripartizione geografica e la tipologia del comune di residenza, la condizione e la posizione professionale, il titolo di studio e l'età vengono ad assumere il ruolo di variabili *proxy* dei suddetti fattori esplicativi "latenti". Con la successiva analisi multivariata di segmentazione si potrà sottoporre a una più stringente verifica l'effettiva capacità esplicativa di tali variabili *proxy*, in modo da individuare quelle maggiormente influenti.

2.2 *Motivi per cui non si è andati in vacanza*

Le considerazioni finali del precedente paragrafo trovano un'indiretta conferma anche nelle risposte fornite dai "non vacanzieri" al quesito sui motivi della mancata fruizione di vacanze. Nelle graduatorie per ordine di importanza, riportate nella Tabella 2, risultano sempre nettamente in testa le ragioni di natura economica, seguite dai motivi di famiglia, dalla mancanza di abitudine e dagli impedimenti legati al lavoro o allo studio, all'età e alle condizioni di salute. Al contrario, non sembrano essere molto rilevanti né la circostanza di risiedere già in un luogo di villeggiatura, né motivi diversi da quelli elencati. E' interessante notare come tali graduatorie non varino apprezzabilmente né quando i rapporti si riferiscono al totale delle risposte fornite (coll. 2.1, 2.2 della Tabella 2) anziché al totale dei "non vacanzieri" (coll. 1.1, 1.2), né quando i casi caratterizzati da una sola risposta (coll. 1.2, 2.2) vengono distinti dall'insieme dei casi (in cui compaiono anche risposte plurime: coll. 1.1, 2.1).

Tabella 2 – Motivi della mancata vacanza per 100 "non vacanzieri", nel caso di risposte plurime o meno (1.1) e di risposta unica (1.2), e per 100 risposte, plurime o meno (2.1) e uniche (2.2). Italia, anno 1997

	Risposte plurime o meno (1.1)	Risposta unica (1.2)	Risposte plurime o meno (2.1)	Risposta unica (2.2)
Motivi economici	42,7%	27,3%	34,1%	36,6%
Motivi di famiglia	25,6%	16,9%	20,5%	22,7%
Mancanza di abitudine	17,2%	10,2%	13,7%	13,6%
Motivi di lavoro o di studio	11,9%	7,4%	9,5%	9,9%
Motivi legati all'età	10,3%	4,8%	8,2%	6,5%
Motivi di salute	10,2%	4,1%	8,2%	5,4%
Residente in località di vacanza	4,8%	2,1%	3,9%	2,8%
Altri motivi	2,4%	1,9%	1,9%	2,5%
Totale	125,1%	74,7%	100,0%	100,0%

Fonte: Nostre elaborazioni del file standard di tipo B relativo a ISTAT (1999).

Ancora una volta, dunque, viene indirettamente confermata l'importanza del reddito quale fattore decisivo, anche se non esclusivo, della possibilità di fare o no vacanza (Costa e Manente, 2000, pp. 62-63 e 136-137). Congiuntamente o meno a tale fattore agiscono poi le condizioni familiari – essendo il turismo un fenomeno fruito prevalentemente a livello familiare – e quelle relative al tipo di lavoro svolto (che, come nel caso dei lavoratori in proprio e/o degli attivi in agricoltura, poco predispone al godimento di “lunghe” vacanze) o, ancora, motivazioni legate direttamente o indirettamente all'età, oltre che alla generazione di appartenenza (quali la mancanza di abitudine e/o lo stato di salute).

Sebbene il sesso e altre variabili demografiche sembrino scarsamente connesse con la probabilità di non fare vacanza, esse verranno considerate ugualmente nella successiva analisi multivariata di segmentazione. Ciò si giustifica per motivi di confronto con quanto precede, soprattutto a fini di verifica dei risultati dell'analisi bivariata di connessione; non è infatti da escludere l'esistenza di connessioni “spurie” – sottovalutate o sopravvalutate – a causa dell'effetto di terzi fattori influenti, debitamente trattabili solo con un approccio multivariato.

3. Procedura di segmentazione e opzioni prescelte

L'analisi di segmentazione è stata successivamente condotta su due blocchi distinti di predittori, il primo dei quali rappresenta un ampio sottoinsieme delle 11 variabili in esame, comprendente quelle rilevate sul maggior numero possibile di individui (compatibilmente con la natura delle variabili stesse). Il secondo blocco, invece, è composto da tutti e 11 i potenziali predittori di cui sopra, a prescindere dall'ammontare delle corrispondenti unità di rilevazione. Nella fattispecie, al primo blocco appartengono tutte le variabili del secondo tranne la “posizione nella professione” e il “ramo di attività economica” che, essendo rilevabili sul solo sottoinsieme degli occupati, riguardano meno della metà della popolazione indagata.

L'analisi dei dati disponibili è stata effettuata con gli algoritmi di segmentazione implementati in *AnswerTree* (SPSS, 1998) e, in particolare, con la procedura *C&RT* o *CART* (Breiman et al., 1984; Steinberg e Colla, 1995). Tale procedura è stata preferita alle altre presenti in *AnswerTree* (*CHAID*: Kass, 1980; *CHAID* esaustivo: Biggs et al., 1991 e *QUEST*: Loh e Shih, 1997) perché ha portato a risultati più stabili e più facilmente interpretabili; ovviamente, si è proceduto alla sua utilizzazione dopo aver verificato il soddisfacimento delle cosiddette “regole di salvaguardia” o condizioni di applicabilità del metodo. Come funzione criterio si è adottata quella basata sull'indice di eterogeneità di Gini, scegliendo per la regola di arresto del processo un valore di soglia più elevato di quello standard (0,001, anziché 0,0001). In tal modo si

è cercato di conciliare le due opposte esigenze, da una parte, di non frazionare eccessivamente la popolazione di interesse e, dall'altra, di ottenere gruppi o "nodi figlio" il più possibile omogenei al loro interno, salvaguardando nel contempo la stabilità e l'interpretabilità dei risultati.

