



Università
Ca'Foscari
Venezia

Corso di Laurea Magistrale
in Marketing e Comunicazione

Tesi di Laurea

Abitudini di lettura e piccoli e medi editori: un'analisi quantitativa

Relatore

Ch. Prof. Isabella Procidano

Laureanda

Sara Ferrarese
842195

Anno Accademico

2017 / 2018

Riassunto

Il mondo editoriale ha risentito in maniera importante della crisi dei consumi legata alla grande recessione degli ultimi 10 anni, la quale continua ad avere un grosso impatto sui ricavi del settore. Oltre alla crisi della domanda si è assistito anche all'ingresso nel mercato di nuovi competitors in ambito digitale, in risposta al mutamento delle tendenze di lettura; numerose le buone risposte di adeguamento da parte dei grandi players, in crisi le piccole e medie case editrici locali le quali spesso non dispongono delle risorse sufficienti per affiancare all'offerta tradizionale la controparte digitale.

La tesi ha come obiettivo l'analisi e lo studio approfondito delle abitudini di lettura e del consumo di libri in un'ottica di comprensione delle preferenze e delle tendenze di comportamento più diffuse dai potenziali clienti di una casa editrice di piccole o medie dimensioni. I dati oggetto dell'indagine sono stati raccolti attraverso la diffusione di un questionario in forma digitale articolato in più sezioni, utili ad esaminare a fondo l'atteggiamento del campione riguardo la frequenza di lettura, i canali informativi utilizzati, le preferenze di genere e di formato e relativi approfondimenti.

Indice

INTRODUZIONE	1
CAPITOLO 1 - PRODUZIONE DI LIBRI ED ABITUDINI DI LETTURA IN ITALIA: LETTERATURA E STATO DELL'ARTE	3
1.1 TENDENZE DEL SETTORE EDITORIALE E DOMANDA EFFETTIVA: DETTAGLIO SULLA NARRATIVA.....	4
1.2 LA DOMANDA DI LIBRI IN ITALIA: DETTAGLIO SOCIODEMOGRAFICO	5
1.2.1 I lettori	5
1.2.2 I non lettori	12
1.2.3 Incidenza del formato digitale	15
CAPITOLO 2 - MATERIALI E METODI PER L'INDAGINE: NOTA METODOLOGICA	19
2.1 OBIETTIVI CONOSCITIVI DELL'INDAGINE ED UTILIZZO PRATICO	19
2.2 STRUTTURA DEL QUESTIONARIO	19
2.3 MODALITÀ DI SOMMINISTRAZIONE E CAMPIONE	19
2.4 SCALE DI MISURA UTILIZZATE.....	20
2.5 DETTAGLIO DEI QUESITI RIVOLTI	21
2.5.1 Sezione 1: Introduzione	21
2.5.2 Sezione 2: Abitudini e preferenze di lettura	22
2.5.3 Sezione 3: Informazioni generali sui lettori	22
2.5.4 Sezione 4: Livello di gradimento dei generi narrativi	24
2.5.5 Sezione 5: Preferenze relative al formato di lettura	24
2.5.6 Sezione 6: Analisi della preferenza per il formato cartaceo	25
2.5.7 Sezione 8: Analisi della preferenza per il formato digitale	26
2.5.8 Sezioni 7 e 9: Choice set per formato cartaceo e digitale.....	26
2.5.9 Sezione 10: Profilo socio-demografico	30

CAPITOLO 3 - RISULTATI DELL'INDAGINE:

CARATTERIZZAZIONE SOCIODEMOGRAFICA DEL CAMPIONE..33

3.1 DETTAGLIO SOCIALE E DEMOGRAFICO DEI LETTORI	33
3.1.1 Caratteristiche sociali e demografiche dei lettori di formato cartaceo.....	37
3.1.2 Caratteristiche sociali e demografiche dei lettori di formato digitale	41
3.1.3 Analisi bivariate sui lettori	44
3.2 DETTAGLIO SOCIALE E DEMOGRAFICO DEI NON LETTORI.....	49

CAPITOLO 4 - RISULTATI DELL'INDAGINE: ANALISI UNIVARIATE

.....53

4.1 ANALISI UNIVARIATE GENERALI.....	53
4.1.1 Quesito q1: Negli ultimi 12 mesi ha letto libri per hobby?.....	53
4.1.2 Quesito q2: Quanti libri legge per hobby, indicativamente, all'anno?	54
4.1.3 Quesito q3: Tra i libri che ha letto, quanti provenivano dalla biblioteca o sono stati presi in prestito da conoscenti?	55
4.1.4 Quesito q4: Quanto spesso si ritaglia un po' di tempo per leggere?	56
4.1.5 Quesito q5: Quanto ritiene importanti i seguenti parametri quando deve scegliere un libro da leggere?.....	56
4.1.6 Quesito q6 Assegna una valutazione che esprima il Suo livello di gradimento per i generi presentati di seguito	64
4.1.7 Quesito q7: In che formato preferisce leggere?	76
4.2 ANALISI UNIVARIATE: QUESITI RISERVATI AI LETTORI DI FORMATO CARTACEO	77
4.2.1 Quesito q8: Assegna una valutazione che esprima l'importanza per i parametri presentati di seguito per la Sua scelta di un libro in formato cartaceo.....	77
4.2.2 Quesito q9: Quali sono i principali motivi per i quali preferisce leggere in formato cartaceo?.....	87
4.2.3 Quesito q10: Quali caratteristiche dei libri elettronici potrebbero convincerLa a cominciare ad utilizzare il formato digitale per leggere?.....	88
4.3 ANALISI UNIVARIATE: QUESITI RISERVATI AI LETTORI DI FORMATO DIGITALE.....	90
4.3.1 Quesito q16: Assegna una valutazione che esprima l'importanza da Lei assegnata ai parametri relativi al libro in formato digitale riportati di seguito	90

4.3.2 Quesito q17: Quali sono i principali motivi per i quali Lei preferisce leggere in formato digitale? (Massimo 3 alternative).....	99
----------------------------------------------------------------------------------------------------------------------------------------	----

CAPITOLO 5 - CANALI INFORMATIVI: ANALISI DELLE COMPONENTI PRINCIPALI ED ANALISI DEI GRUPPI..... 101

5.1 ANALISI DELLE COMPONENTI PRINCIPALI.....	101
5.1.1 Matrice di varianze e covarianze e matrice di correlazione	101
5.1.2 Calcolo degli autovalori e degli autovettori	103
5.1.3 Scelta del numero di componenti ottimale	104
5.1.4 Calcolo degli autovettori	107
5.1.5 Mappa degli individui e cerchio delle correlazioni	109
5.2 ANALISI DEI GRUPPI	115
5.2.1 Prove di clusterizzazione e bontà dei raggruppamenti	116
5.2.2 Importanza delle variabili di clusterizzazione e t test per il confronto tra le medie	118
5.2.3 Dendrogramma: struttura a due cluster	119
5.2.4 Silhouette: struttura a due cluster	120
5.2.5 Boxplot: struttura a due cluster.....	123
5.2.6 Diagramma di dispersione, confronto tra le medie e profilazione dei caratteri: struttura a due cluster.....	126

CAPITOLO 6 - ANALISI LOGIT CONDIZIONALE E PARAMETRI PER LA SCELTA DEL LIBRO 135

6.1 MODELLI A SCELTA DISCRETA E MODELLO LOGIT CONDIZIONALE	135
6.1.1 Assunzioni sulla variabile dipendente e le variabili esplicative	136
6.1.2 La proprietà di indipendenza delle alternative irrilevanti (IIA)	137
6.2 IMPLEMENTAZIONE DEL MODELLO LOGIT CONDIZIONALE STUDIO DELLA PROBABILITÀ DI SCELTA	137
6.2.1 Stima del modello, ottimizzazione della probabilità ed influenza del prezzo	139
6.2.2 Codifica delle variabili esplicative per le elaborazioni	139
6.3 BONTÀ DI ADATTAMENTO: CONFRONTO TRA MODELLI SCELTA-SPECIFICI ED INDIVIDUO-SPECIFICI.....	142

6.3.1 Il criterio di informazione di Akaike (AIC)	142
6.3.2 Il criterio di informazione Bayesiano.....	142
6.3.3 Confronto tra i due criteri di informazione	143

CAPITOLO 7 - REGRESSIONE LOGISTICA: RISULTATI E

CONSIDERAZIONI 145

7.1 SELEZIONE DEL MODELLO: PROVE EMPIRICHE	145
--------------------------------------------------	-----

7.2 LA REGRESSIONE LOGISTICA NEL MODELLO SCELTA-SPECIFICO	147
-----------------------------------------------------------------	-----

7.3 REGRESSIONE OPERATIVA DEI PARAMETRI NEL MODELLO LOGIT CONDIZIONALE E STIMA DEI COEFFICIENTI DELLE VARIABILI SCELTA-SPECIFICHE.....	148
-------------------------------------------------------------------------------------------------------------------------------------------	-----

7.3.1 Stima del modello scelta-specifico	149
------------------------------------------------	-----

7.3.2 Analisi della combinazione ottimale nel modello scelta-specifico	150
------------------------------------------------------------------------------	-----

CONCLUSIONI 155

BIBLIOGRAFIA E SITOGRAFIA..... 157

RINGRAZIAMENTI 159

Introduzione

Il presente elaborato ha l'obiettivo di studiare le abitudini di lettura e di consumo di libri analizzando dunque, attraverso un'analisi quantitativa, quali sono le principali tendenze che accomunano i lettori, per scoprire nel dettaglio le preferenze e le abitudini degli stessi. Il motivo di tale approfondimento è legato al fatto che la maggior parte delle case editrici operanti sul territorio italiano sono di piccole e medie dimensioni, e nella maggior parte dei casi esse si avvalgono limitatamente, o per nulla, di tecniche di pubblicità o di comunicazione e marketing efficaci, poiché dedicano poco tempo allo studio del comportamento dei clienti. Lo studio è stato condotto servendosi di un questionario, creato appositamente per approfondire il tema in modo specifico, il quale è stato somministrato attraverso il canale digitale ad un campione di 281 individui.

L'elaborato si suddivide in sette capitoli. Nel primo capitolo vengono presentate le tendenze generali della lettura in Italia, partendo dalle ultime indagini ed elaborazioni Istat, le quali analizzano a livello generale la situazione nel territorio italiano. Il secondo capitolo consiste nella descrizione dei materiali e dei metodi attraverso i quali è stata condotta l'indagine, specificando la struttura del questionario, le modalità di somministrazione dello stesso e la presentazione dei quesiti rivolti. Nel terzo capitolo viene descritta la caratterizzazione socio-demografica del campione, suddiviso tra lettori (di formato cartaceo e digitale) e non lettori. Nel quarto capitolo vengono esposti i risultati relativi ai quesiti generali quali la frequenza di lettura, la propensione all'acquisto, i generi letterari ed i canali informativi preferiti nonché il formato di lettura prediletto. Nel quinto capitolo vengono presentati i risultati relativi allo studio approfondito dei canali informativi utilizzati più frequentemente dai lettori, ottenuti attraverso l'analisi delle componenti principali e l'analisi dei gruppi. Nel sesto capitolo viene proposta la spiegazione teorica del modello logit di scelta discreta utilizzato per analizzare l'incidenza del prezzo, della disponibilità delle recensioni, della nazionalità dell'autore e della novità del libro sulla scelta di acquisto. Nel settimo ed ultimo capitolo vengono descritti e presentati nel dettaglio i risultati ottenuti attraverso l'implementazione del modello a scelta discreta.

Capitolo 1

Produzione di libri ed abitudini di lettura in Italia: letteratura e stato dell'arte

Il settore librario italiano è composto in larga prevalenza da realtà di piccole e medie dimensioni: in base alle rilevazioni Istat¹, di fatto, essi rappresentano circa l'86% degli operatori sul mercato, di cui il 63,721% sono piccole case editrici che pubblicano in un anno fino ad un massimo di dieci titoli, mentre il 36,744% sono editori di medie dimensioni, i quali ne pubblicano fino a 50.

Le case editrici di grandi dimensioni rappresentano una minoranza in Italia: esse costituiscono appena il 14% degli operatori del settore, ma nonostante la loro inferiorità numerica detengono il primato in termini di produzione libraria in quanto si occupano della pubblicazione del 76,1% dei titoli, contro il 5,436% dei titoli stampati e distribuiti dagli editori di piccole dimensioni.

Per quanto riguarda la tiratura, intesa come numero di copie stampate per singola edizione, il primato rimane alle grandi case editrici, che si occupano di circa l'86% delle copie stampate, mentre i medi e piccoli editori si dividono rispettivamente il 12,285% e del 2,77% della tiratura rimanente.

Il confronto tra la media di titoli pubblicati dai piccoli e dai medi editori ha evidenziato come i primi siano in netto svantaggio rispetto ai secondi; in particolare le grandi case editrici hanno pubblicato una media di 228 opere letterarie mentre i piccoli editori si attestano ad una media di 4 titoli pubblicati in un anno.

A partire dal 1996 si è registrata una riduzione del totale degli operatori attivi di piccole dimensioni: essi costituivano circa il 67% del totale alla fine degli anni '90, mentre nel 2016 la loro quota di presenza nel settore si è attestata attorno al 55%. Una tendenza radicalmente opposta si è osservata per quanto riguarda invece i grandi players, il cui numero è aumentato dal 25% alla fine degli anni '90 fino al 32% nel 2016. Anche i medi editori hanno seguito negli ultimi decenni un andamento di crescita, nell'ordine di 5,8

¹ Dati riferiti all'anno 2016/2017

punti percentuali, passando dall'8,2% di operatori attivi a circa il 14%. Nello stesso arco temporale la produzione di libri ha conosciuto un aumento del 20% del numero di titoli emessi sul mercato; allo stesso tempo si è registrata una tendenza inversa per la tiratura, diminuita di oltre la metà.

Per quanto concerne la distribuzione degli editori sul territorio nazionale, si osserva una robusta polarizzazione: il 53,4% degli editori attivi opera nel Nord Italia, il 30% di essi ha sede nell'Italia centrale mentre circa il 17% di essi opera al Sud e nelle Isole.

Il 30% degli editori operativi sul suolo italiano si concentrano a Roma e Milano, le quali accolgono il 44% delle grandi case editrici.

1.1 Tendenze del settore editoriale e domanda effettiva: dettaglio sulla narrativa

Il mondo editoriale ha risentito in maniera importante della crisi dei consumi legata alla grande recessione degli ultimi 10 anni, la quale continua ad avere un grosso impatto sui ricavi del settore. Oltre alla crisi della domanda si è assistito anche all'ingresso nel mercato di nuovi competitors in ambito digitale, in risposta al mutamento delle tendenze di lettura; numerose le buone risposte di adeguamento da parte dei grandi players, spesso in crisi le piccole e medie case editrici locali le quali spesso non dispongono delle risorse sufficienti per affiancare all'offerta tradizionale la controparte digitale.

La situazione è aggravata anche dalle abitudini di lettura da parte della popolazione italiana, le quali evidenziano una bassa propensione a leggere: nel 2016 le persone di 6 anni o più che hanno letto almeno un libro nei 12 mesi precedenti l'intervista costituivano circa il 40%, per un totale di quasi 23 milioni di individui, in diminuzione di 1,5 punti percentuali rispetto all'anno precedente.

Il "primo rapporto sulla promozione della lettura in Italia", proposto dall'Associazione Forum del libro, ha messo in luce che il tasso di penetrazione delle opere letterarie nelle famiglie italiane, dal 1996 ad oggi, è rimasto pressoché immutato fino agli inizi del nuovo millennio, a partire dal quale ha subito leggere fluttuazioni dell'ordine di un punto percentuale.

L'indagine Istat relativa all'anno 2016/2017 ha inoltre messo in evidenza una drastica diminuzione dei consumi, quantificati come spesa media mensile, in relazione al settore culturale che comprende, tra le altre cose, anche la lettura di libri per motivi non strettamente scolastici; il calo più significativo si è registrato nell'anno 2013, in cui la

spesa media mensile per i libri non scolastici è passata da 4,80€ a 3,12€, riducendosi complessivamente del 35%.

In generale il problema più grave risulta essere la ristretta domanda effettiva per quanto riguarda il settore della lettura, la quale risulta fortemente contenuta e localizzata, nonché fondamentalmente statica.

1.2 La domanda di libri in Italia: dettaglio sociodemografico

Nei seguenti paragrafi verranno presentati i profili sociali e demografici dei lettori e dei non lettori secondo le rilevazioni Istat, corredate di un approfondimento relativo all'incidenza del formato digitale nelle abitudini di lettura.

1.2.1 I lettori

L'indagine Istat sulle abitudini dei lettori ha messo in luce una propensione alla lettura abbastanza stabile per tutte le fasce d'età; nel 2016 coloro che hanno letto almeno un libro per motivi di svago ed intrattenimento nel corso dell'anno considerato costituivano il 40,5% del campione e nel 2017 questa tendenza si è riproposta, comprendendo il 41% dei rispondenti.

Le donne risultano tendenzialmente più interessate alla lettura rispetto agli uomini: i lettori di sesso maschile, infatti, hanno costituito il 33,5% del totale, contro il 47,1% delle lettrici donne, come rappresentato nella Figura 1.1.

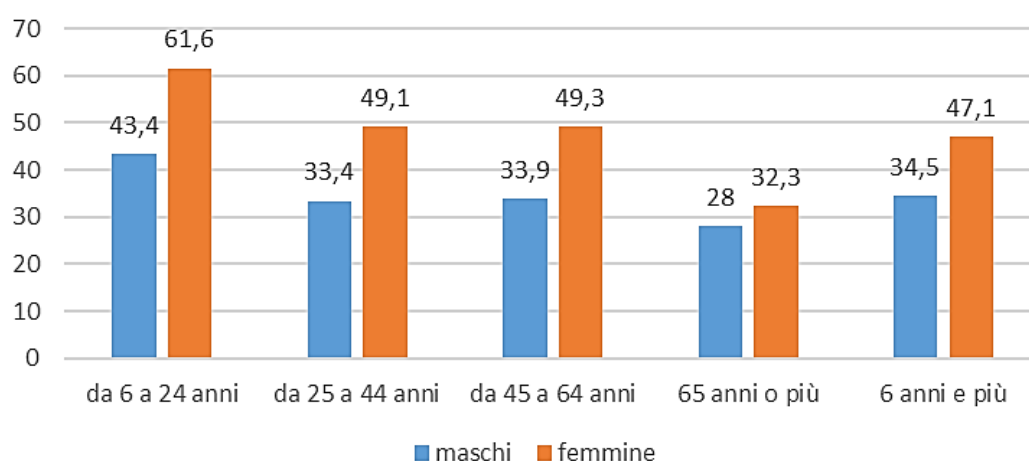


Figura 1.1 Percentuale di persone con più di 6 anni di età che hanno letto almeno un libro nel tempo libero nei 12 mesi precedenti all'intervista, suddivise per sesso e classe di età.

Le disparità a livello geografico, presenti da sempre per quanto riguarda la lettura, persistono. Dalla Figura 1.2 si evidenzia come il Nord Italia detenga il primato per quanto riguarda la percentuale di lettori, i quali costituiscono il 48,4% degli abitanti, il 49%

considerando la circoscrizione al Nord Est. In Italia centrale i lettori si attestano al 44,5% degli abitanti mentre il Mezzogiorno e le Isole registrano una percentuale di lettori sul totale degli abitanti pari rispettivamente al 28,3% ed al 30,4% Figura 1.2.

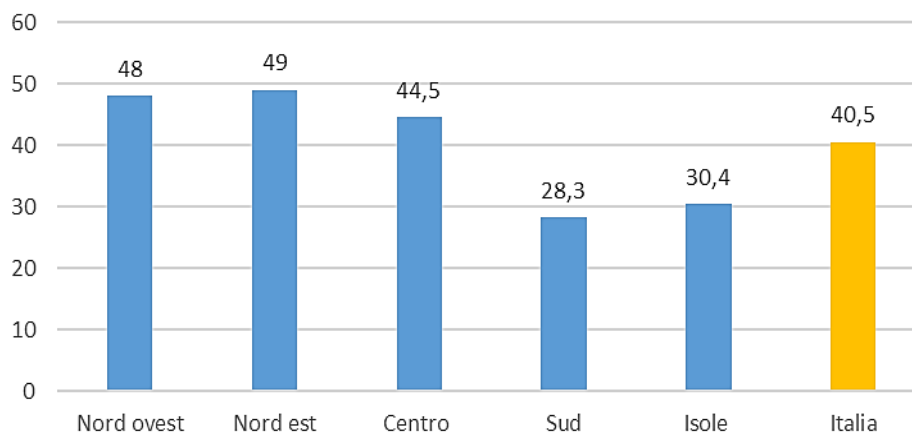


Figura 1.2 Percentuale di persone con più di 6 anni che hanno letto un libro nel tempo libero nei 12 mesi precedenti all'intervista suddivise per area geografica.

Si notano anche delle differenze per quanto riguarda la caratterizzazione urbana: i comuni metropolitani limitrofi al centro città registrano percentuali più alte di lettori (49,1%) rispetto ai comuni situati nelle zone periferiche (circa il 42%). Inoltre le zone che contano un numero di abitanti inferiore a 2000 sono quelle in cui si legge di meno (si dedica alla lettura il 37% degli abitanti) come mostra la Figura 1.3.

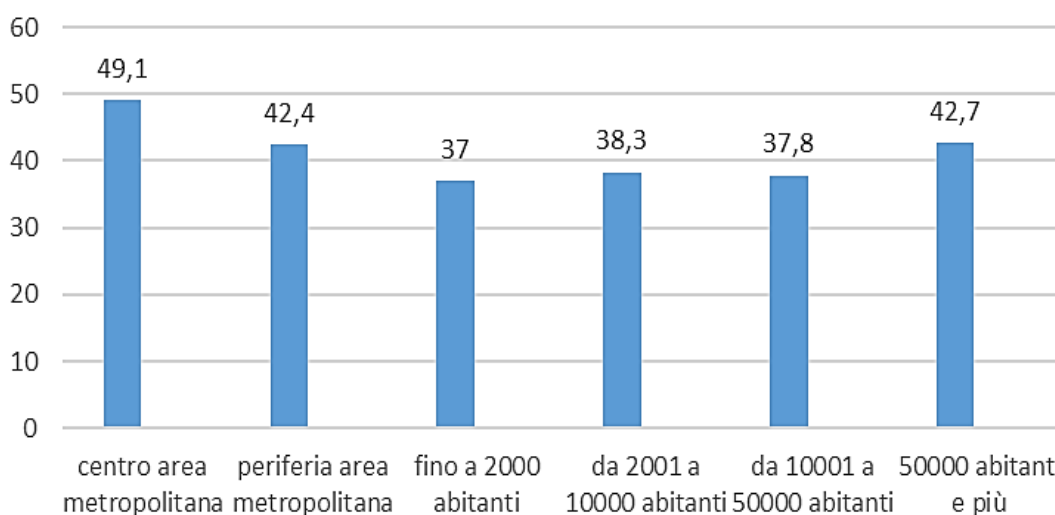


Figura 1.3 Percentuale di persone con più di 6 anni che hanno letto un libro nel tempo libero nei 12 mesi precedenti all'intervista ripartiti per numero di abitanti del comune di residenza e posizione geografica del comune.

Un'altra variabile che incide molto sulle abitudini di lettura è rappresentata dal titolo di studio, infatti gli individui che possiedono un livello di istruzione più elevato leggono mediamente di più rispetto agli individui che possiedono titoli di studio più bassi. Il dettaglio è presentato nella Figura 1.4, in cui si può notare che si dedica alla lettura ben il 71,8% tra i possessori di una laurea, il 49,4% degli individui diplomati ed il 27,6% degli individui che hanno frequentato solamente la scuola primaria.

Un ulteriore dato interessante consiste nella maggioranza di lettori tra le persone che presentano un livello occupazionale di prestigio, oppure che non hanno ancora concluso il loro percorso di studi, come mostrato dalla Figura 1.4.

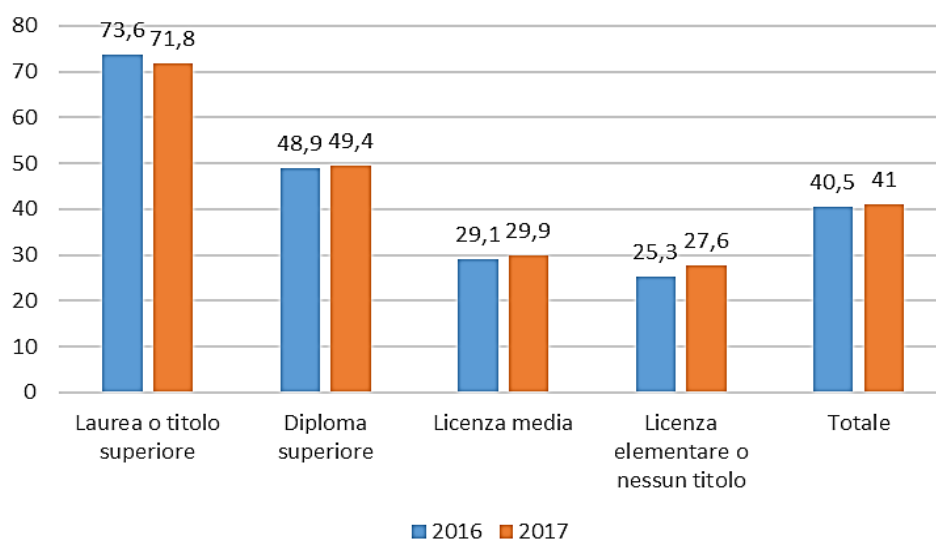


Figura 1.4 Percentuale di persone con più di 6 anni che hanno letto un libro nel tempo libero nei 12 mesi precedenti all'intervista suddivisi per titolo di studio.

Dalla Figura 1.5 si evidenzia come le persone che leggono di più svolgono mansioni di lavoro dipendente impiegatizio o datore di lavoro / membro del direttivo (58,6% e 60,9%) oppure studiano (60,8%).

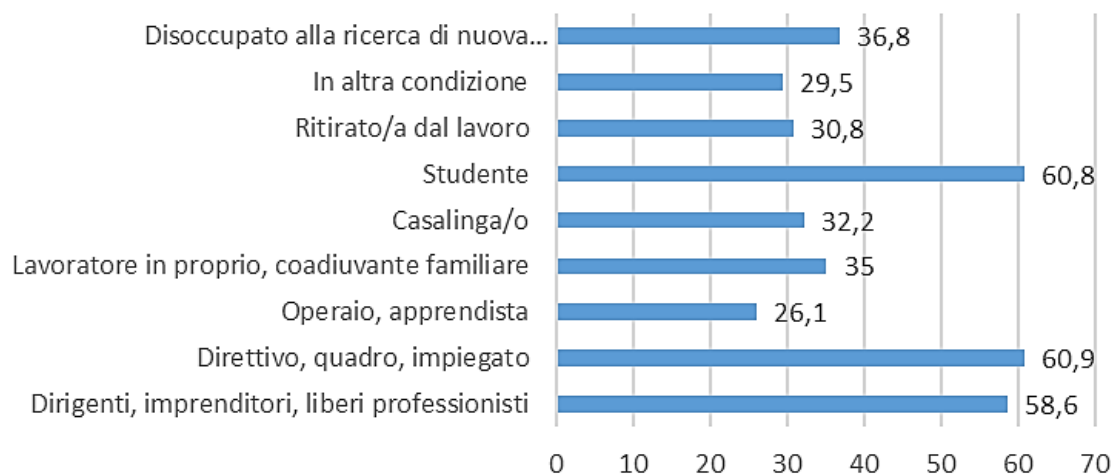


Figura 1.5 Percentuale di persone con più di 6 anni che hanno letto un libro nel tempo libero nei 12 mesi precedenti all'intervista suddivisi per condizione lavorativa.

La categoria che dedica meno tempo alla lettura consiste nella categoria degli operai, con una percentuale totale del 26,1%

Lo studio dei lettori si declina nella suddivisione di essi tra *lettori forti* e *lettori deboli* utilizzata dall'Istat per le rilevazioni:

- sono definiti *lettori forti* coloro che hanno letto più di 12 libri nei 12 mesi precedenti l'intervista;
- sono definiti *lettori deboli* coloro che hanno letto da 1 a 3 libri nei 12 mesi precedenti l'intervista.

La classe intermedia tra le due costituisce la categoria dei *lettori medi*, i quali hanno letto da 4 a 11 libri nel corso dei 12 mesi precedenti l'intervista.

Le rilevazioni operate dall'Istat hanno messo in luce la netta maggioranza di lettori deboli rispetto ai forti: questi ultimi costituiscono il 13,4% dei lettori, rispetto al 47,6% dei lettori deboli. Il dettaglio per sesso indica una percentuale di lettrici forti maggiore rispetto alla percentuale di lettori forti di sesso maschile, per un totale di 14,2% per le prime e 12,2% per i secondi. Per quanto riguarda la classe dei lettori deboli, le donne sono in minoranza (45%) rispetto agli uomini (51,3%), come dimostrato dalla Figura 1.6.

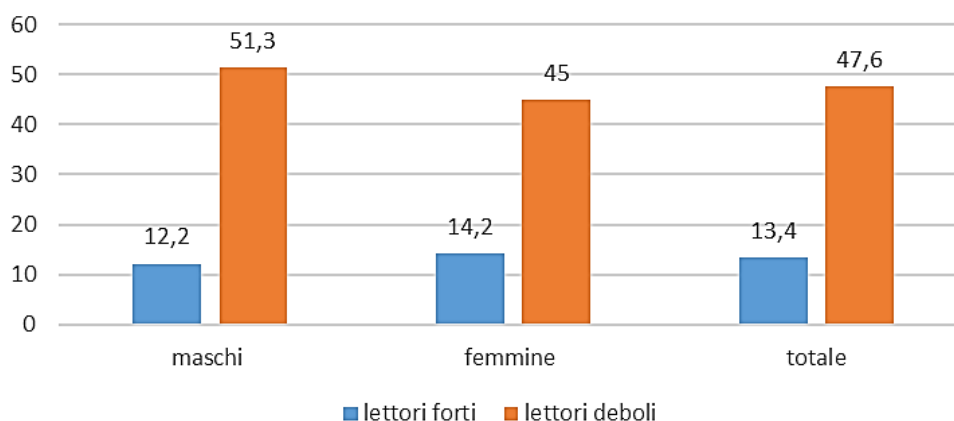


Figura 1.6 Percentuale di persone con più di 6 anni che hanno letto un libro nel tempo libero nei 12 mesi precedenti all'intervista suddivisi per sesso e per lettori "forti" e "deboli".

Il dettaglio relativo al titolo di studio mette in luce il fatto che la maggioranza di lettori di classe debole si colloca in corrispondenza di titoli di studio inferiori al diploma superiore, infatti il 53,3% dei lettori deboli è in possesso della licenza media ed il 54,6% possiede la licenza elementare o nessun titolo di studio. Tra i lettori forti, invece, la maggioranza consiste negli individui che possiedono una laurea od un titolo di studio superiore alla laurea, i quali rappresentano il 19,6%, come si evince dalla Figura 1.7.

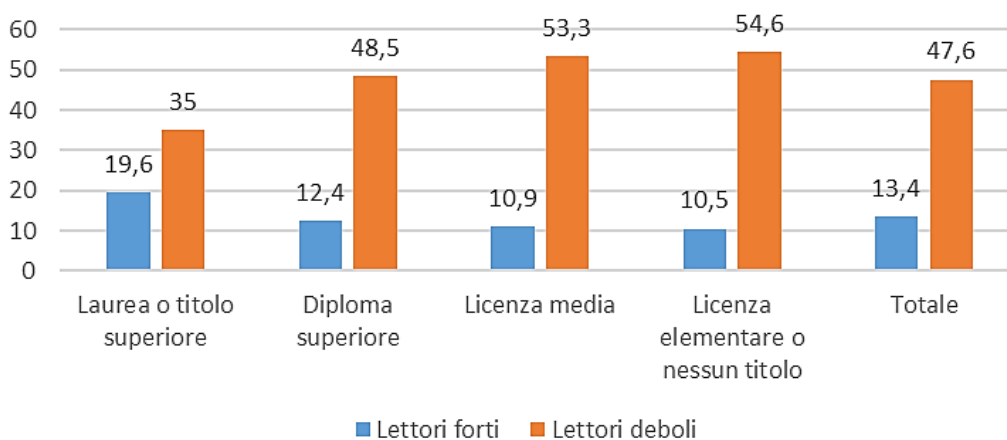


Figura 1.7 Percentuale di persone con più di 6 anni che hanno letto un libro nel tempo libero nei 12 mesi precedenti all'intervista suddivisi per lettori "forti" e "deboli" e per titolo di studio.

Analizzando la collocazione geografica dei lettori forti e deboli sul territorio nazionale, si è evinto che la maggior parte dei lettori deboli, per un totale del 59,5%, risiede nel Sud Italia. La maggioranza dei lettori forti, invece, si colloca nel Nord Ovest italiano costituendo il 16,6%. I dati riportati nella Figura 1.8 mettono dunque in evidenza che al Nord si concentra la maggioranza di persone che leggono più di 12 libri all'anno, mentre al Sud la maggior parte dei lettori tende a leggere da 1 a 3 libri l'anno.

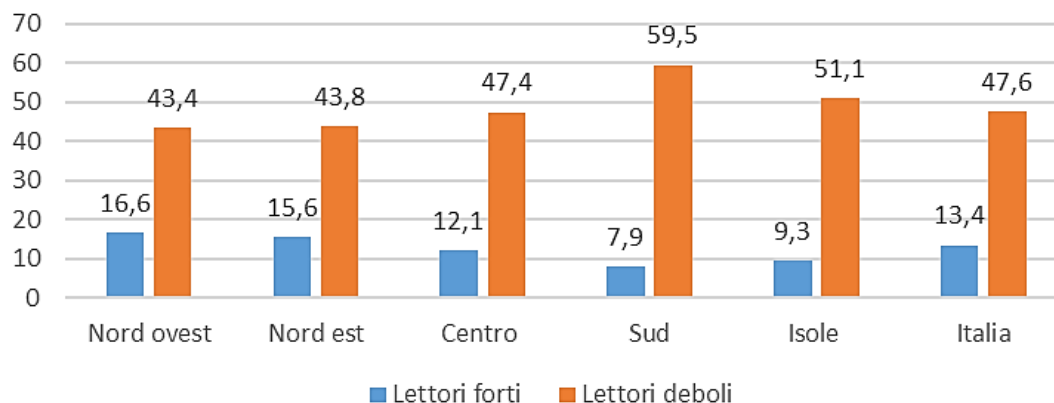


Figura 1.8 Percentuale di persone con più di 6 anni che hanno letto un libro nel tempo libero nei 12 mesi precedenti all'intervista suddivisi per lettori "forti" e "deboli" e per area geografica.

Per ciò che concerne la suddivisione urbana e territoriale, la maggioranza di lettori deboli si colloca nelle aree di periferia, lontane dal centro città (48,3%) mentre la maggior parte dei lettori forti risiede nelle aree urbane nei pressi del centro città (14,9%). Si notano delle differenze anche per quanto riguarda la densità di popolazione dei comuni: la maggior parte dei lettori forti risiede nei comuni popolati da meno di 2000 abitanti (14,4%) mentre i comuni che contano da 2001 a 10000 abitanti ospitano la maggior parte dei lettori deboli (48,9%) come mostrato nella Figura 1.9.

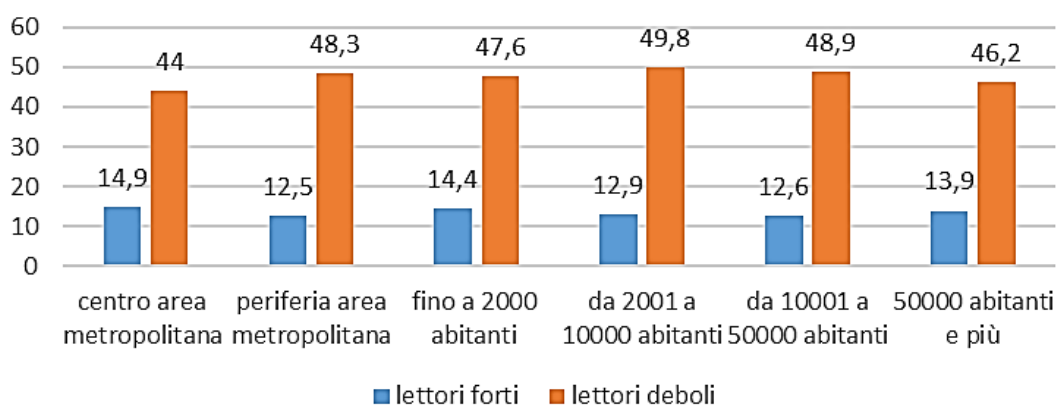


Figura 1.9 Percentuale di persone con più di 6 anni che hanno letto un libro nel tempo libero nei 12 mesi precedenti all'intervista suddivisi per lettori "forti" e "deboli" e per numero di abitanti del comune di residenza o posizione geografica del comune.

L'ambito occupazionale presenta forti differenziazioni tra lettori forti e lettori deboli: tra la classe debole, il 57,2% svolge una mansione lavorativa esplicabile nella definizione di

operaio o apprendista, ed il 55,7% rientra nella categoria di coloro in cerca di una prima occupazione. Tra i lettori forti, invece, il 17,5% rientra nella categoria dei ritirati dal lavoro ed il 17,1% appartiene alla categoria dei dirigenti, imprenditori e liberi professionisti, come rappresentato nella Figura 1.10.

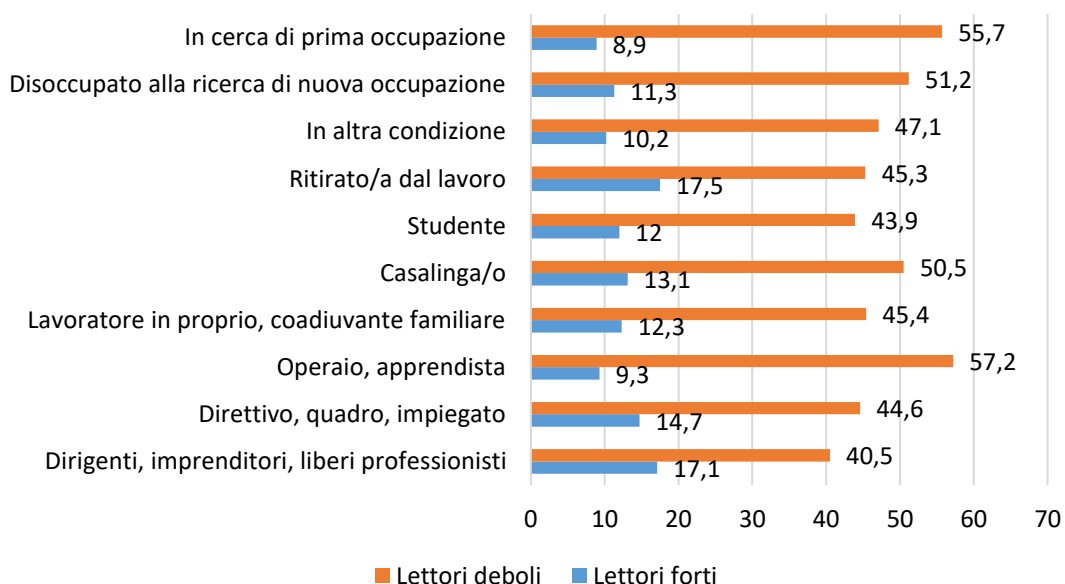


Figura 1.10 Percentuale di persone con più di 6 anni che hanno letto un libro nel tempo libero nei 12 mesi precedenti all'intervista suddivisi per lettori "forti" e "deboli" e per condizione occupazionale.