Così facendo, l'analisi di segmentazione si è interrotta dopo 6 passi o livelli, dando luogo a 14 bipartizioni e a 28 nodi figlio (di cui 15 finali), nel caso del primo blocco di 9 predittori, ovvero a 15 bipartizioni e a 30 nodi figlio (di cui 16 finali), nel caso del secondo blocco di 11 predittori (già oggetto di analisi in Tonini, 2001).

4. Principali risultati ottenuti con l'analisi di segmentazione

Nell'esposizione dei risultati raggiunti si cercherà di rispondere ai quesiti posti nell'introduzione, soffermandosi dapprima sui predittori rilevanti e sulla dicotomizzazione delle loro modalità (par. 4.1), per poi passare all'esame delle categorie sociali più/meno a rischio di non andare in vacanza (par. 4.2).

4.1 Predittori rilevanti e dicotomizzazione delle relative modalità

Delle nove variabili esplicative inizialmente considerate, cinque soltanto entrano in gioco nel processo di segmentazione, mentre quattro non vi contribuiscono affatto, almeno direttamente. Tra i predittori rilevanti rientrano entrambe le variabili di contesto (ripartizione e dominio), due dei tre caratteri di tipo socio-economico (titolo di studio e condizione professionale), ed uno solo di natura demografica (età).

Ciò, se da un lato sembra confermare la scarsa capacità esplicativa del fenomeno "vacanze" da parte di quasi tutti i fattori demografici (in primo luogo il sesso, ma anche lo stato civile e il numero di componenti della famiglia), dall'altro evidenzia la ridondanza della "fonte di reddito principale" rispetto all'altra variabile economica presa inizialmente in esame.

Ordinando i predittori significativi in base al loro apporto alla riduzione dell'eterogeneità nei gruppi⁸ e ai livelli o passi in cui compaiono nel processo di

⁸ Nella procedura *CART* il suddetto decremento viene chiamato "miglioramento" o "riduzione dell'impurità" (eterogeneità) nei gruppi o nodi figlio, dove, per misurare tale "impurità", ci si è basati sull'indice di eterogeneità del Gini. L'arresto dell'analisi è avvenuto al 7° passo del processo di segmentazione, in corrispondenza del quale la "riduzione dell'impurità" rispetto al passo precedente è risultata inferiore al prefissato livello di soglia (pari a 0,001). La Tabella 3 riporta, per ogni predittore significativo dei nove o undici di partenza, il corrispondente valore totale del "miglioramento", opportunamente normalizzato a fini di confronto. Le successive tabelle 3-6 sono state ricavate dalla Figura 1 in Appendice, nel caso dei nove predittori iniziali, e da Tonini (2001), nel caso degli undici.

segmentazione, si ha che il più importante di essi è la variabile di contesto “ripartizione geografica” (vedasi Tabella 3). Più nettamente che in altre circostanze, anche in questo caso trova conferma il tradizionale dualismo tra l'Italia Centro-Settentrionale e quella Meridionale-Insulare, essendo il rischio di non andare in vacanza molto più elevato nel Sud che nel Nord (con, rispettivamente, il 70% di “non vacanzieri” contro il 42%).

All'interno di entrambe queste macroaree, poi, i fattori più discriminanti sono due e cioè, in ordine decrescente di importanza, il titolo di studio e l'età (anche se quest'ultima sembra nettamente meno rilevante del precedente). Peraltro, a parità di livello in cui agisce ciascun predittore, varia la classificazione delle corrispondenti modalità elementari; in particolare, i valori di soglia in base ai quali vengono dicotomizzate le modalità dei caratteri almeno ordinabili risultano più elevati nel Sud-Isole che nel Centro-Nord (vedasi la prima parte della Tabella 4). Ad esempio, nella prima bipartizione secondo il titolo di studio, per appartenere al gruppo dei più istruiti dove la quota di “non vacanzieri” è più bassa, nel Meridione bisogna avere almeno il diploma di scuola superiore quadriennale, mentre nel Settentrione basta possedere almeno la licenza media inferiore.

Tabella 3 – Livelli del processo di segmentazione in cui agisce (anche più volte) ciascun predittore significativo e corrispondente valore del “miglioramento” totale

	Analisi con 9 predittori iniziali		Analisi con 11 predittori iniziali	
	Livelli	Miglioram.	Livelli	Miglioram.
(1) Ripartizione geografica	1°; 5°; 5°	0,0478	1°; 5°; 5°	0,0473
(2) Titolo di studio	2°; 2°; 3°; 4°	0,0414	2°; 2°; 4°	0,0366
(3) Età (in anni compiuti)	3°; 3°; 6°	0,0156	3°; 3°; 6°	0,0156
(4) Tipo di comune	4°; 4°; 4°	0,0062	4°; 4°; 6°	0,0055
(5) Condizione professionale	3°	0,0012	—	—
(6) Posizione professionale	—	—	3°; 3°	0,0143
(7) Ramo di attività economico	—	—	4°	0,0019

Fonte: Nostre elaborazioni del file standard di tipo B relativo a ISTAT (1999).

Tabella 4 – Bipartizioni delle modalità di ciascun predittore significativo secondo l'analisi di segmentazione condotta dapprima con 9 e poi con 11 predittori iniziali

	Ripartizioni centro-settentrionali	Ripartizioni meridionali-insulari
Analisi con 9 predittori iniziali		
(1)	{[Nord-Centro] {[Nord-Ovest o Nord-Est] [Centro]} {[Nord-Ovest o Nord-Est] [Centro]}	{[Sud-Isole]} (non presenta alcuna sottobipartizione)
(2)	{[Almeno licenza media inferiore] [Al più licenza elementare]} {[Almeno diploma sc. superiori (4-5 a.)] [Diploma professionale (2-3 anni) o licenza media inferiore]}	{[Almeno diploma sc. superiori (4-5 a.)] [Al più diploma professionale (2-3 a.)]} {[Diploma professionale (2-3 anni) o licenza media inferiore] [Al più licenza elementare]}
(3)	{[$x \leq 14$ anni] [$x > 14$ anni]} {[$14 < x \leq 71$] [$x > 71$ anni]}	{[$x \leq 18$ anni] [$x > 18$ anni]}
(4)	{[Comuni metropolitani o limitrofi] [Altri comuni]} {Tale bipartizione si ripete altre 2 volte}	(non presenta alcuna bipartizione)
(5)	(non presenta alcuna bipartizione)	{[Occupato o casalinga o studente o ritirato dal lavoro o in altra condizione] [In cerca di nuova o di prima occupaz., in servizio di leva o inabile]}
Analisi con 11 predittori iniziali		
(1)(3) (4)	(come sopra con 9 predittori)	(come sopra con 9 predittori)
(2)	{[Almeno licenza media inferiore] [Al più licenza elementare]}	(come sopra con 9 predittori)
(5)	(come sopra con 9 predittori)	(non presenta alcuna bipartizione)
(6)	{[Almeno impiegato-intermedio o imprenditore o libero professionista] [Altre posizioni professionali]}	{[Dirigente o direttivo-quadro o libero professionista o coadiuvante] [Altre posizioni professionali]}
(7)	{[Agricoltura, caccia e pesca] [Altri rami di attività economica]}	(non presenta alcuna bipartizione)

Fonte: Nostre elaborazioni del file standard di tipo B relativo a ISTAT (1999).

Per quanto riguarda invece l'età, il livello di soglia al di sopra del quale aumenta sensibilmente l'incidenza dei "non vacanzieri" – se non altro tra i meno istruiti – è la maggiore età (18 anni) nel Sud-Isole, e l'inizio dell'età attiva (14 anni) nel Centro-Nord.

Gli altri due predittori rilevanti agiscono invece in una sola delle due macro-ripartizioni individuate e, precisamente, il "tipo di comune" nell'Italia Centro-Settentrionale, e la "condizione professionale" nell'Italia Meridionale-Insulare. In particolare, nel caso della prima macroarea – che comprende le due uniche vere metropoli esistenti in Italia, Milano al Nord e Roma al Centro – si discrimina ripetutamente tra i comuni metropolitani o limitrofi e gli altri comuni (dove il rischio di non fare vacanza risulta più alto). Nel caso del Mezzogiorno, viceversa, le categorie più svantaggiate sono quelle dei "non occupati" (perché in cerca di nuova o di prima occupazione, in servizio di leva o inabili), anche se in possesso di un titolo di studio elevato (dal diploma di scuola superiore quadriennale in poi).

Operando con 11 predittori, anziché con 9, si hanno alcune modifiche nei risultati, limitatamente però alle sole variabili di tipo socio-economico; al contrario, i predittori di contesto e quello demografico non registrano variazioni di sorta (si confronti la parte inferiore della Tabella 4 con quella superiore). In particolare, nell'Italia Meridionale-Insulare la variabile "condizione professionale" viene rimpiazzata dalla "posizione nella professione"; quest'ultimo predittore, assieme al "ramo di attività", agisce ora anche all'interno dell'Italia Centro-Settentrionale.

L'aggiunta di due nuovi predittori al Centro-Nord e la sostituzione di un predittore con un altro (più specifico) nel Sud-Isole non modifica comunque la graduatoria delle variabili esplicative più rilevanti, che vede sempre al primo posto la ripartizione geografica, seguita dal titolo di studio e dall'età (anche se quest'ultima è ora tallonata da vicino dalla "posizione nella professione")⁹. In coda alla classifica troviamo sempre al penultimo posto il "tipo di comune", mentre all'ultimo sta adesso il "ramo di attività"¹⁰, anziché la "condizione professionale". Infine, nel caso del titolo di studio viene a cadere, all'interno della ripartizione Centro-Settentrionale, la distinzione tra il livello intermedio (licenza media o diploma professionale) e quello più elevato (almeno diploma di scuola superiore quadriennale).

⁹ La suddetta graduatoria si basa sui valori del "miglioramento" totale riportati nell'ultima colonna della Tabella 3. E' peraltro interessante notare che, come per l'età ed il titolo di studio, anche nel caso della posizione professionale, il Mezzogiorno si caratterizza per il fatto che la categoria più avvantaggiata parte da posizioni dipendenti relativamente più elevate rispetto al Settentrione e, inoltre, non include gli imprenditori tra le figure autonome (vedasi Tabella 4 in basso).

¹⁰ Con riferimento al "ramo di attività", trova conferma il già evidenziato dualismo tra "agricoltura, caccia e pesca" e gli altri rami, a motivo del fatto che il "primario" è un settore in cui l'attività, oltre a svolgersi spesso senza interruzioni di sorta, si concentra particolarmente nei mesi estivi, lasciando poco tempo libero per eventuali vacanze (specie se lunghe).

4.2 *Categorie sociali più/meno a rischio di non fare vacanza*

Secondo la Tabella 5, tra le sei categorie più a rischio di non andare in vacanza, quattro appartengono al Sud-Isole e due al Centro-Nord; viceversa, tutte le categorie meno a rischio interessano i residenti nella macroripartizione Centro-Settentrionale, mentre in quella Meridionale-Insulare le quote di “non vacanzieri” non scendono mai sensibilmente sotto la percentuale nazionale (pari al 52,2%).