Focalizzandosi sulla categoria dei lettori deboli, suddividendo il campione in base all'età ed al sesso, la percentuale più elevata di lettori di classe debole risulta collocarsi in corrispondenza dei lettori maschi nella fascia di età compresa tra 6 e 24 anni, per un totale del 55,7%. Per quanto riguarda la categoria dei lettori deboli di sesso femminile, la percentuale più elevata si colloca in corrispondenza della fascia di età che descrive le donne dai 25 ai 44 anni, la quale descrive il 47,3% del campione femminile, come mostrato nella Figura 1.11.

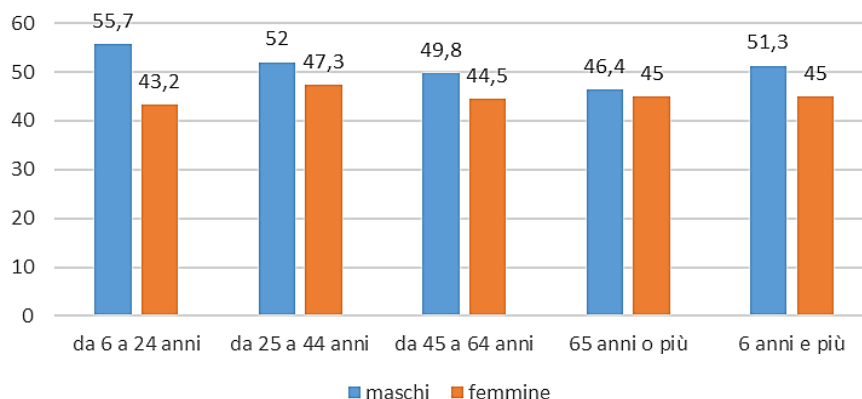


Figura 1.11 Percentuale dei lettori “deboli” suddivisi per classi di età e sesso.

La categoria dei lettori forti è rappresentata nella Figura 1.12 e presenta valori simili per ogni classe di età, fatta eccezione per la fascia comprendente gli individui da 6 a 24 anni, in cui gli uomini si attestano all’8,6% mentre le donne al 13,5%.

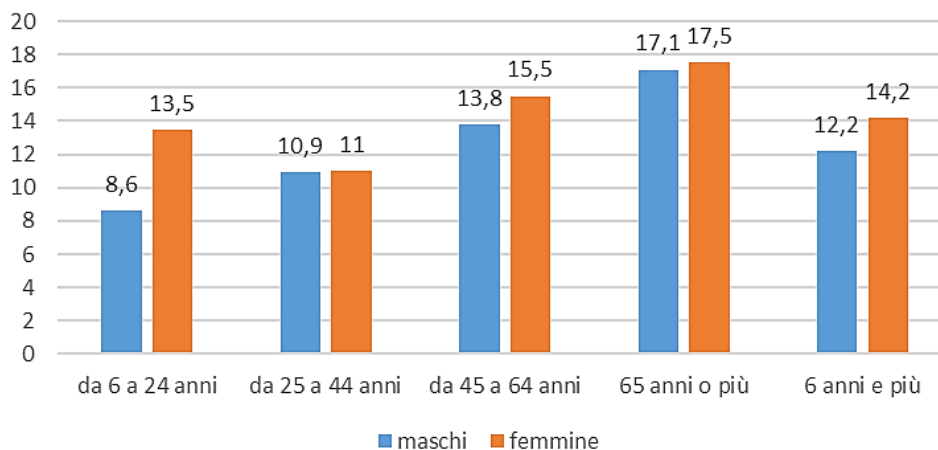


Figura 1.12 Percentuale dei lettori “deboli” suddivisi per classi di età e sesso.

I valori più elevati si registrano per la fascia di età comprendente le persone di 65 anni o più, in cui il 17,1% sono uomini ed il 17,5% sono donne. (Figura 1.12)

1.2.2 I non lettori

La percentuale di non lettori, descritti come coloro i quali hanno dichiarato di non aver letto nemmeno un libro nei 12 mesi precedenti l’intervista, dalla si osserva come dalla Figura 1.13 si attesti al 59%, dato che segnala una situazione allarmante sia a livello sociale che culturale, che persiste insistentemente da oltre quindici anni senza mostrare cenni di ripresa.

Coloro che non si dedicano alla lettura ricoprono percentuali al di sopra del 50% in 18 regioni su 20, con picchi in corrispondenza della regione Calabria, nella quale non legge il 73,9% degli abitanti, e della regione Sicilia, in cui i non lettori arrivano a quota 74,2%.

La regione in cui si legge di più è rappresentata dal Trentino Alto Adige, in cui si registra una percentuale di non lettori minore rispetto a tutte le altre, per un totale del 46,9%.

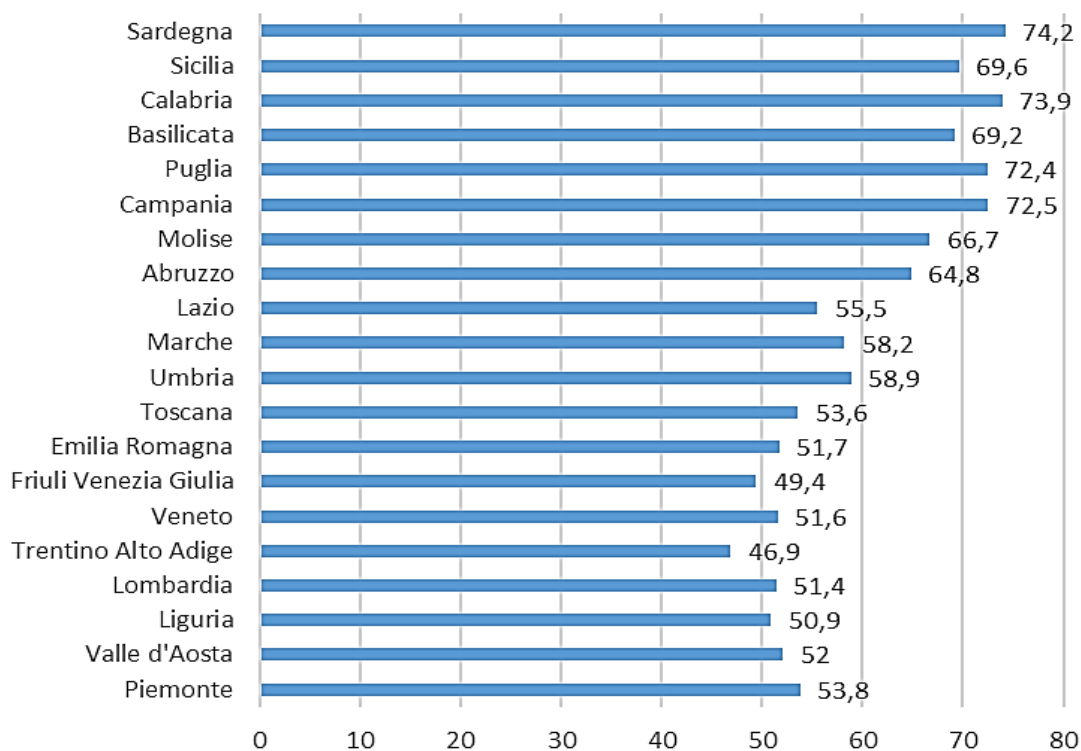


Figura 1.13 Percentuale di persone con più di sei anni che non hanno letto nemmeno un libro nei 12 mesi precedenti all'intervista suddivisi per regione di residenza.

Sulle abitudini di lettura incidono anche i contesti territoriali urbani, infatti nelle aree periferiche della città la lettura è meno diffusa rispetto alle aree del centro: i non lettori residenti in aree lontane dal centro città rappresentano il 57,6%, mentre i non lettori residenti in aree limitrofe al centro urbano rappresentano il 50,9%. Anche la densità di popolazione del comune decreta delle differenze in termini di non lettori: nei comuni meno popolosi (fino a 2000 abitanti) la quota dei non lettori arriva al 63%, mentre nei comuni più popolosi (oltre i 50000 abitanti) essi sono inferiori di numero di 5,7 punti percentuali, come da Figura 1.14.

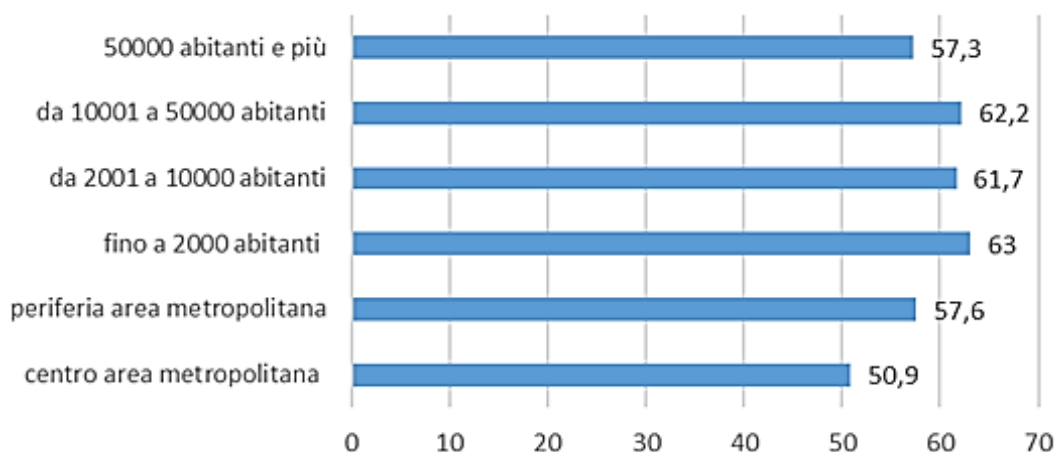


Figura 1.14 Percentuale di persone con più di sei anni che non hanno letto nemmeno un libro nei 12 mesi precedenti all'intervista suddivisi per numero di abitanti del comune di residenza o posizione geografica del comune.

La percentuale più elevata di non lettori si registra in corrispondenza della fascia di età che considera gli individui di 65 anni o più, in cui il 72% sono di sesso maschile mentre il 67,7% sono di sesso femminile. La fascia di età in cui i non lettori sono meno presenti comprende le persone tra i 6 ed i 24 anni, con una maggioranza di non lettori maschi (56,6%) rispetto alle femmine (38,4%), come si vede nella Figura 1.15.

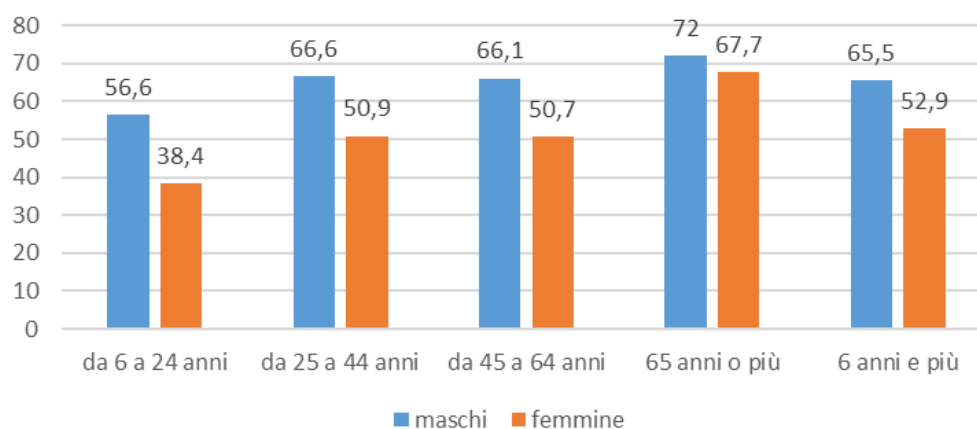


Figura 1.15 Percentuale di persone con più di sei anni che non hanno letto nemmeno un libro nei 12 mesi precedenti all'intervista suddivisi per sesso ed età.

Gli intervistati in possesso di un titolo di studio almeno pari alla licenza media sono quelli che leggono di meno, per i quali il valore percentuale dei non lettori risulta più elevato: in particolare il 72,4% di coloro in possesso della licenza elementare ed il 70,1% di coloro in possesso della licenza media, come si nota dalla Figura 1.16.

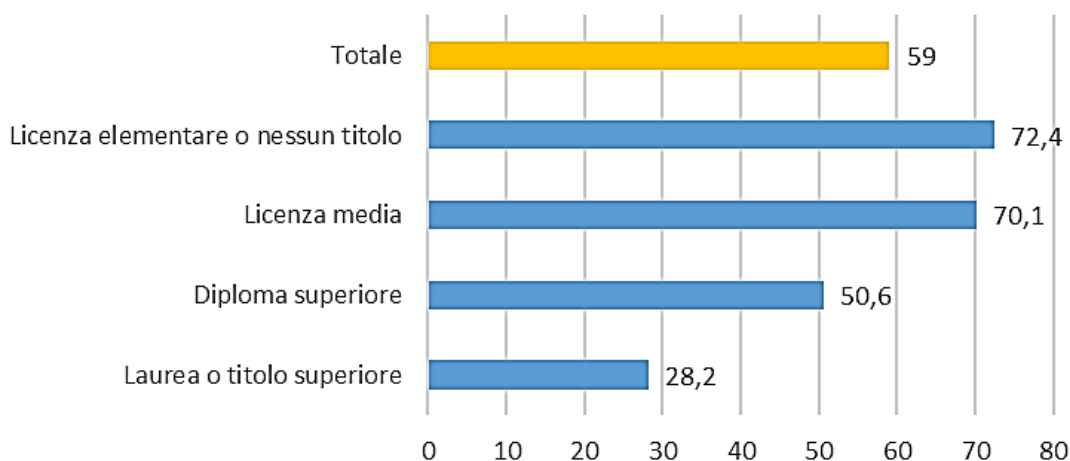


Figura 1.16 Percentuale di persone con più di sei anni che non hanno letto nemmeno un libro nei 12 mesi precedenti all'intervista suddivisi per titolo di studio.

Tra gli individui in possesso di titoli di studio di livello più alto, la quota di non lettori si riduce drasticamente fino ad arrivare al 28,2% nel caso delle persone in possesso di laurea o titoli di livello più elevato, come da Figura 1.16.

Tra gli intervistati, la quota maggiore di non lettori si registra in corrispondenza della condizione di operaio o apprendista, per un totale del 73,9%. Le mansioni in corrispondenza delle quali la percentuale di non lettori risulta più bassa corrispondono agli impieghi come direttivo, quadro o impiegato (39,1%) e come studente (39,2%), come rappresentato nella Figura 1.17.

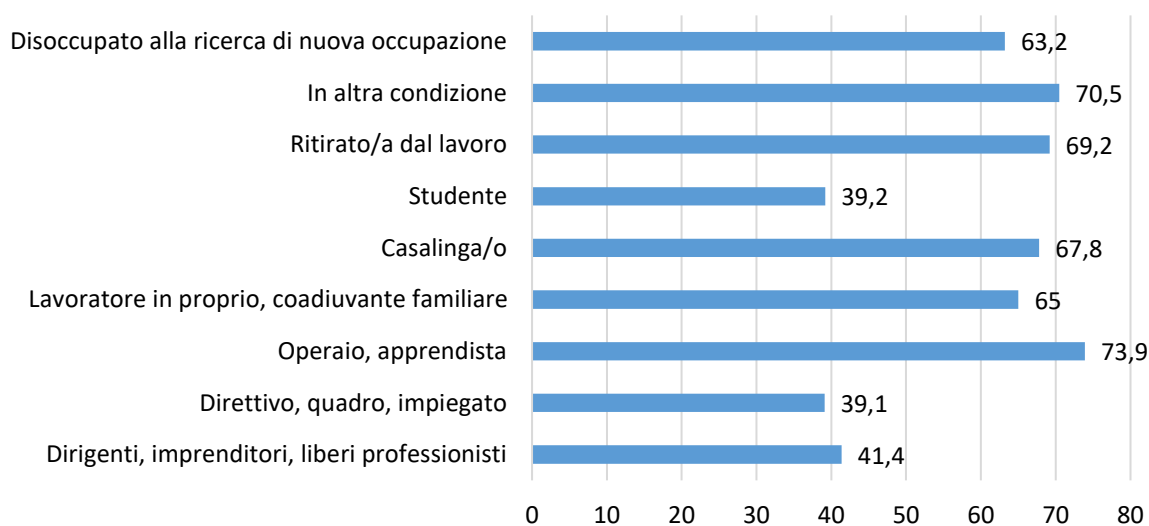


Figura 1.17 Percentuale di persone con più di sei anni che non hanno letto nemmeno un libro nei 12 mesi precedenti all'intervista suddivisi condizione lavorativa.

1.2.3 Incidenza del formato digitale

Per quanto riguarda la lettura digitale e gli e-book, si è registrato negli ultimi anni un continuo e graduale progresso: nell'anno 2016 più di 20000 titoli sono stati offerti alla clientela anche in formato digitale, per un totale del 35,8% dei titoli pubblicati in Italia nell'anno in questione.

Relativamente all'offerta di titoli estesa al formato digitali, i libri per i quali è stata resa disponibile anche la versione e-book rientrano soprattutto nel genere avventura e giallo (l'82,7% dei volumi cartacei).

La maggior parte delle versioni digitali dei libri cartacei è stata pubblicata da grandi case editrici, per una quota dell'88,3%, di cui il 14,3% è caratterizzato dalla presenza di funzionalità supplementari disponibili solo per la versione elettronica, le quali consistono in collegamenti ipertestuali e contenuti multimediali.

Per quanto riguarda le particolarità che riscontrano un maggiore apprezzamento da parte dei lettori secondo le case editrici, sono stati riportati i seguenti risultati in ordine di importanza conferita:

- il prezzo (73,9% delle segnalazioni);
- comodità di trasporto ed archiviazione (52% delle segnalazioni);
- presenza di contenuti interattivi (19% delle segnalazioni);
- semplicità di reperimento ed acquisto dei libri (17,5% delle segnalazioni);
- contenuti audio e video multimediali (12,2% delle risposte).

Per ciò che concerne le particolarità che, secondo le case editrici, scoraggiano l'acquisto dei libri digitali da parte dei lettori sono stati riportati i seguenti risultati, ordinati secondo l'importanza conferita:

- immaterialità dell'e-book (46% delle segnalazioni);
- insufficiente alfabetizzazione informatica (44,1% delle segnalazioni);
- lettura non confortevole alla vista (30% delle segnalazioni);
- scarsa diffusione dei lettori forti (25% delle segnalazioni);
- costo degli e-reader (19,4% delle segnalazioni);

Nell'anno 2016 il 25% delle case editrici si è occupata della digitalizzazione dei titoli di listino, della quale si sono occupati per la maggior parte i grandi editori (48,5%) e per una parte residua le case editrici di piccole dimensioni (18,2%).

Nonostante l'implementazione da parte dei grandi editori, l'offerta di libri digitali detiene tuttavia un ruolo secondario, ciò argomentato dal fatto che, per il 94% degli operatori del

settore, il ricavato delle vendite di formato digitale sul totale del fatturato non va oltre il 10%.

Le rilevazioni Istat sulla fruizione degli e-book hanno messo in luce che il 7,3% dei lettori si sono dedicati alla lettura dell'e-book, e l'11,1% ha scaricato libri in formato digitale, percentuale in aumento rispetto all'anno precedente di 2,9 punti percentuali.

È stata evidenziata una corrispondenza tra la fruizione di libri cartacei e di libri digitali, poiché le rilevazioni hanno dimostrato che la quota di persone che si dedica alla lettura di libri elettronici accresce in proporzione alla quantità di libri presenti in casa. La percentuale di non lettori che hanno fruito di titoli digitali è pari al 4%, mentre tra i lettori di titoli in formato cartaceo, il 21,6% ha affermato di avere scaricato o letto libri in formato elettronico, il cui numero cresce in proporzione all'aumento della quantità di libri letti nell'arco dell'anno in corso (per i lettori deboli la percentuale si attesta al 15%, mentre per i lettori forti è pari al 33,7%).

Si evince dunque dalle elaborazioni Istat che la lettura di libri in formato cartaceo ed in formato digitale sono legate da una corrispondenza positiva, che non le rende esclusive: i fruitori del formato elettronico sono, in larga parte, anche fruitori del formato stampato. Per quanto riguarda il dettaglio sociodemografico dei fruitori, si sono evidenziate le seguenti peculiarità:

- la fruizione di libri in formato digitale ha più larga diffusione tra i giovani, infatti il 69,4% dei ragazzi tra i 15 ed i 24 anni ha dichiarato di servirsi del formato e-book;
- la fruizione di titoli digitali è più diffusa nel Nord Ovest italiano (14%) e meno nel Sud Italia (8,1%);
- la fruizione di titoli in formato e-book è più radicata nei centri urbani (15,3%) rispetto ai territori periferici meno densamente popolati (8,7% nelle zone fino a 10000 abitanti).

Capitolo 2

Materiali e metodi per l'indagine: nota metodologica

2.1 Obiettivi conoscitivi dell'indagine ed utilizzo pratico

Il sondaggio si pone come obiettivo principale quello di tracciare una panoramica sul consumo del prodotto libro, in riferimento al campione raggiunto attraverso modalità online, per verificare eventuali corrispondenze con le indagini svolte dall'Istituto Nazionale di Statistica e riportate nel Capitolo 1 della presente tesi.

2.2 Struttura del questionario

I dati analizzati in questa sede sono stati raccolti attraverso la somministrazione di un questionario creato ad hoc per lo studio di particolari variabili e tendenze che riguardano il comportamento dei lettori. Il questionario si divideva in due parti:

- prima parte, che conteneva domande generiche sul comportamento di lettura, domande per le quali erano previste modalità canoniche di risposta (scelta multipla, griglia, caselle di controllo);
- seconda parte, che consisteva nella predisposizione dei choice set, cartaceo e digitale, le cui risposte sono state utilizzate per la predisposizione del modello a scelta discreta; la struttura dei quesiti e le elaborazioni in questo frangente verranno trattate nel Capitolo 6 e nel Capitolo 7.

2.3 Modalità di somministrazione e campione

Il questionario è stato somministrato secondo la modalità di auto-compilazione online, nel periodo intercorrente tra il 29 maggio 2018 ed il 21 giugno 2018, raggiungendo un totale di 281 persone. La somministrazione è avvenuta attraverso 4 diversi canali:

- Facebook;
- Whatsapp;
- Facebook Messenger;
- Instagram.

La scelta dei suddetti canali di somministrazione è stata guidata dalla consultazione del report Global Digital 2018, sondaggio condotto con cadenza annuale da We Are Social Italia, agenzia internazionale che si occupa di ricerca e studio dei trend nell'ambito dei social media, e Hootsuite, importante piattaforma di social media management. L'indagine Global Digital 2018 ha segnalato come piattaforme social e applicazioni di messaggistica istantanea maggiormente attive, in ordine decrescente di utilizzo, Youtube, Facebook, Whatsapp, Facebook Messenger ed Instagram. Sono stati dunque scelti i 4 canali soggetti a più traffico e ad accesso più agevole, fatta eccezione per Youtube in quanto la somministrazione di un questionario sotto forma di video avrebbe presentato notevoli difficoltà nella collezione delle risposte nonché nella consultazione del questionario stesso da parte degli utenti, non potendo godere di ampia visibilità.

La compilazione online risulta utile per la raccolta di dati in modo rapido e per la qualità delle informazioni collezionate, in quanto le risposte non coerenti od incompatibili vengono tempestivamente segnalate al rispondente, il quale può procedere alla correzione durante la compilazione stessa.

Il questionario è stato condiviso sulla bacheca personale della sottoscritta e su alcune pagine e gruppi dedicati alla lettura ed alle recensioni di libri.

2.4 Scale di misura utilizzate

Le scale di misura utilizzate nel questionario sono di 3 tipi:

- la *scala nominale*, o *categorica*. Questa consiste in una modalità di rilevamento che associa ad una domanda delle scelte qualitative distinte l'una dall'altra, ed in nessun modo ordinabili. Un esempio di scala nominale è rappresentato, per esempio, dal sesso, dall'occupazione ecc.
- la *scala ordinale*, che esprime delle alternative ordinabili secondo un determinato criterio. Ad esempio il titolo di studio è una variabile che presenta svariate possibilità di risposta, ognuna delle quali è collocabile all'interno di una scala crescente (ad esempio la laurea triennale verrà sicuramente dopo il diploma di maturità e sicuramente prima della laurea specialistica).
- Per permettere ai rispondenti di esprimere il loro giudizio personale su vari aspetti legati al mondo del libro è stato scelto di utilizzare anche delle *scale intervallari*, nello specifico delle *scale Likert*. Il livello intervallare permette agli intervistati non solo di scegliere tra alternative esclusive ed ordinali, ma le alternative hanno anche la caratteristica di essere equidistanti l'una dall'altra; nella *scala Likert*,

infatti, la distanza tra la valutazione 2 e la valutazione 4 è la stessa di quella tra la valutazione 5 e la valutazione 7. Le *scale Likert* utilizzate comprendono tutte un range di valutazione da 1 a 7. È stato scelto questo intervallo per scongiurare la possibilità che i rispondenti, trovandosi in dubbio su qualche domanda, scegliessero disinteressatamente di assegnare una valutazione a metà della scala. Predisponendo una scala che presentasse un numero dispari di valutazioni disponibili, l'intervistato è indotto a prendere una posizione, considerando che il voto a metà scala esatta non esiste: se sceglierà di assegnare 3 come valutazione, avrà deciso di considerare il parametro come insufficiente, se deciderà di assegnare 4 come valutazione, avrà deciso di considerare il parametro come sufficiente.

I rispondenti sono dunque chiamati a manifestare il livello di valutazione all'importanza o al gradimento, dovendo scegliere fra 7 modalità di risposta:

1. per nulla importante/gradito;
2. minimamente importante/gradito;
3. poco importante/gradito;
4. sufficientemente importante/gradito;
5. abbastanza importante/gradito;
6. molto importante/gradito;
7. massimamente importante/gradito.

2.5 Dettaglio dei quesiti rivolti

L'indagine è stata svolta attraverso la somministrazione di un questionario creato attraverso il software open source "Google Moduli". I risultati delle analisi univariate e bivariate, nonché delle elaborazioni più specifiche, verranno elencati e commentati nei capitoli successivi. Il questionario creato è stato dotato di 10 sezioni, che verranno elencate e spiegate in maniera estesa di seguito.

2.5.1 Sezione 1: Introduzione

In questa prima sezione è stata svolta la presentazione dell'obiettivo e delle finalità del questionario, introducendone brevemente il contenuto e la struttura; infine è stato dichiarato il completo anonimato per ciò che riguarda le risposte e la conformità del questionario alla normativa del D.Lgs.m196/2003 sulla privacy, dopo averne

espressamente dichiarato lo scopo di ricerca, avendone precisato la natura non commerciale e concludendo con il ringraziamento ai rispondenti.

2.5.2 Sezione 2: Abitudini e preferenze di lettura

Nella suddetta sezione sono state presentate due domande, di cui una facoltativa, per indagare sulla frequenza di lettura degli intervistati:

- la prima domanda rappresenta una domanda filtro, dicotomica, dalla cui risposta dipende il prosieguo del questionario. La domanda in questione è la seguente: “Negli ultimi 12 mesi ha letto libri per hobby?”: ad una risposta affermativa si lega la continuazione alla sezione successiva, ad una risposta negativa si viene indirizzati all’ultima sezione, relativa alle caratteristiche demografiche. Questa domanda filtro è necessaria per selezionare e dunque separare i lettori (coloro che hanno letto almeno un libro nell’ultimo anno) dai non lettori, ai quali verrà riservato un paragrafo a parte nel Capitolo 3.
- la seconda domanda consiste in una semplice striscia di testo breve in cui, facoltativamente, si può inserire il titolo dell’ultimo libro letto; questa domanda serve come riferimento per verificare quanti dei lettori, dichiaratisi tali, effettivamente ricordavano il titolo dell’ultimo libro letto.

2.5.3 Sezione 3: Informazioni generali sui lettori

A questa sezione hanno accesso solo i rispondenti dichiaratisi lettori. Una volta operata la prima scrematura, avendo suddiviso il campione in lettori e non lettori, si sono volute approfondire alcune caratteristiche relative alle abitudini di lettura in senso stretto. Questa sezione è stata suddivisa in 4 domande, ognuna delle quali indaga su un aspetto saliente relativo alle consuetudini personali di lettura dei rispondenti:

- la prima domanda ha come obiettivo la conoscenza del numero di libri letti per hobby pro capite, indicativamente, in un anno. Le alternative di risposta proposte, nella modalità scelta multipla, comprendono 3 fasce in cui potersi collocare:
 - da 1 a 3 libri (lettore debole);
 - da 4 a 11 libri (lettore medio);
 - più di 12 libri (lettore forte).

È stata operata questa specifica suddivisione sulla base della definizione di lettori deboli, lettori medi e lettori forti fornita dall’Istat.

- la seconda domanda indaga, considerando tutti i libri letti in un anno, quanti sono stati effettivamente acquistati di tasca propria dai rispondenti. Il quesito mira a tracciare un quadro indicativo del numero di libri effettivamente reperiti da venditori autorizzati, depurando le dichiarazioni di lettura degli intervistati da eventuali prestiti da biblioteche o regali da parte di conoscenti. Anche in questo caso le 4 alternative sono state proposte secondo la modalità scelta multipla, nella quale è stato richiesto quanti dei libri letti sono stati presi in prestito dalla biblioteca o da conoscenti. Le risposte sono state suddivise in fasce:
 - nessuno;
 - meno della metà;
 - più della metà;
 - tutti.
- la terza domanda riguarda la frequenza di lettura in senso stretto, espressa in numero di volte al mese in cui si riserva del tempo per la lettura. Le alternative di risposta presentate ammontano a 5, suddividono la frequenza di lettura in fasce:
 - più di 10 volte al mese;
 - dalle 5 alle 10 volte al mese;
 - dalle 3 alle 5 volte al mese;
 - 1 o 2 volte al mese;
 - meno di una volta al mese.

Sono state scelte queste specifiche fasce stabilendo un limite massimo di oltre 10 volte al mese, fino ad arrivare al limite minimo inferiore ad una volta al mese, comprendendo in questa espressione la lettura saltuaria (nel caso in cui, per esempio, in uno o più mesi specifici il rispondente non avesse riservato per nulla del tempo alla lettura).

- la quarta nonché ultima domanda di questa sezione prevede una lista di parametri, proposti attraverso una griglia a scelta multipla, ai quali i rispondenti sono chiamati ad assegnare una valutazione da 1 a 7 in termini di importanza, codificata come descritto sopra, riferendosi al momento in cui si trovano a scegliere un libro da leggere. Questi parametri comprendono consigli, segnalazioni, interviste e classifiche di varia natura, per i quali agli intervistati è stato richiesto di esprimere, in termini quantitativi, l'influenza nella scelta. Gli attributi in questione sono i seguenti:

- consigli di amici e parenti;
- consigli dalla casa editrice (sul sito o sui social network);
- segnalazioni su pagine o gruppi specializzati e community sui social;
- consigli di bibliotecari/insegnanti/esperti nel settore del libro;
- consigli di persone famose (attori, ballerini, cantanti, influencer);
- interviste e presenza dell'autore sui media (TV, radio, Internet);
- classifiche sui giornali/riviste/internet.

2.5.4 Sezione 4: Livello di gradimento dei generi narrativi

La sezione numero 4, comprendente un'unica domanda, succede immediatamente alla terza e ne hanno accesso gli stessi intervistati. In questa sezione è stato dato peso ai gusti personali del campione in fatto di generi narrativi. Nel quesito proposto si chiede infatti di valutare, utilizzando ancora una volta una scala Likert da 1 a 7 codificata sempre alla stessa maniera, il livello di gradimento di tutti i generi narrativi esistenti in ambito di narrativa:

- storico;
- fantascienza;
- fantasy;
- avventura/azione;
- per ragazzi, horror;
- thriller/crimine/giallo;
- erotico;
- umoristico;
- drammatico;
- poesie/opere teatrali/favole/fiabe;
- sentimentale/amore.

2.5.5 Sezione 5: Preferenze relative al formato di lettura

In questa sezione viene richiesto agli intervistati quale sia la loro preferenza in ambito di formato di lettura. In particolare vengono proposti due formati, cartaceo e digitale; il formato digitale comprende i libri in formato e-book nonché gli audiolibri. La suddetta rappresenta una domanda filtro con l'obiettivo di scremare i rispondenti dichiaratisi lettori in lettori di formato cartaceo e lettori di formato digitale, per poter analizzare le caratteristiche di entrambe le modalità di lettura in maniera più approfondita e specifica.

2.5.6 Sezione 6: Analisi della preferenza per il formato cartaceo

I rispondenti che hanno accesso a questa sezione sono coloro che hanno espresso la preferenza per il formato cartaceo nella domanda filtro della sezione 5. Questa sezione si suddivide in 3 domande che permettono di indagare su 3 diversi aspetti che riguardano i lettori di formato cartaceo:

- la prima domanda consiste in una griglia a scelta multipla, attraverso la quale agli intervistati viene richiesto di valutare attraverso la scala Likert da 1 a 7 l'importanza di alcune caratteristiche relative ai libri cartacei. Le caratteristiche in questione sono:
 - prezzo;
 - copertina;
 - disponibilità immediata;
 - autore;
 - trasportabilità;
 - recensioni/raccomandazioni;
 - genere;
 - novità.
- la seconda domanda ricerca i punti di forza del formato cartaceo secondo i rispondenti, espressi con modalità di risposta “caselle di controllo”, la quale permette all'intervistato di scegliere tra un elenco di possibili risposte le 3 per lui più significative. Le alternative proposte sono:
 - collezionismo;
 - arredamento (il libro può anche costituire un oggetto decorativo)
 - lettura rilassata (senza passare per uno schermo);
 - attaccamento emotivo alla fisicità del libro;
 - abitudine.
- la terza domanda ha come scopo l'analisi delle impressioni sul formato digitale da parte dei lettori del formato cartaceo. La domanda chiede quali siano le caratteristiche dei libri elettronici che potrebbero convincere i lettori di libri cartacei ad utilizzare invece il formato digitale; anche in questo caso gli intervistati possono scegliere fino a 3 alternative tra quelle proposte:
 - disponibilità a prezzo ridotto;
 - comodità di trasporto;

- disponibilità di contenuti digitali integrativi;
- risparmio di carta (questione ambientale);
- possibilità di leggere attraverso i dispositivi mobili personali;
- possibilità di portare con sé più libri alla volta;
- nessuna delle precedenti.

2.5.7 Sezione 8: Analisi della preferenza per il formato digitale

Potevano accedere alla sezione 8 solo i rispondenti che avevano espresso la loro preferenza per la lettura in formato digitale nella sezione 5; essi non accedevano alla sezione 6 e 7, in quanto relative al formato cartaceo. La sezione è stata divisa in 2 domande:

- la prima chiedeva di assegnare un livello di importanza quantificabile da 1 a 7 ai seguenti aspetti relativi ai libri in formato digitale: prezzo, recensioni o raccomandazioni, disponibilità a prezzo scontato, novità, autore, genere, copertina.
- la seconda domanda chiarisce, attraverso una selezione di massimo 3 alternative, quali sono i principali motivi per i quali i lettori di formato digitale preferiscono questa modalità rispetto al cartaceo. Le possibili scelte sono:
 - prezzo ridotto;
 - comodità di trasporto;
 - disponibilità di contenuti digitali integrativi;
 - risparmio di carta (questione ambientale);
 - possibilità di leggere attraverso i dispositivi mobili personali;
 - possibilità di portare con sé più libri alla volta;
 - possibilità di accedere al libro dovunque ci si trovi;
 - abitudine.

2.5.8 Sezioni 7 e 9: Choice set per formato cartaceo e digitale

Le sezioni 7 e 9 contengono, rispettivamente, il choice set relativo al formato di lettura cartacea ed il choice set riferito al formato digitale. Alla sezione 7 hanno accesso esclusivamente i rispondenti che hanno preferito la modalità cartacea nella sezione 5, mentre alla sezione 9 accedono solamente gli intervistati che hanno dichiarato la loro preferenza per il formato digitale nella sezione 5. In ciascuna di queste due sezioni sono

proposti dei *cartellini*, ovvero degli insiemi combinati di alternative raggruppate di 4 in 4; per ogni gruppo i rispondenti devono selezionare una delle alternative dal choice set, quella che presenta per loro un valore aggiunto maggiore. Il choice set (set di scelta) rappresenta dunque un insieme di alternative a disposizione del rispondente, e presenta i seguenti elementi:

- un numero di alternative ipotetiche: alla fine di ciascun set di alternative è stata aggiunta la scelta “*nessuna di queste*” come quarta preferenza;
- attributi: gli attributi delle alternative idealmente si escludono a vicenda e sono indipendenti.
- livelli: ogni attributo ha un numero di possibili livelli che lo specificano. I livelli specifici mostrati sono stati stabiliti attraverso la creazione del disegno frazionario, ed essi sono discreti, anche nel caso in cui l'attributo sia scalare, come nel caso del prezzo. In questo caso, i livelli vengono discretizzati in modo uniforme lungo un intervallo di valori consentiti.

In un modello a scelta discreta, il set di scelta deve soddisfare 3 requisiti:

- l'insieme di alternative deve essere collettivamente esaustivo, il che significa che l'insieme comprende tutte le possibili alternative. Questo requisito implica che la persona scelga necessariamente un'alternativa dal set.
- le alternative devono essere mutuamente esclusive, il che significa che scegliere un'alternativa significa non scegliere altre alternative. Questo requisito implica che la persona può scegliere solo un'alternativa dal set.
- il set deve contenere un numero finito di alternative. Questo terzo requisito distingue l'analisi della scelta discreta da forme di analisi di regressione in cui la variabile dipendente può (teoricamente) assumere un numero infinito di valori.

➤ **Cartellini sezione 7 (formato cartaceo)**

I cartellini proposti in questa sezione implicano una scelta unica per ogni gruppo di alternative. Gli attributi scelti per la sezione dei cartellini sono:

- prezzo;
- recensioni;
- autore;
- novità.

Ogni attributo, come detto, è stato suddiviso in 2 livelli:

- il prezzo è stato fissato a 12€ e a 25€ avendo individuato in questi una buona approssimazione dell'intera offerta di libri delle case editrici;
- per le recensioni, è stato scelto di proporre come scelta l'accesso o meno ad esse;
- per quanto riguarda l'autore, i livelli consistono nella nazionalità italiana, quindi autore più vicino al lettore in termini di fama, e nella nazionalità straniera, comprendendo nel novero la totalità degli autori esistenti;
- la novità è stata declinata come presente o assente, indicando come "nuovo" un libro pubblicato entro l'anno corrente e "non nuovo" un libro pubblicato più di un anno fa, comprendendo in quest'ultima definizione anche i classici.

I 3 cartellini proposti nella sezione relativa al formato cartaceo sono così composti:

✓ Domanda 1:

- *Alternativa 1* - Prezzo 12€, Recensioni non disponibili, Autore italiano, Pubblicato nell'ultimo anno
- *Alternativa 2* - Prezzo 12€, Recensioni disponibili, Autore straniero, Pubblicato nell'ultimo anno
- *Alternativa 3* - Prezzo 12€, Recensioni disponibili, Autore italiano, Pubblicato più di un anno fa

✓ Domanda 2:

- *Alternativa 4* - Prezzo 25€, Recensioni disponibili, Autore italiano, Pubblicato nell'ultimo anno
- *Alternativa 5* - Prezzo 12€, Recensioni non disponibili, Autore italiano, Pubblicato più di un anno fa
- *Alternativa 6* - Prezzo 12€, Recensioni non disponibili, Autore straniero, Pubblicato più di un anno fa

✓ Domanda 3:

- *Alternativa 7* - Prezzo 25€, Recensioni non disponibili, Autore italiano, Pubblicato nell'ultimo anno
- *Alternativa 8* - Prezzo 25€, Recensioni disponibili, Autore straniero, Pubblicato più di un anno fa

- *Alternativa 9* - Prezzo 25€, Recensioni non disponibili, Autore straniero, Pubblicato più di un anno fa

➤ **Cartellini sezione 9 (formato digitale)**

In questa sezione, come nella sezione numero 7, sono state proposte, a gruppi di 4, 12 combinazioni di attributi, comprendenti anche la facoltà di non operare una scelta tra le alternative proposte, attraverso l'opzione "nessuna delle precedenti". Le variabili proposte nel caso del formato digitale sono:

- prezzo;
- recensioni;
- disponibilità di contenuti digitali integrativi;
- disponibilità del libro in formato PDF.