In particolare, le tre categorie più svantaggiate sono tutte contraddistinte da un basso titolo di studio (al più la licenza elementare), e riguardano altresì: le sole persone maggiorenni del Sud-Isole, ovvero gli ultraquattordicenni del Centro o gli ultrasessantunenni del Nord, purché residenti entrambi in comuni non metropolitani né limitrofi. Al contrario, le categorie più avvantaggiate sono quelle dei residenti in comuni metropolitani o limitrofi del Centro-Nord, a condizione che siano in possesso di un livello di istruzione elevato o intermedio (almeno il diploma di scuola superiore quadriennale, oppure la licenza media o un diploma di scuole professionali).

Un discorso a parte merita la subpopolazione in età preattiva ($x \leq 14$ anni) dell'Italia Centro-Settentrionale, la quale, pur essendo contraddistinta da un basso livello di istruzione (al più la licenza elementare), presenta una scarsa percentuale di “non vacanzieri” (in sintonia con il minimo relativo assunto da tale percentuale in corrispondenza della classe d'età 6-14 anni). Si tratta di bambini e ragazzi che, essendo ancora molto legati alla famiglia, partecipano della spiccata propensione ad andare in vacanza dei loro genitori.

Operando con 11 predittori, anziché con 9, non mutano le caratteristiche delle tre categorie più a rischio di non effettuare vacanze (vedasi Tabella 6, ricavata da Tonini, 2001). Peraltro, fra le categorie più svantaggiate, una soltanto cambia radicalmente dopo l'aggiunta delle due variabili economiche alle nove iniziali; il gruppo [11] infatti, che riguardava il Sud-Isole, viene ora sostituito dal [7], composto da residenti nel Centro-Nord che, pur essendo in possesso almeno della licenza media, hanno una bassa propensione ad andare in vacanza perché ricoprono posizioni non impiegate (al più di capo operaio o apprendista o lavorante a domicilio), ovvero sono lavoratori in proprio o soci di cooperativa o coadiuvanti che risultano attivi nel settore primario (agricoltura, caccia e pesca).

Viceversa, le categorie meno a rischio si modificano profondamente, oltre a crescere da cinque a sei; di queste ultime, infatti, solo la [9] coincide perfettamente con una delle precedenti, riguardando come sopra i residenti nell'Italia Centro-Settentrionale in età preattiva e con un titolo di studio ovviamente basso (al più la licenza elementare).

Per quanto concerne invece gli altri casi, l'introduzione del predittore “posizione nella professione” fa sì che bastino titoli di studio relativamente più bassi (dalla

Tabella 5 – Categorie sociali (nodi) più e meno a rischio di non fare vacanza, e corrispondenti percentuali di “non vacanzieri”. Caso con 9 predittori iniziali

Analisi di segmentazione con 9 predittori iniziali					
Nodo	Categorie più a rischio del		Nodo	Categorie meno a rischio del	
	Nord-Centro	Sud-Isole		Nord-Centro	Sud-Isole
[22]	—	88,7%	[23]	39,3%	—
[26]	82,4%	—	[16]	30,6%	—
[28]	79,2%	—	[17]	29,5%	—
[21]	—	75,3%	[9]	27,5%	—
[11]	—	67,0%	[15]	17,3%	—
[13]	—	64,8%		—	—
[IT]	52,2%		[IT]	52,2%	

Legenda relativa ai nodi o gruppi di individui:

- [22] Maggiorenni ($x > 18$ anni) e con al più la licenza elementare.
- [26] In età attiva o postattiva o anziana ($x > 14$ anni), residenti in comuni non metropolitani né limitrofi del Centro e con al più la licenza elementare.
- [28] In età anziana ($x > 71$ anni), residenti in comuni non metropolitani né limitrofi del Nord e con al più la licenza elementare.
- [21] Maggiorenni ($x > 18$ anni) e con la licenza media inferiore o un diploma di scuole professionali (2-3 anni).
- [11] Con almeno il diploma di scuola superiore (4-5 anni) e in condizione di non occupato (perché in cerca di nuova o di prima occupazione, in servizio di leva o inabili).
- [13] Minorenni ($x \leq 18$ anni) e con al più un diploma di scuole professionali (2-3 anni).

[IT] *Nodo genitore con la percentuale di “non vacanzieri” in Italia.*

- [23] Residenti in comuni non metropolitani né limitrofi del Nord e con la licenza media inferiore o un diploma di scuole professionali (2-3 anni).
- [16] Residenti in comuni non metropolitani né limitrofi e con almeno il diploma di scuola superiore (4-5 anni).
- [17] Residenti in comuni metropolitani o limitrofi e con la licenza media inferiore o un diploma di scuole professionali (2-3 anni).
- [9] In età preattiva ($x \leq 14$ anni) e con al più la licenza elementare.
- [15] Residenti in comuni metropolitani o limitrofi e con almeno il diploma di scuola superiore (4-5 anni).

Fonte: Figura 1 in Appendice.

Tabella 6 – Categorie sociali (nodi) più e meno a rischio di non fare vacanza, e corrispondenti percentuali di “non vacanzieri”. Caso con 11 predittori iniziali

Analisi di segmentazione con 11 predittori iniziali					
Nodo	Categorie più a rischio del		Nodo	Categorie meno a rischio del	
	Nord-Centro	Sud-Isole		Nord-Centro	Sud-Isole
[22]	—	88,7%	[12]	—	36,4%
[26]	82,4%	—	[23]	36,1%	—
[28]	79,2%	—	[29]	35,4%	—
[21]	—	75,3%	[16]	30,9%	—
[7]	64,8%	—	[9]	27,5%	—
[13]	—	64,8%	[15]	17,7%	—
[IT]	52,2%		[IT]	52,2%	

Legenda relativa ai nodi o gruppi di individui:

- [22] Maggiorenni ($x > 18$ anni) e con al più la licenza elementare.
- [26] In età attiva o postattiva o anziana ($x > 14$ anni), residenti in comuni non metropolitani né limitrofi del Centro e con al più la licenza elementare.
- [28] In età anziana ($x > 71$ anni), residenti in comuni non metropolitani né limitrofi del Nord e con al più la licenza elementare.
- [21] Maggiorenni ($x > 18$ anni) e con la licenza media inferiore o un diploma di scuole professionali (2-3 anni).
- [7] Con almeno la licenza media inferiore, in posizione di: al più capo operaio o apprendista o lavorante a domicilio o lavoratore in proprio o socio di cooperativa o coadiuvante, e attivi nel ramo “agricoltura, caccia e pesca”.
- [13] Minorenni ($x \leq 18$ anni) e con al più un diploma di scuole professionali (2-3 anni).
- [IT] *Nodo genitore con la percentuale di “non vacanzieri” in Italia.*
- [12] Con almeno il diploma di scuola superiore (4-5 anni) e in posizione di dirigente o direttivo-quadro o libero professionista o coadiuvante.
- [23] Residenti nel Nord, con almeno la licenza media inferiore e in posizione di: al più capo operaio o apprendista o lavorante a domicilio o lavoratore in proprio o socio di cooperativa o coadiuvante, e attivi in rami diversi da “agricoltura, caccia e pesca”.
- [29] Residenti in comuni metropolitani o limitrofi del Centro, con almeno la licenza media inferiore, in posizione di: al più capo operaio o apprendista o lavorante a domicilio o lavoratore in proprio o socio di cooperativa o coadiuvante, e attivi in rami diversi da “agricoltura, caccia e pesca”.
- [16] Residenti in comuni non metropolitani né limitrofi, con almeno la licenza media inferiore e in posizione di: almeno impiegato-intermedio o imprenditore o libero professionista.
- [9] In età preattiva ($x \leq 14$ anni) e con al più la licenza elementare.
- [15] Residenti in comuni metropolitani o limitrofi, con almeno la licenza media inferiore e in posizione di: almeno impiegato-intermedio o imprenditore o libero professionista.

licenza media in poi, anziché dal diploma di scuola superiore quadriennale) per godere dello stesso basso rischio di non fare vacanza, purché tali livelli di istruzione siano associati a posizioni professionali almeno impiegatizie, tra i dipendenti, o di imprenditore o libero professionista, tra i lavoratori autonomi (si confrontino i gruppi [15] e [16] della Tabella 6 con quelli corrispondenti della Tabella 5).

Rispetto al primo blocco di nove predittori, tra le categorie meno a rischio ne compare ora anche una del Sud-Isole (la [12]); essa riguarda individui per cui risulta elevato sia il titolo di studio (almeno il diploma di scuola superiore quadriennale), sia la posizione nella professione (almeno di quadro, fra i lavoratori dipendenti). Per converso, nel Centro-Nord bastano livelli meno elevati sia di istruzione (almeno la licenza media) che di posizione professionale (almeno impiegatizia, tra i dipendenti) per poter godere di una più spiccata propensione ad andare in vacanza. Ciò dipende, tra l'altro, dai differenziali di reddito esistenti di fatto in Italia dove, com'è noto, il reddito pro capite è più alto nel Settentrione che nel Mezzogiorno.

5. Conclusioni

L'analisi di segmentazione, oltre a confermare in gran parte i risultati ottenuti con l'approccio univariato e con gli indici di connessione, ha fornito in più le combinazioni tra le modalità dei predittori rilevanti, le quali combinazioni hanno permesso di identificare le categorie sociali più/meno a rischio di non fare vacanza.

Le conferme hanno riguardato sia il novero dei caratteri non significativi (tutti di natura demografica: sesso, stato civile e numero di componenti della famiglia), sia quello delle variabili maggiormente influenti. Tra queste ultime, accanto alla variabile di contesto "ripartizione geografica", un ruolo fondamentale nel processo di segmentazione viene svolto dal predittore "titolo di studio"; esso presenta una relazione di tipo inverso con la "probabilità" di non fare vacanza, nel senso che al crescere del livello di istruzione diminuisce la quota di non vacanzieri, e viceversa. Ciò si verifica, fra l'altro, perché a livelli di istruzione più elevati corrispondono, solitamente, non solo posizioni professionali più qualificate e maggiori disponibilità di tempo libero e di reddito da destinare alle vacanze, ma anche una maggiore propensione alla mobilità per motivi culturali, di svago o di divertimento.

Pur trattandosi di elaborazioni iniziali dei dati di interesse, i risultati ottenuti sembrano incoraggianti, tanto per la loro stabilità (verificata anche con un opportuno processo di validazione), quanto per la loro interpretabilità. Tutto ciò è di stimolo a proseguire la ricerca, ad esempio, effettuando confronti tra i precedenti risultati e quelli ottenibili, *ceteris paribus*, sia operando con altre tecniche di segmentazione non implementate in *AnswerTree*, sia forzando l'ordine di entrata dei predittori nella

procedura, sia utilizzando dati di fonte diversa ma omogenei rispetto a quelli impiegati finora.