Gli attributi sono suddivisi in 2 livelli:

- per quanto riguarda il prezzo sono stati individuati, come buona approssimazione, i prezzi 9€ e 20€ (comprendendo in quest'ultimo livello di prezzo anche i libri più rari e specializzati);
- le recensioni sono state divise in 2 livelli: la disponibilità o meno di esse;
- la presenza od assenza di disponibilità è stata la declinazione utilizzata anche nel caso dei contenuti digitali integrativi compresi nel prezzo del libro, in quanto questi rappresentano un valore aggiunto di cui si può godere solamente scegliendo la lettura in formato digitale;
- l'ultimo attributo proposto, la disponibilità o meno del formato in PDF, è stato inserito in queste due specificazioni di livello nell'ottica di considerare anche i lettori digitali che non si servono di supporti elettronici portatili per leggere i loro libri (Kindle, e-reader ed altri) ma che preferiscono invece leggerli attraverso tablet o personal computer.

I 3 cartellini relativi alla sezione digitale sono stati impostati nel seguente modo:

✓ Domanda I:

- *Alternativa A* - Prezzo 9€, Recensioni non disponibili, No Contenuti integrativi, No formato PDF
- *Alternativa B* - Prezzo 20€, Recensioni non disponibili, Sì Contenuti integrativi, Sì formato PDF
- *Alternativa C* - Prezzo 20€, Recensioni disponibili, No Contenuti digitali integrativi, No formato PDF

- *Alternativa D* - Nessuna delle precedenti
- ✓ Domanda II:
- *Alternativa E* - Prezzo 20€, Recensioni disponibili, Sì Contenuti integrativi, Sì formato PDF
 - *Alternativa F* - Prezzo 9€, Recensioni disponibili, No Contenuti integrativi, Sì formato PDF
 - *Alternativa G* - Prezzo 20€, Recensioni non disponibili, No Contenuti integrativi, No formato PDF
 - *Alternativa H* - Nessuna delle precedenti
- ✓ Domanda III:
- *Alternativa I* - Prezzo 9€, Recensioni non disponibili, Sì Contenuti integrativi, Sì formato PDF
 - *Alternativa L* - Prezzo 9€, Recensioni disponibili, Sì Contenuti integrativi, No formato PDF
 - *Alternativa M* - Prezzo 9€, Recensioni non disponibili, Sì Contenuti integrativi, No formato PDF
 - *Alternativa N* - Nessuna delle precedenti

2.5.9 Sezione 10: Profilo socio-demografico

Risulta di grande utilità per l'indagine l'individuazione delle caratteristiche socio-demografiche del campione rispondente al questionario. In primo luogo, queste informazioni sono utili per tracciare il profilo tipico degli intervistati, ma sono state utilizzate anche per cercare una possibile corrispondenza tra le risposte fornite ed un determinato tipo di caratteri socio-demografici. Se sussistesse questa corrispondenza, le case editrici potrebbero diversificare l'offerta di libri in base al profilo sociodemografico del cliente, migliorando l'efficienza delle campagne pubblicitarie ed innalzando il tasso di conversione. Sono dunque presentati qui di seguito gli aspetti socio-demografici che si sono voluti analizzare:

- *Sesso*
- *Età*, suddivisa nelle seguenti fasce:
 - ✓ fino a 17 anni;
 - ✓ da 18 a 24 anni;

- ✓ da 25 a 29 anni;
 - ✓ da 30 a 34 anni;
 - ✓ da 35 a 39 anni;
 - ✓ da 40 a 44 anni;
 - ✓ da 45 a 49 anni;
 - ✓ da 50 a 54 anni;
 - ✓ da 55 a 59 anni;
 - ✓ da 60 a 64 anni;
 - ✓ 65 anni o più.
- *Titolo di studio*, suddiviso nelle seguenti modalità:
- ✓ nessun titolo di studio;
 - ✓ licenza di scuola elementare;
 - ✓ licenza di scuola media;
 - ✓ diploma di scuola superiore (3 anni);
 - ✓ diploma di maturità scuola superiore (5 anni);
 - ✓ laurea triennale;
 - ✓ laurea magistrale;
 - ✓ istruzione post universitaria (dottorato, master...).
- *Occupazione*, suddivisa in:
- ✓ dirigente/ quadro;
 - ✓ impiegato/ operaio;
 - ✓ libero professionista;
 - ✓ lavoratore in proprio;
 - ✓ studente;
 - ✓ non occupato;
 - ✓ pensionato;
 - ✓ casalingo/a.
- *Regione di residenza*;
- *Provincia di residenza*, per verificare la coerenza con la risposta relativa alla regione;

-
- *Numero di componenti del nucleo familiare* (intervistato compreso), suddivisi in 1, 2, 3, 4 o più di 4, per poterlo confrontare con il numero di persone che percepiscono reddito in famiglia;
 - *Presenza o meno di anziani e/o bambini all'interno del nucleo familiare* (per individuare se la lettura assidua di libri fosse o meno influenzata dalla presenza in famiglia di soggetti con necessità di assistenza);
 - *Numero di persone che percepiscono reddito all'interno della famiglia*, suddivisi in: suddivisi in 1, 2, 3, 4 o più di 4, per individuare una possibile corrispondenza tra il livello di ricchezza della famiglia e la lettura di libri.

Capitolo 3

Risultati dell'indagine: caratterizzazione sociodemografica del campione

In questo capitolo verranno presentati i risultati relativi alle caratteristiche socio-demografiche degli intervistati, suddivisi tra la categoria dei lettori, la quale conta un totale di 254 rispondenti suddivisi a loro volta tra lettori di formato cartaceo (222 rispondenti) e di formato digitale (32 rispondenti), e la categoria dei non lettori, la quale comprende 27 tra i rispondenti totali.

Al fine di identificare le caratteristiche socio-demografiche dei rispondenti al questionario, è stata riservata una sezione rivolta a tutti i rispondenti, fossero essi lettori o non lettori. Si riportano in seguito i risultati dell'analisi socio-demografica.

3.1 Dettaglio sociale e demografico dei lettori

I rispondenti che si sono dichiarati lettori, come mostrato in Tabella 3.1, sono risultati in netta prevalenza donne (81,102%); questo risultato può dipendere dai canali attraverso i quali si è deciso di diffondere il questionario, in riferimento soprattutto ai gruppi di lettura su Facebook nei quali la maggior parte dei partecipanti è di sesso femminile.

Sesso			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Maschio	48	18,898%	18,898%
Femmina	206	81,102%	100,000%
Totale	254	100,000%	

Tabella 3.1: Lettori: frequenze relative e cumulate della variabile sesso.

Per quanto riguarda l'età, il cui dettaglio è mostrato nella Tabella 3.2, il campione dei lettori si è dimostrato in prevalenza composto da giovani: il 73,622% di loro ha dichiarato di avere meno di 30 anni. Analizzando nel dettaglio la distribuzione, le frequenze più elevate si sono registrate in corrispondenza della fascia di età compresa tra 18 e 29 e, più precisamente:

- il 44,882% dei lettori ha dichiarato di rientrare nella fascia di età 18 -24 anni;

- il 26,378% dei lettori ha dichiarato di rientrare nella fascia di età 25 – 29 anni.

Anche in questo caso è possibile che il dato sia influenzato dal vettore attraverso il quale il questionario è stato diffuso, in quanto i social network sono popolati in prevalenza da giovani, considerata anche l'età della sottoscritta e la relativa cerchia di amicizie.

Età			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Fino a 17 anni	6	2,362%	2,362%
Da 18 a 24 anni	114	44,882%	47,244%
Da 25 a 29 anni	67	26,378%	73,622%
Da 30 a 34 anni	14	5,512%	79,134%
Da 35 a 39 anni	14	5,512%	84,646%
Da 40 a 44 anni	12	4,724%	89,370%
Da 45 a 49 anni	12	4,724%	94,094%
Da 50 a 54 anni	8	3,150%	97,244%
Da 55 a 59 anni	5	1,969%	99,213%
65 anni o più	2	0,787%	100,000%
Totale	254	100,000%	

Tabella 3.2: Lettori: frequenze relative e cumulate della variabile età.

In relazione al titolo di studio, mostrato nella Tabella 3.3, la maggioranza dei lettori (37,008%) ha dichiarato di aver conseguito il diploma di maturità superiore, il 25,591% ha dichiarato di aver conseguito la laurea triennale ed il 20,866% ha dichiarato di aver ottenuto la laurea magistrale. Il campione dei lettori risulta dunque, in media, abbastanza acculturato, poiché:

- il 46,457% dei lettori ha dichiarato di avere ottenuto una laurea;
- il 37,008% ha dichiarato di avere conseguito il diploma di maturità superiore.

Titolo di studio			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Nessun titolo di studio	1	0,394%	0,394%
Licenza di scuola media	16	6,299%	6,693%
Diploma di scuola superiore (3 anni)	11	4,331%	11,024%
Diploma di maturità scuola superiore (5 anni)	94	37,008%	48,031%
Laurea triennale	65	25,591%	73,622%
Laurea magistrale	53	20,866%	94,488%
Istruzione post universitaria (dottorato, master...)	14	5,512%	100,000%
Totale	254	100,000%	

Tabella 3.3: Lettori: frequenze relative e cumulate della variabile titolo di studio.

La maggior parte dei lettori, il 52,362% sul totale, non ha ancora concluso il suo percorso di studi mentre il 20,866% svolge un lavoro impiegatizio o come operaio, come si evince dalla Tabella 3.4. In linea generale, dunque, più di una persona su cinque tra gli intervistati

che si sono dichiarati lettori sono studenti i quali, nel tempo libero, si dedicano anche alla lettura slegatamente dai motivi scolastici.

Occupazione			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Dirigente/ Quadro	8	3,150%	3,150%
Impiegato/ Operaio	53	20,866%	24,016%
Libero professionista	22	8,661%	32,677%
Lavoratore in proprio	8	3,150%	35,827%
Studente	133	52,362%	88,189%
Non occupato	22	8,661%	96,850%
Pensionato	2	0,787%	97,638%
Casalingo/a	6	2,362%	100,000%
Totale	254	100,000%	

Tabella 3.4: Lettori: frequenze relative e cumulate della variabile occupazione.

Per quanto concerne invece la localizzazione dei lettori sul territorio nazionale, presentata nella Tabella 3.5, la maggioranza ha dichiarato di risiedere in Veneto, precisamente il 35,433%; questo dato può essere influenzato dalle modalità di somministrazione del questionario, che hanno privilegiato l'accesso ad esso agli individui più geograficamente vicini.

Regione di residenza			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Abruzzo	5	1,969%	1,969%
Basilicata	1	0,394%	2,362%
Calabria	6	2,362%	4,724%
Campania	13	5,118%	9,843%
Emilia - Romagna	14	5,512%	15,354%
Friuli - Venezia Giulia	5	1,969%	17,323%
Lazio	14	5,512%	22,835%
Liguria	4	1,575%	24,409%
Lombardia	30	11,811%	36,220%
Marche	8	3,150%	39,370%
Molise	2	0,787%	40,157%
Piemonte	17	6,693%	46,850%
Puglia	15	5,906%	52,756%
Sardegna	6	2,362%	55,118%
Sicilia	9	3,543%	58,661%
Toscana	10	3,937%	62,598%
Umbria	4	1,575%	64,173%
Valle d'Aosta	1	0,394%	64,567%
Veneto	90	35,433%	100,000%
Totale	254	100,000%	

Tabella 3.5: Lettori: frequenze relative e cumulate della variabile regione di residenza.

Il 46,850% del campione ha dichiarato di vivere in un nucleo familiare composto da 3 persone oltre a sé stessi, mentre il 27,559% ha affermato di vivere in una famiglia

composta, in tutto, da 3 persone, come si vede nella Tabella 3.6. I dati evidenziano che la maggior parte dei lettori vivono in una famiglia composta da almeno 3 persone, mentre tra le coppie sposate o conviventi e tra coloro che vivono da soli si conta un numero inferiore di persone che si dedicano alla lettura.

Componenti del nucleo familiare			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata
1	7	2,756%	2,756%
2	30	11,811%	14,567%
3	70	27,559%	42,126%
4	119	46,850%	88,976%
Più di 4	28	11,024%	100,000%
Totale	254	100,000%	

Tabella 3.6: Lettori: frequenze relative e cumulate della variabile componenti del nucleo familiare.

Il 74,016% del campione dei lettori ha dichiarato che, all'interno della propria famiglia, non ci sono anziani o bambini (Tabella 3.7); questo dato, se confrontato con quelli tratti dalla figura precedente, mette in luce che più di 7 famiglie su 10, composte in prevalenza da 3 o 4 persone, non ospitano bambini: si può dunque dedurre che i rispondenti vivano, per la maggior parte, ancora con i genitori, asserzione confermata anche dal fatto che sono in prevalenza studenti e che sono, in prevalenza, giovani.

Il Suo nucleo familiare comprende anche anziani e/o bambini?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Sì	66	25,984%	25,984%
No	188	74,016%	100,000%
Totale	254	100,000%	

Tabella 3.7: Lettori: frequenze relative e cumulate della variabile presenza di anziani o bambini nel nucleo familiare.

Per quanto riguarda il numero di persone che percepiscono un reddito all'interno del nucleo familiare, la maggior parte delle dichiarazioni riguarda famiglie con due persone percepenti reddito (53,937%) o una persona percepente reddito (29,528%) come mostrato nella Tabella 3.7. Si può presumere che, nella maggior parte dei casi, le persone che percepiscono un reddito all'interno della famiglia del rispondente siano i genitori.

Quante persone percepiscono un reddito all'interno del Suo nucleo familiare?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
1	75	29,528%	29,528%
2	137	53,937%	83,465%
3	35	13,780%	97,244%
4	6	2,362%	99,606%
Più di 4	1	0,394%	100,000%
Totale	254	100,000%	

Tabella 3.7: Lettori: frequenze relative e cumulate della variabile percettori di reddito.

3.1.1 Caratteristiche sociali e demografiche dei lettori di formato cartaceo

Considerando che il campione dei lettori di libri stampati consiste nell'87,40% di tutto il campione dei lettori, in molti casi i risultati relativi alle caratteristiche sociali e demografiche nel caso dei lettori e dei lettori di formato cartaceo coincidono.

La prevalenza di lettrici sui lettori si manifesta infatti anche nel caso del formato cartaceo, infatti l'81,081% dei lettori di libri stampati dichiara di essere donna, come si evince dalla Tabella 3.8.

Sesso			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Maschio	42	18,919%	18,919%
Femmina	180	81,081%	100,000%
Totale	222	100,000%	

Tabella 3.8: Lettori di formato cartaceo: frequenze relative e cumulate della variabile sesso.

Le fasce di età più largamente ricoperte rimangono quelle che comprendono le persone dai 18 ai 24 anni (45,495% dei lettori di libri stampati) e dai 25 ai 29 anni (28,829% dei lettori di libri stampati), come mostrato nella Tabella 3.9.

Età			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Fino a 17 anni	6	2,703%	2,703%
Da 18 a 24 anni	101	45,495%	48,198%
Da 25 a 29 anni	64	28,829%	77,027%
Da 30 a 34 anni	12	5,405%	82,432%
Da 35 a 39 anni	11	4,955%	87,387%
Da 40 a 44 anni	8	3,604%	90,991%
Da 45 a 49 anni	8	3,604%	94,595%
Da 50 a 54 anni	5	2,252%	96,847%
Da 55 a 59 anni	5	2,252%	99,099%
65 anni o più	2	0,901%	100,000%
Totale	222	100,000%	

Tabella 3.9: Lettori di formato cartaceo: frequenze relative e cumulate della variabile età.

Per quanto riguarda il titolo di studio, descritto nella Tabella 3.10, la maggioranza delle risposte si è collocata in corrispondenza del diploma di maturità superiore, nel 34,685% dei casi, il 28,829% delle risposte in corrispondenza della laurea triennale ed il 20,721% delle risposte in corrispondenza della laurea magistrale.

Titolo di studio			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Nessun titolo di studio	1	0,450%	0,450%
Licenza di scuola media	14	6,306%	6,757%
Diploma di scuola superiore (3 anni)	10	4,505%	11,261%
Diploma di maturità scuola superiore	77	34,685%	45,946%
Laurea triennale	64	28,829%	74,775%
Laurea magistrale	46	20,721%	95,495%
Istruzione post universitaria	10	4,505%	100,000%
Totale	222	100,000%	

Tabella 3.10: Lettori di formato cartaceo: frequenze relative e cumulate della variabile titolo di studio.

Anche nel caso dei fruitori di libri in formato cartaceo, la maggioranza non ha ancora concluso il proprio percorso di studi, in particolare il 54,955% dei rispondenti, ed il 18,468% del campione dei lettori cartacei lavora come impiegato od operaio, come si evince dalla Tabella 3.11.

Occupazione			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Dirigente/ Quadro	6	2,703%	2,703%
Impiegato/ Operaio	41	18,468%	21,171%
Libero professionista	18	8,108%	29,279%
Lavoratore in proprio	7	3,153%	32,432%
Studente	122	54,955%	87,387%
Non occupato	20	9,009%	96,396%
Pensionato	2	0,901%	97,297%
Casalingo/a	6	2,703%	100,000%
Totale	222	100,000%	

Tabella 3.11: Lettori di formato cartaceo: frequenze relative e cumulate della variabile occupazione.

Tra il campione dei lettori di formato cartaceo il 35,135% dichiara di essere veneto ed il 12,162% dichiara di risiedere in Lombardia, come mostrato nella Tabella 3.12.

Regione di residenza			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Abruzzo	4	1,802%	1,802%
Basilicata	1	0,450%	2,252%
Calabria	6	2,703%	4,955%
Campania	13	5,856%	10,811%
Emilia - Romagna	10	4,505%	15,315%
Friuli - Venezia Giulia	5	2,252%	17,568%
Lazio	12	5,405%	22,973%
Liguria	4	1,802%	24,775%
Lombardia	27	12,162%	36,937%
Marche	8	3,604%	40,541%
Molise	2	0,901%	41,441%
Piemonte	14	6,306%	47,748%
Puglia	14	6,306%	54,054%
Sardegna	5	2,252%	56,306%
Sicilia	9	4,054%	60,360%
Toscana	6	2,703%	63,063%
Umbria	0	0,000%	63,063%
Valle d'Aosta	4	1,802%	64,865%
Veneto	78	35,135%	100,000%
Totale	222	100,000%	

Tabella 3.12: Lettori di formato cartaceo: frequenze relative e cumulate della variabile regione di residenza.

Riguardo l'ampiezza del nucleo familiare, il 46,396% dei rispondenti che leggono i libri stampati dichiara di vivere in una famiglia comprendente 4 persone in totale, il 27,477% di vivere in un nucleo familiare comprendente in tutto 3 persone, come si legge nella Tabella 3.13.

Componenti del nucleo familiare			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
1	4	1,802%	1,802%
2	27	12,162%	13,964%
3	61	27,477%	41,441%
4	103	46,396%	87,838%
Più di 4	27	12,162%	100,000%
Totale	222	100,000%	

Tabella 3.13: Lettori di formato cartaceo: frequenze relative e cumulate della variabile componenti del nucleo familiare.

L'assenza di anziani o bambini nel nucleo familiare viene segnalata dal 74,324% del campione, il restante 25,676% ne dichiara invece la presenza, come si vede nella Tabella 3.14.

Il Suo nucleo familiare comprende anche anziani e/o bambini?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Sì	57	25,676%	25,676%
No	165	74,324%	100,000%
Totale	222	100,000%	

Tabella 3.14: Lettori di formato cartaceo: frequenze relative e cumulate della variabile presenza di anziani o bambini nel nucleo familiare.

All'interno della famiglia dei lettori di formato cartaceo i percettori di reddito sono 2 nella maggioranza dei casi (52,703%) ed anche la quota di dichiaranti un solo percettore di reddito spicca per numerosità (30,180%) come si evince dalla Tabella 3.15. Appare chiaro che, alla luce delle risposte, le famiglie dei rispondenti siano nuclei composti in prevalenza dai genitori o fratelli dei rispondenti.

Quante persone percepiscono un reddito all'interno del Suo nucleo familiare?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
1	67	30,180%	30,180%
2	117	52,703%	82,883%
3	32	14,414%	97,297%
4	5	2,252%	99,550%
Più di 4	1	0,450%	100,000%
Totale	222	100,000%	

Tabella 3.15: Lettori di formato cartaceo: frequenze relative e cumulate della variabile percettori di reddito in famiglia.

3.1.2 Caratteristiche sociali e demografiche dei lettori di formato digitale

I lettori di formato digitale costituiscono la netta minoranza, esattamente il 12,60%. Essi si dimostrano in prevalenza donne, per un totale dell'81,250%, mentre gli uomini sono circa il 19%, come si vede nella Tabella 3.16.

Sesso			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Maschio	6	18,750%	18,750%
Femmina	26	81,250%	100,000%
Totale	32	100,000%	

Tabella 3.16: Lettori di formato digitale: frequenze relative e cumulate della variabile sesso.

La maggior parte di essi rientrano nella fascia di età compresa tra 18 e 24 anni, precisamente il 40,625%, come mostrato nella Tabella 3.17.

Età			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Da 18 a 24 anni	13	40,625%	40,625%
Da 25 a 29 anni	3	9,375%	50,000%
Da 30 a 34 anni	2	6,250%	56,250%
Da 35 a 39 anni	3	9,375%	65,625%
Da 40 a 44 anni	4	12,500%	78,125%
Da 45 a 49 anni	4	12,500%	90,625%
Da 50 a 54 anni	3	9,375%	100,000%
Totale	32	100,000%	

Tabella 3.17: Lettori di formato digitale: frequenze relative e cumulate della variabile età.

Il 53,125% dei lettori di libri in formato elettronico ha conseguito il diploma di maturità superiore, ed il 21,875% dichiara di avere conseguito la laurea magistrale, come si vede nella Tabella 3.18.

Titolo di studio			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Licenza di scuola media	2	6,250%	6,250%
Diploma di scuola superiore (3 anni)	1	3,125%	9,375%
Diploma di maturità scuola	17	53,125%	62,500%
Laurea triennale	1	3,125%	65,625%
Laurea magistrale	7	21,875%	87,500%
Istruzione post universitaria	4	12,500%	100,000%
Totale	32	100,000%	

Tabella 3.18: Lettori di formato digitale: frequenze relative e cumulate della variabile titolo di studio.

Il 37,5% del campione in questione asserisce di praticare una mansione di tipo impiegatizio o operaio, ed il 34,375% dichiara di essere uno studente, come mostrato nella Tabella 3.19.

Occupazione			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Dirigente/ Quadro	2	6,250%	6,250%
Impiegato/ Operaio	12	37,500%	43,750%
Libero professionista	4	12,500%	56,250%
Lavoratore in proprio	1	3,125%	59,375%
Studente	11	34,375%	93,750%
Non occupato	2	6,250%	100,000%
Totale	32	100,000%	

Tabella 3.19: Lettori di formato digitale: frequenze relative e cumulate della variabile occupazione.

La maggioranza del campione risiede in Veneto, precisamente il 37,5%, un 12,5% asserisce di abitare in Emilia Romagna ed un altro 12,5% dichiara di risiedere in Toscana, come si vede dalla Tabella 3.20.

Regione di residenza			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Abruzzo	1	3,125%	3,125%
Emilia - Romagna	4	12,500%	15,625%
Lazio	2	6,250%	21,875%
Lombardia	3	9,375%	31,250%
Piemonte	3	9,375%	40,625%
Puglia	1	3,125%	43,750%
Sardegna	1	3,125%	46,875%
Toscana	4	12,500%	59,375%
Valle d'Aosta	1	3,125%	62,500%
Veneto	12	37,500%	100,000%
Totale	32	100,000%	

Tabella 3.20: Lettori di formato digitale: frequenze relative e cumulate della variabile regione di residenza.

Come mostrato nella Tabella 3.21, anche per il campione di lettori digitali si registra una maggioranza di famiglie composte da 3 o 4 persone, rispettivamente per il 28,125% e per il 50%.

Componenti del nucleo familiare			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
1	3	9,375%	9,375%
2	3	9,375%	18,750%
3	9	28,125%	46,875%
4	16	50,000%	96,875%
Più di 4	1	3,125%	100,000%
Totale	32	100,000%	

Tabella 3.21: Lettori di formato digitale: frequenze relative e cumulate della variabile componenti del nucleo familiare.

I rispondenti che dichiarano che il loro nucleo familiare non comprende anziani o bambini formano il 71,875% del campione, il restante 28,125% dichiara di vivere in una famiglia comprendente anche anziani e bambini, come si nota dalla Tabella 3.22.

Il Suo nucleo familiare comprende anche anziani e/o bambini?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Sì	9	28,125%	28,125%
No	23	71,875%	100,000%
Totale	32	100,000%	

Tabella 3.22: Lettori di formato digitale: frequenze relative e cumulate della variabile presenza di anziani o bambini nel nucleo familiare.

La numerosità dei percipienti reddito all'interno del nucleo familiare riprende la tendenza evidenziata anche per i lettori cartacei, infatti il 62,5% dei rispondenti dichiara di vivere in una famiglia in cui lavorano due persone, mentre il 25% dichiara che, nella sua famiglia, lavora una sola persona, come si nota nella Tabella 3.23.

Quante persone percepiscono un reddito all'interno del Suo nucleo familiare?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
1	8	25,000%	25,000%
2	20	62,500%	87,500%
3	3	9,375%	96,875%
4	1	3,125%	100,000%
Totale	32	100,000%	

Tabella 3.23: Lettori di formato digitale: frequenze relative e cumulate della variabile percettori di reddito.

3.1.3 Analisi bivariate sui lettori

Per accertare la tendenza del consumo di libri in base alle condizioni sociali e demografiche ed alla frequenza di lettura, la quale decreta l'appartenenza dell'individuo al gruppo dei lettori forti o deboli, si sono svolte delle analisi bivariate incrociando nella stessa tabella la frequenza di lettura con la variabile socio-demografica che si vuole studiare.

Partendo dalla variabile sesso, si è notato come la suddetta variabile incida nella frequenza di lettura, come viene mostrato nella Tabella 3.24: la maggior parte delle donne (97 individui) infatti rientra nella categoria dei lettori forti con più di 12 libri letti all'anno, mentre la maggior parte degli uomini rientra invece nella categoria dei lettori medi (21 individui), che comprende gli individui che leggono da 4 a 11 libri all'anno.

Sesso	Frequenza di lettura			Totale
	Da 1 a 3 libri	Da 4 a 11 libri	Più di 12 libri	
Femmina	26	83	97	206
Maschio	9	21	18	48
Totale	35	104	115	254

Tabella 3.24: Studio bivariato delle variabili sesso e frequenza di lettura.

Per quanto riguarda l'età, la maggior parte dei lettori si colloca nella fascia comprendente gli individui da 18 a 24 anni, per un totale di 114 individui, e di questi la maggior parte si dichiara lettore forte (55 individui), mentre i lettori deboli sono la minoranza (15 individui). La seconda fascia d'età per numerosità è rappresentata da quella comprendente le persone da 25 a 29 anni: in questo caso la maggioranza è rappresentata dai lettori medi (35 individui) e la minoranza dai lettori deboli (6 individui), come mostrato nella Tabella 3.25.

Età	Frequenza di lettura			Totale
	Da 1 a 3 libri	Da 4 a 11 libri	Più di 12 libri	
65 anni o più	0	0	2	2
Da 18 a 24 anni	15	44	55	114
Da 25 a 29 anni	6	35	26	67
Da 30 a 34 anni	1	6	7	14
Da 35 a 39 anni	1	5	8	14
Da 40 a 44 anni	4	5	3	12
Da 45 a 49 anni	4	1	7	12
Da 50 a 54 anni	0	5	3	8
Da 55 a 59 anni	3	2	0	5
Fino a 17 anni	1	1	4	6
Totale	35	104	115	254

Tabella 3.25: Studio bivariato delle variabili età e frequenza di lettura.

Attraverso l'analisi del titolo di studio, si evince che la maggior parte dei lettori forti (44 individui) possiede come titolo di studio il diploma di maturità superiore, come anche nel caso dei lettori medi (38 individui) e dei lettori deboli (12 individui), come mostrato nella Figura 3.26.

Titolo di studio	Frequenza di lettura			Totale
	Da 1 a 3 libri	Da 4 a 11 libri	Più di 12 libri	
Diploma di maturità superiore	12	38	44	94
Diploma di scuola superiore (3 anni)	1	1	9	11
Istruzione post universitaria	0	4	10	14
Laurea magistrale	7	25	21	53
Laurea triennale	10	32	23	65
Licenza di scuola media	5	4	7	16
Nessun titolo di studio	0	0	1	1
Totale	35	104	115	254

Tabella 3.26: Studio bivariato delle variabili titolo di studio e frequenza di lettura.

In tema di occupazione, la maggior parte dei lettori è rappresentata dagli studenti (133 individui) e sia tra i lettori deboli, sia tra i lettori medi sia tra i lettori forti costituiscono la maggioranza (rispettivamente 15, 61 e 57 individui), come si nota nella Tabella 3.27.

Occupazione	Frequenza di lettura			Totale
	Da 1 a 3 libri	Da 4 a 11 libri	Più di 12 libri	
Casalingo/a	2	2	2	6
Dirigente/ Quadro	2	1	5	8
Impiegato/ Operaio	9	22	22	53
Lavoratore in proprio	1	1	6	8
Libero professionista	3	8	11	22
Non occupato	3	9	10	22
Pensionato	0	0	2	2
Studente	15	61	57	133
Totale	35	104	115	254

Tabella 3.27: Studio bivariato delle variabili occupazione e frequenza di lettura.

La maggioranza dei lettori risiede in Veneto sia per quanto riguarda i lettori deboli (27 individui), sia per i lettori medi (37 individui) sia per i lettori forti (26 individui). La seconda regione per numerosità di lettori è la Lombardia, con un totale di 15 individui tra i lettori medi e 15 individui tra i lettori forti, come mostrato nella Tabella 3.28.

Regione di residenza	Da 1 a 3 libri	Da 4 a 11 libri	Più di 12 libri	Totale
Abruzzo	0	1	4	5
Basilicata	0	0	1	1
Calabria	0	3	3	6
Campania	0	6	7	13
Emilia - Romagna	2	3	9	14
Friuli - Venezia Giulia	2	2	1	5
Lazio	0	6	8	14
Liguria	0	1	3	4
Lombardia	0	15	15	30
Marche	1	4	3	8
Molise	0	1	1	2
Piemonte	0	5	12	17
Puglia	0	11	4	15
Sardegna	0	3	3	6
Sicilia	1	3	5	9
Toscana	0	3	7	10
Umbria	1	0	3	4
Valle d'Aosta	1	0	0	1
Veneto	27	37	26	90
Totale	35	104	115	254

Tabella 3.28: Studio bivariato delle variabili regione di residenza e frequenza di lettura.

Per quanto concerne la numerosità del nucleo familiare la maggior parte dei lettori forti vivono in una famiglia comprendente 4 componenti (62 individui), ed anche la maggior parte dei lettori medi (45 individui), mentre la maggioranza dei lettori deboli vive in una famiglia comprendente 3 componenti (15 individui), come mostra la Tabella 3.29.

Familiari	Frequenza di lettura			Totale
	Da 1 a 3 libri	Da 4 a 11 libri	Più di 12 libri	
1	1	2	4	7
2	2	13	15	30
3	15	38	17	70
4	12	45	62	119
Più di 4	5	6	17	28
Totale	35	104	115	254

Tabella 3.29: Studio bivariato delle variabili componenti del nucleo familiare e frequenza di lettura.

La presenza di anziani o bambini nel nucleo familiare, rappresentata nella Tabella 3.30, è smentita dalla maggioranza dei lettori forti (80 individui), dei lettori medi (85 individui) e dei lettori deboli (23 individui). In proporzione, nella categoria dei lettori medi i dichiaranti l'assenza di anziani o bambini nel nucleo familiare sono in numero maggiore rispetto che nella categoria dei lettori forti.

Presenza anziani o bambini	Frequenza di lettura			Totale
	Da 1 a 3 libri	Da 4 a 11 libri	Più di 12 libri	
No	23	85	80	188
Sì	12	19	35	66
Totale	35	104	115	254

Tabella 3.30: Studio bivariato delle variabili presenza di anziani o bambini nel nucleo familiare e frequenza di lettura.

In tema di componenti del nucleo familiare che percepiscono reddito, la maggioranza (137 individui) ha dichiarato di contare, all'interno della propria famiglia, 2 componenti. Le famiglie con 2 percettori di reddito sono in maggioranza nel caso dei lettori forti (69 individui), mentre per ciò che riguarda le famiglie con 1 solo percettore di reddito, i dichiaranti sono in maggioranza lettori medi (38 individui), come si evince dalla Tabella 3.31.

Percettori reddito	Frequenza di lettura			Totale
	Da 1 a 3 libri	Da 4 a 11 libri	Più di 12 libri	
1	6	38	31	75
2	22	46	69	137
3	7	17	11	35
4	0	2	4	6
Più di 4	0	1	0	1
Totale	35	104	115	254

Tabella 3.31: Studio bivariato delle variabili percettori di reddito e frequenza di lettura.

3.2 Dettaglio sociale e demografico dei non lettori

Analizzando il dettaglio dei non lettori la differenza in termini di genere rimane, anche se risulta meno marcata, infatti il 56% riferisce di essere donna ed il 44% riferisce di essere uomo. Tra i non lettori, dunque, si collocano in proporzione più uomini rispetto che tra i lettori, confermando la tendenza che a leggere per svago ed intrattenimento sono soprattutto le donne, come mostra la Tabella 3.32.

Sesso			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Maschio	12	44%	44%
Femmina	15	56%	100%
Totale	27	100%	

Tabella 3.32: Non lettori: frequenze relative e cumulate della variabile sesso.

La maggior parte dei non lettori appartiene inoltre alle prime due fasce di età proposte: per il 44,444% si tratta di persone dai 18 ai 24 anni e per il 29,630% si tratta di persone dai 25 ai 29 anni. Il campione dei non lettori si dichiara dunque giovane, non mostrando evidenti differenze con il campione dei lettori, questo perché la totalità degli intervistati rappresentava in maggioranza un pubblico giovane, come si evince dalla Tabella 3.33.

Età			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Da 18 a 24 anni	12	44,444%	44,444%
Da 25 a 29 anni	8	29,630%	74,074%
Da 35 a 39 anni	2	7,407%	81,481%
Da 40 a 44 anni	1	3,704%	85,185%
Da 50 a 54 anni	1	3,704%	88,889%
Da 55 a 59 anni	1	3,704%	92,593%
65 anni o più	2	7,407%	100,000%
Totale	27	100,000%	

Tabella 3.33: Non lettori: frequenze relative e cumulate della variabile età.

Il 55,556% dei non lettori dichiara di possedere la laurea triennale; dunque più di cinque persone su dieci, pur essendo laureate, non leggono per svago, come viene riportato nella Tabella 3.34.

Titolo di studio			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Nessun titolo di studio	0	0,000%	0,000%
Licenza di scuola elementare	1	3,704%	3,704%
Licenza di scuola media	3	11,111%	14,815%
Diploma di scuola superiore (3 anni)	0	0,000%	14,815%
Diploma di maturità scuola superiore (5 anni)	4	14,815%	29,630%
Laurea triennale	15	55,556%	85,185%
Laurea magistrale	2	7,407%	92,593%
Istruzione post universitaria (dottorato, master...)	2	7,407%	100,000%
Totale	27	100,000%	

Tabella 3.34: Non lettori: frequenze relative e cumulate della variabile titolo di studio.

In termini di condizioni lavorative, il 44,444% del campione dei non lettori dichiara di essere studente ed il 33,333% dichiara di svolgere una mansione come impiegato od operaio, come illustrato nella Tabella 3.35.

Occupazione			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Dirigente/ Quadro	0	0,000%	0,000%
Impiegato/ Operaio	9	33,333%	33,333%
Libero professionista	0	0,000%	33,333%
Lavoratore in proprio	2	7,407%	40,741%
Studente	12	44,444%	85,185%
Non occupato	2	7,407%	92,593%
Pensionato	1	3,704%	96,296%
Casalingo/a	1	3,704%	100,000%
Totale	27	100,000%	

Tabella 3.35: Non lettori: frequenze relative e cumulate della variabile occupazione.

Come nel caso dei lettori, anche per i non lettori la maggioranza risiede in Veneto, per un totale dell'81,481% dei rispondenti, come mostrato nella Tabella 3.36; le ragioni sono da ricercare nel fatto che la totalità dei rispondenti è per la maggior parte veneta, per via delle modalità di somministrazione del questionario.

Regione di residenza			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Abruzzo	1	3,704%	3,704%
Lazio	1	3,704%	7,407%
Marche	1	3,704%	11,111%
Puglia	2	7,407%	18,519%
Veneto	22	81,481%	100,000%
Totale	27	100,000%	

Tabella 3.36: Non lettori: frequenze relative e cumulate della variabile regione di residenza.

Anche nel caso dei non lettori, la maggioranza vive in una famiglia comprendente in tutto quattro persone (44,444% dei non lettori) ed il 33,333% di essi vive in una famiglia con tre componenti.

Componenti del nucleo familiare			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
1	1	3,704%	3,704%
2	2	7,407%	11,111%
3	9	33,333%	44,444%
4	12	44,444%	88,889%
Più di 4	3	11,111%	100,000%
Totale	27	100,000%	

Tabella 3.37: Non lettori: frequenze relative e cumulate della variabile componenti del nucleo familiare.

La maggior parte dei nuclei familiari in questione non ospita né anziani né bambini, situazione dichiarata dal 70,370% dei non lettori, come si evince dalla Tabella 3.38. Il dato è presumibilmente legato al fatto che la maggioranza degli intervistati totali è giovane, e vive in casa con i genitori.

Il Suo nucleo familiare comprende anche anziani e/o bambini?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Sì	8	29,630%	29,630%
No	19	70,370%	100,000%
Totale	27	100,000%	

Tabella 3.38: Non lettori: frequenze relative e cumulate della variabile presenza di anziani o bambini nel nucleo familiare.

I componenti della famiglia che percepiscono reddito sono due nel 55,556% dei casi e sono tre nel 25,926% dei casi; la differenza con il campione dei lettori risiede nel fatto che, nel caso di questi ultimi, le famiglie con una sola persona lavoratrice erano numericamente maggiori rispetto alle famiglie con tre persone lavoratrici, come si evince dalla Tabella 3.39. La conclusione è che, nelle famiglie che contano un numero maggiore di persone lavoratrici, è più alta la presenza di non lettori: in altre parole l'assenza di lettura si lega in maniera direttamente proporzionale alla presenza più marcata di una condizione lavorativa.

Quante persone percepiscono un reddito all'interno del Suo nucleo familiare?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
1	2	7,407%	7,407%
2	15	55,556%	62,963%
3	7	25,926%	88,889%
4	1	3,704%	92,593%
Più di 4	2	7,407%	100,000%
Totale	27	100,000%	

Tabella 3.39: Non lettori: frequenze relative e cumulate della variabile percettori di reddito.