Sembra infine opportuno, da una parte, approfondire il tema delle interazioni tra i predittori del rischio di non fare vacanza e, dall'altra, passare da un'analisi esplorativa ad una confermativa, basata ad esempio sul modello di regressione logistica; ciò consentirebbe di valutare non solo la significatività, ma anche l'entità dell'impatto che le variabili influenti esercitano sul rischio di non effettuare vacanze. In tal modo si potrebbero ottenere nuove indicazioni che, assieme a quelle ricavabili dal presente lavoro, risulterebbero utili per orientare opportunamente una politica socio-economica del turismo finalizzata, tra l'altro, ad allargare il più possibile la partecipazione al mercato delle vacanze da parte degli italiani.

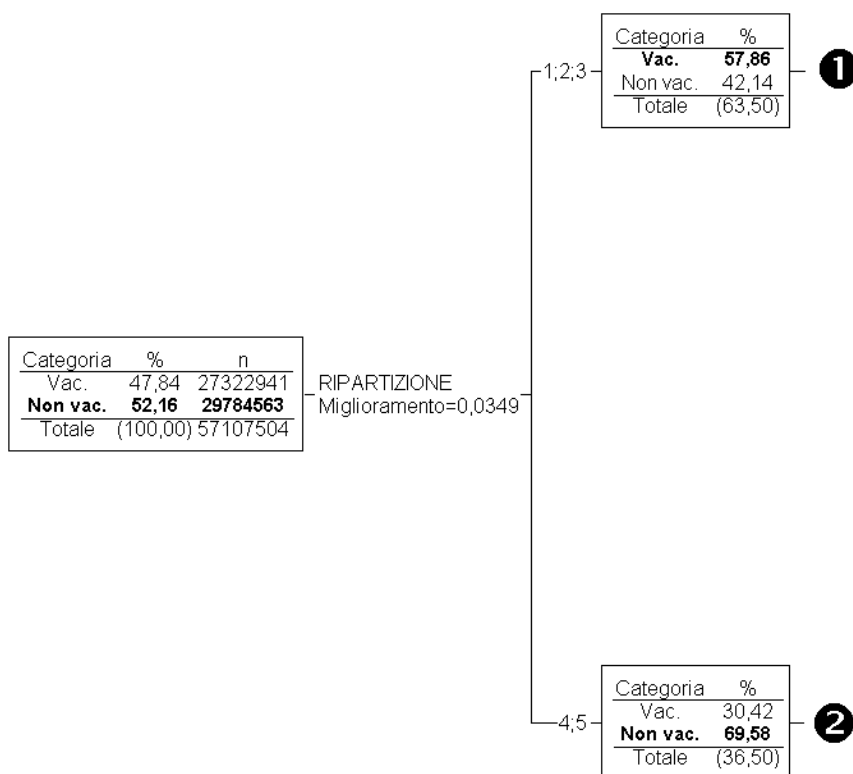
Riferimenti bibliografici

- AA.VV. (1998) *Ottavo rapporto sul turismo italiano, 1998*, Mercury, Firenze.
- BERNINI C. (1998) I consumi di beni e servizi turistici delle famiglie italiane. In: GARDINI A. (a cura di) *L'analisi della domanda e della produzione di servizi turistici*, CLUEB, Bologna: 111-132.
- BIGGS D., de VILLE B., SUEN E. (1991) A method of choosing multiway partitions for classification and decision trees, *Journal of Applied Statistics*, **18**: 49-62.
- BREIMAN L., FRIEDMAN J.H., OLSHEN R.A., STONE C.J. (1984) *Classification and Regression Trees*, Wadsworth, Belmont, CA.
- CANDELA G. (1996) *Manuale di economia del turismo*, CLUEB, Bologna.
- CLARK L.A., PREGIBON D. (1992) Tree-based models. In CHAMBERS J.M., HASTIE T.J. (eds.) *Statistical Models in S*, Wadsworth, CA: 377-420.
- COSTA P., MANENTE M. (2000) *Economia del turismo. Modelli di analisi e misura delle dimensioni economiche del turismo*, Touring University Press, Touring Club Italiano, Milano.
- FABBRIS L. (1997) *Statistica multivariata. Analisi esplorativa dei dati*, McGraw-Hill, Milano.
- FIELDING A., O'MUIRCHEARTAIGH C.A. (1977) Binary segmentation in survey analysis, with particular reference to AID, *The Statistician*, **26**: 17-28.
- GISMONDI R. (2000) Le presenze turistiche in Italia nel 1997: problemi di integrazione tra fonti, *Turistica*, **9**: 13-60.
- ISTAT (1998) *I viaggi in Italia e all'estero nel 1997*, Collana *Informazioni*, n. 100.
- ISTAT (1999) *La vita quotidiana nel 1997 (Indagine Multiscopo sulle Famiglie. Aspetti della vita quotidiana. Anno 1997)*, Collana *Informazioni*, n. 12, Roma.