Capitolo 4

Risultati dell'indagine: analisi univariate

4.1 Analisi univariate generali

I primi risultati presentati consistono nei resoconti delle analisi univariate, ovvero lo studio degli attributi del questionario considerati singolarmente; in altre parole, per ogni domanda del questionario sono state analizzate nel dettaglio le risposte degli intervistati, svolgendo una conta del numero di risposte date per ciascuna scelta proposta. Per l'analisi univariata sono state raccolte in una tabella le distribuzioni di frequenza del carattere in oggetto, avendo dunque misurato la frequenza di appartenenza dei dati alle diverse categorie. Le frequenze riportate in tabella sono di tre tipi:

- frequenza assoluta: il numero di risposte per ogni alternativa presentata, considerando i quesiti uno per volta;
- frequenza relativa percentuale: la percentuale di risposte per ogni alternativa presentata, fatto 100 il totale delle risposte del campione;
- frequenza cumulata percentuale: la percentuale cumulata di risposte conferite all'alternativa considerata ed a quelle precedenti, con soluzione di continuità; in altre parole la somma tra la frequenza relativa percentuale dell'alternativa in esame e la frequenza relativa percentuale delle alternative ad essa precedenti.

L'analisi monovariata dei quesiti è utile per studiare in maniera generale la tendenza delle risposte a ciascun quesito, individuando la modalità di risposta che si presenta più spesso, dando modo al ricercatore di individuare i comportamenti più diffusi tra i lettori.

In seguito vengono dunque riportati i risultati delle analisi univariate, raccolti in tabelle.

4.1.1 Quesito q1: Negli ultimi 12 mesi ha letto libri per hobby?

La prima domanda del questionario consisteva in una domanda filtro, che aveva il compito di suddividere la totalità del campione in due gruppi principali: il gruppo dei lettori ed il gruppo dei non lettori.

q1 - Negli ultimi 12 mesi ha letto libri per hobby?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
sì	254	90%	90%
no	27	10%	100%
Totale	281	100%	

Tabella 4.1: *Frequenze di lettura di almeno un libro negli ultimi 12 mesi*

Dai risultati riportati nella Tabella 4.1, si evince che 254 rispondenti su un totale di 281 hanno dichiarato di aver letto almeno un libro, per hobby, nell'ultimo anno; il campione analizzato è composto dunque al 90% da persone che non sono state attive nell'ambito della lettura non obbligatoria nell'ultimo anno. Il restante 10% che ha invece dichiarato di non aver letto nemmeno un libro nell'ultimo anno non ha avuto accesso alle domande riservate ai lettori ed è stato indirizzato alla compilazione del profilo socio-demografico.

4.1.2 Quesito q2: Quanti libri legge per hobby, indicativamente, all'anno?

Il secondo quesito, le cui risposte sono riportate nella Tabella 4.2, indagava la frequenza di lettura, per comprendere in che fascia si collocano coloro che si sono dichiarati lettori, facendo riferimento alla suddivisione Istat:

- è considerato “lettore debole” un individuo che legge massimo 3 libri all'anno;
- è considerato “lettore medio” un individuo che legge da 4 ad 11 libri all'anno;
- è considerato “lettore forte” un individuo che legge più di 12 libri all'anno.

La strategia di promozione e pubblicità deve essere mirata per lo specifico target per poter raggiungere il massimo risultato; questa suddivisione permette di valutare diverse strategie di marketing sulla base dei diversi comportamenti di lettura, a cominciare dal target più numeroso.

q2 - Quanti libri legge per hobby, indicativamente, all'anno?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Da 1 a 3 libri	35	13,780%	13,780%
Da 4 a 11 libri	104	40,945%	54,724%
Più di 12 libri	115	45,276%	100,000%
Totale	254	100,000%	

Tabella 4.2: *Frequenze relative alla quantità di libri letti per la suddivisione tra lettori forti, medi e deboli.*

Il campione dei lettori, il quale contava 254 individui, è stato suddiviso nei 3 macro-gruppi descritti in precedenza: la maggior parte dei rispondenti, per un totale di 115 individui su 254, rientra nella categoria dei “lettori forti”, segue per numerosità la categoria dei “lettori medi”, che conta 104 individui, mentre la categoria del “lettori deboli” racchiude la restante minoranza, dell’ordine di 35 individui. La situazione si riassume dunque come segue:

- i “lettori forti” rappresentano il 45,276% del campione;
- i “lettori medi” rappresentano il 40,945% del campione;
- i “lettori deboli” rappresentano il 13,780% del campione.

I dati risultano dunque in controtendenza rispetto quanto studiato nella rassegna della letteratura, in cui la maggior parte dei lettori si dichiarava debole.

4.1.3 Quesito q3: Tra i libri che ha letto, quanti provenivano dalla biblioteca o sono stati presi in prestito da conoscenti?

È stato proposto un quesito sulla fruizione di biblioteche o prestiti, con lo scopo di chiarire il comportamento di riferimento dei lettori in veste di acquirenti; in particolare si è voluta indagare la tendenza all’acquisto di tasca propria od al prestito, evidenziando la quantità di libri che sono stati acquistati da rivenditori autorizzati e la quantità di quelli presi in prestito, e dunque non acquistati direttamente. Le risposte a questo quesito sono riportate di seguito, nella Tabella 4.3.

q3 - Tra i libri che ha letto, quanti provenivano dalla biblioteca o sono stati presi in prestito da conoscenti?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Nessuno	125	49,213%	49,213%
Meno della metà	98	38,583%	87,795%
Più della metà	26	10,236%	98,031%
Tutti	5	1,969%	100,000%
Totale	254	100,000%	

Tabella 4.3: *Frequenze relative alla quantità di libri presi in prestito da biblioteche o conoscenti.*

La tendenza principale è quella di acquistare i libri piuttosto che prenderli in prestito infatti, osservando le frequenze cumulate, osserviamo che circa l'88% dei rispondenti ha dichiarato di aver preso in prestito meno della metà dei libri che ha letto. Il dato è rassicurante dal punto di vista della casa editrice, in quanto il 98,031% degli individui acquista almeno uno dei libri che legge.

4.1.4 Quesito q4: Quanto spesso si ritaglia un po' di tempo per leggere?

Questo quesito aveva lo scopo di indagare la frequenza di lettura, dato che si lega con la quantità di libri letti all'anno. I risultati sono presentati nella Tabella 4.4.

q4 - Quanto spesso si ritaglia un po' di tempo per leggere?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Più di 10 volte al mese	146	57,480%	57,480%
Dalle 5 alle 10 volte al mese	45	17,717%	75,197%
Dalle 3 alle 5 volte al mese	38	14,961%	90,157%
1 o 2 volte al mese	17	6,693%	96,850%
Meno di 1 volta al mese	8	3,150%	100,000%
Totale	254	100,000%	

Tabella 4.4: Frequenze relative al tempo dedicato alla lettura.

Si osserva che il 90,157% dei rispondenti (229 individui) ha dichiarato di dedicare del tempo alla lettura come passatempo dalle 3 volte al mese in su. Considerando la quantità di libri letti dichiarata in precedenza, la tendenza è rispettata. Tuttavia, il 57,480% ha dichiarato di dedicarsi più di 10 volte al mese alla lettura, bensì in precedenza si erano dichiarati lettori "forti" una quantità minore di rispondenti, circa il 45% del campione. Si può dunque presupporre che il restante 12% che legge più di 10 volte al mese ma non rientra nella categoria dei lettori "forti" sia soggetto ad una di questa variabili:

- si dedichi alla lettura di libri con un numero di pagine maggiore;
- dedichi minore tempo ad ogni seduta di lettura.

4.1.5 Quesito q5: Quanto ritiene importanti i seguenti parametri quando deve scegliere un libro da leggere?

La domanda q5 è di fondamentale importanza nell'analisi del comportamento del lettore, perché indaga su quali aspetti i lettori fanno affidamento per leggere (e quindi in larga

parte anche comprare) un libro, e su quali siano i canali informativi da loro utilizzati più di sovente. Queste informazioni risultano preziose dal punto di vista della presente tesi, che mira a tracciare delle linee guida per l'implementazione di una strategia di marketing efficace da parte delle case editrici, partendo dalle dirette dichiarazioni della clientela.

Le risposte al presente quesito saranno studiate minuziosamente nel capitolo successivo. La domanda q5 del questionario chiedeva ai rispondenti di valutare, attraverso una scala Likert con range di punteggio da 1 a 7, quanto essi ritenessero importanti i parametri elencati nella scelta di un libro da leggere. I parametri sono stati codificati, ai fini dell'analisi, come di seguito:

- q5a: Consigli di amici e parenti
- q5b: Consigli dalla casa editrice (sul sito o sui social network)
- q5c: Segnalazioni su pagine o gruppi specializzati e community sui social
- q5d: Consigli di bibliotecari/insegnanti/esperti nel settore del libro
- q5e: Consigli di persone famose (attori, ballerini, cantanti, influencer)
- q5f: Interviste e presenza dell'autore sui media (TV, radio, Internet)
- q5g: Classifiche sui giornali/riviste/internet

Di seguito verrà analizzato, su diversi livelli di dettaglio, ciascuno dei parametri proposti. Anche in questo caso, per ciascun parametro analizzato sono state calcolate le distribuzioni di frequenza, in particolare la frequenza assoluta, quella relativa e quella cumulata. Per quanto riguarda le statistiche descrittive dei singoli parametri, si è scelto di concentrarsi sulla valutazione della media e della moda; trattandosi di scale Likert a 7 valori, la media esprime il voto medio che è stato assegnato dai rispondenti all'importanza del parametro in questione. La moda esprime invece il valore di importanza che si è presentato con maggior frequenza. In alternativa al calcolo della mediana è stato deciso di affidarsi alle frequenze cumulate, in quanto permettono di evidenziare in maniera chiara la distribuzione delle valutazioni in termini percentuali, suddivise consecutivamente per punteggio, tenendo conto ad ogni step delle frequenze relative riferite ai punteggi precedenti.

Per quanto riguarda la variabile q5a che descrive la valutazione conferita all'importanza dei consigli di amici e parenti nella scelta di un libro da leggere, si è registrato il punteggio "5" come valore modale, seguito a poca distanza dal punteggio "4"; il calcolo delle

frequenze ha evidenziato che il 22,8% dei rispondenti (58 unità statistiche) considerano dunque il suddetto parametro “abbastanza importante”, ed il 20,5% dei rispondenti (52 unità statistiche) lo considerano “sufficientemente importante”. La media dei punteggi assegnati conferma quanto suddetto, attestandosi al valore 4,421. I risultati sono presentati nella Tabella 4.5. Si può dunque asserire che il campione, mediamente, considera abbastanza o sufficientemente importanti i consigli di amici e parenti quando deve scegliere un libro.

q5a - Quanto ritiene importanti i consigli di amici e parenti quando deve scegliere un libro da leggere?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	17	6.693%	6.693%
Minimamente importante	27	10.630%	17.323%
Poco importante	26	10.236%	27.559%
Sufficientemente importante	52	20.472%	48.031%
Abbastanza importante	58	22.835%	70.866%
Molto importante	42	16.535%	87.402%
Massimamente importante	32	12.598%	100.000%
Totale	254	100.000%	
Media	4.421		

Tabella 4.5: *Frequenze relative alle valutazioni assegnate ai consigli di amici e parenti nella scelta del libro.*

Per quanto riguarda invece il parametro q5b, che considerava i consigli della casa editrice, si è evinto che il 53% dei rispondenti ha conferito valutazioni insufficienti all'importanza di tale carattere. Questo è descritto in particolar modo dall'analisi delle frequenze cumulate, oltre che dalla media delle valutazioni, pari a 3,5, e dalla moda, corrispondente alla valutazione “3” (21,7% dei rispondenti, per un totale di 55 unità statistiche). Si evince da questi risultati, presentati nella Tabella 4.6, che il campione, in media, non è sufficientemente interessato ai consigli online della casa editrice quando deve scegliere un libro da leggere.

q5b - Quanto ritiene importanti i consigli dalla casa editrice (sul sito o sui social network) quando deve scegliere un libro da leggere?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	40	15.748%	15.748%
Minimamente importante	40	15.748%	31.496%
Poco importante	55	21.654%	53.150%
Sufficientemente importante	44	17.323%	70.472%
Abbastanza importante	39	15.354%	85.827%
Molto importante	19	7.480%	93.307%
Massimamente importante	17	6.693%	100.000%
Totale	254	100.000%	
Media	3.500		

Tabella 4.6: Frequenze relative alle valutazioni assegnate ai consigli della casa editrice (sul sito o sui social network) nella scelta del libro.

Il parametro q5c misurava il livello di importanza assegnato alle segnalazioni sui social network, in particolare su pagine, gruppi o communities specializzati nel settore. Il punteggio che è stato assegnato dai rispondenti con maggior frequenza all'importanza di tale variabile è stato pari a "5", corrispondente ad "abbastanza importante" (20,9% dei rispondenti, per un totale di 53 unità) seguito dalla valutazione "6", corrispondente a "molto importante" (18,9% dei rispondenti, 48 unità statistiche). La media delle valutazioni è dunque positiva, e si attesta al valore di 4,378, confermando la prevalenza di valutazioni positive da parte del campione. Il terzo punteggio in ordine di numerosità è rappresentato dal valore "3", corrispondente alla valutazione "poco importante" (44 record, 17,3% del totale dei rispondenti). Si può dunque affermare che, in media, il campione considera abbastanza, od addirittura molto importanti, i consigli di pagine o gruppi del settore librario sui social network, sebbene vi sia un gruppo significativo di rispondenti (17,3%) che li considera poco importanti. I dati raccolti sono riportati nella Tabella 4.7.

q5c - Quanto ritiene importanti le segnalazioni su pagine o gruppi specializzati e community sui social quando deve scegliere un libro da leggere?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	22	8.661%	8.661%
Minimamente importante	24	9.449%	18.110%
Poco importante	44	17.323%	35.433%
Sufficientemente importante	28	11.024%	46.457%

Abbastanza importante	53	20.866%	67.323%
Molto importante	48	18.898%	86.220%
Massimamente importante	35	13.780%	100.000%
Totale	254	100.000%	
Media	4.378		

Tabella 4.7: *Frequenze relative alle valutazioni assegnate alle segnalazioni su pagine o gruppi specializzati e community sui social nella scelta del libro.*

La variabile q5d, che esprimeva le valutazioni di importanza ai consigli degli esperti nel settore del libro (ad esempio bibliotecari ed insegnanti), ha mostrato una distribuzione di valutazioni per molti aspetti sovrapponibile a quella di q5c: il valore modale corrisponde al punteggio di “5”, codificato come “abbastanza importante” (52 unità statistiche, corrispondenti al 20,5% dei rispondenti), seguito dalla valutazione di “6”, codificata come “molto importante” (50 unità statistiche, corrispondenti al 19,7% dei rispondenti). Il 70,9% del campione ha conferito valutazioni considerate positive alla variabile in questione. La media è di 4,405 e conferma, dunque, che la maggioranza del campione considera abbastanza o molto importanti i consigli degli esperti del libro nella scelta di un libro da leggere. I dati sono riportati nella Tabella 4.8.

q5d - Quanto ritiene importanti i consigli di bibliotecari/insegnanti/esperti nel settore del libro quando deve scegliere un libro da leggere?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	33	12.992%	12.992%
Minimamente importante	17	6.693%	19.685%
Poco importante	24	9.449%	29.134%
Sufficientemente importante	42	16.535%	45.669%
Abbastanza importante	52	20.472%	66.142%
Molto importante	50	19.685%	85.827%
Massimamente importante	36	14.173%	100.000%
Totale	254	100.000%	
Media	4.406		

Tabella 4.8: *Frequenze relative alle valutazioni assegnate ai consigli di bibliotecari/insegnanti/esperti nel settore del libro nella scelta del libro.*

Per ciò che riguarda le valutazioni ai consigli di persone famose (attori, ballerini, cantanti, influencer), codificati attraverso il parametro q5e, la posizione del campione risulta molto chiara: il 48,8% dei rispondenti (per un totale di 123 unità statistiche su 254) afferma con

sicurezza di considerare per nulla importanti (voto “1”) le segnalazioni di tali personalità del mondo dello spettacolo. Il punteggio medio infatti è 2,154 su un massimo di 7, a conferma del netto disinteresse del campione nei confronti di suggerimenti e indicazioni sui libri da scegliere da parte delle celebrità, come si evince dalla Tabella 4.9.

q5e - Quanto ritiene importanti i consigli di persone famose (attori, ballerini, cantanti, influencer) quando deve scegliere un libro da leggere?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	123	48.425%	48.425%
Minimamente importante	48	18.898%	67.323%
Poco importante	34	13.386%	80.709%
Sufficientemente importante	27	10.630%	91.339%
Abbastanza importante	15	5.906%	97.244%
Molto importante	6	2.362%	99.606%
Massimamente importante	1	0.394%	100.000%
Totale	254	100.000%	
Media	2.154		

Tabella 4.9: Frequenze relative alle valutazioni assegnate ai consigli di persone famose (attori, ballerini, cantanti, influencer) nella scelta del libro.

La presenza dell'autore del libro sui principali media, sia tradizionali che new media, viene valutata per nulla importante dalla maggioranza degli intervistati, che le hanno assegnato un punteggio pari ad “1” (61 record, 24% del campione). Un'altra importante percentuale, la seconda per numerosità, considera la visibilità dell'autore minimamente importante, codificata con la valutazione di “2” (46 unità, il 18,2% dei rispondenti). Anche la media delle valutazioni si attesta ad un livello non sufficiente, 3,244. La maggioranza del campione considera quindi per nulla o minimamente importanti le interviste e la presenza dell'autore del libro sui media. I risultati sono riportati nella Tabella 4.10.

q5f - Quanto ritiene importanti interviste e presenza dell'autore sui media (TV, radio, Internet) quando deve scegliere un libro da leggere?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	61	24.016%	24.016%
Minimamente importante	46	18.110%	42.126%
Poco importante	38	14.961%	57.087%
Sufficientemente importante	34	13.386%	70.472%
Abbastanza importante	40	15.748%	86.220%

Molto importante	24	9.449%	95.669%
Massimamente importante	11	4.331%	100.000%
Totale	254	100.000%	
Media	3.244		

Tabella 4.10: *Frequenze relative alle valutazioni assegnate alle interviste e presenza dell'autore sui media (TV, radio, Internet) nella scelta del libro.*

Le classifiche su giornali, riviste o internet (q5g) sono state considerate in modi diversi, in particolare il 21,7% degli intervistati, per un totale di 55 rispondenti, ha conferito loro la valutazione minima, “per nulla importante” corrispondente al punteggio “1”. Il secondo voto in classifica per numerosità è, tuttavia, il voto “4” che esprime la sufficienza all’importanza del parametro (scelto dal 17,7% degli intervistati, per un totale di 45 rispondenti). La media dei voti si attesta comunque a 3,472, valutazione che non raggiunge la sufficienza. Considerando infine la frequenza cumulata si evince che il 51,2% sul totale dei rispondenti ha considerato insufficiente l’importanza di tale parametro. Possiamo concludere che la maggioranza degli intervistati considera per nulla importanti le classifiche dei libri più letti sia su carta che su internet. Si segnala tuttavia la presenza di una significativa fetta di intervistati che considerano invece tale caratteristica sufficientemente influente nella scelta del libro. I risultati sono presentati nella Tabella 4.11.

q5g - Quanto ritiene importanti le classifiche sui giornali/riviste/internet quando deve scegliere un libro da leggere?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	55	21.654%	21.654%
Minimamente importante	33	12.992%	34.646%
Poco importante	42	16.535%	51.181%
Sufficientemente importante	45	17.717%	68.898%
Abbastanza importante	35	13.780%	82.677%
Molto importante	28	11.024%	93.701%
Massimamente importante	16	6.299%	100.000%
Totale	254	100.000%	
Media	3.472		

Tabella 4.11: *Frequenze relative alle valutazioni assegnate alle classifiche sui giornali/riviste/internet nella scelta del libro.*

Si è operato inoltre un confronto tra le medie delle singole variabili componenti la variabile q5. Attraverso tale confronto è possibile fare una comparazione tra le caratteristiche alle quali il campione ha assegnato le valutazioni più alte e le caratteristiche che invece sono state segnalate come meno importanti dal campione nella scelta del libro. In particolare, le medie considerate sono state rappresentate attraverso un grafico a barre, rappresentato nella Figura 4.1.

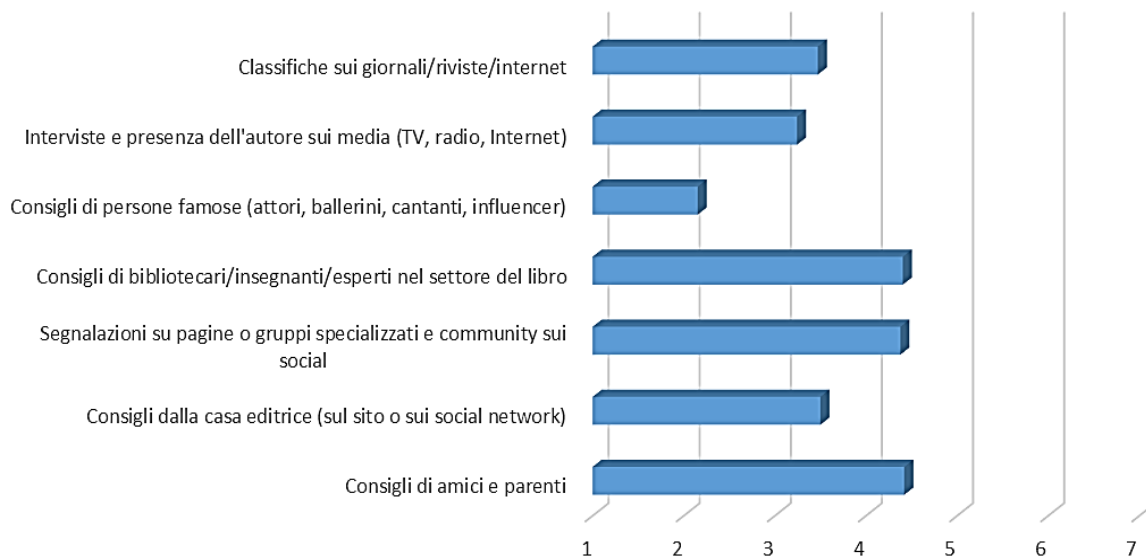


Figura 4.1: Confronto tra le medie delle singole variabili componenti il quesito q5.

Dal grafico si evince chiaramente che gli intervistati hanno assegnato valutazioni di importanza più che sufficienti a 3 variabili su 7, in ordine decrescente sono: consigli di amici e parenti, consigli di bibliotecari, insegnanti ed esperti del settore e segnalazioni su pagine o gruppi specializzati nel settore librario. Il parametro che è stato considerato meno importante si riferisce ai consigli di persone famose nel mondo dello spettacolo. Da questa analisi risulta evidente il fatto che i rispondenti tendono a valutare come massimamente importanti nella scelta di un libro i consigli e i suggerimenti che derivano da entità a loro care (familiari, parenti o amici), oppure da entità che ritengono indiscutibilmente esperte nel settore, che siano comunque a loro vicine: bibliotecari e insegnanti in quanto persone esperte con le quali poter avere un dialogo faccia a faccia, pagine o gruppi specializzati sui social network in quanto communities informali nelle quali poter condividere informazioni e proposte letterarie come in un gruppo di amici. I consigli della casa editrice si classificano al quarto posto come livello di importanze,

probabilmente per via del rapporto formale e della distanza percepita con l'istituzione "casa editrice". Gli intervistati preferiscono quindi avere la possibilità di confrontarsi in modo rapido e soprattutto informale, esigenza presumibilmente dettata dalla natura del prodotto stesso, la cui preferenza è in buona parte frutto del personale gusto letterario legato alla posizione sociale che i singoli individui vogliono occupare. Questo senso di appartenenza ad un gruppo con i medesimi interessi, mostrato nel momento della scelta di un libro da leggere, spinge i lettori a rivolgersi a quelle entità che essi ritengono umanamente più vicine a loro e con le quali possono confidarsi apertamente e tranquillamente, come se stessero parlando a degli amici.

4.1.6 Quesito q6 Assegna una valutazione che esprima il Suo livello di gradimento per i generi presentati di seguito

Il quesito codificato con la denominazione q6 intimava agli intervistati di dare una valutazione da 1 a 7, secondo il modello della scala Likert, che dichiarasse il livello di gradimento associato ai generi narrativi proposti.

Gli attributi sono stati codificati, ai fini dell'analisi, come di seguito:

- q6a: Storico
- q6b: Fantascienza
- q6c: Fantasy
- q6d: Avventura/Azione
- q6e: Per ragazzi
- q6f: Horror
- q6g: Thriller/Crimine/Giallo
- q6h: Erotico
- q6i: Uморistico
- q6l: Drammatico
- q6m: Poesie/Opere teatrali/Favole/Fiabe
- q6n: Sentimentale/Amore

In seguito verrà presentato dettaglio delle analisi univariate relative ai generi narrativi.

Le valutazioni superiori al “4”, ovvero quelle che superano la “sufficienza”, indicano una propensione crescente all’acquisto di un libro di tale genere narrativo da parte del lettore, partendo dalla intenzione minima di acquisto corrispondente alla valutazione appena sufficiente.

I dati sono dunque stati interpretati suddividendoli nelle categorie di valutazione sufficiente ed insufficiente, che assumono il seguente significato:

- se un genere narrativo viene valutato con valori rientranti nella soglia della sufficienza, il rispondente presenta una propensione, anche minima, all’acquisto;
- se un genere narrativo viene valutato con valori inferiori alla soglia della sufficienza, il rispondente non presenta propensione all’acquisto.

Come per il quesito q5, anche per l’attributo q6 si sono considerate la media e la moda come statistiche descrittive di riferimento, e per ciascun parametro sono riportate le frequenze assolute, quelle relative percentuali e quelle cumulate percentuali.

Si è considerato per primo il genere storico, codificato con la denominazione q6a, come narrazione ambientata nel passato che descrive fatti realistici mantenendo vive condizioni sociali, usi e costumi dell’epoca. I risultati sono rappresentati nella Tabella 4.12.

q6a - Assegna una valutazione che esprima il Suo livello di gradimento per il genere storico			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla gradito	29	11.417%	11.417%
Minimamente gradito	25	9.843%	21.260%
Poco gradito	42	16.535%	37.795%
Sufficientemente gradito	30	11.811%	49.606%
Abbastanza gradito	35	13.780%	63.386%
Molto gradito	43	16.929%	80.315%
Massimamente gradito	50	19.685%	100.000%
Totale	254	100.000%	
Media	4.362		

Tabella 4.12: Frequenze relative alle valutazioni assegnate al genere storico.

Le valutazioni a questo genere letterario si distribuiscono abbastanza uniformemente per tutta la scala. Il valore modale è pari a 50 risposte e si è manifestato in corrispondenza della valutazione “massimamente gradito”. Le valutazioni positive superano di gran lunga

quelle negative, infatti solo il 37,795% ha dichiarato di non ritenere sufficiente in termini di gradimento il suddetto genere, mentre il restante 62,205% ha assegnato valori almeno sufficienti. La media delle valutazioni si colloca al centro della scala, attestandosi al 4,362 e confermando il medio gradimento per le opere di tipo storico.

Alla luce dei dati, si può dunque dichiarare con certezza che il genere storico risulta largamente apprezzato dal campione, nell'ordine di 3 individui su 5.

Il genere descritto successivamente consiste nella fantascienza, che annovera tutte quelle opere caratterizzate da vicende ed intuizioni di carattere scientifico. I risultati sono presentati nella Tabella 4.13.

q6b - Assegna una valutazione che esprima il Suo livello di gradimento per il genere fantascienza			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla gradito	57	22.441%	22.441%
Minimamente gradito	57	22.441%	44.882%
Poco gradito	34	13.386%	58.268%
Sufficientemente gradito	25	9.843%	68.110%
Abbastanza gradito	34	13.386%	81.496%
Molto gradito	18	7.087%	88.583%
Massimamente gradito	29	11.417%	100.000%
Totale	254	100.000%	
Media	3.362		

Tabella 4.13: *Frequenze relative alle valutazioni assegnate al genere fantascientifico.*

Il genere letterario fantascientifico ha registrato valutazioni in maggioranza negative: il 58,268% dei rispondenti hanno asserito di non gradirlo sufficientemente ed il dato modale pari a 57 individui si è manifestato sia per la valutazione minima della scala, “per nulla gradito”, sia per la valutazione appena superiore ma pur sempre negativa, “minimamente gradito”. Il 41,733% del campione dei lettori ha affermato, invece, di apprezzare sufficientemente il genere. La media dei voti ammonta a 3,362 e documenta la tendenza di soddisfazione negativa.

Dai dati si evince dunque che il genere fantascientifico non risulta sufficientemente apprezzato da 3 individui su 5.

Il terzo genere narrativo al quale era richiesto di dare una valutazione di gradimento è rappresentato dal novero fantasy, genere fantastico che descrive vicende di personaggi

dotati di capacità sovranaturali inserite in un ambiente magico e paesaggi fantastici. I dati sono riportati nella Tabella 4.14.

q6c - Assegna una valutazione che esprima il Suo livello di gradimento per il genere fantasy			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla gradito	43	16.929%	16.929%
Minimamente gradito	47	18.504%	35.433%
Poco gradito	30	11.811%	47.244%
Sufficientemente gradito	26	10.236%	57.480%
Abbastanza gradito	40	15.748%	73.228%
Molto gradito	22	8.661%	81.890%
Massimamente gradito	46	18.110%	100.000%
Totale	254	100.000%	
Media	3.878		

Tabella 4.14: Frequenze relative alle valutazioni assegnate al genere fantasy.

Le valutazioni assegnate a questo genere non mostrano una tendenza chiara ed univoca, al contrario i voti sono distribuiti in modo quasi uniforme in tutti i gradi della scala; l'orientamento risulta alquanto ambiguo in quanto le espressioni "massimamente gradito" e "minimamente gradito", ai lati opposti della scala, registrano pressoché stesso numero di risposte (46 e 47 record), ed anche le espressioni "abbastanza gradito" e "per nulla gradito" differiscono di appena 3 record (40 e 43 risposte).

In questa condizione risulta impossibile segnalare un andamento preciso delle risposte, poiché il campione si è scisso in due metà similmente consistenti: il 47,244% del campione ha assegnato valutazioni negative al genere in questione ed il 52,755% ha assegnato valutazioni positive. La media assume il valore di 3,878 che si colloca al centro della scala.

Si può dunque asserire che il campione si divide a metà per quanto concerne il gradimento del genere fantasy:

- una metà lo gradisce sufficientemente;
- una metà non lo gradisce sufficientemente.

Il quarto genere per il quale è stato richiesto di esprimere una valutazione comprende le opere di avventura e azione. Queste due categorie sono state agglomerate per il fatto che

spesso si presentano l'una come sfaccettatura dell'altra e, per certi versi, si sovrappongono. I risultati sono riportati nella Tabella 4.15.

q6d - Assegna una valutazione che esprima il Suo livello di gradimento per il genere avventura/azione

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla gradito	27	10.630%	10.630%
Minimamente gradito	30	11.811%	22.441%
Poco gradito	35	13.780%	36.220%
Sufficientemente gradito	47	18.504%	54.724%
Abbastanza gradito	57	22.441%	77.165%
Molto gradito	35	13.780%	90.945%
Massimamente gradito	23	9.055%	100.000%
Totale	254	100.000%	
Media	4.079		

Tabella 4.15: Frequenze relative alle valutazioni assegnate al genere avventura/azione.

Il dato modale si registra in corrispondenza della valutazione “abbastanza gradito” (per un totale di 57 risposte), seguita dall’espressione “sufficientemente gradito” (che contava 47 risposte). Nel complesso si può tracciare un andamento di carattere positivo per questa variabile, considerando che la percentuale di risposte negative risulta contenuta (36,220%).

Si può quindi affermare che la maggioranza del campione gradisce abbastanza (nell’ordine di valutazione pari a 4) il genere avventuroso; il che è attestato inoltre dal valore della media, di 4,079.

In seguito, nella Tabella 4.16, si presentano i risultati relativi al genere per ragazzi, il quale comprende tutte quelle opere dedicate ad un pubblico giovane, dai 10 anni in su, fino agli adolescenti.

q6e - Assegna una valutazione che esprima il Suo livello di gradimento per il genere per ragazzi

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla gradito	55	21.654%	21.654%
Minimamente gradito	59	23.228%	44.882%
Poco gradito	34	13.386%	58.268%
Sufficientemente gradito	42	16.535%	74.803%
Abbastanza gradito	29	11.417%	86.220%

Molto gradito	21	8.268%	94.488%
Massimamente gradito	14	5.512%	100.000%
Totale	254	100.000%	
Media	3.197		

Tabella 4.16: Frequenze relative alle valutazioni assegnate al genere per ragazzi.

La tendenza di apprezzamento si dimostra negativa: il 58,268% degli intervistati hanno conferito valutazioni che denotano un livello di gradimento non sufficiente. A conferma del suddetto andamento si dimostra il dato modale, pari a 59 individui e relativo all'espressione "minimamente gradito" corrispondente alla valutazione pari a 2 e seguito dall'espressione "per nulla gradito" corrispondente al voto "1" della scala, in cui si contano 55 individui.

L'andamento risulta negativo: quasi 6 rispondenti su 10 non apprezzano il genere per ragazzi in maniera sufficiente, dimostrando l'assenza di propensione all'acquisto. Si può consciamente presupporre che il dato raccolto sia influenzato dalla fascia d'età in cui il campione dei lettori si attesta, superiore ai 17 anni. Le valutazioni basse relative al genere si possono dunque spiegare con il fatto che il campione presenta, in media, un'età superiore a quella del target a cui si riferisce codesto genere narrativo e, dunque, esso non risulta nel complesso apprezzato.

Il genere narrativo successivamente analizzato consiste nell'orrore, che comprende tutte le opere che suscitano nel lettore sentimenti di spavento e ribrezzo. Come si evince dalla Tabella 4.17, il genere non risulta essere apprezzato, in media, dal campione considerato: ben 106 individui, corrispondenti al 41,732% dei rispondenti l'hanno classificato con la valutazione più bassa della scala, "per nulla gradito".

I voti insufficienti assegnati al genere orrifico raggiungono una frequenza cumulata percentuale del 69,685%, il che denota che 7 individui su 10 non lo apprezzano abbastanza; la tendenza è confermata dal valore assunto dalla media delle risposte per questo genere narrativo, che si è mantenuta ad un valore molto basso pari a 2,677.

q6f - Assegna una valutazione che esprima il Suo livello di gradimento per il genere horror			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla gradito	106	41.732%	41.732%
Minimamente gradito	39	15.354%	57.087%
Poco gradito	32	12.598%	69.685%

Sufficientemente gradito	28	11.024%	80.709%
Abbastanza gradito	18	7.087%	87.795%
Molto gradito	19	7.480%	95.276%
Massimamente gradito	12	4.724%	100.000%
Totale	254	100.000%	
Media	2.677		

Tabella 4.17: *Frequenze relative alle valutazioni assegnate al genere horror.*

Un genere con caratteristiche simili al tipo orrorifico, sebbene non sovrapponibili, consiste nel genere thriller che, a seconda delle sfumature narrative che assume, può sfacciarsi in giallo, crimine e poliziesco. I risultati relativi all'apprezzamento di questo genere vengono presentati nella Tabella 4.18.

q6g - Assegna una valutazione che esprima il Suo livello di gradimento per il genere thriller/crimine/giallo			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla gradito	32	12.598%	12.598%
Minimamente gradito	30	11.811%	24.409%
Poco gradito	28	11.024%	35.433%
Sufficientemente gradito	31	12.205%	47.638%
Abbastanza gradito	34	13.386%	61.024%
Molto gradito	43	16.929%	77.953%
Massimamente gradito	56	22.047%	100.000%
Totale	254	100.000%	
Media	4.409		

Tabella 4.18: *Frequenze relative alle valutazioni assegnate al genere thriller/crimine/giallo.*

Analizzando l'andamento delle risposte si nota come questo genere, il quale suscita nel lettore una versione più morbida e tollerabile dei sentimenti che provoca il genere horror, con l'aggiunta di interesse e curiosità per la risoluzione di enigmi o casi da risolvere, sia largamente più apprezzato. Il dato modale si riferisce al 22,047% del campione considerato, il quale ha valutato il genere assegnandogli il massimo voto della scala, "massimamente gradito". Il 64,567% del campione ha asserito di gradire il suddetto genere in maniera soddisfacente, e la media delle risposte assume il valore di 4,409 collocandosi di una valutazione e mezza al di sopra della soglia di non gradimento. Il

genere thriller/crimine/giallo gode di una certa popolarità tra gli individui del campione, infatti almeno 6 persone su 10 lo apprezzano in maniera sufficiente.

Il genere valutato in seguito consiste nel tipo erotico che include tutti gli scritti che trattano di temi legati alla sessualità ed all'amore corporeo. I dati raccolti sono presentati nella Tabella 4.19.

q6h - Assegna una valutazione che esprima il Suo livello di gradimento per il genere erotico			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla gradito	102	40.157%	40.157%
Minimamente gradito	51	20.079%	60.236%
Poco gradito	31	12.205%	72.441%
Sufficientemente gradito	30	11.811%	84.252%
Abbastanza gradito	26	10.236%	94.488%
Molto gradito	6	2.362%	96.850%
Massimamente gradito	8	3.150%	100.000%
Totale	254	100.000%	
Media	2.516		

Tabella 4.19: Frequenze relative alle valutazioni assegnate al genere erotico.

Nonostante la netta maggioranza di lettrici di sesso femminile, il genere non risulta mediamente amato; ben il 72,441% dei lettori lo hanno considerato non sufficiente in termini di gradimento, ed il dato modale è riferito alla valutazione più bassa di tutta la scala, la quale dimostra un interesse al di sotto del minimo da parte del campione (102 risposte in totale) e la media di 2,516 conferma questa tendenza negativa.

Possiamo asserire con certezza che più di 7 individui su 10 del campione non apprezzano sufficientemente gli scritti erotici.

Le valutazioni presentate di seguito, nella Tabella 4.20, si riferiscono al genere letterario umoristico, il quale suscita nel lettore divertimento e risate.

q6i - Assegna una valutazione che esprima il Suo livello di gradimento per il genere umoristico			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla gradito	51	20.079%	20.079%
Minimamente gradito	34	13.386%	33.465%
Poco gradito	56	22.047%	55.512%

Sufficientemente gradito	39	15.354%	70.866%
Abbastanza gradito	31	12.205%	83.071%
Molto gradito	28	11.024%	94.094%
Massimamente gradito	15	5.906%	100.000%
Totale	254	100.000%	
Media	3.429		

Tabella 4.20: *Frequenze relative alle valutazioni assegnate al genere umoristico.*

Anche in questo caso più della metà dei lettori (precisamente il 55,512% del campione) non gradisce il genere, assegnandogli dunque delle valutazioni al di sotto della soglia di sufficienza. La moda dei dati corrisponde a 56 individui ed alla valutazione “poco gradito”, corrispondente al valore “3” della scala Likert. Le valutazioni risultano dunque pressoché negative ma la media di 3,429 si avvicina alla sufficienza.

Si è dunque evinto che il campione si sia diviso quasi a metà per quanto riguarda l’apprezzamento del genere: circa 5 persone su 10 lo considerano interessante, le altre 5 non lo trovano allettante.

Il genere successivo consiste nel genere drammatico il quale tratta di temi ad alto impatto emotivo, raccontati attraverso lo sviluppo di un rapporto tra due o più persone. Le frequenze relative a tale genere sono riportate nella Tabella 4.21.

q6l - Assegna una valutazione che esprima il Suo livello di gradimento per il genere drammatico			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla gradito	24	9.449%	9.449%
Minimamente gradito	15	5.906%	15.354%
Poco gradito	26	10.236%	25.591%
Sufficientemente gradito	28	11.024%	36.614%
Abbastanza gradito	49	19.291%	55.906%
Molto gradito	64	25.197%	81.102%
Massimamente gradito	48	18.898%	100.000%
Totale	254	100.000%	
Media	4.760		

Tabella 4.21: *Frequenze relative alle valutazioni assegnate al genere drammatico.*

Il genere trova largo apprezzamento tra i lettori, infatti solo il 25,591% dei rispondenti dichiara di non apprezzarlo; il restante 74,409% lo gradisce, con una maggioranza pari al 25,197% (per un totale di 64 rispondenti) che ha assegnato una valutazione pari a “6”, corrispondente all’espressione “molto gradito”. La media, pari a 4,760, denota una tendenza positiva delle risposte, collocando la valutazione media tra “sufficientemente” ed “abbastanza” gradito.

L’andamento risulta dunque molto positivo, più di 7 individui su 10 affermano di apprezzare il dramma.

Il genere di scritti trattato in seguito fa riferimento a tutte quelle opere, anche non in prosa, che sono tipicamente oggetto di rappresentazioni teatrali: in particolare si è concentrato il focus su poesie, opere teatrali in senso stretto, favole e fiabe. Le frequenze sono riportate nella Tabella 4.22.

q6m - Assegna una valutazione che esprima il Suo livello di gradimento per poesie/ opere teatrali/ favole/ fiabe			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla gradito	29	11.417%	11.417%
Minimamente gradito	42	16.535%	27.953%
Poco gradito	27	10.630%	38.583%
Sufficientemente gradito	37	14.567%	53.150%
Abbastanza gradito	45	17.717%	70.866%
Molto gradito	29	11.417%	82.283%
Massimamente gradito	45	17.717%	100.000%
Totale	254	100.000%	
Media	4.157		

Tabella 4.22: Frequenze relative alle valutazioni assegnate a poesie/opere teatrali/favole/fiabe.

La risposta più frequente in relazione al suddetto genere consiste nella valutazione massima della scala con il punteggio di “7” e traslitterata nell’espressione “massimamente gradito” (45 individui pari al 17,717%), sebbene a distanza di sole 3 risposte si attesti la seconda valutazione più frequente pari a “minimamente gradito”,

corrispondente alla valutazione di “2” (42 individui pari al 16,535% del campione). Il confronto fra le risposte positive e quelle negative ha evidenziato i seguenti risultati: il 38,583% ha considerato non sufficientemente apprezzato il genere, contro il 61,417% che l’ha considerato sufficientemente gradito. La media si attesta ad un valore pari a 4,157, a riconoscimento del fatto che il genere risulta, in media, sufficientemente apprezzato; si desume dunque dai dati raccolti che più di 6 individui su 10 considerano il genere letterario interessante.

L’ultimo genere narrativo studiato consiste nel genere rosa anche detto sentimentale che racconta vicende legate a temi di tipo romantico ed affettivo. I dati relativi al quesito vengono riportati nella Tabella 4.23.

q6n - Assegna una valutazione che esprima il Suo livello di gradimento per il genere sentimentale			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla gradito	37	14.567%	14.567%
Minimamente gradito	39	15.354%	29.921%
Poco gradito	31	12.205%	42.126%
Sufficientemente gradito	38	14.961%	57.087%
Abbastanza gradito	36	14.173%	71.260%
Molto gradito	39	15.354%	86.614%
Massimamente gradito	34	13.386%	100.000%
Totale	254	100.000%	
Media	3.984		

Tabella 4.23: Frequenze relative alle valutazioni assegnate al genere sentimentale.

Le valutazioni non evidenziano una tendenza in particolare, poiché la numerosità delle risposte è molto simile per le varie valutazioni. Il 42,126% dei rispondenti ha conferito valutazioni sotto la soglia della sufficienza mentre il restante 57,874% l’ha reputato un genere sufficientemente apprezzato. Il dato modale in questa situazione risulta essere poco significativo poiché le quantità di risposte per ogni voto differiscono di poche unità. La media si attesta al 3,984, al centro della scala sebbene non sia ancora una valutazione sufficiente.

La tendenza risulta, per i motivi elencati, di difficile definizione ma si può affermare che più di 4 individui su 10 non apprezzano le opere sentimentali, mentre più di 5 individui su 10 dichiarano il contrario.

Per comprendere quali sono stati i generi narrativi più apprezzati e quali i meno graditi si è proceduto operando un confronto di grandezza tra le medie osservate relative ad ogni sotto-variabile componente q6. I risultati sono presentati nella Figura 4.2.

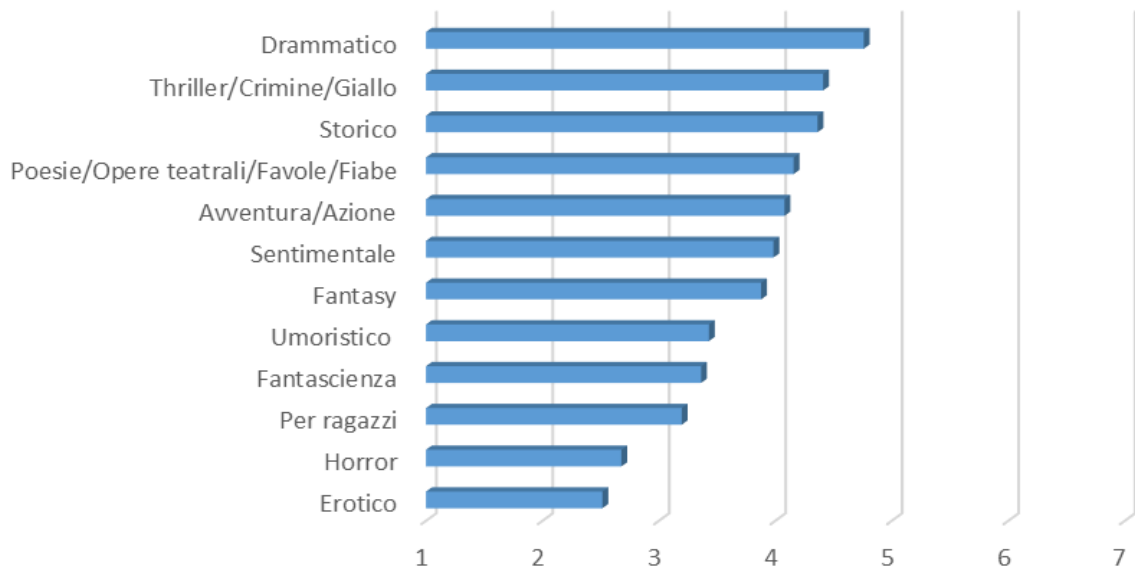


Figura 4.2: Confronto tra le medie delle singole variabili componenti il quesito q6.

I generi ammirati in maniera più che sufficiente sono risultati essere, in ordine dal più apprezzato al meno apprezzato:

- Drammatico;
- Thriller/Crimine/Giallo;
- Storico;
- Poesie/Opere teatrali/Favole/Fiabe;
- Avventura/Azione.

I generi narrativi che hanno collezionato, in media, valutazioni non sufficienti sono elencati di seguito, partendo dal meno gradito:

- Erotico;
- Horror;
- Per ragazzi;
- Fantascienza;
- Umoristico;
- Fantasy;

- Sentimentale.

Ragionando sui risultati ottenuti e concentrando l'attenzione sui primi 3 generi classificati (drammatico, thriller/crimine/giallo, storico) appare manifesto che i generi considerati maggiormente interessanti e piacevoli hanno in comune tra loro alcune caratteristiche, tra cui:

- La complessità della trama e l'impegno di lettura maggiore;
- La vicinanza delle vicende alla realtà;
- La serietà del tono della narrazione.

I lettori preferiscono dunque leggere dei libri che trattino di temi concreti e realistici, in alcuni casi che presentino vere e proprie esperienze di vita (per esempio la narrazione drammatica), raccontati con tono riflessivo e serio e che riescano a stuzzicare l'interesse per l'accattivante e l'enigma ma anche le emozioni ed i sentimenti più profondi. Le letture che rientrano nei generi primi classificati sono in media più impegnative rispetto alle tipologie di narrazione che hanno collezionato i punteggi più bassi, toccando temi per nulla astratti e presentando nel finale risvolti spesso commoventi o angosciosi.

4.1.7 Quesito q7: In che formato preferisce leggere?

La domanda sul formato di lettura prediletto consiste in una domanda filtro impiegata sia per contare il numero di lettori di formato cartaceo in relazione a quelli in formato digitale, sia per separare le due categorie, in modo da poter rivolgere ad ogni gruppo delle domande specifiche. I risultati del quesito sono presentati nella Tabella 4.24.

q7 - In che formato preferisce leggere?			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Cartaceo	222	87.402%	87.402%
Digitale	32	12.598%	100.000%
Totale	254	100.000%	

Tabella 4.24: Frequenze relative alla preferenza tra formato cartaceo e formato digitale.

Il formato cartaceo risulta largamente più utilizzato, per un totale di 222 preferenze su 254, che corrisponde quasi al 90% del campione. Il formato digitale conta, invece, solo 32 dichiarazioni di preferenza, il 12,598% del campione.

La grande maggioranza degli intervistati, dell'ordine di circa 9 persone su 10, preferisce leggere libri stampati piuttosto che fare affidamento a dispositivi digitali o elettronici come e-reader, tablet o personal computer. Sono stati indagati in seguito anche i motivi di questa scelta, che verranno esposti successivamente.

4.2 Analisi univariate: quesiti riservati ai lettori di formato cartaceo

Tutti i quesiti presentati in questo paragrafo sono stati sottoposti esclusivamente alla parte del campione dichiaratasi lettrice di formato cartaceo, composta da 222 individui tra u 254 dichiaratisi lettori.

La suddivisione tra genere di lettori permette di identificare delle tendenze di comportamento in base alle differenti abitudini di lettura.

4.2.1 Quesito q8: Assegna una valutazione che esprima l'importanza per i parametri presentati di seguito per la Sua scelta di un libro in formato cartaceo

La domanda aveva come obiettivo la costruzione di un grafico riassuntivo dei parametri a cui i lettori di libri cartacei danno più peso nella selezione di un libro da leggere che, come si è visto dai risultati precedenti, si può esprimere, nel 98,031% dei casi, con la domanda “che cosa conta maggiormente per un lettore durante l'acquisto di un libro cartaceo?”.

Anche questa domanda prevedeva una scala di valutazione Likert da 1 a 7 in cui la valutazione “1” identificava l'espressione “per nulla importante” e la valutazione “7” identificava l'espressione “massimamente importante”.

I parametri sono stati codificati, ai fini dell'analisi, in questo modo:

- q8a: Prezzo
- q8b: Copertina
- q8c: Disponibilità immediata
- q8d: Autore

- q8e: Trasportabilità
- q8f: Recensioni
- q8g: Genere narrativo
- q8h: Novità

Di seguito viene proposto lo studio dei risultati raccolti per le singole variabili comprese nell'intero quesito. La suddivisione tra valutazioni sufficienti e non sufficienti (rispettivamente superiori ed inferiori a "4") è stata operata alla stessa maniera della domanda relativa ai generi narrativi, in particolare:

- le valutazioni superiori alla sufficienza denotano una propensione all'acquisto del libro, anche minima, da parte del rispondente;
- le valutazioni inferiori alla sufficienza non denotano propensione all'acquisto del libro da parte del rispondente.

Anche per il quesito in questione si sono calcolate la media e la moda, le quali permettono di studiare l'andamento delle risposte e di classificare in maniera oggettiva le preferenze. Per ciascuna delle alternative si presentano le frequenze assolute, quelle relative percentuali e quelle cumulate percentuali. Il campione considerato è formato da coloro che hanno segnalato come preferenza di lettura il formato cartaceo, per un totale di 222 individui.

Il primo attributo analizzato consisteva nel prezzo. Si è dunque chiesto agli intervistati quale valutazione si sentissero di assegnare all'importanza del prezzo nel momento della scelta di un libro fisico. I risultati sono riportati nella Tabella 4.25.

q8a - Assegna una valutazione che esprima l'importanza del prezzo per la Sua scelta di un libro in formato cartaceo

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	14	6.306%	6.306%
Minimamente importante	21	9.459%	15.766%
Poco importante	32	14.414%	30.180%
Sufficientemente importante	39	17.568%	47.748%
Abbastanza importante	42	18.919%	66.667%
Molto importante	45	20.270%	86.937%

Massimamente importante	29	13.063%	100.000%
Totale	222	100.000%	
Media	4.464		

Tabella 4.25: Frequenze relative alle valutazioni sull'importanza del prezzo nella scelta di un libro in formato cartaceo.

La maggioranza dei rispondenti ha considerato il livello di prezzo del libro come sufficientemente importante nella scelta, in particolare il 69,82% del campione di lettori di libro cartaceo ritengono di assegnare un'importanza sufficientemente vincolante a questo parametro. La maggioranza di risposte si è registrata in corrispondenza delle valutazioni "5" e "6", rispettivamente "abbastanza importante" e "molto importante" (42 risposte per la prima, 45 risposte per la seconda). La media, pari a 4,464 conferma la tendenza.

Si può dunque affermare che il 52,252% del campione considerato, ovvero più di 5 persone su 10, assegna al prezzo un'importanza più che sufficiente nella scelta di un'opera cartacea.

Il parametro analizzato successivamente indagava l'influenza della copertina sulla scelta di un libro. I risultati sono presentati nella Tabella 4.26.

q8b - Assegna una valutazione che esprima l'importanza della copertina per la Sua scelta di un libro in formato cartaceo

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	17	7.658%	7.658%
Minimamente importante	19	8.559%	16.216%
Poco importante	20	9.009%	25.225%
Sufficientemente importante	43	19.369%	44.595%
Abbastanza importante	49	22.072%	66.667%
Molto importante	40	18.018%	84.685%
Massimamente importante	34	15.315%	100.000%
Totale	222	100.000%	
Media	4.550		

Tabella 4.26: Frequenze relative alle valutazioni sull'importanza della copertina nella scelta di un libro in formato cartaceo.

Anche in questo caso la tendenza è verso l'alto della scala; solo il 25,225% degli intervistati non reputa la copertina un parametro sufficientemente discriminante nella scelta. Le risposte con frequenza più alta si concentrano nella parte prominente della scala, con una maggiore concentrazione in corrispondenza delle espressioni "sufficientemente", "abbastanza" e "molto importante". La valutazione "6", trascritta attraverso l'idioma "abbastanza importante" ha collezionato il maggior numero di risposte, per un totale di 49 (il 22,072% dei rispondenti).

Si è evinto dunque che il 55,405% dei rispondenti, corrispondenti a più di 5 persone su 10, considerano la copertina una caratteristica più che sufficientemente discriminante nella scelta di un libro in formato cartaceo.

Il parametro successivo per il quale era richiesto di esprimere una valutazione di importanza era la disponibilità immediata del libro, ovvero la possibilità di ottenerlo direttamente in sede di acquisto. I dati raccolti sono riportati nella Tabella 4.28.

q8c - Assegna una valutazione che esprima l'importanza della disponibilità immediata per la Sua scelta di un libro in formato cartaceo

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	12	5.405%	5.405%
Minimamente importante	13	5.856%	11.261%
Poco importante	23	10.360%	21.622%
Sufficientemente importante	29	13.063%	34.685%
Abbastanza importante	42	18.919%	53.604%
Molto importante	56	25.225%	78.829%
Massimamente importante	47	21.171%	100.000%
Totale	222	100.000%	
Media	4.946		

Tabella 4.27: Frequenze relative alle valutazioni sull'importanza della disponibilità immediata nella scelta di un libro in formato cartaceo.

La disponibilità immediata gioca un ruolo discriminante nella scelta di un libro per il 78,378% dei rispondenti, la maggioranza dei quali lo ritiene un parametro "molto importante" (il 25,225% per un totale di 56 individui). Le risposte crescono

proporzionalmente di numero al crescere dell'importanza conferita al parametro. La media è pari a 4,946 e conferma questa tendenza di valutazione positiva, attestandosi quasi al livello della valutazione "abbastanza importante".

Si può affermare che circa 8 persone su 10 considerano la disponibilità immediata come sufficientemente vincolante nella scelta di un libro da leggere.

La conoscenza e l'apprezzamento dell'autore rappresenta l'attributo successivo, le cui valutazioni sono analizzate in seguito, a partire dai dati raccolti nella Tabella 4.28.

q8d - Assegna una valutazione che esprima l'importanza dell'autore per la Sua scelta di un libro in formato cartaceo

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	10	4.505%	4.505%
Minimamente importante	13	5.856%	10.360%
Poco importante	12	5.405%	15.766%
Sufficientemente importante	26	11.712%	27.477%
Abbastanza importante	45	20.270%	47.748%
Molto importante	44	19.820%	67.568%
Massimamente importante	72	32.432%	100.000%
Totale	222	100.000%	
Media	5.266		

Tabella 4.28: Frequenze relative alle valutazioni sull'importanza della copertina nella scelta di un libro in formato cartaceo.

Il suddetto parametro risulta del tutto vincolante nella scelta di molti dei rispondenti, infatti solo il 15,766% di essi l'ha dichiarato di importanza non sufficiente; il restante 84,234% l'ha ritenuto sufficientemente discriminante della scelta, con una netta maggioranza di risposte conferite al punteggio più alto della scala di importanza (32,432% per un totale di 72 individui) e per questo motivo la media risulta essere di valore così elevato, dell'ordine di 5,266.

L'analisi permette di asserire che l'autore rappresenta per più di 8 rispondenti su 10 un parametro sufficientemente significativo nella scelta di un libro, e per più di 3 rispondenti questo attributo risulta massimamente importante.

È stato richiesto ai rispondenti di valutare, in seguito, anche l'importanza della trasportabilità del libro, intesa come la tascabilità o comunque la sua taglia ridotta,

considerando in questo frangente che i libri più facilmente trasportabili sono quelli formati da un numero più ridotto di pagine. I dati sono riportati nella Tabella 4.29.

q8e - Assegna una valutazione che esprima l'importanza della trasportabilità per la Sua scelta di un libro in formato cartaceo

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	40	18.018%	18.018%
Minimamente importante	30	13.514%	31.532%
Poco importante	26	11.712%	43.243%
Sufficientemente importante	38	17.117%	60.360%
Abbastanza importante	40	18.018%	78.378%
Molto importante	28	12.613%	90.991%
Massimamente importante	20	9.009%	100.000%
Totale	222	100.000%	
Media	3.775		

Tabella 4.29: *Frequenze relative alle valutazioni sull'importanza della trasportabilità nella scelta di un libro in formato cartaceo.*

La valutazione media assegnata a questo parametro risulta inferiore alla sufficienza (3,775). La percentuale di risposte per ogni valutazione risulta pressoché simile: il dato modale è duplice e corrisponde a due valutazioni distanti tra loro nella scala, precisamente alla valutazione minima della scala (“per nulla importante”) ed alla valutazione “5” (“abbastanza importante”). Il 43,243% dei rispondenti hanno affermato di non ritenere sufficientemente importante la trasportabilità nella scelta del libro, il restante 56,757% invece dichiara di prestare sufficientemente attenzione alla trasportabilità nel momento della scelta.

L'andamento non è chiaramente definibile in quanto il campione considerato si divide quasi a metà, circa 5 persone su 10 affermano di far caso alla trasportabilità nella scelta di un libro, le altre circa 5 persone affermano di non curarsene.

Un altro parametro da valutare consisteva nelle recensioni del libro, intendendole come consigli sia da parte di esperti, persone conosciute, famose o membri della famiglia, sia disponibili in rete. I risultati sono presentati nella Tabella 4.30.

q8f - Assegna una valutazione che esprima l'importanza delle recensioni per la Sua scelta di un libro in formato cartaceo

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	14	6.306%	6.306%
Minimamente importante	21	9.459%	15.766%
Poco importante	27	12.162%	27.928%
Sufficientemente importante	45	20.270%	48.198%
Abbastanza importante	50	22.523%	70.721%
Molto importante	39	17.568%	88.288%
Massimamente importante	26	11.712%	100.000%
Totale	222	100.000%	
Media	4.428		

Tabella 4.30: Frequenze relative alle valutazioni sull'importanza delle recensioni nella scelta di un libro in formato cartaceo.

I rispondenti che hanno assegnato la non sufficienza a questo parametro costituiscono la minoranza, il 27,928%. Il dato modale si attesta in corrispondenza della valutazione “5”, “abbastanza importante”, la quale contava 50 risposte (22,523% del campione considerato). Il 72,072% degli intervistati hanno dunque considerato sufficientemente importante il genere narrativo quando si trovano a dover scegliere un libro da leggere. La media delle risposte si attesta ad un valore più che sufficiente in termini di livello di importanza, confermando l'andamento positivo della suddetta variabile.

La tendenza che si evince dai dati è dunque la seguente: più di 7 persone su 10 affermano di considerare discriminante le recensioni o le argomentazioni altrui circa il libro cartaceo che andranno a scegliere, le quali determineranno in più di 7 casi su 10 la decisione di acquisto dello stesso.

La seguente variabile alla quale è stato richiesto di assegnare una valutazione consiste nell'importanza del genere narrativo del libro in un'ottica di acquisto dello stesso. I risultati sono presentati nella Tabella 4.31.

q8g - Assegna una valutazione che esprima l'importanza del genere narrativo per la Sua scelta di un libro in formato cartaceo

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	1	0.450%	0.450%
Minimamente importante	7	3.153%	3.604%
Poco importante	6	2.703%	6.306%

Sufficientemente importante	13	5.856%	12.162%
Abbastanza importante	35	15.766%	27.928%
Molto importante	59	26.577%	54.505%
Massimamente importante	101	45.495%	100.000%
Totale	222	100.000%	
Media	5.950		

Tabella 4.31: *Frequenze relative alle valutazioni sull'importanza del genere narrativo nella scelta di un libro in formato cartaceo.*

La tendenza risulta estremamente positiva: solo il 6,306% del campione di riferimento ha asserito la non sufficienza del parametro a condizionare la scelta di un libro, mentre ben il 93,694% degli intervistati hanno invece dichiarato di porre sufficiente attenzione al genere narrativo del libro in questione quando si trovano a dover scegliere se acquistarlo o no. Il dato modale si presenta, con ben 101 risposte su 222 (per un totale di quasi la metà delle risposte, il 45,495%), in corrispondenza della più elevata valutazione della scala di importanza, etichettata con l'espressione "massimamente importante". La media risulta in effetti molto elevata, attestandosi attorno al valore 6 e confermando la tendenza oltremodo positiva a considerare il genere narrativo come discriminante nella scelta del libro.

Risulta dunque dai dati che più di 9 persone su 10 considerano sufficientemente discriminante il genere narrativo del libro cartaceo quando si tratta di scegliere se acquistarlo oppure no; in particolare, 4 su 10 lo ritengono massimamente importante nella scelta, ponendolo sopra ad ogni altro parametro ed assegnandogli dunque il massimo dell'importanza (valutazione "7").

L'ultima variabile della quale è stato chiesto di valutare l'importanza consiste nella novità del libro, ovvero se il fatto di essere stato pubblicato meno di un anno fa risulta discriminante o meno nella scelta di acquistarlo. I risultati relativi a questo parametro sono riportati nella Tabella 4.32.

q8h - Assegna una valutazione che esprima l'importanza della novità per la Sua scelta di un libro in formato cartaceo

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	33	14.865%	14.865%
Minimamente importante	29	13.063%	27.928%

Poco importante	31	13.964%	41.892%
Sufficientemente importante	54	24.324%	66.216%
Abbastanza importante	44	19.820%	86.036%
Molto importante	19	8.559%	94.595%
Massimamente importante	12	5.405%	100.000%
Totale	222	100.000%	
Media	3.685		

Tabella 4.32: Frequenze relative alle valutazioni sull'importanza della novità nella scelta di un libro in formato cartaceo.

Il 41,892% dei rispondenti ha dichiarato di non ritenere il parametro della novità sufficientemente importante nella scelta di un libro cartaceo da leggere, mentre il 58,108% lo ritiene un parametro sufficientemente interessante. Il dato modale si attesta in corrispondenza della valutazione “4”, appena sufficiente, con un totale di 54 risposte su 222 (il 24,324%). Le due valutazioni più alte della scala, “molto importante” e “massimamente importante”, registrano il minor numero di risposte, precisamente 19 per quanto riguarda il “molto importante” e 12 per quanto riguarda il “massimamente importante”, per un totale complessivo del 13,964% delle risposte. La media delle valutazioni assegnate si attesta al valore 3,685 a prova del fatto che il rispondente medio considera la novità del libro non sufficiente per la decisione di acquistarlo oppure no.

La tendenza evidenziata dai dati mette in luce una tendenza mista: circa 4 persone su 10 affermano di non ritenere la novità un parametro sufficientemente discriminante nella scelta, mentre quasi 6 persone su 10 affermano di affidarsi sufficientemente nella scelta del libro cartaceo da leggere. Tuttavia la maggioranza delle risposte registrate si attesta attorno alla valutazione appena sufficiente.

Si può dunque dedurre che il fatto che il libro sia stato pubblicato meno di un anno fa, nella maggior parte dei casi, incide in maniera appena sufficiente nella scelta di acquisto. Per riuscire a stilare una classifica di importanza dei parametri di scelta di un libro cartaceo appena descritti si è operato un confronto diretto tra le medie delle variabili componenti il quesito q8.

I parametri che hanno collezionato più consensi positivi in termini di importanza sono riportati nella parte superiore del grafico rappresentato in Figura 4.3, a scendere quelli considerati progressivamente meno importanti.

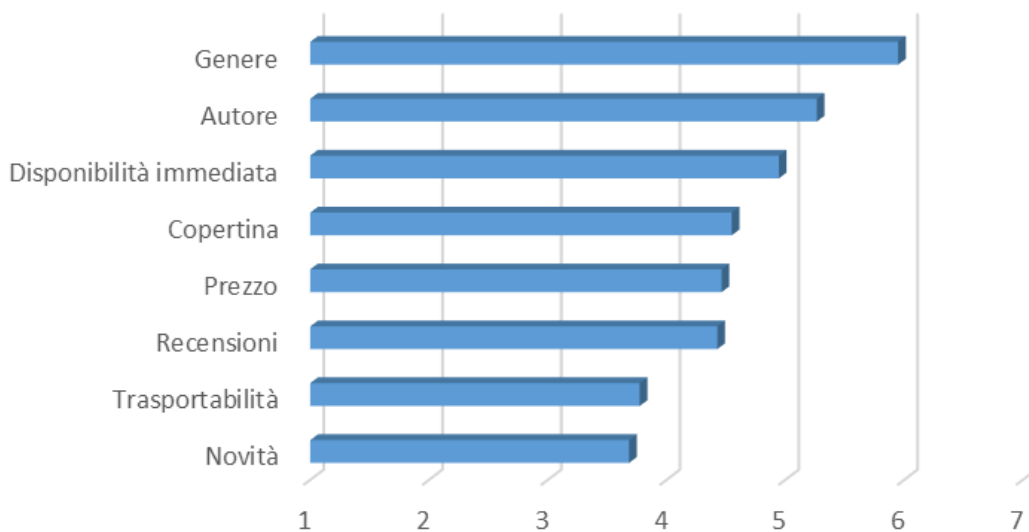


Figura 4.3: Confronto fra le medie delle variabili componenti il quesito q8.

Le variabili riferite al libro cartaceo che hanno collezionato una media di risposte che fosse almeno sufficiente sono state le seguenti, riportate in ordine da quella valutata più importante a quella valutata meno importante:

- Genere narrativo;
- Autore;
- Disponibilità immediata;
- Copertina;
- Prezzo;
- Recensioni.

I parametri che non hanno mediamente raggiunto la sufficienza in termini di influenza nella scelta di acquisto sono stati i seguenti, presentati a partire da quello considerato meno importante:

- Novità;
- Trasportabilità.

Analizzando i risultati ottenuti nel dettaglio si è evinto che i primi 2 parametri classificati a livello di importanza hanno a che vedere con le caratteristiche intrinseche del libro, in quanto consistono nella tipologia e nel tono di narrazione nonché nei temi toccati (genere narrativo) e nello scrittore. Queste due variabili concorrono in maniera molto importante alla scelta positiva di acquisto di un libro da parte del cliente. Le preferenze di genere,

utili alla decisione da parte della casa editrice dei testi da seguire nella pubblicazione, sono state descritte in precedenza; predisponendo un'offerta mirata che tenga conto delle preferenze dei consumatori, l'editore può beneficiare di ritorni economici maggiori e ad una maggiore fidelizzazione dei clienti che acquistano i libri da esso pubblicati.

4.2.2 Quesito q9: Quali sono i principali motivi per i quali preferisce leggere in formato cartaceo?

Il quesito in questione mirava a comprendere quali fossero le ragioni per le quali coloro che avevano selezionato come preferenza il libro di tipo cartaceo lo favorissero rispetto al formato digitale.

La domanda permetteva, attraverso la modalità di risposta multipla, di selezionare da una a tre risposte, per non costringere i rispondenti a dover scegliere tra due o più risposte ugualmente significative. I risultati sono riportati nella Tabella 4.33.

q9 - Quali sono i principali motivi per i quali preferisce leggere in formato cartaceo? (Massimo 3 alternative)			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Collezionismo	119	20.988%	20.988%
Arredamento (il libro può anche costituire un oggetto decorativo)	58	10.229%	31.217%
Lettura rilassata (senza passare per uno schermo)	154	27.160%	58.377%
Attaccamento emotivo alla fisicità del libro	182	32.099%	90.476%
Abitudine	54	9.524%	100.000%
Totale	567	100.000%	

Tabella 4.33: Frequenze relative alle motivazioni legate alla preferenza del formato cartaceo.

Le tre alternative che hanno registrato maggior riscontro da parte degli intervistati, in ordine dalla più selezionata, sono state:

- Attaccamento emotivo alla fisicità del libro;
- Lettura rilassata (senza passare per uno schermo);
- Collezionismo.

Il 32,099% del campione ha considerato molto importante nella scelta del formato cartaceo l'attaccamento emotivo alla fisicità del libro, intendendo con questo le sensazioni suscitate dallo strisciare delle pagine, il profumo del libro ed i ricordi che esso rievoca. Il 27,160% del campione ha compreso nella scelta delle motivazioni più importanti la lettura rilassata che il formato cartaceo permette, cioè la possibilità di non appesantire la vista di fronte ad uno schermo. Il collezionismo è stato compreso nella scelta nel 20,988% dei casi, confermandosi al terzo posto nella classifica delle risposte preferite.

Le tre alternative più scelte dimostrano che la lettura rappresenta per la maggior parte dei lettori di formato cartaceo un'esperienza riposante, un momento di rilassamento e di svago in cui riservarsi il piacere di godersi il libro di carta stampata nella sua consistenza, forma e colore, assaporandone con comodo il profumo e la materialità. Il piacere di godersi la fisicità del libro sfocia in molti casi nel collezionismo, inteso come raccolta fisica di volumi cartacei con caratteristiche piacevoli, raccolti in scaffali o mensole all'interno dell'ambiente di casa.

4.2.3 Quesito q10: Quali caratteristiche dei libri elettronici potrebbero convincerLa a cominciare ad utilizzare il formato digitale per leggere?

Ai lettori di volumi in formato cartaceo è stata riservata anche una domanda che indagasse le qualità dei libri elettronici da essi percepite, per essere in grado di trarre delle leve guida nel caso in cui si volesse pianificare un'offerta di tipo elettronico che risponda alle predilezioni dei lettori tradizionali, evidenziando gli aspetti che maggiormente stanno loro a cuore ed i vantaggi pratici che il formato digitale riserva. I risultati delle risposte sono riportati nella Tabella 4.34.

q10 - Quali tra le seguenti caratteristiche dei libri elettronici potrebbero convincerLa a cominciare ad utilizzare il formato digitale per leggere? (Massimo 3 alternative)

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Disponibilità a prezzo ridotto	116	22.437%	22.437%
Comodità di trasporto	121	23.404%	45.841%
Disponibilità di contenuti digitali integrativi	15	2.901%	48.743%
Risparmio di carta (questione ambientale)	94	18.182%	66.925%

Possibilità di leggere attraverso i dispositivi mobili personali	28	5.416%	72.340%
Possibilità di portare con sé più libri alla volta	121	23.404%	95.745%
Nessuna di queste	22	4.255%	100.000%
Totale	517	100.000%	

Tabella 4.34: *Frequenze relative alle caratteristiche considerate interessanti del libro in formato digitale per i lettori di formato cartaceo.*

Solamente il 4,255% dei rispondenti ha escluso del tutto la possibilità di dedicarsi alla lettura di tipo digitale, indice di una duplice possibilità:

- il restante 95,745% del campione intervistato legge già in quel formato, nonostante preferisca il cartaceo;
- il restante 95,745% del campione non si è ancora dedicato al tipo di lettura digitale, ma non esclude di dedicarglisi a determinate condizioni, espresse dalle altre alternative del quesito.

In particolare, le peculiarità del formato digitale che si sono rivelate più apprezzate dal pubblico di rispondenti sono state le seguenti, in ordine dalle più apprezzate:

- Comodità di trasporto e possibilità di trasportare più libri alla volta (23,404%);
- Disponibilità a prezzo ridotto (22,437%);
- Risparmio di carta, questione ambientale (18,182%).

Le peculiarità che sono state segnalate con maggior frequenza per quanto riguarda i vantaggi della lettura digitale secondo i lettori cartacei sono legate alla praticità del formato, ovvero alla possibilità di trasportare agilmente i libri servendosi degli apparecchi elettronici quali e-reader o tablet, nonché la possibilità di usufruire della loro memoria interna per raccogliere più testi nello stesso dispositivo, ed avere dunque la possibilità di consultare in qualsiasi momento qualunque libro disponibile nella propria “libreria digitale”.

La seconda motivazione in termini di importanza è relativa alla disponibilità dell’acquisto dell’e-book a prezzo ridotto rispetto alla versione cartacea, a riprova del fatto che i lettori di libro cartaceo danno un livello di importanza quantomeno sufficiente al prezzo del

libro. Il prezzo è una delle leve più importanti sul quale si basa la maggior parte delle conversioni ottenute negli store online di libri.

Al terzo posto della classifica si posiziona la questione ambientale, ossia il risparmio di carta reso possibile dai contenuti in forma digitale.

4.3 Analisi univariate: quesiti riservati ai lettori di formato digitale

In questo paragrafo si presentano i risultati dell'analisi univariata delle risposte sulle abitudini e preferenze di lettura da parte del pubblico digitale.

4.3.1 Quesito q16: Assegna una valutazione che esprima l'importanza da Lei assegnata ai parametri relativi al libro in formato digitale riportati di seguito

Il quesito q16 è stato rivolto esclusivamente ai lettori di tipo digitale, 32 tra i 254 dichiaratisi lettori, indagando l'importanza assegnata ai parametri presentati, prerogativa dei libri in formato elettronico. La scala di valutazione, ancora una volta, è stata predisposta sotto forma di scala Likert a 7 parametri, codificati come in precedenza.

Ai fini dell'analisi, il quesito è stato suddiviso nei seguenti attributi:

- q16a: Prezzo
- q16b: Recensioni / Raccomandazioni
- q16c: Disponibilità a prezzo scontato
- q16d: Novità
- q16e: Autore
- q16f: Genere narrativo
- q16g: Copertina

La scala di valutazione è suddivisa in due parti: le valutazioni non sufficienti all'importanza del parametro sono le prime 3 della scala, mentre le valutazioni ritenute sufficienti all'interno della scala sono rappresentate dalle valutazioni dalla 4 alla 7. Gli intervalli di valutazione esprimono i seguenti significati:

- Se al parametro viene assegnata una valutazione inferiore alla sufficienza, il rispondente non presenta intenzione di acquisto;
- Se al parametro viene assegnata una valutazione superiore alla sufficienza, il rispondente presenta intenzione di acquisto.

Il primo parametro per il quale è stato richiesto di dare una valutazione in termini di importanza è rappresentato dal prezzo, il quale rappresenta una leva molto importante nella discriminazione o meno dell'acquisto. I risultati sono riportati nella Tabella 4.35.

q16a - Assegna una valutazione che esprima quanto è importante per Lei il prezzo nella scelta di un libro in formato digitale			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	3	9.375%	9.375%
Minimamente importante	3	9.375%	18.750%
Poco importante	3	9.375%	28.125%
Sufficientemente importante	4	12.500%	40.625%
Abbastanza importante	7	21.875%	62.500%
Molto importante	8	25.000%	87.500%
Massimamente importante	4	12.500%	100.000%
Totale	32	100.000%	
Media	4.531		

Tabella 4.35: *Frequenze relative alle valutazioni sull'importanza del prezzo nella scelta di un libro in formato digitale.*

Per quanto riguarda il prezzo, il 28,125% dei rispondenti al quesito hanno dichiarato di non farsi influenzare in maniera sufficientemente forte da questo parametro nella scelta, mentre il restante 71,875% ha espresso una valutazione di importanza sufficiente alla discriminazione di scelta. Le risposte più numerose si sono contate in corrispondenza delle valutazioni 5 e 6, rispettivamente “abbastanza importante” (7 risposte su 32, pari al 21,875%) e “molto importante” (8 risposte su 32, pari al 25%). La media delle valutazioni si attesta al valore di 4,531 il quale rispecchia in media una valutazione sufficiente al parametro prezzo in un'ottica di importanza nella scelta di acquisto.

Si può dunque affermare che il prezzo rappresenta un parametro sufficientemente discriminante nella scelta di acquisto di un libro per circa 7 persone su 10 tra coloro che preferiscono leggere in formato digitale. Le valutazioni più numerose si sono contate in corrispondenza delle espressioni “abbastanza” e “molto importante”, rendendo possibile la collocazione del prezzo nella parte alta della scala di importanza.

Il secondo parametro da valutare per ciò che concerne le preferenze dei lettori di libri in formato elettronico si riferisce alla presenza in rete di recensioni del libro in questione oppure di raccomandazioni da parte di esperti, persone care, conoscenti. I dati raccolti vengono riportati nella Tabella 4.36.

q16b - Assegna una valutazione che esprima quanto è importante per Lei le recensioni/ raccomandazioni nella scelta di un libro in formato digitale			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	2	6.250%	6.250%
Minimamente importante	0	0.000%	6.250%
Poco importante	4	12.500%	18.750%
Sufficientemente importante	5	15.625%	34.375%
Abbastanza importante	8	25.000%	59.375%
Molto importante	7	21.875%	81.250%
Massimamente importante	6	18.750%	100.000%
Totale	32	100.000%	
Media	4.938		

Tabella 4.36: *Frequenze relative alle valutazioni sull'importanza delle recensioni nella scelta di un libro in formato digitale.*

Anche in questo caso, ed anzi in maniera più netta, le valutazioni non sufficienti rappresentano la minoranza, il 18,750% contro l'81,25% delle valutazioni positive. La maggioranza di risposte si colloca anche in questo caso nella parte alta della scala di valutazione, con una predilezione per la valutazione numero 5, "abbastanza importante" che conta 8 risposte su 32 per un totale del 25% delle preferenze. La seconda valutazione per numerosità di risposte è la numero 6, "molto importante" con 7 risposte su 32 (21,975%). Segue la valutazione numero 7, la più alta di tutta la scala, codificata con l'espressione "massimamente importante" che conta 6 risposte su 32, il 18,750% del totale. La media raggiunge il valore di 4,938 che denota una importanza media attribuita al campione approssimativamente pari alla valutazione "abbastanza importante", di un punto più in alto rispetto alla mera sufficienza.

Le recensioni e le raccomandazioni del libro sono considerate abbastanza importanti da circa 8 persone su 10 tra coloro che leggono abitualmente in formato digitale. Ne consegue che la tendenza si conferma positiva e che recensioni e raccomandazioni si

rivelano nella maggior parte dei casi delle variabili più che sufficientemente importanti nella scelta di acquisto.