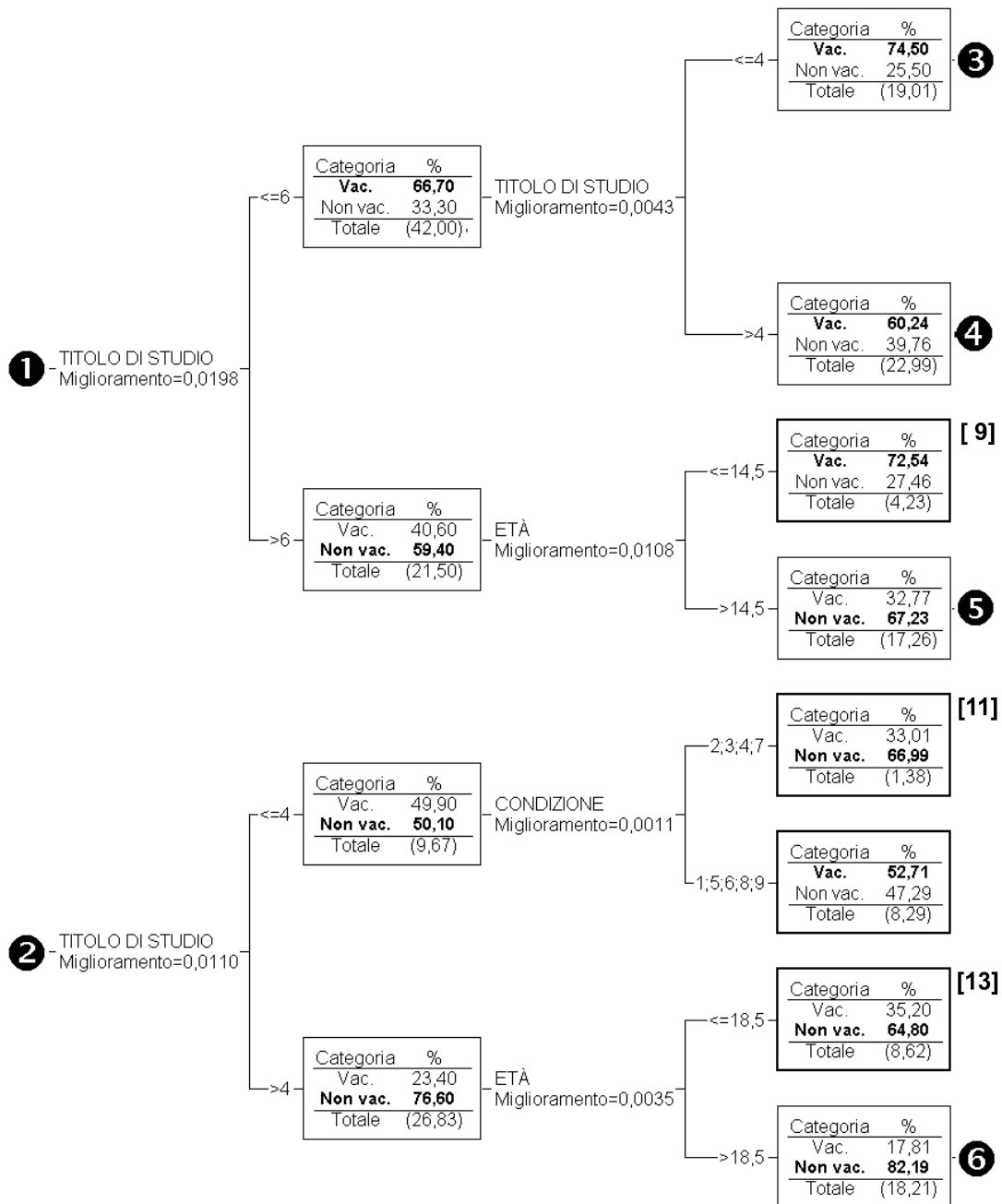
- KASS G.V. (1980) An exploratory technique for investigating large quantities of categorical data, *Applied Statistics*, **29**: 119-127.
- KIM H., LOH W.-Y. (2001) Classification trees with unbiased multiway splits, *Journal of the American Statistical Association*, **96**: 589-604.
- LETI G. (1983) *Statistica descrittiva*, Il Mulino, Bologna.
- LIM T.-S., LOH W.-Y., SHIH Y.-S. (2000) A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms, *Machine Learning*, **40**: 203-229.
- LOH W.-Y., SHIH Y.-S. (1997) Split selection methods for classification trees, *Statistica Sinica*, **7**: 815-840.
- O'MUIRCHEARTAIGH C.A., PAYNE C. (a cura di) (1977) *The Analysis of Survey Data. Volume 1: Exploring Data Structures*, Wiley, London.
- SMITH S.L.J. (1989) *Tourism Analysis: A Handbook*, Longman Group, UK.
- SONG H., WITT S.F. (2000) *Tourism Demand Modelling and Forecasting: Modern Econometric Approaches*, Elsevier Science, UK.
- SONQUIST J.A., DUNKELBERG W.C. (1977) *Survey and Opinion Research: Procedures for Processing and Analysis*, Prentice-Hall, Englewood Cliffs, NJ.
- SPSS (1998) *AnswerTree 2.0. Guida per l'utente*, SPSS Inc., Chicago.
- STEINBERG D., COLLA P. L. (1995) *CART: Tree-Structured Nonparametric Data Analysis*, Salford Systems, San Diego, CA.
- TONINI G. (2001) *Analisi di segmentazione e categorie a rischio di non fare vacanza*, Rapporto n. 37, Dipartimento di Metodi quantitativi per l'Economia, Facoltà di Economia, Università degli Studi di Milano-Bicocca, Milano.
- VACH W. (1995) Classification trees, *Computational Statistics*, **10**: 9-14.