Il parametro in questione riguarda la disponibilità del libro a prezzo scontato, pratica molto comunemente utilizzata nell'ambito degli store online, per via del risparmio legato ai costi di trasporto ma anche a stampa, rilegatura e, in generale, alle attività di predisposizione del libro in formato cartaceo. I risultati vengono presentati nella Tabella 4.37.

q16c - Assegna una valutazione che esprima quanto è importante per Lei la disponibilità a prezzo scontato nella scelta di un libro in formato digitale			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	2	6.250%	6.250%
Minimamente importante	2	6.250%	12.500%
Poco importante	4	12.500%	25.000%
Sufficientemente importante	4	12.500%	37.500%
Abbastanza importante	5	15.625%	53.125%
Molto importante	10	31.250%	84.375%
Massimamente importante	5	15.625%	100.000%
Totale	32	100.000%	
Moda	10		
Media	4.813		

Tabella 4.37: Frequenze relative alle valutazioni sull'importanza della disponibilità a prezzo scontato nella scelta di un libro in formato digitale.

La disponibilità a prezzo scontato è “molto importante” per la maggior parte dei rispondenti (10 risposte su 32, per un totale del 31,250%) e, in generale, il 75% dei rispondenti lo considera un parametro sufficientemente importante da influenzare la scelta di acquisto o meno del libro. Il 62,5% dei rispondenti ha assegnato a questa variabile una valutazione superiore alla semplice sufficienza, a prova del fatto che la possibilità di sconto del formato digitale rappresenta una condizione vincolante nella scelta di acquisto per più della metà dei rispondenti. Anche la media delle risposte conferma la tendenza appena descritta, attestandosi al valore di 4,813.

La disponibilità a prezzo scontato del libro in formato digitale rappresenta dunque, per più di 7 lettori digitali su 10, una leva discriminante nella decisione di acquisto. La

maggior parte di essi lo considera un parametro molto importante su cui fare affidamento per decidere se effettuare o meno l'acquisto.

La novità del libro rappresenta un altro parametro per il quale è stata richiesta una valutazione da parte dei lettori digitali. Il concetto di “nuovo”, in questa sede, dipende dalla data di pubblicazione del libro in questione: si ritiene “nuovo” un libro che sia stato pubblicato meno di un anno fa, rispetto alla data di pubblicazione del questionario (29 maggio 2018). I risultati sono riportati nella Tabella 4.38.

q16d - Assegna una valutazione che esprima quanto è importante per Lei la novità nella scelta di un libro in formato digitale

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	8	25.000%	25.000%
Minimamente importante	4	12.500%	37.500%
Poco importante	1	3.125%	40.625%
Sufficientemente importante	4	12.500%	53.125%
Abbastanza importante	3	9.375%	62.500%
Molto importante	7	21.875%	84.375%
Massimamente importante	5	15.625%	100.000%
Totale	32	100.000%	
Media	3.969		

Tabella 4.38: *Frequenze relative alle valutazioni sull'importanza della novità nella scelta di un libro in formato digitale.*

Il dato modale è stato registrato in corrispondenza della valutazione minima della scala, ossia l'espressione “per nulla importante” corrispondente alla valutazione 1 (8 risposte su 32, pari al 25%). La seconda valutazione in ordine di importanza assegnata è la valutazione pari a 6, codificata con l'espressione “molto importante”, la quale ha collezionato 7 risposte su 32, rispecchiando la preferenza del 21,875% dei lettori di formato elettronico. I rispondenti al quesito che hanno considerato la novità un parametro non sufficiente ad incidere sulla scelta o meno del libro corrispondono al 40,625% del campione di lettori digitali, mentre il 59,375% del suddetto campione ha considerato sufficientemente influente la novità nel momento della scelta. La media dei voti è approssimativamente sufficiente, anche se la tendenza non è di facile interpretazione per via della dispersione delle valutazioni assegnate.

In generale si può dedurre che la novità risulta almeno tanto importante da influenzare l'acquisto del libro per quasi 6 persone su 10. Questo significa che, anche se le valutazioni sono variamente distribuite lungo la scala dei valori, più della metà dei lettori digitali conferiscono un'importanza sufficiente da discriminare l'acquisto del libro al fattore novità, ovvero circa 6 persone su 10 preferiscono acquistare un libro nuovo piuttosto uno pubblicato più di un anno fa.

La variabile successiva consiste nell'autore dell'opera, e fornisce informazioni riguardanti l'importanza di esso nella scelta di un libro digitale. Con la variabile "autore" si intende dunque la considerazione dell'autore da parte del rispondente, in termini di riconoscimento, di apprezzamento, di nazionalità e qualunque altra caratteristica intrinseca dello scrittore che ne influenzi la conoscenza ed il prestigio agli occhi dell'intervistato. Le risposte a questo quesito sono riassunte nella Tabella 4.39.

q16e - Assegna una valutazione che esprima quanto è importante per Lei l'autore nella scelta di un libro in formato digitale			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	2	6.250%	6.250%
Minimamente importante	2	6.250%	12.500%
Poco importante	4	12.500%	25.000%
Sufficientemente importante	2	6.250%	31.250%
Abbastanza importante	3	9.375%	40.625%
Molto importante	9	28.125%	68.750%
Massimamente importante	10	31.250%	100.000%
Totale	32	100.000%	
Media	5.156		

Tabella 4.39: Frequenze relative alle valutazioni sull'importanza dell'autore nella scelta di un libro in formato digitale.

Il 25% dei lettori cartacei non considera sufficientemente importante la figura dell'autore nella scelta di un libro digitale, ovvero non ritiene che lo scrittore possa influenzare in maniera significativa la decisione di compiere l'acquisto del libro. Il restante 75% dei lettori di formato elettronico considerano invece sufficientemente importante nella scelta l'autore del libro; in particolare il 28,125% di essi lo considera "molto importante" assegnandogli una valutazione pari a 6 (9 risposte su 32 totali) ed il 31,250% lo considera

“massimamente importante”, assegnandogli la valutazione più elevata dell’intera scala e costituendo il dato modale (10 risposte su 32). La media del valore di 5,156 dimostra la tendenza positiva descritta delle valutazioni.

Dunque l’autore rappresenta un carattere sufficientemente importante per decidere di acquistare un libro digitale per circa 7 lettori digitali su 10; inoltre, quasi 3 persone su 10 lo considerano un parametro molto importante e più di 3 persone su 10 lo considerano un parametro che influisce nella decisione positiva di acquisto in maniera massimamente importante.

Si è in seguito indagato sulla rilevanza della tipologia di genere narrativo nella decisione di acquisto di un libro in formato elettronico. I dati sono riportati nella Tabella 4.40.

q16f - Assegna una valutazione che esprima quanto è importante per Lei il genere nella scelta di un libro in formato digitale

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	2	6.250%	6.250%
Minimamente importante	0	0.000%	6.250%
Poco importante	3	9.375%	15.625%
Sufficientemente importante	1	3.125%	18.750%
Abbastanza importante	3	9.375%	28.125%
Molto importante	10	31.250%	59.375%
Massimamente importante	13	40.625%	100.000%
Totale	32	100.000%	
Media	5.656		

Tabella 4.40: *Frequenze relative alle valutazioni sull’importanza del genere narrativo nella scelta di un libro in formato digitale.*

Il genere narrativo non è stato considerato sufficientemente importante dal 15,625% dei rispondenti, i quali rappresentano la netta minoranza, il rimanente 84,375% dei rispondenti ha dichiarato invece di considerare il genere narrativo discriminante nella scelta del libro elettronico da leggere. Il dato modale si è registrato in corrispondenza della valutazione più alta della scala; il genere narrativo del libro è stato considerato “massimamente importante” da 13 rispondenti su 32, dunque dal 40,625% dei lettori digitali. Una numerosità di risposte molto alta è stata registrata anche dalla valutazione 6 della scala, corrispondente all’espressione “molto importante”, che ha contato il 31,250%

delle preferenze dei lettori digitali (per un totale di 10 risposte su 32). Il valore medio delle risposte si attesta al valore 5,656, a dimostrazione dell'attenzione posta dal campione in questione a questo parametro.

In particolare si è evinto che più di 8 persone su 10 considerano il genere narrativo un parametro vincolante nella scelta di comprare un libro, e circa 4 persone su 10 considerano il genere narrativo massimamente influente nel momento in cui si trovano a dover scegliere di acquistare un libro in formato digitale.

Anche i libri in formato digitale sono dotati di una copertina, la quale può contenere illustrazioni o simboli grafici, ma può anche consistere nella mera esplicitazione del titolo e dell'autore dell'opera. Di seguito, nella Tabella 4.41, si riportano i risultati riguardanti le valutazioni conferite all'importanza della copertina di un libro digitale, facendo riferimento alle immagini e, in generale, alle copertine dotate di una qualità grafica apprezzabile.

q16g - Assegna una valutazione che esprima quanto è importante per Lei la copertina nella scelta di un libro in formato digitale

	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Per nulla importante	11	34.375%	34.375%
Minimamente importante	4	12.500%	46.875%
Poco importante	5	15.625%	62.500%
Sufficientemente importante	4	12.500%	75.000%
Abbastanza importante	4	12.500%	87.500%
Molto importante	2	6.250%	93.750%
Massimamente importante	2	6.250%	100.000%
Totale	32	100.000%	
Media	3.000		

Tabella 4.41: Frequenze relative alle valutazioni sull'importanza della copertina nella scelta di un libro in formato digitale.

La copertina non è stata considerata sufficientemente importante dalla maggioranza dei rispondenti, precisamente dal 62,5% dei lettori digitali. Il dato rappresentante la moda delle preferenze si è registrato in corrispondenza della valutazione più bassa della scala, "per nulla importante" con una percentuale di rispondenti pari a 34,375% (11 risposte su 32). Mano a mano che si procede verso la parte alta della scala di valutazione, le

preferenze si riducono di numero; le frequenze si distribuiscono dunque in maniera decrescente partendo dalle valutazioni più basse della scala. La media delle valutazioni denota una condizione di non sufficienza, presentandosi con il valore 3.

Si evince dai dati che più di 6 persone su 10 non hanno considerato sufficientemente importante la copertina come parametro di scelta di un libro digitale; dunque la decisione di acquisto di un libro digitale non dipende dalla sua copertina nella maggior parte dei casi (6 casi su 10).

Confrontando le medie dei singoli attributi componenti il quesito q16 è possibile comporre un ranking delle scelte più frequentemente segnalate da parte del campione dei lettori digitali sulle caratteristiche del libro digitale più apprezzate e ritenute più incidenti sulla scelta d'acquisto. Il confronto tra le medie è rappresentato attraverso il grafico a barre proposto nella Figura 4.4.



Figura 4.4: Confronto tra le medie delle singole variabili componenti il quesito q16.

Le caratteristiche alle quali i rispondenti hanno dimostrato di prestare più attenzione ed alle quali hanno assegnato valutazioni superiori alla sufficienza, sono state, in ordine dalla più importante alla meno importante:

- Genere narrativo;
- Autore;
- Recensioni / Raccomandazioni;
- Disponibilità a prezzo scontato;

- Prezzo.

I due parametri che invece non hanno ottenuto la sufficienza sono stati:

- Novità;
- Copertina.

I lettori dimostrano dunque che la scelta di acquistare un libro digitale viene effettuata in prevalenza in base al genere narrativo, all'autore ed alle recensioni o raccomandazioni. I primi due parametri classificati sono gli stessi che sono stati segnalati anche nel caso dei lettori cartacei. Il genere narrativo e l'autore rappresentano dunque le leve principali nella scelta di un libro da leggere sia per quanto riguarda i lettori cartacei sia per quanto riguarda i lettori digitali; questi ultimi hanno dimostrato invece di dare più peso alle recensioni online ed alle raccomandazioni, tendenza probabilmente legata alla loro familiarità con la rete.

4.3.2 *Quesito q17: Quali sono i principali motivi per i quali Lei preferisce leggere in formato digitale? (Massimo 3 alternative)*

Ai lettori di formato digitale è stata infine rivolta un'ulteriore domanda, che esplorava quali fossero le motivazioni per le quali essi preferissero leggere in formato digitale. Il quesito è stato strutturato in modo che ai rispondenti fosse permesso di esprimere fino a 3 alternative, in modo da lasciare loro più libertà di scelta, senza costringerli ad escludere qualche alternativa ugualmente importante. I risultati sono riportati nella Tabella 4.42.

q17 - Quali sono i principali motivi per i quali preferisce leggere in formato digitale? (Massimo 3 alternative)			
	Frequenza assoluta	Frequenza relativa (%)	Frequenza cumulata (%)
Prezzo ridotto	14	15.385%	15.385%
Comodità di trasporto	23	25.275%	40.659%
Disponibilità di contenuti digitali integrativi	1	1.099%	41.758%
Risparmio di carta (questione ambientale)	9	9.890%	51.648%
Possibilità di leggere sui dispositivi mobili personali	10	10.989%	62.637%
Possibilità di portare con sè più libri alla volta	23	25.275%	87.912%
Possibilità di accedere al libro dovunque ci si trovi	9	9.890%	97.802%
Abitudine	2	2.198%	100.000%
Totale	91	100.000%	

Tabella 4.42: *Frequenze relative alle motivazioni dietro la preferenza per il formato digitale.*

Le risposte hanno mostrato che le tre alternative maggiormente segnalate, in ordine dalla più importante alla meno importante, sono state:

- Possibilità di portare con sé più libri alla volta e comodità di trasporto;
- Prezzo ridotto;
- Possibilità di leggere attraverso i dispositivi mobili personali.

Il 25,275% del campione ha scelto come motivazione la comodità di trasporto e la possibilità di portare con sé più libri alla volta, due caratteristiche che identificano il formato digitale in quanto il libro può essere trasportato all'interno di appositi dispositivi della dimensione di uno smartphone, e viene salvato in un archivio che può contenere numerosi altri libri, rendendo dunque possibile il trasporto congiunto di più letture nello stesso dispositivo. Il 15,385% delle risposte esaltava il prezzo ridotto del libro digitale rispetto al formato cartaceo, possibile poiché la pubblicazione digitale non necessita di alcuni passaggi ai quali, invece, il formato cartaceo deve essere inevitabilmente sottoposto, permettendo di risparmiare in termini di costi e dunque di abbassare il prezzo di vendita. La terza scelta in ordine di numerosità delle risposte è stata la possibilità di leggere attraverso i propri dispositivi mobili personali, i quali essendo sempre a portata di mano in ogni momento della giornata, e sempre connessi, permettono la fruizione dei testi in ogni momento.

Il campione dei lettori digitali dimostra dunque di preferire il formato elettronico principalmente per la sua caratteristica di essere immateriale, la quale esprime i suoi effetti in termini di praticità e trasportabilità, sia perché i dispositivi dedicati contengono una libreria virtuale in cui collocare molti libri, alla quale si può accedere con facilità, anche attraverso i dispositivi mobili personali. Inoltre le risposte mettono in evidenza anche l'interesse dei lettori digitali per il prezzo ridotto, prerogativa esclusiva del libro in formato elettronico.

Capitolo 5

Canali informativi: analisi delle componenti principali ed analisi dei gruppi

Come già dichiarato in precedenza, si è ritenuto di approfondire l'analisi delle risposte al quesito q5 relativo alle valutazioni di una serie di parametri come approssimazione dei canali informativi più seguiti da parte del campione. Lo studio approfondito di questa variabile permette di individuare i canali informativi più trafficati per agire su di essi attraverso un piano di comunicazione ben strutturato e targettizzato. In seguito vengono presentate nel dettaglio l'*analisi delle componenti principali* e l'*analisi dei gruppi*, o *cluster analysis*. Per la realizzazione di queste ultime si è considerata la variabile nella sua interezza, esaminando i singoli parametri (da q5a a q5g) come caratteri singoli. Il campione considerato nell'analisi delle risposte alla domanda q5 corrisponde al campione dei lettori, che conta una numerosità di 254 unità statistiche.

5.1 Analisi delle componenti principali

L'analisi delle componenti principali rappresenta un metodo di sintesi dei dati multivariati applicabile a caratteri di tipo quantitativo che, attraverso una serie di tecniche statistiche, permette di ridurre la quantità di variabili che caratterizzano un fenomeno, riducendo dunque la dimensionalità dell'insieme dei dati iniziali. Il metodo consiste nell'individuazione di un certo numero di attributi latenti non correlati tra loro che siano combinazioni lineari delle variabili originarie e che riassumano in maniera coerente e concisa il fenomeno iniziale, riducendo al minimo la perdita di informazioni.

5.1.1 Matrice di varianze e covarianze e matrice di correlazione

La *matrice di varianze e covarianze* riportata in Tabella 5.1 e denominata S è una matrice quadrata e simmetrica, che contiene le varianze tra i vari caratteri nella diagonale e le covarianze tra i vari caratteri nelle restanti posizioni. La matrice S è calcolata considerando gli attributi da q5a a q5g, dopo aver operato la standardizzazione dei dati; questo passaggio fa sì che la *matrice di varianze e covarianze* del dataset selezionato corrisponda alla *matrice di correlazione*. La traccia della matrice S rappresenta la

variabilità complessiva dei dati la quale, essendo riferita ad un sistema di dati standard con medie uguali a 0 e varianze uguali ad 1, corrisponde alla numerosità delle variabili considerate, nel nostro caso 7.

	q5a	q5b	q5c	q5d	q5e	q5f	q5g
q5a	1	0,326	0,379	0,366	0,31	0,328	0,243
q5b	0,326	1	0,541	0,485	0,283	0,474	0,393
q5c	0,379	0,541	1	0,552	0,304	0,42	0,379
q5d	0,366	0,485	0,552	1	0,367	0,44	0,393
q5e	0,31	0,283	0,304	0,367	1	0,459	0,311
q5f	0,328	0,474	0,42	0,44	0,459	1	0,586
q5g	0,243	0,393	0,379	0,393	0,311	0,586	1

Tabella 5.1 Matrice di varianze e covarianze dei caratteri componenti il quesito q5.

La *matrice di correlazione* riporta quindi tutti i valori di correlazione tra le diverse variabili: in diagonale appare costantemente il valore 1, in quanto la correlazione tra una variabile e sé stessa è massima. Per accertare la sussistenza o meno di dipendenza lineare effettiva tra le diverse variabili è stata svolta una verifica di ipotesi. Considerando come riferimento due variabili generiche, x e y , esprimiamo con ρ_{xy} la correlazione tra la variabile x e la variabile y . Il sistema di ipotesi viene disposto come segue:

$$\begin{cases} H_0: \rho_{xy} = 0 \\ H_1: \rho_{xy} \neq 0 \end{cases}$$

in cui l'ipotesi nulla consiste nel fatto che la correlazione tra le due variabili sia nulla, ovvero che non vi sia dipendenza lineare tra le due variabili mentre l'ipotesi alternativa consiste nel fatto che esista una correlazione, positiva o negativa, tra le due variabili, esprimendo una dipendenza lineare tra di esse. Durante il test di ipotesi l'obiettivo da raggiungere è quello di rifiutare l'ipotesi nulla H_0 , poiché la condizione ideale che si ricerca è la dipendenza lineare tra le variabili considerate.

Come primo step è stato calcolato il valore critico dell'intervallo di accettazione di H_0 . Considerando un livello di errore $\alpha = 0,05$, ovvero con una confidenza del 95%, è stato calcolato il valore critico che rappresenta gli estremi dell'intervallo di accettazione:

$$z_{1,2} = \pm \frac{1,96}{\sqrt{n}}$$

dove con n si intende la numerosità campionaria, nel nostro caso pari a 254 rispondenti. Calcolando dunque i valori di z_1 e di z_2 si ottiene:

$$z_1 = -0,123$$

$$z_2 = 0,123$$

Gli intervalli di accettazione e rifiuto dell'ipotesi nulla sono, dunque, i seguenti:

$$\begin{cases} z_1 \leq \rho_{xy} \leq z_2 \rightarrow \text{Accetto } H_0 \\ \rho_{xy} \leq z_1, \rho_{xy} \geq z_2 \rightarrow \text{Rifiuto } H_0 \end{cases}$$

Osservando la Tabella 5.1 delle correlazioni si può notare come esse abbiano tutte segno positivo; ciò che vogliamo provare è che esista effettivamente una relazione di dipendenza lineare tra le variabili. Come si può notare dai valori in tabella, tutte le correlazioni entrano nell'intervallo di rifiuto dell'ipotesi nulla, essendo tutti valori maggiori di z_2 . Questo significa che tutti coefficienti di correlazione tra le diverse variabili sono statisticamente significativi, cioè tra loro sussiste una dipendenza lineare, anche se non possiamo affermare con certezza la presenza di una relazione di causalità tra di esse. Essendo tutte di segno positivo anche le covarianze avranno segno positivo; ciò significa che all'aumentare delle valutazioni che i rispondenti hanno assegnato ad una variabile aumentano anche le valutazioni che essi hanno assegnato all'altra variabile considerata e viceversa, al diminuire delle valutazioni assegnate all'una diminuiscono anche le valutazioni conferite all'altra.

5.1.2 Calcolo degli autovalori e degli autovettori

L'analisi delle componenti principali si basa sul calcolo di *autovalori* ed *autovettori*. Data la matrice S di dimensione $n \times n$, si definisce λ *autovalore* della matrice S il numero reale associato ad x , *autovettore* della matrice S , ovvero vettore non nullo $n \times 1$, se viene soddisfatta la seguente relazione:

$$Sx = \lambda x$$

dove S è la *matrice di varianze e covarianze* delle variabili componenti il quesito q5. Esistono tanti *autovalori* quante sono le variabili considerate; essendo che la domanda q5 è stata scomposta in 7 parametri, il numero degli *autovalori* sarà pari a 7. Ad ogni *autovalore* viene associato un *autovettore*, secondo la relazione:

$$n\lambda \Rightarrow nx$$

Autovettori associati ad *autovalori* distinti sono linearmente indipendenti. Infine, la somma degli *autovalori* è pari al numero delle variabili considerate, nel nostro caso 7, visto che i dati utilizzati per calcolare la *matrice di varianze e covarianze* sono stati standardizzati; questo significa che gli *autovalori* sono stati calcolati partendo dalla *matrice di correlazione*, ed essi sono stati ordinati in maniera decrescente.

Gli *autovalori*, calcolati attraverso il pacchetto `FactoMineR` del software R, sono riportati nella Tabella 5.2.

	Autovalori (λ)
Componente 1	3,409
Componente 2	0,853
Componente 3	0,796
Componente 4	0,631
Componente 5	0,512
Componente 6	0,429
Componente 7	0,37

Tabella 5.2 Autovalori relativi alla variabile *q5*.

5.1.3 Scelta del numero di componenti ottimale

Per il calcolo delle componenti principali è stato scelto il numero delle componenti da utilizzare in base a tre diversi criteri:

- **Criterio della variabilità cumulata**

Il *criterio della variabilità cumulata* per la determinazione del numero di componenti ottimo da utilizzare si basa sul calcolo della variabilità spiegata da parte di ciascuno degli autovalori, mentre il calcolo della variabilità cumulata indica percentuale di variabilità che è possibile spiegare attraverso utilizzo simultaneo di diversi *autovalori*, in maniera aggregata (prima si valuta la variabilità spiegata dal singolo primo autovalore, successivamente dal primo assieme al secondo, in seguito dal primo assieme al secondo ed al terzo, ecc.). Il livello di variabilità cumulata considerato sufficiente per asserire la significatività dell'analisi è pari al 60% (più si avvicina al 100%, più variabilità dei dati viene spiegata attraverso il modello). Osservando la varianza cumulata riportata in Tabella 5.3 nella terza colonna, si evince che l'utilizzo

dei primi due autovalori è in grado di spiegare il 60,894% della variabilità dei dati, valore appena superiore al minimo consentito per preservare la significatività del metodo.

	Autovalori	Varianza spiegata (%)	Varianza cumulata (%)
Componente 1	3,409	48,704	48,704
Componente 2	0,853	12,19	60,894
Componente 3	0,796	11,373	72,266
Componente 4	0,631	9,013	81,28
Componente 5	0,512	7,31	88,589
Componente 6	0,429	6,123	94,713
Componente 7	0,37	5,287	100

Tabella 5.3 Autovalori per la variabile q5 e relative varianze spiegate e cumulate.

Considerando i primi tre autovalori, i quali determinano a loro volta l'utilizzo di tre variabili latenti nell'*analisi delle componenti principali*, è possibile spiegare ben il 72,266% della variabilità dei dati. Il *criterio della variabilità cumulata* suggerisce dunque di considerare i primi tre autovalori, riportati nella Tabella 5.4.

λ_1	λ_2	λ_3
3,409	0,853	0,796

Tabella 5.4 Autovalori scelti per l'analisi delle componenti principali.

- **Criterio dell'autovalore medio o criterio di Kaiser**

Il *criterio dell'autovalore medio*, chiamato anche *criterio di Kaiser*, rappresenta un altro metodo per identificare il numero di componenti da utilizzare nel modello. Questo criterio consiste nel calcolo di un *autovalore medio*, denominato $\bar{\lambda}$, che rappresenta un *autovalore* di riferimento per verificare la significatività di ciascuno degli *autovalori* calcolati. Esso viene calcolato come media degli *autovalori*:

$$\bar{\lambda} = \frac{\sum_{i=1}^n \lambda_i}{n} = \frac{7}{7} = 1$$

L'*autovalore medio* assume sempre valore pari ad 1 quando gli autovalori sono stati calcolati prendendo come riferimento la matrice di correlazione.

Il *criterio di Kaiser* conta quanti tra gli autovalori calcolati sono di valore superiore all'*autovalore medio*; tutti gli *autovalori* che superano il $\bar{\lambda}$ saranno considerati utili ai fini dell'analisi delle componenti, decretando la quantità di variabili latenti da utilizzare per la sintesi dei dati. Sarà dunque sufficiente contare quanti *autovalori* superano il valore 1 (essendo essi ordinati in ordine decrescente, i λ maggiori di 1 saranno tra loro consecutivi).

Nel nostro caso appare solamente 1 *autovalore* maggiore rispetto all'*autovalore medio*. Questo significa che l'applicazione del *criterio di Kaiser* ci suggerisce di considerare una sola componente come approssimazione del modello di partenza.

- **Criterio dello Scree Plot**

Lo Scree Plot consiste in un criterio grafico che si concretizza nella creazione di una spezzata in cui vengono rappresentati gli *autovalori*. Il numero di componenti da utilizzare è suggerito dal “gomito” che si viene a creare graficamente nel punto della spezzata in cui il *gap* tra gli *autovalori* è massimo in quanto, fino a quel valore, i λ sono consistenti. Lo Scree Plot relativo alla variabile q5 è riportato nella Figura 5.1.

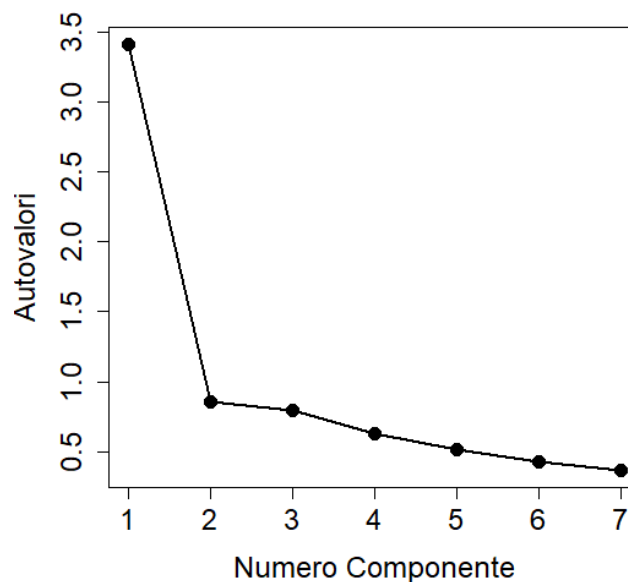


Figura 5.1 Scree Plot relativo alla variabile q5.

Alla luce della rappresentazione grafica degli *autovalori* proposta in figura, la spezzata forma un angolo evidente tra la prima e la terza componente, in corrispondenza della

seconda; due è il numero di componenti che il criterio dello Scree Plot ci suggerisce di adoperare nella creazione del nuovo modello.

Si è scelto di considerare i risultati del primo *criterio della variabilità cumulata*, dato che tiene conto con maggior precisione dell'entità di ciascun autovalore e del suo contributo alla spiegazione dei dati: il numero di componenti da utilizzare è dunque pari a 3.

5.1.4 Calcolo degli autovettori

Dopo aver calcolato gli *autovalori* si procede al calcolo dei rispettivi *autovettori*. La Tabella 5.5 riporta i *loadings* degli *autovettori* (considerati come colonne) che esprimono, confrontati elemento per elemento in ordine di posizione, quale variabile ha contribuito maggiormente nella creazione delle varie componenti. Le componenti considerate sono evidenziate di colore rosso in tabella.

	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6	Componente 7
q5a	-0,315	0,426	-0,575	0,619	-0,015	-0,056	0,047
q5b	-0,395	0,229	0,349	-0,099	-0,702	-0,298	0,280
q5c	-0,403	0,385	0,220	-0,184	0,107	0,758	-0,147
q5d	-0,406	0,245	0,084	-0,327	0,607	-0,539	-0,048
q5e	-0,327	-0,357	-0,632	-0,518	-0,129	0,132	0,252
q5f	-0,417	-0,430	0,037	0,162	-0,170	-0,098	-0,758
q5g	-0,370	-0,496	0,302	0,414	0,286	0,126	0,507

Tabella 5.5 *Loadings* relativi alla variabile q5.

Il passo successivo consiste nel confrontare i pesi, espressi da ciascun elemento di ogni autovettore, con l'elemento che occupa la stessa posizione negli altri *autovettori*, a partire dal primo elemento dei 3 *autovettori* considerati; esaminando gli elementi degli *autovettori* considerati, partendo dalla prima riga della tabella, si evidenzia in colore rosso l'elemento che presenta il valore maggiore, in valore assoluto, rispetto agli altri; si ripete l'operazione fino alla settima riga (che identifica la settima ed ultima variabile da considerare). I valori più elevati per ogni riga sono evidenziati in colore rosso nella Tabella 5.6.

	Componente 1	Componente 2	Componente 3
q5a	-0,315	0,426	-0,575
q5b	-0,395	0,229	0,349
q5c	-0,403	0,385	0,220
q5d	-0,406	0,245	0,084
q5e	-0,327	-0,357	-0,632
q5f	-0,417	-0,430	0,037
q5g	-0,370	-0,496	0,302

Tabella 5.6 Loadings di valore maggiore per ogni variabile considerata.

Dal confronto dei loadings si nota che nella prima colonna le variabili che hanno contribuito in modo maggiore alla determinazione della prima variabile latente sono i caratteri q5b, q5c, q5d che rappresentano rispettivamente i consigli della casa editrice (sul sito o sui social network), le segnalazioni su pagine o gruppi specializzati e community sui social ed infine i consigli di bibliotecari, insegnanti, esperti nel settore del libro. Valutando le caratteristiche comuni a tutte le variabili suddette, la prima componente esprime i “consigli di enti, gruppi e persone esperti nel settore del libro”. La seconda dimensione, riportata nella seconda colonna, è determinata dai caratteri q5f e q5g, che esprimono rispettivamente le interviste e la presenza dell’autore sui media e le classifiche su media e riviste; volendo identificare la seconda variabile latente, si è dunque scelto di nominarla “informazioni ottenute dai media tradizionali”. La terza ed ultima componente è stata determinata dai caratteri che valutano i consigli di amici e parenti e i consigli di persone famose nel mondo dello spettacolo. Tenendo conto che ai consigli delle persone famose sono state accostate valutazioni molto negative, si è concentrata l’attenzione sui consigli dei familiari o dei conoscenti: la terza variabile latente descrive, dunque, i “consigli di persone non specializzate nel settore del libro”, che spesso hanno un legame affettivo o di parentela con il rispondente. Il resoconto delle denominazioni delle varie componenti è riportato nella Tabella 5.7.

Numero componente	Denominazione delle componenti
Prima componente	Consigli di enti, gruppi e persone esperte nel settore del libro
Seconda componente	Informazioni ottenute dai media tradizionali
Terza componente	Consigli di persone non specializzate nel settore del libro

Tabella 5.7 Denominazione delle tre componenti considerate.

5.1.5 Mappa degli individui e cerchio delle correlazioni

Una volta determinate le componenti principali si sono calcolate le nuove coordinate. Le coordinate delle nuove componenti sono rappresentate come individui nella Figura 5.2 che rappresenta la mappa degli individui in cui i due assi che identificano la prima e la seconda componente.

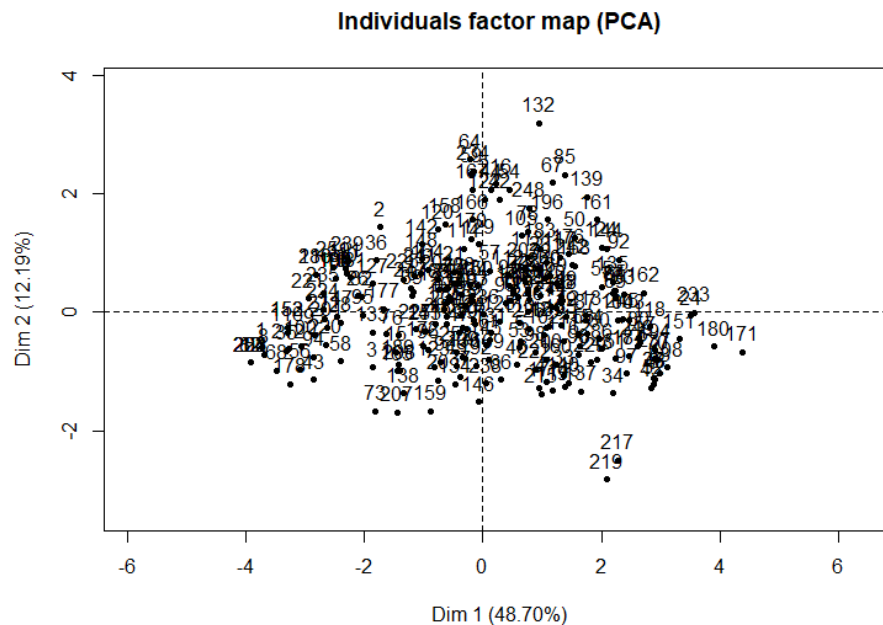


Figura 5.2 Mappa degli individui rispetto a prima e seconda componente

La mappa degli individui è centrata in 0, poiché i calcoli sono stati effettuati partendo dalla *matrice di correlazione*, dunque considerando i dati standardizzati; l'origine degli assi rappresenta dunque l'individuo medio. Si può osservare come tutti gli individui si distribuiscono in maniera abbastanza omogenea intorno all'individuo medio, anche se si registrano degli outliers significativi, in particolare nel primo e nel quarto quadrante (a titolo di esempio si segnalano le unità statistiche più significative, la numero 132, 180, 171, 217 e 219). Ogni quadrante raccoglie gli individui in base alle valutazioni di importanza:

- Il primo quadrante, quello in alto a destra, contiene tutti gli individui che hanno espresso valutazioni di importanza superiori alla media sia nel caso della dimensione 1 (“Consigli di enti, gruppi e persone esperti nel settore del libro”) sia nel caso della dimensione 2 (“Informazioni ottenute dai media tradizionali”);
- Il secondo quadrante, quello in alto a sinistra, rappresenta tutti gli individui che hanno espresso valutazioni di importanza sotto la media per quanto concerne la

prima dimensione (“Consigli di enti, gruppi e persone esperti nel settore del libro”) e hanno contemporaneamente assegnato valutazioni di importanza sopra la media per quanto riguarda la seconda dimensione (“Informazioni ottenute dai media tradizionali”);

- Il terzo quadrante, quello collocato in basso a sinistra, presenta gli individui che hanno espresso valutazioni di importanza sotto la media sia per quanto riguarda la prima dimensione (“Consigli di enti, gruppi e persone esperti nel settore del libro”), sia per quanto riguarda la seconda dimensione (“Informazioni ottenute dai media tradizionali”);
- Il quarto ed ultimo quadrante, quello posizionato in basso a destra, contiene invece gli individui che hanno espresso valutazioni di importanza sopra la media per quanto riguarda la dimensione 1 (“Consigli di enti, gruppi e persone esperti nel settore del libro”) mentre hanno espresso valutazioni sotto la media per quanto riguarda la dimensione 2 (“Informazioni ottenute dai media tradizionali”).

La mappa degli individui relativa alla seconda ed alla terza componente, rappresentata dalla Figura 5.3, presenta un numero più elevato di outliers rispetto ai valori assegnati dall’individuo medio, anche se gli individui sono caratterizzati da una minore dispersione; essi sono inoltre più omogeneamente distribuiti attorno all’origine, a differenza della mappa precedente, in cui gli individui si distribuivano in maniera più dispersa rispetto alla dimensione 1 ed in maniera più concentrata rispetto alla dimensione 2. Anche nella mappa degli individui in cui si considerano la prima e la terza componente, presentata in Figura 5.4, sono presenti degli outliers, concentrati soprattutto nel quarto quadrante.

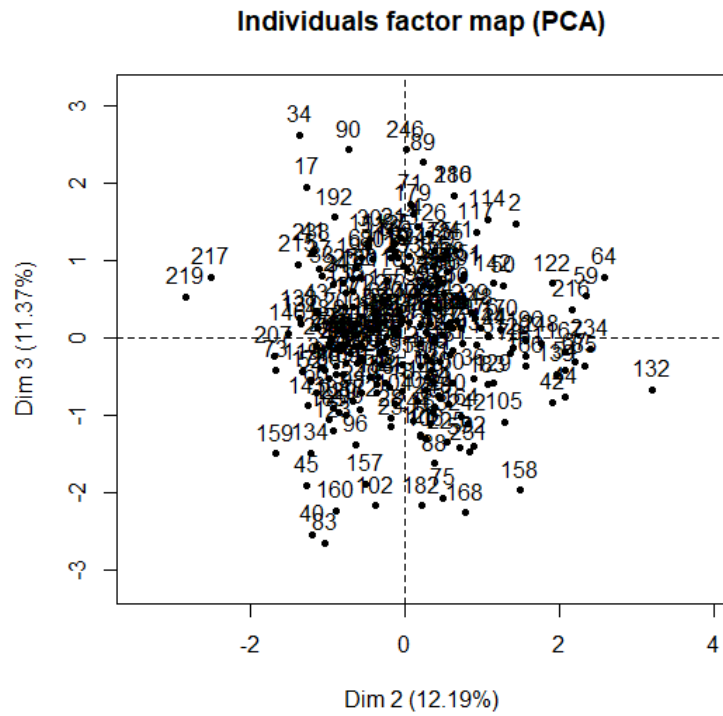


Figura 5.3 Mappa degli individui rispetto a seconda e terza componente.

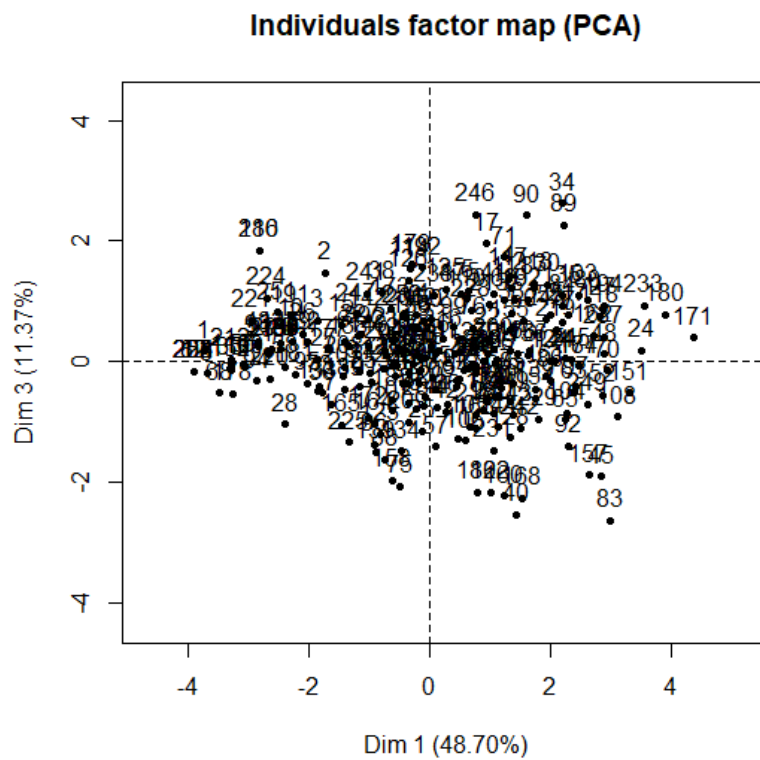


Figura 5.4 Mappa degli individui rispetto a prima e terza componente.

Il cerchio delle correlazioni consiste invece in una circonferenza di raggio unitario centrata sull'origine degli assi che contiene i vettori che rappresentano le correlazioni tra le variabili originarie e le nuove componenti. Gli assi rappresentano le due dimensioni considerate; per confrontare i dati in maniera significativa facciamo ora riferimento al cerchio costruito tenendo della prima e seconda componente. Vi sono 2 dettagli fondamentali da tenere in considerazione:

- La norma del vettore: tanto più la norma è elevata, tanto più la correlazione è forte, in quanto il vettore è tanto più vicino alla circonferenza, dunque la norma (e quindi la correlazione) è tanto più vicina ad 1;
- Gli angoli θ_1 e θ_2 formati dal vettore con i due assi, Y_1 e Y_2 , che rappresentano rispettivamente la prima e la seconda componente. L'angolo minore tra i due indica che la variabile originaria in questione ha contribuito maggiormente alla determinazione di quella componente, rappresentata da uno dei due assi. Se θ_1 è minore di θ_2 allora il vettore è più vicino all'asse Y_1 che rappresenta la prima componente; ciò significa che la variabile originaria in questione è stata più importante nella formazione della prima dimensione piuttosto che della seconda, dunque la correlazione con Y_1 è più elevata. Questo peso delle variabili originarie nella determinazione delle componenti si può osservare alla stessa maniera osservando i pesi degli autovettori.

Analizzando il cerchio delle correlazioni riportato nella Figura 5.5, che riguarda la prima e la seconda componente, si apprezzino le norme dei vettori che rappresentano le correlazioni tra le variabili originarie e le nuove componenti: la norma maggiore è rappresentata dal vettore che si riferisce alla variabile “consigli di amici e parenti” (q5a); esso si avvicina molto alla circonferenza goniometrica, di conseguenza la norma è apprezzabilmente vicina al valore 1, questo identifica una correlazione molto forte, stante a significare una elevata contribuzione da parte della variabile “consigli di amici e parenti” (q5a) alla determinazione della seconda componente. Il secondo elemento fondamentale consiste nel confronto tra gli angoli θ_1 e θ_2 . L'angolo formato con l'asse y che rappresenta la seconda componente, denominato θ_1 , è minore dell'angolo formato dal vettore con l'asse x, che rappresenta la prima componente, denominato θ_2 ; ciò significa che la variabile “consigli di amici e parenti” ha contribuito in modo più consistente alla determinazione della seconda componente.

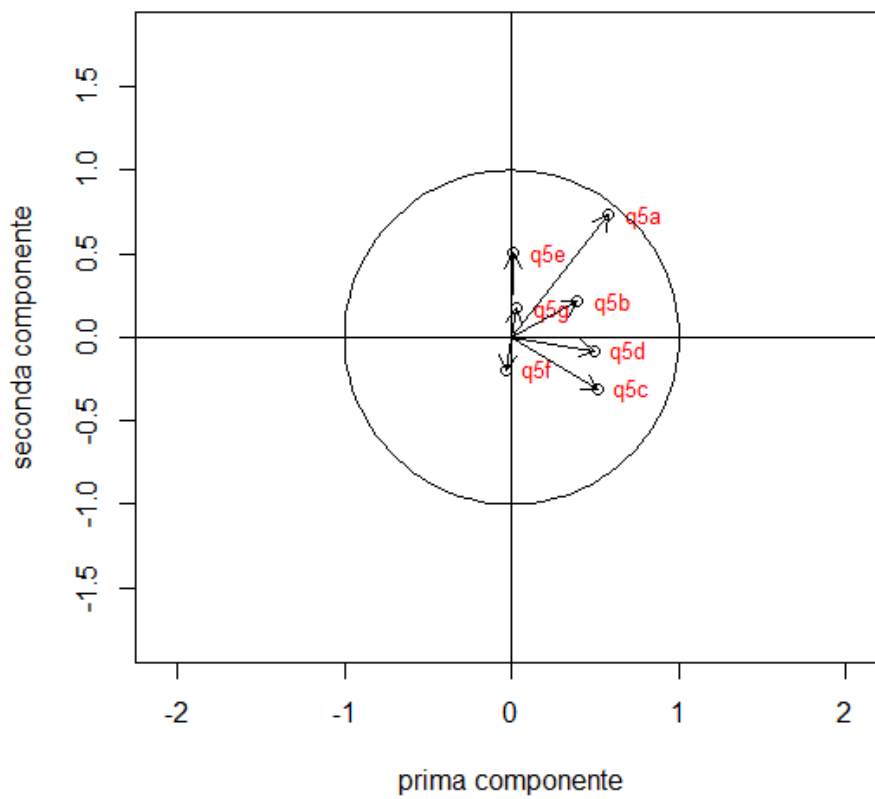


Figura 5.5 Cerchio delle correlazioni tra la prima e la seconda componente.

Alla stessa maniera si può ragionare sul contributo di ciascuna delle altre variabili, ottenendo le seguenti considerazioni:

- le variabili “consigli di persone famose tra cui attori, ballerini, cantanti, influencer” (q5e), “interviste e presenza dell'autore sui media tra cui TV, radio, Internet” (q5f) e “classifiche sui giornali/riviste/internet” (q5g) hanno contribuito maggiormente alla determinazione della prima componente: “consigli di enti, gruppi e persone esperti nel settore del libro”.
- le variabili “consigli dalla casa editrice sul sito o sui social network” (q5b), “segnalazioni su pagine o gruppi specializzati e community sui social” (q5c) e “consigli di bibliotecari/insegnanti/esperti nel settore del libro” (q5d) hanno contribuito maggiormente alla determinazione della seconda componente: “informazioni ottenute dai media tradizionali”.

Allo stesso modo si può ragionare sul contributo delle variabili nel caso della seconda e terza componente, il cui cerchio delle correlazioni è riportato nella Figura 5.6, in particolare:

- le variabili “segnalazioni su pagine o gruppi specializzati e community sui social” (q5c), “consigli di persone famose tra cui attori, ballerini, cantanti, influencer” (q5e) e “classifiche sui giornali/riviste/internet” (q5g) hanno concorso maggiormente alla determinazione della seconda componente: “informazioni ottenute dai media tradizionali”.
- le variabili “consigli di amici e parenti” (q5a), “consigli dalla casa editrice sul sito o sui social network” (q5b), “consigli di bibliotecari/insegnanti/esperti nel settore del libro” (q5d) e “interviste e presenza dell'autore sui media tra cui TV, radio, Internet” (q5f) hanno concorso maggiormente nella determinazione della terza componente: “consigli di persone non specializzate nel settore del libro”.

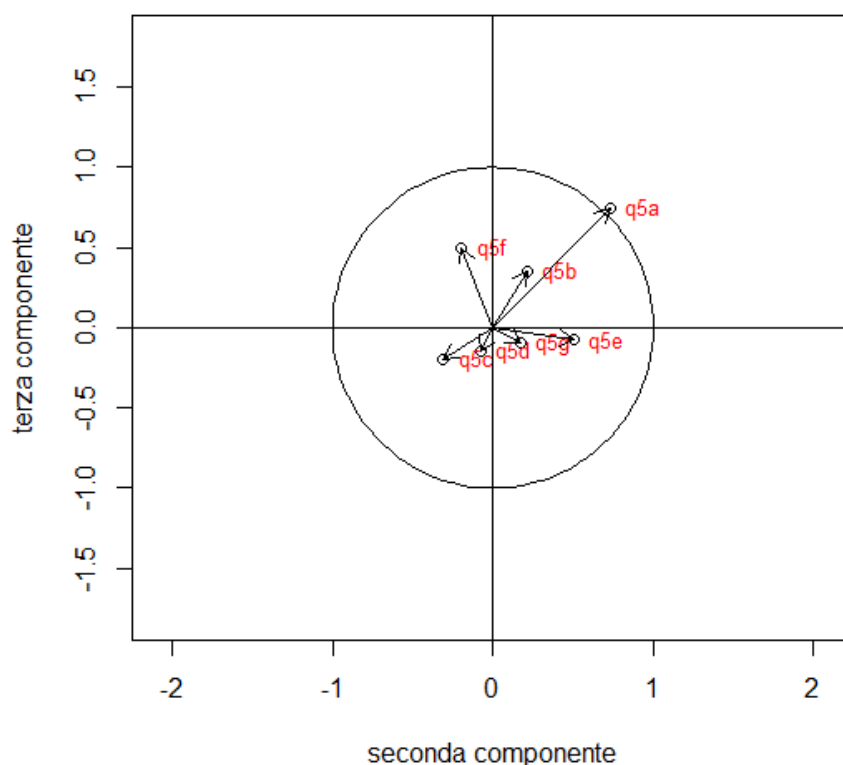


Figura 5.6 Cerchio delle correlazioni tra seconda e terza componente.

Analizzando il cerchio delle correlazioni relativo alla prima ed alla terza componente, riportato nella Figura 5.7, si evince che:

- le variabili “consigli dalla casa editrice sul sito o sui social network” (q5b), “segnalazioni su pagine o gruppi specializzati e community sui social” (q5c) e “consigli di bibliotecari/insegnanti/esperti nel settore del libro” (q5d) hanno contribuito maggiormente alla definizione della prima componente: “consigli di enti, gruppi e persone esperte nel settore del libro”.
- le variabili “consigli di amici e parenti” (q5a), “consigli di persone famose tra cui attori, ballerini, cantanti, influencer” (q5e), “interviste e presenza dell'autore sui media tra cui TV, radio, Internet” (q5f) e “classifiche sui giornali/riviste/internet” (q5g) hanno contribuito maggiormente alla definizione della terza componente: “consigli di persone non specializzate nel settore del libro”.

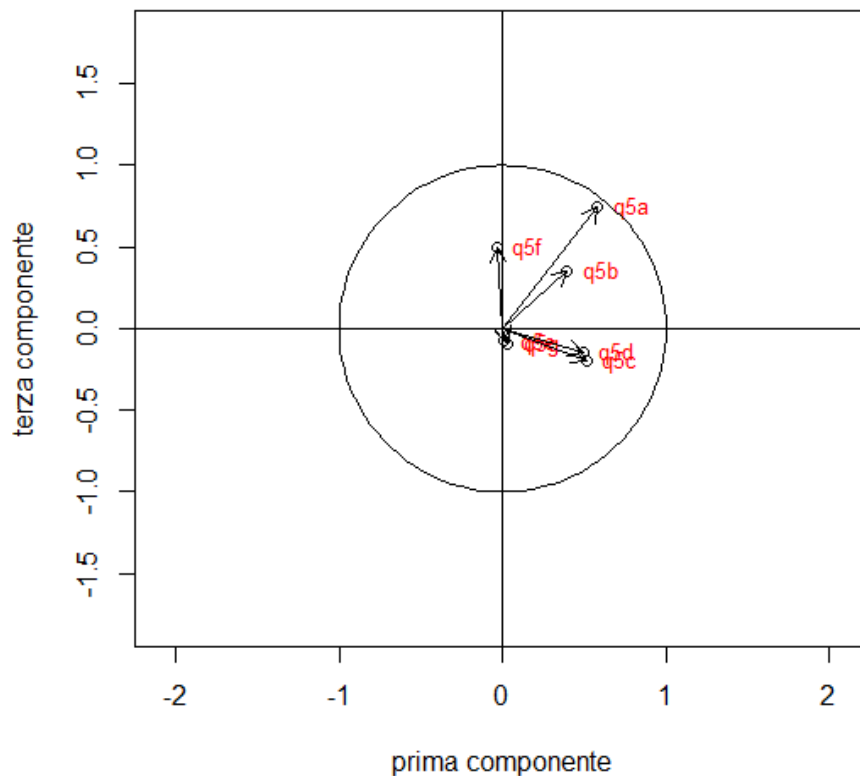


Figura 5.7 Cerchio delle correlazioni tra la prima e la terza componente.

5.2 Analisi dei gruppi

L'*analisi dei gruppi cluster*, detta anche *analisi dei cluster*, è lo strumento che permette di scoprire ed identificare le associazioni e la struttura all'interno dello sviluppo dei dati e delle varie tipologie di essi. L'*analisi dei cluster* fornisce informazioni sui dati dividendo il dataset in gruppi (*cluster*) di oggetti, in modo tale che gli elementi in un *cluster* siano più simili tra loro rispetto agli elementi di altri cluster. Al momento esistono molti diversi algoritmi di *clustering* che sono categorizzati in base al loro modello.

Per questo studio è stato scelto, in base alle prove effettuate ed ai risultati ottenuti, di utilizzare il *metodo del legame completo*, che rientra nel novero degli algoritmi gerarchici agglomerativi ovvero che, partendo da *cluster* formati da una sola unità statistica, riunisce gli elementi in *cluster* sempre più numerosi fino ad arrivare ad un insieme unico contenente tutti gli elementi. Il *metodo del legame completo (Complete-Linkage)* considera la maggiore delle distanze, calcolate tra tutte le coppie di elementi. Questo metodo permette di agglomerare le unità tra loro od inserirle nei cluster tenendo conto della distanza massima tra tutti gli elementi. A titolo di esempio, ipotizziamo di avere 4 elementi x, y, z, k : le distanze tra x e le altre unità vengono chiamate d_{xy}, d_{xz}, d_{xk} . Presumendo che le unità x e y vengano agglomerate in un unico gruppo, la distanza tra l'insieme (x, y) e l'elemento z è definita come:

$$d[(x, y)z] = \max(d_{xz}, d_{yz})$$

e la distanza tra il gruppo (x, y) ed il gruppo (z, k) è definita nel seguente modo:

$$d[(x, y), (z, k)] = \max(d_{xz}, d_{xk}, d_{yz}, d_{yk})$$

È stato scelto di utilizzare il *legame completo* perché permette di concentrare l'attenzione sulle dissimilarità tra le unità statistiche, in base alle quali vengono formati i *cluster*, oltre al fatto che è stato segnalato come algoritmo consigliato dalla funzione "optimal scores" messa a disposizione dal software R, la quale suggerisce l'algoritmo da utilizzare ed il numero ottimale di *cluster* per una più valida clusterizzazione.

5.2.1 Prove di clusterizzazione e bontà dei raggruppamenti

Come già accennato in precedenza, la *cluster analysis* è stata condotta considerando le tre componenti scaturite dall'*analisi delle componenti principali* come variabili di

clusterizzazione. Sono state effettuate diverse prove di raggruppamento, combinando i diversi algoritmi ed i vari metodi di calcolo della distanza tra le unità statistiche, per poter scegliere alla fine la combinazione che assicurasse la massima omogeneità possibile tra le unità all'interno dei gruppi e, contemporaneamente, la massima eterogeneità raggiungibile tra i *cluster*; si sono dunque selezionati l'algoritmo del *legame completo*, come detto in precedenza, ed il *metodo della distanza euclidea*, poiché permettevano meglio degli altri di raccogliere nello stesso *cluster* le unità con caratteristiche più simili tra loro ed, allo stesso tempo, più dissimili rispetto alle unità statistiche appartenenti agli altri *cluster*.

L'omogeneità all'interno del *cluster* è rappresentata dal valore dell'indice della *silhouette*, che varia da -1 ad 1, e che valuta la consistenza tra i vari gruppi nonché verifica quanto una singola unità statistica è adatta a far parte del *cluster* a cui è stata assegnata, ovvero quanto una unità è simile alle altre unità presenti nel suo *cluster*, rispetto a quelle al di fuori del suo *cluster* di appartenenza.

Una buona approssimazione della qualità della clusterizzazione è data dallo studio del *dendrogramma*, che consiste nella rappresentazione grafica dei passaggi di agglomerazione gerarchici eseguiti attraverso l'algoritmo prescelto. Esso rappresenta la struttura dei *cluster*, figlia del processo di agglomerazione dipendente dal metodo considerato, partendo dalle prime agglomerazioni fino ad arrivare a quelle finali. Sull'asse delle ascisse vengono riportate le unità statistiche mentre sull'asse delle ordinate vengono riportate le distanze. Ciascun ramo rappresenta un *cluster* e la linea orizzontale che li congiunge indica il livello di fusione in corrispondenza del quale si uniscono i gruppi. Livelli di fusione bassi indicano che le unità statistiche che sono state agglomerate avevano caratteristiche simili; dunque un *dendrogramma* che rappresenta una buona omogeneità all'interno dei gruppi presenterà delle linee di congiunzione tra i *cluster*, parallele all'asse delle ascisse, che hanno origine in corrispondenza di bassi valori sull'asse delle ordinate e, dunque, più vicine all'asse delle ascisse, per via dei bassi livelli di fusione. L'eterogeneità tra i due *cluster* finali, infine, si nota dall'unione dei due ultimi *cluster*, in corrispondenza del più alto livello di fusione segnalato sull'asse delle ordinate; quanto più questo livello di fusione si distanzia dal livello di fusione appena precedente, tanto più i *cluster* sono eterogenei tra loro, e dunque la clusterizzazione è tanto più soddisfacente.

5.2.2 Importanza delle variabili di clusterizzazione e *t* test per il confronto tra le medie

Per raggiungere risultati significativi in termini di descrizione del campione dei lettori, sono state condotte varie prove di clusterizzazione. Ciascuno dei raggruppamenti è stato valutato servendosi del confronto tra le medie dei cluster e le medie delle componenti dalle quali sono stati formati, attraverso i *test t* di Student a due code. Attraverso il test *t* si può valutare l'utilità delle variabili nell'operazione di clusterizzazione, attraverso la valutazione del *p-value*. Il test di ipotesi è strutturato come segue:

$$\begin{cases} H_0: \mu_{cluster} = \mu_{campione} \\ H_1: \mu_{cluster} \neq \mu_{campione} \end{cases}$$

dove $\mu_{cluster}$ è la media del cluster e $\mu_{campione}$ è la media del campione. Il test è stato ripetuto per ogni *cluster* e per ogni componente considerata. L'ipotesi nulla asserisce la sovrapposibilità tra la media del cluster per una data componente e la media del campione per la stessa componente. L'utilità della variabile considerata sussisteva nel caso in cui si fosse rifiutata l'ipotesi nulla, ovvero se le due medie non fossero coincise significativamente, cioè se il *p-value* fosse inferiore di 0,05, o al più inferiore a 0,10. Nel caso in cui le medie fossero coincise, e dunque se si fosse accettata l'ipotesi nulla, il test avrebbe asserito l'inutilità di tale variabile ai fini della clusterizzazione, poiché la media del *cluster* ottenuto non sarebbe differita da quella del campione.

La prima clusterizzazione è stata effettuata partendo dalle tre componenti principali ottenute dall'*analisi delle componenti principali*. Questa prima clusterizzazione ha restituito una suddivisione sulla base dell'utilizzo delle suddette tre variabili; è stata dunque effettuata la verifica dell'utilità delle singole variabili nella creazione dei gruppi, per verificare se ce ne fosse qualcuna di non vincolante. Dal *test t* di confronto tra le medie è risultato che, nel primo *cluster* si è potuto rifiutare l'ipotesi nulla per quanto riguardava la prima e la seconda componente, in quanto il calcolo del *p-value* restituiva, rispettivamente, i valori di 0,0038 e 0,0633, entrambi inferiori a 0,10. Per quanto riguarda la terza componente, il *t test* ha permesso di esprimere l'inutilità della stessa nella clusterizzazione, in quanto il valore del *p-value* era pari a 0,8147, superiore a 0,10. Applicando il *t test* anche sul secondo *cluster*, è risultato anche in questo caso che le prime due componenti concorrevano alla creazione dei gruppi (*p-value* = 6,666e-15 per la prima componente, *p-value* = 1,469e-07 per la seconda componente) mentre la terza componente non aveva contribuito (*p-value* = 0,5545).

Dato che la terza componente non è stata utile nella formazione di nessuno dei due *cluster*, si è proceduto con una seconda clusterizzazione che escludesse l'utilizzo della terza dimensione nella procedura di suddivisione delle unità statistiche in gruppi. Dopo aver operato questa seconda ripartizione, utilizzando sempre il legame completo e la distanza euclidea, si è proceduto nuovamente con la verifica dell'utilità delle variabili considerate, prima e seconda componente. In riferimento al primo *cluster*, il risultato del *t test* sulla prima componente ha portato a dover accettare l'ipotesi nulla, avendo riscontrato un *p-value* pari a 0,7561, superiore a 0,10. Il test sulla seconda componente ha invece restituito un *p-value* pari a 0,03493 permettendo di rifiutare l'ipotesi nulla; la prima componente non è dunque risultata utile ai fini della composizione del primo *cluster*, mentre la seconda componente ha concorso a determinarlo. Lo stesso risultato si è ottenuto anche nella verifica di ipotesi *t test* relativa al secondo cluster; in particolare, la prima componente non ha concorso alla formazione del suddetto *cluster* (*p-value* = 0,1277) mentre la seconda componente è stata determinante (*p-value* < 2,2e-16).

In seguito si è dunque operata una terza clusterizzazione, che considerasse esclusivamente la seconda componente come variabile, poiché la prima è risultata inutile sia nella determinazione del primo che nella determinazione del secondo *cluster*. Eseguendo il *t test* per il primo e secondo *cluster*, il risultato ha permesso di rifiutare l'ipotesi nulla in entrambi i casi e con valori del *p-value* di molto inferiori sia a 0,10 che a 0,05 (rispettivamente, *p-value* < 2,2e-16 e *p-value* = 4,374e-09).

Questa terza clusterizzazione costituisce il frazionamento definitivo, che è stato poi considerato per effettuare la profilazione. In seguito verranno descritti i dettagli della clusterizzazione definitiva ed i risultati ottenuti.

5.2.3 Dendrogramma: struttura a due cluster

Si è scelto di considerare il numero di *cluster* ottimale pari a 2 confrontando i vari livelli di fusione e non prima di aver consultato il resoconto disponibile nel pacchetto *cValid* fruibile in R e le sue funzioni, le quali combinano vari algoritmi di clusterizzazione ed identificano quali sono gli algoritmi migliori e il numero di *cluster* ottimale per ogni algoritmo. I due *cluster* che sono stati formati, mostrati nella Figura 5.8 (incorniciati in due riquadri di colore rosso) contavano rispettivamente 89 individui e 165 individui; la numerosità differente tra i due cluster suggerisce una buona clusterizzazione, che verrà poi verificata in seguito.

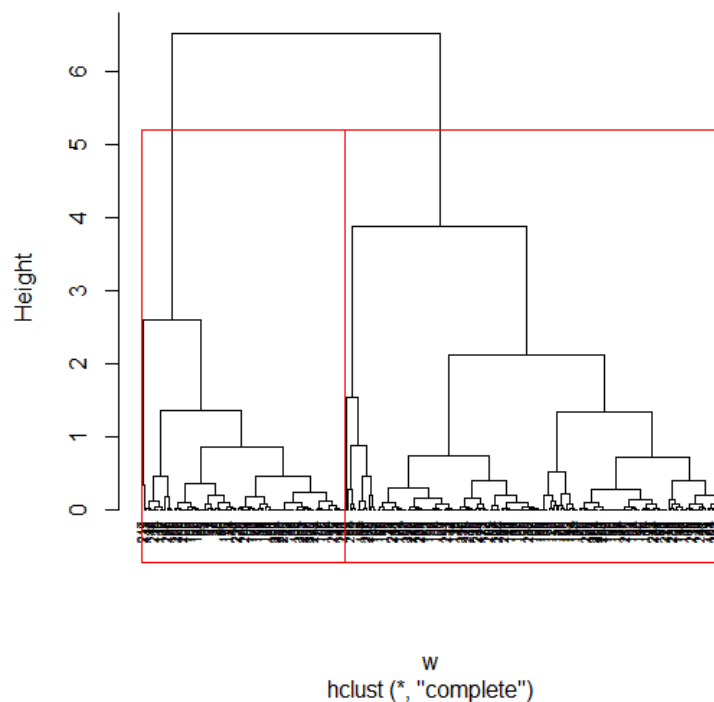


Figura 5.8 Dendrogramma a due cluster relativo alla seconda componente.

Come si vede dal *dendrogramma* riportato in Figura 5.8, il livello di fusione più elevato lo si osserva in corrispondenza dell'ultima fusione, quella che unisce i due *cluster* ottimi; esso assume il valore 6,5128. Il massimo livello di fusione suggerisce di operare il taglio del *dendrogramma* in corrispondenza del segmento che identifica la differenza tra il massimo livello di fusione suddetto e quello immediatamente precedente. I bassi livelli di fusione a cui si sono agglomerate la maggior parte delle unità statistiche indicano che la clusterizzazione è stata netta, ovvero che si sono individuate similarità significative tra gli individui che sono stati assegnati allo stesso *cluster*, e dissimilarità altrettanto consistenti tra questi ultimi e gli individui che sono invece stati assegnati all'altro *cluster*. Questa partizione così decisa dovrebbe riflettersi in una soddisfacente omogeneità interna ai cluster, la quale è possibile verificare attraverso lo studio della *silhouette*.

5.2.4 Silhouette: struttura a due cluster

La *silhouette*, come già accennato in precedenza, misura la qualità della partizione dal punto di vista dell'omogeneità *intra-cluster* e, nello specifico, associa ad ogni unità statistica all'interno del *cluster* un valore compreso tra -1 e 1 che esprima il livello di adeguatezza dell'unità statistica in questione all'interno di quel *cluster*: più il valore si

avvicina ad 1 più la partizione è ben fatta. Le *silhouette* di primo secondo e terzo livello vengono rappresentate nella Figura 5.9.

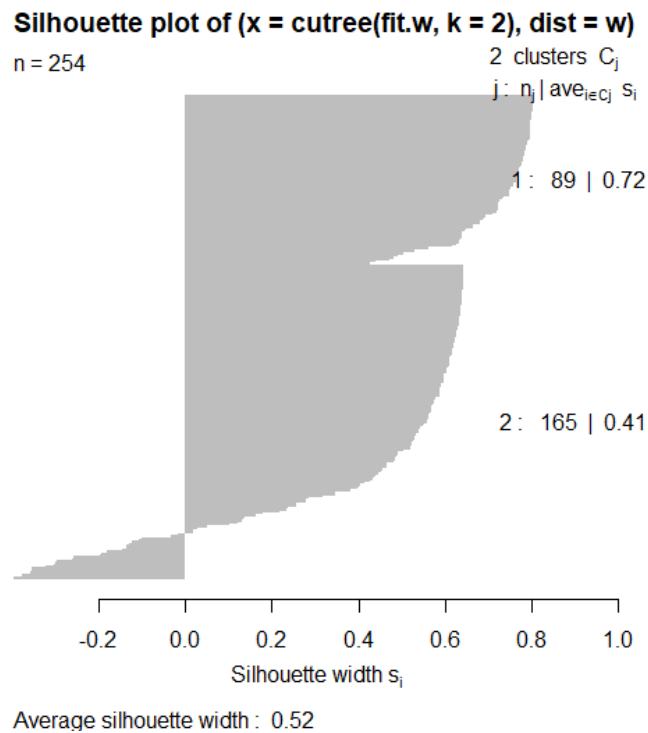


Figura 5.9 Dettaglio complessivo delle silhouette.

La *silhouette* di primo livello viene calcolata nel seguente modo:

$$s(i) = \frac{b(i) - a(i)}{\max[b(i), a(i)]}$$

dove:

- i rappresenta la generica unità statistica;
- $b(i)$ rappresenta la distanza minima tra l'unità statistica i e le unità statistiche al di fuori di quel *cluster*, è dunque una misura dell'eterogeneità tra gli individui: quanto più è elevata, tanto più la clusterizzazione sarà soddisfacente;
- $a(i)$ rappresenta la distanza media tra l'unità statistica i e le unità statistiche che appartengono allo stesso *cluster*, è dunque una misura dell'omogeneità fra individui: quanto più è vicina a 0, tanto più la clusterizzazione sarà soddisfacente.

Inoltre, l'intervallo di definizione della *silhouette* è il seguente:

$$-1 \leq s(i) \leq 1$$

Le *silhouette* delle singole unità statistiche sono rappresentate dai listelli di colore grigio posizionati in orizzontale nella figura. Come si può notare, le unità statistiche del primo *cluster* registrano tutte una *silhouette* molto positiva, in tutti i casi superiore al valore di 0,4: questo indica che le unità sono ben posizionate all'interno del primo *cluster*. Nel secondo *cluster* la maggioranza delle unità statistiche presenta una *silhouette* positiva, ma ci sono anche alcune unità statistiche per le quali il valore della *silhouette* di primo grado è risultato negativo.

Si può ora ragionare sulla *silhouette* di secondo livello, che consiste nella media tra le *silhouette* delle singole unità statistiche all'interno di ogni *cluster*.

Per ogni gruppo, la *silhouette* di secondo livello viene calcolata come segue:

$$S(g_j) = \frac{\sum_{i=1}^{n_{gj}} S(i)}{n_g}$$

dove:

- $S(i)$ rappresenta le silhouette di primo grado di una generica unità statistica;
- n_g rappresenta la numerosità del *cluster* g ;
- j rappresenta il numero identificativo del *cluster*.

I valori delle *silhouette* di secondo livello sono presentati nella parte destra della figura, preceduti dalla numerosità del *cluster*. Entrambe hanno valore positivo, per il primo *cluster* ha valore pari a 0,72 mentre per il secondo *cluster* ha un valore pari a 0,41. La partizione è stata effettuata adeguatamente; la *silhouette* migliore è quella relativa al primo *cluster*, essa è molto vicina ad 1 e dunque identifica un'ottima partizione, ed anche la *silhouette* del secondo cluster attesta una buona partizione.

L'ultima informazione interessante riguardo la *silhouette* è data dal valore della *silhouette* di terzo livello, o *silhouette media*; essa rappresenta una sintesi delle *silhouette* dei due cluster, calcolata come media pesata delle *silhouette* di ciascun gruppo:

$$\bar{S} = \frac{S(g_1) * n_1 + S(g_2) * n_2}{n_1 + n_2}$$

dove:

- $S(g_1)$ identifica la silhouette di secondo livello relativa al *cluster* 1;
- $S(g_2)$ identifica la silhouette di secondo livello relativa al *cluster* 2;
- n_1 è la numerosità del primo *cluster*;

- n_2 è la numerosità del secondo *cluster*.

La *silhouette* di terzo livello denota un'ottima partizione; lo si evince dal valore decisamente alto, pari a 0,52.

5.2.5 Boxplot: struttura a due cluster

Il *boxplot* consiste in un grafico che descrive in maniera sintetica la distribuzione dei dati della variabile quantitativa considerata e permette di fare delle considerazioni in merito alla sua simmetria od asimmetria; si rivela particolarmente utile per confrontare due o più distribuzioni. La creazione del *boxplot* viene posta in essere a partire da 5 quantità:

- *Mediana*, denominata M_e , la quale fornisce indicazioni sulla disposizione centrale dei dati;
- *Primo e terzo quartile*, denominati Q_1 e Q_3 , i quali notificano la variabilità dei dati attraverso lo scarto interquartile;
- Il *massimo*, che identifica il valore più alto;
- Il *minimo*, che identifica il valore più basso.

La costruzione del *boxplot* avviene collocando su un asse verticale le quantità di sintesi suddette; si verranno così a creare la “scatola” del *boxplot* ed i “baffi” dello stesso:

- Il *primo quartile* delimita inferiormente la scatola, mentre il *terzo quartile* la delimita superiormente;
- La *mediana* è la quantità che la divide in due parti;
- La linea di congiunzione tra il valore *minimo* ed il *primo quartile* forma il baffo inferiore, mentre il segmento che congiunge il *terzo quartile* al valore *massimo* crea il baffo superiore.

Le unità appartenenti alla variabile considerata sono suddivise, all'interno del grafico, nel modo mostrato nella Figura 5.10.

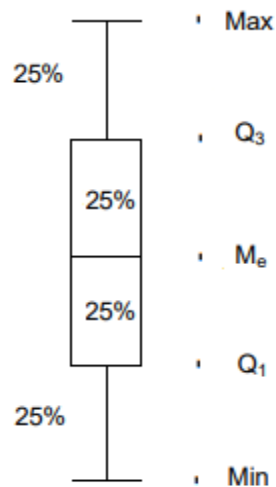


Figura 5.10 Struttura del boxplot.

In particolare:

- Tra il *minimo* ed il *primo quartile* Q_1 ricade il 25% delle unità;
- Tra il *primo quartile* Q_1 e la *mediana* ricade il 25% delle unità;
- Tra la *mediana* ed il *terzo quartile* Q_3 ricade il 25% delle unità;
- Tra il *terzo quartile* Q_3 ed il *massimo* ricade il 25% delle unità.

Dunque, all'interno della scatola ricade il 50% delle unità, ed il restante 50% si colloca nei baffi.

Per quanto riguarda la simmetria della distribuzione, si può affermare che essa sussiste se l'attributo assume valori che presentano, con egual frequenza, valori egualmente distanti dal centro di simmetria, in altre parole se la distanza tra il *primo quartile* e la *mediana* è uguale alla distanza tra la *mediana* ed il *terzo quartile*; se invece una distribuzione presenta una quantità maggiore di unità che presenta caratteristiche più alte o più basse, essa deve considerarsi asimmetrica.

Per l'analisi dei grafici *boxplot* si è proceduto con il confronto tra il *boxplot* rappresentativo di ciascun *cluster* ed il *boxplot* relativo alla variabile considerata, ovvero la seconda componente. Calcolando per ciascun gruppo e per il campione i 5 parametri di riferimento, si sono ottenuti i risultati riportati nella Tabella 5.8.

	Cluster 1	Cluster 2	Campione
Minimo	-3,059	-0,424	-3,059
Primo quartile	-1,231	-0,047	-0,765
Media	-1,015	0,547	0
Mediana	-0,915	0,414	-0,070
Terzo quartile	-0,728	0,805	0,595
Massimo	-0,458	3,454	3,454

Tabella 5.8 Valori per la costruzione dei boxplot.

Sulla base di questi parametri si è poi proceduto alla costruzione dei *boxplot*, che sono riportati di seguito nella Figura 5.11. Sono stati inseriti a due a due nel grafico, di modo che il confronto tra la distribuzione di ciascun gruppo con la distribuzione del campione fosse quanto più intuitiva possibile.

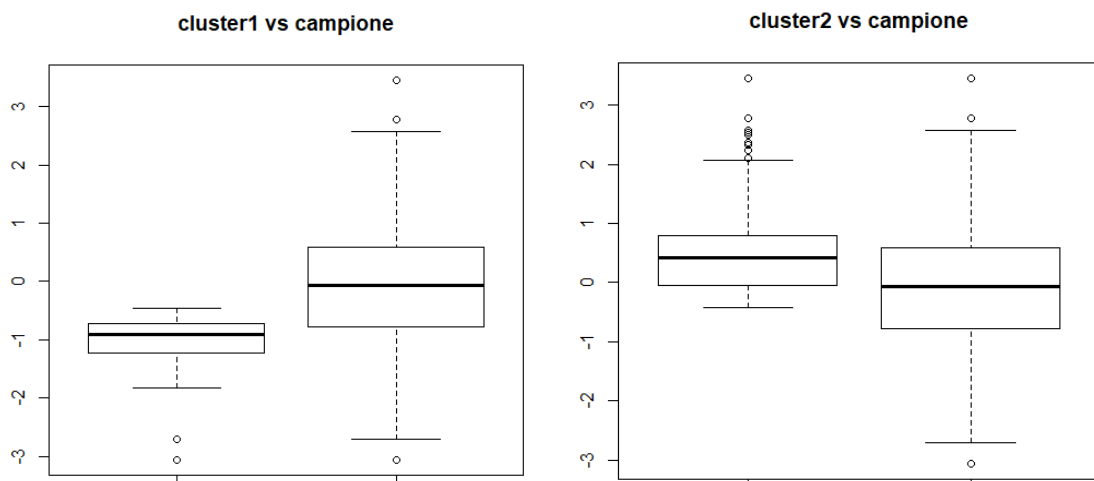


Figura 5.11 Boxplot del primo gruppo con il campione (a sinistra) e del secondo gruppo con il campione (a destra).

La *mediana* divide in due parti la distribuzione, e ciascuna delle due parti ospita il 50% delle unità. Per valutare la sussistenza di simmetria nella distribuzione è stato calcolato

un indice in grado di valutare la “deviazione” dalla condizione di simmetria. Questo indice è stato chiamato i , ed è strutturato come segue:

$$i = (M_e - Q_1) - (Q_3 - M_e)$$

Se la distribuzione risulta simmetrica, il range tra la *mediana* ed il *primo quartile* contiene esattamente lo stesso numero di unità contenuto nel range tra *mediana* e *terzo quartile*. Nel caso in cui il valore dell'indice si discosti da 0, la distribuzione in questione rientra nei canoni di asimmetria. Si è proceduto calcolando l'indice i per ciascun gruppo e per il campione; i risultati sono riportati nella Tabella 5.9:

	Cluster 1	Cluster 2	Campione
$M_e - Q_1$	0,316	0,461	0,695
$Q_3 - M_e$	0,187	0,390	0,665
i	0,129	0,071	0,030

Tabella 5.9 Risultati dell'indice di valutazione della simmetria della distribuzione.

Per quanto riguarda il primo *cluster*, la concentrazione di unità statistiche è maggiore tra il *primo quartile* e la *mediana*, dunque la distribuzione assume una forma asimmetrica caratterizzata da una maggiore quantità di elementi con valori inferiori alla *mediana*; si è in presenza dunque di asimmetria positiva ed il grafico presenterà una conformazione allungata della coda destra. Si è registrata una condizione di asimmetria positiva anche nel caso del secondo *cluster*, sebbene la differenza di concentrazione sia meno ingente rispetto a quella nel primo *cluster*. Il campione presenta pressoché la stessa concentrazione di unità statistiche, a destra ed a sinistra della *mediana*, il che denota una condizione di simmetria della distribuzione.

5.2.6 Diagramma di dispersione, confronto tra le medie e profilazione dei caratteri: struttura a due cluster

Si può rappresentare graficamente la struttura dei due *cluster* considerati attraverso un diagramma di dispersione che evidenzia la disposizione delle unità statistiche della seconda componente, per verificare la consistenza del *cluster* e la coesione tra le varie unità appartenenti allo stesso *cluster*. Dalla Figura 5.12 si può facilmente notare che la suddivisione in due *cluster* risulta significativa, in quanto i record del campione formano due gruppi ben distinguibili anche a prima vista.