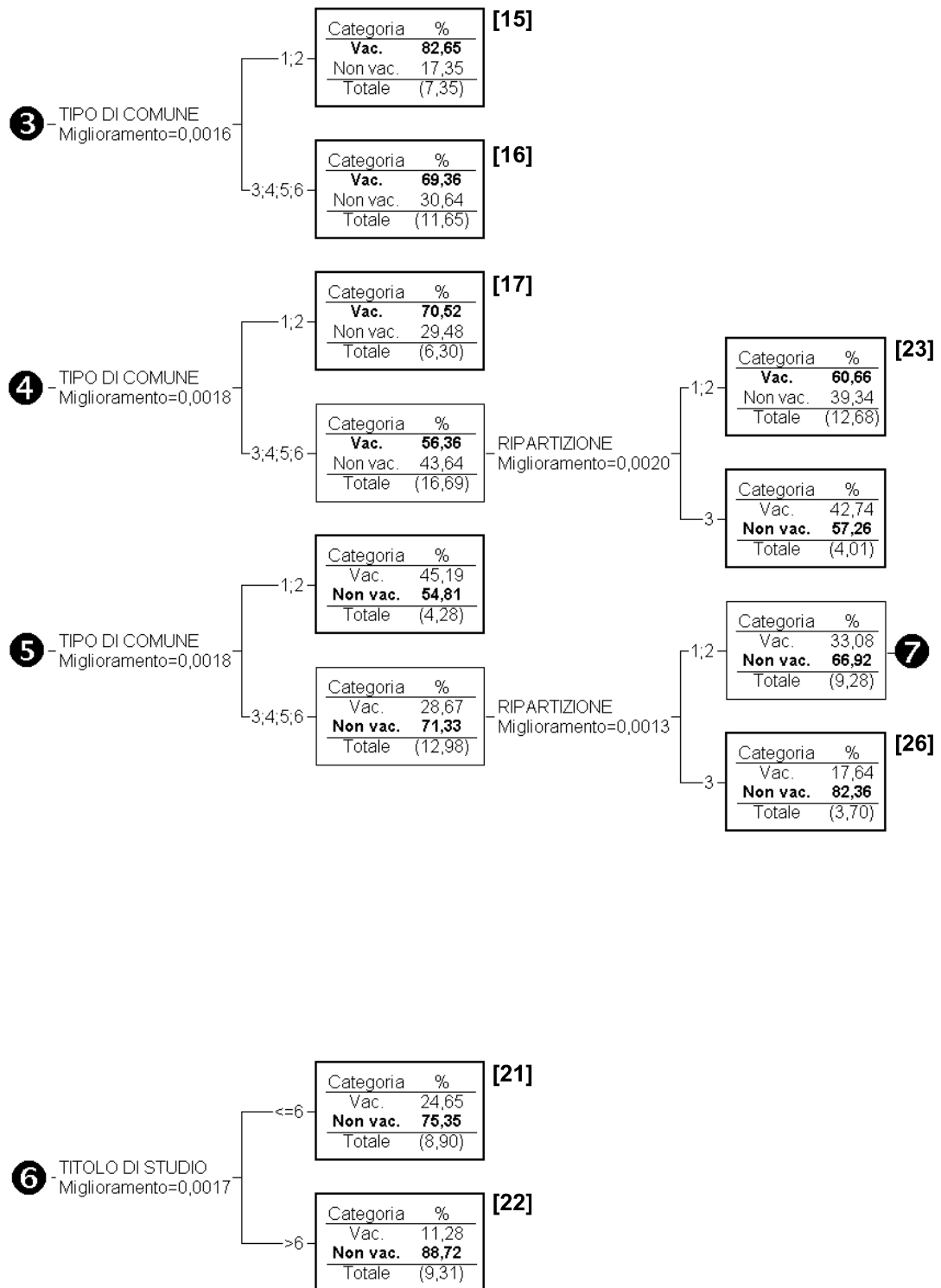
APPENDICE

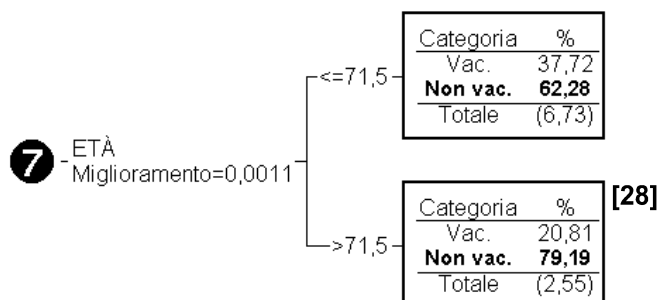
Figura 1 – Albero di segmentazione ottenuto operando con 9 predittori iniziali



Fonte: Nostre elaborazioni del file standard di tipo B relativo a ISTAT (1999).







Legenda:

RIPARTIZIONE GEOGRAFICA

- 1 Italia Nord-Occidentale
- 2 Italia Nord-Orientale
- 3 Italia Centrale
- 4 Italia Meridionale
- 5 Italia Insulare

TIPO DI COMUNE o DOMINIO

- 1 Comuni metropolitani (centro di area metropolitana)
- 2 Comuni limitrofi al centro di area metropolitana
- 3 Comuni con oltre 50.000 abitanti
- 4 Comuni con 10.001-50.000 abitanti
- 5 Comuni con 2001-10.000 abitanti
- 6 Comuni con al più 2000 abitanti

TITOLO DI STUDIO

- 1 Dottorato di ricerca o specializzazione post-laurea
- 2 Laurea
- 3 Diploma universitario
- 4 Diploma di scuola media superiore (4-5 anni)
- 5 Diploma di scuola media superiore professionale (2-3 anni)
- 6 Licenza di scuola media inferiore
- 7 Licenza elementare
- 8 Nessun titolo (sa leggere e scrivere)
- 9 Nessun titolo (non sa leggere e/o scrivere)

CONDIZIONE PROFESSIONALE

- 1 Occupato
- 2 In cerca di nuova occupazione
- 3 In cerca di prima occupazione
- 4 In servizio di leva o civile sostitutivo
- 5 Casalinga
- 6 Studente
- 7 Inabile
- 8 Ritirato dal lavoro
- 9 Altra condizione

***Social categories out and in holidays' market:
an application of classification trees to Italian case***

Summary: *In this paper the methods of classification trees are applied to the holidays' market of Italian population. The aims are the followings: 1) to select the most important factors of chance to not make holiday; 2) to classify the factors items in two clusters, the former with high rate of "non holiday-makers", and the latter with the low one; 3) to identify the social categories with the greatest or the smallest chance to not make holiday. The results of this analysis are interesting and useful not only in itself, but also for further developments of tourism studies.*

Keywords: *"Non holiday-makers"; classification trees; CART method.*