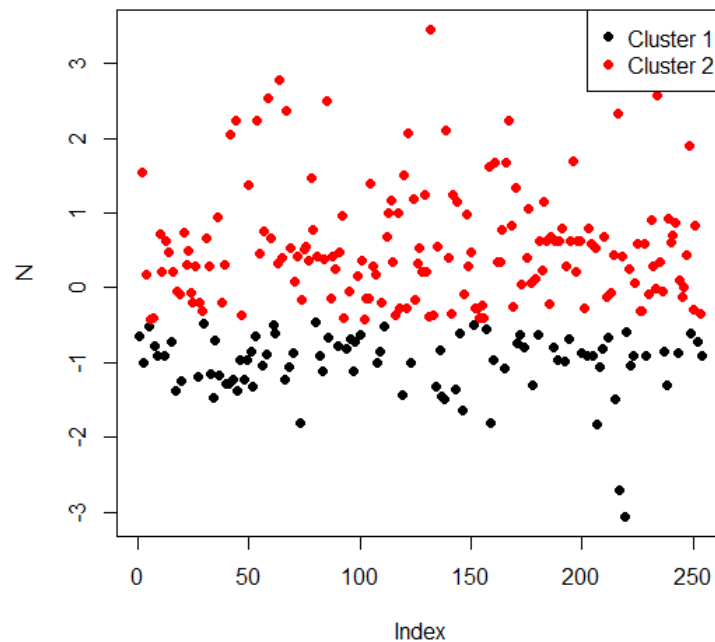


Figura 5.12 Diagramma di dispersione delle unità statistiche suddiviso per cluster.

La figura mette in luce come nel primo *cluster*, i cui elementi sono rappresentati in nero, siano contenute unità statistiche con caratteristiche più univoche, in quanto esse non sono largamente disperse; al contrario si dispongono, fatta eccezione per qualche sporadico caso, tutte molto vicine tra loro. Le unità appartenenti al secondo *cluster* risultano invece più disperse, soprattutto quelle che presentano i valori della componente più elevati. Per identificare in termini di risposte e caratteristiche l'entità della suddivisione dei record, si è proceduto ad indentificare quali fossero le peculiarità degli individui racchiusi in ciascun *cluster*.

Si partiti dalla media delle valutazioni conferite dall'intero campione alla richiesta di quantificare l'importanza attribuita alla seconda componente, ovvero alle informazioni ottenute attraverso i media tradizionali; in seguito si è svolto il confronto tra la suddetta media e le medie delle valutazioni degli individui contenuti nei rispettivi *cluster*. La media del campione è stata denominata μ_c , mentre le medie del primo e secondo cluster sono state denominate rispettivamente μ_1 e μ_2 . Nella Figura 5.13 viene riportato il confronto grafico, riportando come estremi il *massimo* ed il *minimo* valore del campione.

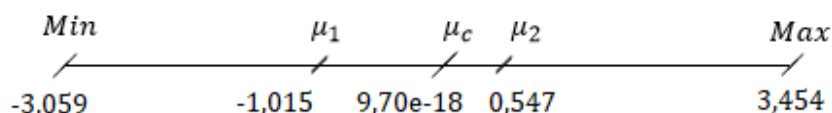


Figura 5.13 *Suddivisione valutazioni in base al cluster di appartenenza.*

Dall'analisi delle medie si evince quanto segue:

- Il primo *cluster* contiene le unità statistiche che hanno conferito alle informazioni ottenute dai media tradizionali delle valutazioni di importanza sotto la media campionaria: il primo *cluster* contiene dunque gli individui che hanno considerato mediamente meno importanti le informazioni derivanti dal mondo dei media classici;
- Il secondo *cluster* contiene le unità statistiche che hanno invece assegnato alle indicazioni provenienti dai media tradizionali delle valutazioni sopra la media campionaria: il secondo gruppo contiene dunque le unità statistiche che hanno considerato mediamente più importanti le informazioni provenienti dai classici media.

Sono state, infine, analizzate nel dettaglio le risposte generali e le caratteristiche socio-demografiche degli individui appartenenti ai due *cluster*, per individuare quali fossero le peculiarità più diffuse e condivise tra le persone assegnate allo stesso raggruppamento. Tra le risposte generali si sono considerate la quantità di libri letti pro capite in un anno e la frequenza di lettura; i risultati sono riportati di seguito, nelle Tabelle 5.10 e 5.11.

Quanti libri all'anno?	Frequenze cluster 1	Frequenze cluster 2
Da 1 a 3 libri	19	16
Da 4 a 11 libri	34	70
Più di 12 libri	36	79

Tabella 5.10 *Quantità di libri letti per ogni cluster.*

Quante volte legge?	Frequenze cluster 1	Frequenze cluster 2
---------------------	---------------------	---------------------

Più di 10 volte al mese	46	100
Dalle 5 alle 10 volte al mese	18	27
Dalle 3 alle 5 volte al mese	12	26
1 o 2 volte al mese	9	8
Meno di 1 volta al mese	4	4

Tabella 5.11 Frequenza di lettura per ogni cluster.

In entrambi i gruppi, i dati modali per ciascun attributo evidenziano:

- Maggioranza di persone che leggono più di 12 libri all'anno;
- Maggioranza di persone che leggono con una frequenza superiore alle 10 volte al mese.

Come si può osservare, la distribuzione di frequenze è simile tra il primo ed il secondo gruppo per entrambe le domande anche se, riguardo la quantità dei libri letti annualmente, nel secondo *cluster* si può notare che le persone che leggono più di 4 libri all'anno sono in quantità proporzionalmente maggiore rispetto al primo *cluster*. Il campione analizzato è formato dunque, in maggioranza, da lettori assidui (più di 12 libri l'anno) che si ritagliano del tempo per leggere più di 10 volte al mese.

Si presentano di seguito i risultati delle analisi socio-demografiche degli intervistati, nell'ottica di individuare le differenze salienti tra il primo ed il secondo gruppo. Le peculiarità sociali e demografiche sono riportate nelle Tabelle 5.12, 5.13, 5.14, 5.15, 5.16.

Sesso	Frequenze cluster 1	Frequenze cluster 2
Maschio	14	34
Femmina	75	131

Tabella 5.12 Genere dei lettori per ogni cluster.

Età	Frequenze cluster 1	Frequenze cluster 2
-----	---------------------	---------------------

Fino a 17 anni	3	3
Da 18 a 24 anni	39	75
Da 25 a 29 anni	22	45
Da 30 a 34 anni	5	9
Da 35 a 39 anni	4	10
Da 40 a 44 anni	3	9
Da 45 a 49 anni	6	6
Da 50 a 54 anni	4	4
Da 55 a 59 anni	3	2
Da 60 a 64 anni	0	0
65 anni o più	0	2

Tabella 5.13 Età dei lettori per ogni cluster.

Titolo di studio	Frequenze cluster 1	Frequenze cluster 2
Nessun titolo di studio	0	1
Licenza di scuola elementare	0	0
Licenza di scuola media	8	8
Diploma breve di scuola superiore	2	9
Diploma di maturità superiore	34	60
Laurea triennale	18	47
Laurea magistrale	22	31
Istruzione post universitaria	5	9

Tabella 5.14 Titolo di studio dei lettori per ogni cluster.

Occupazione	Frequenze cluster 1	Frequenze cluster 2
-------------	---------------------	---------------------

Dirigente/ Quadro	4	4
Impiegato/ Operaio	24	29
Libero professionista	8	14
Lavoratore in proprio	2	6
Studente	41	92
Non occupato	8	14
Pensionato	0	2
Casalengo/a	2	4

Tabella 5.15 Occupazione dei lettori per ogni cluster.

Regione di residenza	Frequenze cluster 1	Frequenze cluster 2
Abruzzo	1	4
Basilicata	1	0
Calabria	2	4
Campania	4	9
Emilia - Romagna	3	11
Friuli - Venezia Giulia	2	3
Lazio	7	7
Liguria	2	2
Lombardia	11	19
Marche	3	5
Molise	0	2
Piemonte	3	14
Puglia	3	12
Sardegna	2	4

Sicilia	2	7
Toscana	4	6
Trentino - Alto Adige	0	0
Umbria	1	3
Valle d'Aosta	1	0
Veneto	37	53

Tabella 5.16 Regione di residenza dei lettori per ogni cluster.

Il campione considerato per l'analisi delle componenti principali e per l'analisi dei cluster corrispondeva al campione dei lettori ed era caratterizzato dalle peculiarità socio-demografiche descritte nel Capitolo 3.

La formazione dei gruppi ha suddiviso il campione in due parti; dopo aver confrontato i dati modali per ciascun attributo socio-demografico si è desunto che i due gruppi presentavano, approssimativamente, le stesse caratteristiche ovvero:

- Maggioranza di persone di sesso femminile;
- Maggioranza di rispondenti con un'età compresa tra 18 e 24 anni;
- Maggioranza di persone possedenti il diploma di maturità superiore;
- Maggioranza di studenti;
- Maggioranza di persone residenti nella regione Veneto.

Per quanto riguarda le risposte generali collegate a tali profili si è registrata una netta maggioranza di lettori assidui o medi (da 4 a più di 12 libri all'anno) per un totale dell'86,22% dell'intero campione (suddivisi in 40,944% di lettori medi e 45,276% di lettori assidui), nonché il 57,48% dei rispondenti che ha dichiarato di leggere con una frequenza che supera le 10 volte al mese.

Per quanto riguarda la suddivisione del campione in gruppi, essa ha evidenziato delle differenze lievi, leggermente più marcate per quanto riguardava il titolo di studio, l'occupazione e la regione di residenza:

- Il diploma di maturità, la laurea triennale e la laurea magistrale rappresentavano i tre titoli di studio che descrivevano l'83,464% dei rispondenti. Di questo 83,464%

si sono analizzate l'appartenenza al *cluster* e la suddivisione tra diplomati e laureati triennali, messa a confronto con gli studenti magistrali. Nel primo *cluster*, considerando solo il gruppo di rispondenti che ha segnalato diploma di maturità e laurea (triennale o magistrale) come titolo conseguito, il 70,73% degli intervistati ha dichiarato di avere conseguito il diploma di maturità o la laurea triennale, mentre nel secondo *cluster* la percentuale saliva quasi all'80%. Dunque, sotto il profilo dell'istruzione, il secondo *cluster* conteneva una percentuale maggiore di intervistati che possedevano un'istruzione inferiore al diploma di laurea magistrale rispetto al primo *cluster*; il primo gruppo dunque comprendeva, in proporzione, più persone con un'istruzione più alta.

- Per quanto riguarda l'occupazione, si è deciso di considerare nel confronto tra *cluster* le persone che possiedono un lavoro (dirigente/quadro, impiegato/operaio, libero professionista, lavoratore in proprio) con le persone che devono ancora completare gli studi (studente). I dichiaranti totali con le suddette caratteristiche sono risultati essere l'88,189% dell'intero campione. Considerando dunque questo 88,189%, si è voluta studiare l'appartenenza dei rispondenti dei rispettivi *cluster* alla categoria degli studenti od alla categoria dei lavoratori. All'interno del primo *cluster* il 48,101% degli intervistati ha dichiarato di avere un lavoro, mentre il restante 51,899% ha dichiarato di essere uno studente. Nel secondo *cluster* il 36,551% ha dichiarato di esercitare una professione, mentre il 63,448% ha dichiarato di non aver ancora completato il percorso di studi. Si evince dunque che, sotto il profilo dell'occupazione, il secondo *cluster* raggruppa più persone che stanno ancora studiando rispetto al primo *cluster*.
- Per ciò che concerne la regione di residenza, sebbene in entrambi i gruppi ci fosse una netta maggioranza di individui provenienti dalla regione Veneto, nel primo *cluster* la percentuale di veneti sul totale era pari al 41,57% mentre nel secondo *cluster* era pari al 20,866%. Dunque il primo *cluster* contava un numero di veneti maggiore, in proporzione, rispetto al secondo gruppo.

Capitolo 6

Analisi logit condizionale e parametri per la scelta del libro

In questo capitolo viene descritta l'analisi logit implementata per lo studio della scelta del libro da parte dei lettori di formato cartaceo.

6.1 Modelli a scelta discreta e modello logit condizionale

I modelli a scelta discreta si basano sulla massimizzazione dell'utilità. Nel modello di scelta dell'alternativa, l'alternativa scelta deve conferire all'individuo una maggiore utilità rispetto ad altre alternative. Se l'utilità dell'individuo i che sceglie l'alternativa j è rappresentato come U_{ij} , allora l'alternativa j sarà scelta se e solo se $U_{ij} > U_{il}$ per $j \neq l$.

Poiché i ricercatori non conoscono U_{ij} , la vera utilità dell'individuo, non possono dire con certezza quale destinazione un individuo sceglierà. U_{ij} consiste di due componenti, i componenti osservabili e non osservabili:

$$U_{ij} = V_{ij} + \varepsilon_{ij}.$$

U_{ij} consiste in un'utilità prevista, V_{ij} in una utilità osservabile in base agli attributi della scelta, e una componente casuale non osservata, ε_{ij} . Se si conoscessero ε_{ij} , i ricercatori potrebbero conoscere U_{ij} potendo affermare con certezza quale alternativa sarebbe stata scelta. Dato che gli ε_{ij} non sono osservabili, il meglio che possono fare è predire il risultato finale in termini di probabilità.

La probabilità dell'individuo i che sceglie l'alternativa j può essere descritta come:

$$\begin{aligned} P_{ij} &= P(U_{ij} > U_{il}) \\ &= P((V_{ij} + \varepsilon_{ij}) > (V_{il} + \varepsilon_{il})) \\ &= P((\varepsilon_{il} - \varepsilon_{ij}) < (V_{ij} - V_{il})) \end{aligned}$$

per ogni $j \neq l$. Per risolvere l'equazione si deve imporre una funzione di densità di probabilità per gli ε_{ij} . Ogni tipo di distribuzione di probabilità imposta a ε_{ij} porta a un diverso modello di scelta discreta, in questo studio si presuppone che gli ε_{ij} mostrino una distribuzione del valore estremo

Il modello logit condizionale presuppone che ε_{ij} mostri la distribuzione dei valori estremi. La funzione di densità di probabilità assume la forma seguente:

$$f(\varepsilon_{ij}) = e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon_{ij}}}$$

e la sua funzione di densità cumulativa è espressa come:

$$F(\varepsilon_{ij}) = e^{-e^{-\varepsilon_{ij}}}$$

Inoltre questo modello limita tutti gli ε_{ij} ad essere indipendenti e identicamente distribuiti (iid). La probabilità che l'individuo scelga la destinazione j può essere risolta come l'espressione:

$$P_{ij} = \frac{e^{V_{ij}}}{\sum_j e^{V_{ij}}} = \frac{e^{\alpha'Z_{ij}}}{\sum_j e^{\alpha'Z_{ij}}}$$

Z_{ij} rappresenta tutti i fattori osservati o le variabili esplicative, ed α rappresenta i parametri ottenuti dal modello.

Essendo ε_{ij} indipendenti e identicamente distribuiti, l'equazione impone l'assunzione IIA.

Considera la probabilità che l'individuo i scelga l'alternativa j rispetto all'alternativa l :

$$P_{ij} = \frac{e^{\alpha'Z_{ij}}}{\sum_j e^{\alpha'Z_{ij}}} \quad ; \quad \text{and} \quad P_{il} = \frac{e^{\alpha'Z_{il}}}{\sum_j e^{\alpha'Z_{ij}}}$$

Il rapporto della probabilità di scelta tra j e l è:

$$\frac{P_{ij}}{P_{il}} = \frac{\sum_j e^{\alpha'Z_{ij}}}{\sum_j e^{\alpha'Z_{ij}}} / \frac{e^{\alpha'Z_{il}}}{e^{\alpha'Z_{il}}} = \frac{e^{\alpha'Z_{ij}}}{e^{\alpha'Z_{il}}}$$

Il rapporto tra le probabilità dipende solo dagli attributi di j e l , e non dipende dagli attributi di altre scelte.

La regressione logistica condizionale si riferisce all'applicazione individuale del modello logistico a ciascuna delle condizioni. I coefficienti dei predittori del modello logistico sono modellati condizionatamente in base all'appartenenza dei casi a una particolare condizione.

6.1.1 Assunzioni sulla variabile dipendente e le variabili esplicative

La variabile dipendente nella nostra elaborazione è la scelta del libro, in termini di acquisto e si esplica come variabile dicotomica.

La variabile dipendente del sistema presenta altre caratteristiche, che vengono elencate di seguito:

- è una “*unordered categorical variable*”, ovvero il valore da lei assunto, oltre ad essere dicotomica, denota l'appartenenza ad una categoria; infine è “*unordered*” perché le scelte non rispettano un ordine di grandezza.
- I valori assunti dalla variabile dipendente sono mutuamente esclusivi, ovvero se la variabile assume valore 1 non può assumere valore 0 e viceversa.

Le variabili esplicative dello studio sono quattro: prezzo, recensioni, autore e novità denominate x_1 , x_2 , x_3 , x_4 e consistono in regressori “*alternative specific*”, ovvero regressori x_{ij} che variano sia rispetto all'individuo i che rispetto all'alternativa j .

6.1.2 La proprietà di indipendenza delle alternative irrilevanti (IIA)

Il modello logit condizionale esamina la scelta di un libro da parte degli individui come una scelta discreta multinomiale, ma l'aspetto problematico del modello risiede nella proprietà di indipendenza delle alternative irrilevanti (IIA).

Questa ipotesi implica che il rapporto di probabilità tra gli individui che scelgono tra due alternative non dipende dalla disponibilità o dagli attributi delle altre alternative. La suddetta ipotesi può essere realistica in alcune situazioni anche se in generale, tuttavia, risulta troppo restrittiva, soprattutto quando il numero di alternative nel set di scelta è ampio, come nel caso di questo studio.

L'assunzione di indipendenza delle alternative irrilevanti viene dunque violata, e ciò può indurre il modello a prevedere in modo errato la probabilità di scelta: in particolare, il modello può sovrastimare la probabilità di scegliere un'alternativa, mentre allo stesso tempo sottostima la probabilità di sceglierne un'altra.

6.2 Implementazione del modello logit condizionale studio della probabilità di scelta

L'obiettivo di questa analisi consiste dunque nello sviluppo di un modello in grado di spiegare il comportamento di scelta del campione relativamente ad un set di alternative composte da determinate variabili esplicative, relativamente a delle proposte di libro cartaceo diversamente caratterizzate. Il modello permette di descrivere l'influenza delle suddette variabili esplicative sulla probabilità di scelta.

Come già accennato nel Capitolo 2, le variabili considerate per l'analisi, le quali componevano le diverse alternative proposte, sono state suddivise ciascuna in due categorie e la loro formulazione è presentata di seguito:

- Prezzo, declinato in due possibili determinazioni quantitative mutuamente esclusive: 12€ e 25€;
- Recensioni, declinata in due possibili determinazioni qualitative mutuamente esclusive: disponibili e non disponibili;
- Autore, declinata in due possibili determinazioni qualitative mutuamente esclusive: italiano e straniero;
- Novità, declinata in due possibili determinazioni qualitative mutuamente esclusive: presente o assente (la novità presente qualifica i libri pubblicati da meno di un anno, la novità assente caratterizza i libri pubblicati da più di un anno).

La diversa combinazione tra le suddette variabili ed i rispettivi livelli ha dato luogo alla creazione di 9 diverse alternative di scelta, proposte al campione rispondente attraverso la formula del *cartellino*, la cui modalità consiste nella presentazione di tre alternative alla volta per le quali gli intervistati sono stati chiamati ad esprimere una singola preferenza, individuata come combinazione migliore tra le variabili esplicative di volta in volta proposte.

Il modello logit si rivela utile alla spiegazione delle scelte discrete di una tra diverse alternative mutuamente esclusive. Esistono molte utili applicazioni di modellazione a scelta discreta in diversi campi dell'econometria applicata.

Nel caso in questione, i dati individuali raccolti sono del tipo *stated preferences data*, relativi dunque alle preferenze dichiarate, tre in totale per ogni individuo. Le alternative proposte, tra le quali è stato richiesto di esprimere una preferenza, erano così composte:

- Alternativa 1: prezzo 12€, recensioni non disponibili, autore italiano, presenza di novità;
- Alternativa 2: prezzo 12€, recensioni disponibili, autore straniero, presenza di novità;
- Alternativa 3: prezzo 12€, recensioni disponibili, autore italiano, assenza di novità;
- Alternativa 4: prezzo 25€, recensioni disponibili, autore italiano, presenza di novità;
- Alternativa 5: prezzo 12€, recensioni non disponibili, autore italiano, assenza di novità;
- Alternativa 6: prezzo 12€, recensioni non disponibili, autore straniero, assenza di novità;
- Alternativa 7: prezzo 25€, recensioni non disponibili, autore italiano, presenza di novità;
- Alternativa 8: prezzo 25€, recensioni disponibili, autore straniero, assenza di novità;
- Alternativa 9: prezzo 25€, recensioni non disponibili, autore straniero, assenza di novità.

In ogni cartellino era compresa la possibilità di astenersi dall'esprimere una preferenza tra le alternative proposte, selezionando la casella proponente l'alternativa "*nessuna delle precedenti*". Nelle elaborazioni è stato deciso di non considerare le suddette risposte in quanto non numericamente rilevanti, inoltre le elaborazioni sono state condotte solo per quanto riguarda il campione lettore di libri cartacei, poiché i rispondenti della sezione dedicata al digitale non sono risultati numericamente sufficienti per condurre un'analisi adeguata.

6.2.1 Stima del modello, ottimizzazione della probabilità ed influenza del prezzo

Attraverso l'impiego del logit condizionale si è voluto ottenere un modello di regressione nelle variabili esplicative (prezzo, recensioni, autore e novità) il quale rendesse possibile la stima della probabilità di scelta di un libro a seconda delle sue caratteristiche. In secondo luogo si è indagato su quale fosse l'alternativa tra quelle proposte che massimizzasse la probabilità di scelta, per poterne studiare le caratteristiche. Infine si è eseguita un'analisi di esame del peso della variabile prezzo nella determinazione della scelta, in relazione con le altre variabili esplicative.

6.2.2 Codifica delle variabili esplicative per le elaborazioni

Per compiere la codifica dei valori assunti dalle variabili esplicative sono state scelte delle *categorie di riferimento* in base alle tendenze evidenziate dalle analisi dei dati raccolti nella prima parte del questionario, che individuassero per ogni variabile la categoria che, secondo le previsioni, avrebbe ottenuto maggiore riscontro.

Le preferenze relative alle variabili prezzo, recensioni, autore e novità evidenziate nella prima parte del questionario hanno dimostrato le seguenti tendenze:

- preferenza per un prezzo basso;
- preferenza per la disponibilità delle recensioni;
- preferenza per gli autori stranieri;
- preferenza per la presenza di novità.

Le *categorie di riferimento* selezionate consistevano in quelle che rispecchiavano al meglio gli andamenti evidenziati, ovvero:

- prezzo = 12€;
- recensioni disponibili;
- autore straniero;

- presenza di novità.

Per la codifica delle risposte individuali alle variabili qualitative (recensioni, autore, novità) si è proceduto attraverso l'assegnazione del valore “-1” o “1” secondo la seguente modalità:

- nel caso in cui la risposta dell'individuo corrispondesse alla categoria di riferimento prescelta per quella variabile esplicativa, essa è stata codificata assegnandole il valore “-1”;
- nel caso in cui la risposta dell'individuo non corrispondesse alla categoria di riferimento prescelta per quella variabile esplicativa, essa è stata codificata assegnandole il valore “1”;

La codifica del valore assunto dalle variabili dicotomiche qualitative è stata operata con l'utilizzo dei valori “-1” e “1” in quanto equidistanti da 0 che, per questo motivo, conferiscono una direzionalità alla scelta individuando una crescita o decrescita della variabile.

Le variabili sociodemografiche che sono state considerate per la stima dei modelli definiti sia dalle variabili scelta-specifiche che individuo-specifiche sono le seguenti:

- sesso;
- età;
- titolo di studio;
- condizione occupazionale;
- regione di residenza;
- componenti del nucleo familiare;
- presenza di anziani o bambini nel nucleo familiare;
- percettori di reddito.

Le variabili sesso, condizione occupazionale e presenza di anziani o bambini nel nucleo familiare sono state codificate dicotomicamente con i valori 0 e 1, in particolare:

- per il sesso, 0 se l'individuo è maschio, 1 se l'individuo è femmina;
- per la condizione occupazionale, 0 se l'individuo non è occupato, 1 se l'individuo è occupato;
- per la presenza di anziani o bambini nel nucleo familiare, 0 se sono presenti, 1 se sono assenti.

Tutte le altre variabili, essendo descritte attraverso delle fasce, hanno subito una codifica differente.

La variabile età è stata suddivisa in tre specificazioni:

- fino a 34 anni;
- da 35 a 54 anni;
- oltre i 54 anni.

La variabile che identificava il titolo di studio è stata suddivisa in tre specificazioni:

- fino alla licenza media;
- fino alla maturità;
- possesso di una laurea o titolo superiore.

La variabile che identificava la regione di appartenenza è stata suddivisa in tre specificazioni:

- Nord Italia;
- Italia centrale;
- Sud Italia e Isole.

La variabile che identificava i componenti del nucleo familiare e la variabile che identificava il numero dei percettori di reddito all'interno della famiglia sono state suddivise alla stessa maniera, in cinque specificazioni, rispettivamente:

- una persona nel nucleo familiare;
- due persone nel nucleo familiare;
- tre persone nel nucleo familiare;
- quattro persone nel nucleo familiare;
- più di quattro persone nel nucleo familiare.

e:

- una persona percepente reddito;
- due persone percepenti reddito;
- tre persone percepenti reddito;
- quattro persone percepenti reddito;
- più di quattro persone percepenti reddito.

6.3 Bontà di adattamento: confronto tra modelli scelta-specifici ed individuo-specifici

Si è proceduto operando delle prove di stima di modelli che fossero specificati sia sulla base delle alternative sia sulla base delle caratteristiche individuali. Per valutare se i modelli stimassero in maniera verosimile il comportamento dei dati sono stati utilizzati due criteri di selezione del modello, che verranno descritti di seguito.

6.3.1 Il criterio di informazione di Akaike (AIC)

Il criterio di informazione di Akaike (AIC) rappresenta un metodo per confrontare la qualità dei modelli statistici, che considera il numero di osservazioni e di parametri esplicativi in relazione ad un determinato insieme di dati.

Il criterio AIC è definito come segue:

$$AIC = 2k - 2\ln(\hat{L})$$

Dove k rappresenta il numero di parametri del modello e \hat{L} rappresenta il valore massimo della funzione di verosimiglianza, che costituisce una misura della bontà di adattamento. Dato un insieme di modelli, AIC stima la qualità di ciascun modello in relazione agli altri, pertanto fornisce un mezzo per la selezione del modello che stima in maniera più verosimile la relazione tra la variabile dipendente ed i parametri esplicativi.

L'AIC trova fondamento sulla teoria dell'informazione, infatti esso stima la perdita di informazioni relativamente all'utilizzo di un dato modello di stima: il modello che permette di perdere meno informazioni rappresenta il modello migliore, considerando un compromesso tra la bontà dell'adattamento del modello e la semplicità dello stesso in termini di numero di parametri.

Il criterio lega dunque la quantità di parametri alla bontà di adattamento in termini di perdita di informazioni: sceglieremo quindi come migliore quel modello che minimizza la perdita di informazioni, ovvero quello a cui corrisponde il valore di AIC più basso.

6.3.2 Il criterio di informazione Bayesiano

Il criterio di informazione Bayesiano rappresenta un secondo criterio per la selezione del modello che meglio stima il comportamento dei dati tra un insieme finito di modelli.

Il criterio BIC è definito come segue:

$$BIC = k\ln(n) - 2\ln(\hat{L})$$

La differenza con il criterio di Akaike è che nel criterio di informazione Bayesiano si tiene conto direttamente sia del numero di parametri sia del numero di osservazioni n . Dunque il valore di BIC aumenta all'aumentare del numero di parametri e di osservazioni, il che penalizza i modelli caratterizzati da una più elevata complessità.

Anche in questo caso il modello considerato migliore risulta quello che presenta il valore di BIC più basso.

6.3.3 Confronto tra i due criteri di informazione

La formula per il criterio di informazione Bayesiano (BIC) è simile alla formula per il criterio di informazione di Akaike (AIC), ma il primo assegna una penalità maggiore per il numero di parametri, infatti nel caso di AIC la penalità è $2k$, mentre nel caso di BIC la penalità è $k \ln(n)$. Un punto sollevato da diversi ricercatori è che AIC e BIC sono appropriati per diversi compiti. In particolare, si ritiene che BIC sia più appropriato per selezionare il *vero modello* (cioè il processo che ha generato i dati) dall'insieme di modelli candidati.

Se il *vero modello* è nell'insieme dei modelli candidati, l'indice BIC lo selezionerà con probabilità 1, con $n \rightarrow \infty$; al contrario, quando la selezione viene effettuata tramite l'indice AIC, la probabilità può essere inferiore a 1. Tuttavia i ricercatori sostengono che questo problema risulta spesso trascurabile, poiché il *vero modello* non si trova praticamente mai nel set candidato.

Capitolo 7

Regressione logistica: risultati e considerazioni

7.1 Selezione del modello: prove empiriche

Per la valutazione del modello da considerare ai fini delle analisi, sulla base della verosimiglianza con il processo di generazione dei dati, si è proceduto con il calcolo degli indici AIC e BIC per tutti i modelli generati, i quali comprendevano:

- il modello esclusivamente scelta-specifico;
- il modello scelta-specifico ed individuo-specifico, caratterizzato da tutte le variabili socio-demografiche;
- i modelli scelta-specifici con una sola variabile individuo-specifica alla volta.

I risultati del calcolo dei due criteri di informazione sono presentati nella Tabella 7.1.

	AIC	BIC
Modello scelta-specifico	1278,202	1293,398
Modello scelta ed individuo specifico completo	1364,713	1714,229
Modello scelta-specifico e sesso	1261,905	1307,494
Modello scelta-specifico e età	1261,86	1337,841
Modello scelta-specifico e titolo di studio	1288,488	1364,47
Modello scelta-specifico e occupazione	1261,983	1307,572
Modello scelta-specifico e regione	1270,677	1346,659
Modello scelta-specifico e componenti della famiglia	1251,571	1297,16
Modello scelta-specifico e presenza di anziani o bambini	1264,605	1310,194
Modello scelta-specifico e percettori di reddito	1256,477	1302,066

Tabella 7.1 Codifica delle alternative di scelta.

Considerando il criterio d'informazione di Aikake, il modello che stima in maniera più verosimile il processo generatore dei dati risulta essere il modello scelta-specifico corredato della variabile individuo-specifica relativa ai componenti del nucleo familiare. Nonostante questo modello presenti il valore migliore di AIC, analizzando nel dettaglio la regressione si è notato come la variabile individuo-specifica relativa al numero di componenti del nucleo familiare risulta strettamente significativa solo in relazione all'alternativa 5 ed all'alternativa 6. Inoltre, l'aggiunta della suddetta variabile individuo-specifica ai parametri del modello iniziale ha portato le variabili scelta-specifiche prezzo, autore e novità a perdere la loro significatività statistica, come mostrato nella Tabella 7.2.

Variabili scelta-specifiche	Coefficienti	Errore standard	z-value	Pr(> z)
Prezzo	0,028	0,037	0,753	0,451
Recensioni	-0,732	0,290	-2,524	0,012
Autore	-2,889	0,242	-1,195	0,232
Novità	-0,190	0,254	-0,748	0,454
Variabile componenti del nucleo familiare	Coefficienti	Errore standard	z-value	Pr(> z)
Alternativa 2	0,417	0,293	1,420	0,155
Alternativa 3	0,539	0,326	1,654	0,098
Alternativa 4	0,283	0,278	1,018	0,309
Alternativa 5	0,796	0,256	3,111	0,002
Alternativa 6	0,827	0,277	2,982	0,003
Alternativa 7	0,134	0,279	0,478	0,632
Alternativa 8	0,510	0,335	1,522	0,128
Alternativa 9	0,347	0,303	1,146	0,252

Tabella 7.2 Output del modello scelta-specifico ed individuo-specifico.

Si è dunque proceduto all'analisi del criterio di informazione Bayesiano, il cui valore ottimale, e dunque indice del modello più verosimile, si riferisce al modello scelta-specifico iniziale senza specificazioni individuo-specifiche.

È stato dunque fatto affidamento sulla stima dell'indice BIC, assumendo come modello ottimale il modello solamente scelta-specifico. L'analisi dei risultati di tale modello è presentata di seguito.

7.2 La regressione logistica nel modello scelta-specifico

Per la stima del modello scelta-specifico si è definito un *indice di gradimento* V_j che esprimesse il grado di soddisfazione degli individui per ogni alternativa j . L'*indice di gradimento* dipende linearmente dal prezzo e dal valore assunto dalle altre variabili esplicative (recensioni, autore, novità).

Secondo il modello lineare:

$$V_j = \beta_1 x_{j,1} + \beta_2 x_{j,2} + \beta_3 x_{j,3} + \beta_4 x_{j,4}$$

in cui, per ogni $j = 1, 2, \dots, 9$, sono definite le specificazioni delle variabili x_1, x_2, x_3, x_4 riportati nella Tabella 7.3.

j	$x_{j,1}$	$x_{j,2}$	$x_{j,3}$	$x_{j,4}$
choiceid	Prezzo	Recensioni	Autore	Novità
1	12	non disponibili	italiano	presente
2	12	disponibili	straniero	presente
3	12	disponibili	italiano	assente
4	25	disponibili	italiano	presente
5	12	non disponibili	italiano	assente
6	12	non disponibili	straniero	assente
7	25	non disponibili	italiano	presente
8	25	disponibili	straniero	assente
9	25	non disponibili	straniero	assente

Tabella 7.3 Specificazione delle alternative.

Si ha dunque il seguente set di equazioni:

$$V_1 = \beta_1 x_{1,1} + \beta_2 x_{1,2} + \beta_3 x_{1,3} + \beta_4 x_{1,4}$$

$$V_2 = \beta_1 x_{2,1} + \beta_2 x_{2,2} + \beta_3 x_{2,3} + \beta_4 x_{2,4}$$

$$V_3 = \beta_1 x_{3,1} + \beta_2 x_{3,2} + \beta_3 x_{3,3} + \beta_4 x_{3,4}$$

$$V_4 = \beta_1 x_{4,1} + \beta_2 x_{4,2} + \beta_3 x_{4,3} + \beta_4 x_{4,4}$$

$$V_5 = \beta_1 x_{5,1} + \beta_2 x_{5,2} + \beta_3 x_{5,3} + \beta_4 x_{5,4}$$

$$V_6 = \beta_1 x_{6,1} + \beta_2 x_{6,2} + \beta_3 x_{6,3} + \beta_4 x_{6,4}$$

$$V_7 = \beta_1 x_{7,1} + \beta_2 x_{7,2} + \beta_3 x_{7,3} + \beta_4 x_{7,4}$$

$$V_8 = \beta_1 x_{8,1} + \beta_2 x_{8,2} + \beta_3 x_{8,3} + \beta_4 x_{8,4}$$

$$V_9 = \beta_1 x_{9,1} + \beta_2 x_{9,2} + \beta_3 x_{9,3} + \beta_4 x_{9,4}$$

Dove V_j , definito per $j = 1, 2, \dots, 9$ può assumere qualsiasi valore.

Si è proceduto trasformando tale valore V_j in un parametro numerico compreso tra 0 e 1, secondo la trasformazione logit:

$$P_j = \frac{e^{V_j}}{e^{V_1} + e^{V_2} + e^{V_3} + e^{V_4} + e^{V_5} + e^{V_6} + e^{V_7} + e^{V_8} + e^{V_9}}$$

resa in forma compatta come:

$$P_j = \frac{e^{V_j}}{\sum_{j=1}^9 V_j}$$

in cui P_j rappresenta il valore della probabilità di scelta del libro che contiene le caratteristiche delle variabili esplicative corrispondenti alla j -esima alternativa.

Risulta evidente, in questo frangente, che all'aumentare del valore di V_j aumenta il valore della probabilità di scelta del libro proposto dalla j -esima alternativa.

Essendo P_j una probabilità, essa gode di due proprietà tipiche:

$$\left\{ \begin{array}{l} 0 \leq P_j \leq 1, \forall j = 1, 2, \dots, 9 \\ \sum_{j=1}^9 P_j = 1 \end{array} \right.$$

Per arrivare alla determinazione di un modello per la stima delle probabilità di scelta delle diverse alternative si è proceduto regredendo i coefficienti $\beta_1, \beta_2, \beta_3, \beta_4$.

7.3 Regressione operativa dei parametri nel modello logit condizionale e stima dei coefficienti delle variabili scelta-specifiche

È stato stimato il modello logit condizionale tramite il software R utilizzando la libreria *mlogit*, attraverso l'utilizzo delle variabili esplicative scelta-specifiche ed individuo-specifiche, considerato j il numero di alternative ed i il numero degli individui.

La stima dei modelli è operata dal software attraverso il metodo di Newton-Raphson utilizzando il gradiente e l'hessiano, metodo numerico multivariato derivato dal metodo di Newton classico per la ricerca degli zeri ($x_{k \rightarrow 0}$) di una funzione f in una variabile x_k . Secondo il metodo, dunque:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

7.3.1 Stima del modello scelta-specifico

Si è partiti operando la stima del modello logit condizionale definito dalle sole variabili esplicative proprie delle alternative, in particolare il prezzo, le recensioni, l'autore e la novità.

I risultati della regressione vengono riportati nella Tabella 7.4.

Variabili	Coefficienti regrediti	Errore standard	z-value	Pr(> z)
Prezzo	-0,023	0,009	-2,498	0,012
Recensioni	-0,621	0,064	-9,728	0
Autore	-0,416	0,060	-6,891	0
Novità	0,261	0,062	4,192	0

Tabella 7.4 Rappresentazione dei risultati ottenuti dall'analisi attraverso il software R.

In particolare:

- il valore negativo di un coefficiente indica che l'utilità decresce all'aumentare del valore del parametro;
- il valore positivo di un coefficiente indica che l'utilità aumenta all'aumentare del valore del parametro.

Dai risultati si evince quanto segue:

- un prezzo più elevato diminuisce l'utilità;
- la non disponibilità delle recensioni diminuisce l'utilità;
- il fatto che l'autore sia italiano piuttosto che straniero diminuisce l'utilità;

- l'assenza di novità aumenta l'utilità.

Le frequenze delle scelte in termini di probabilità sono invece riportate nella Tabella 7.5.

Alternative	P_j
Alternativa 1	0,027
Alternativa 2	0,218
Alternativa 3	0,160
Alternativa 4	0,070
Alternativa 5	0,046
Alternativa 6	0,106
Alternativa 7	0,020
Alternativa 8	0,273
Alternativa 9	0,079

Tabella 7.5 Probabilità di scelta delle varie alternative.

Dai risultati si evince che l'alternativa per la quale si registra la più alta probabilità di scelta è l'alternativa 8, mentre le alternative alle quali sono associate le probabilità di scelta peggiori sono, in ordine di decrescente, l'alternativa 3, l'alternativa 4, l'alternativa 1 e l'alternativa 7.

7.3.2 Analisi della combinazione ottimale nel modello scelta-specifico

Per analizzare la combinazione di variabili che massimizza la probabilità di scelta si è proceduto codificando le variabili dicotomiche con i valori di riferimento “-1” e “1”, facendo in modo che il rispettivo coefficiente fornisca indicazioni sulla scelta della variabile, in particolare:

- valori del coefficiente minori di 0 implicano la scelta della categoria codificata con “-1”;
- valori del coefficiente maggiori di 0 implicano la scelta della categoria codificata con “1”.

Per la massimizzazione di V_j si verificano le corrispondenze mostrate nella Tabella 7.6.

β	x_j
---------	-------

> 0	> 0
< 0	< 0

Tabella 7.6 *Corrispondenze tra i coefficienti ed i valori delle variabili esplicative.*

Dai risultati ottenuti si ottiene che, relativamente alle variabili recensioni, autore e novità, l'opzione che massimizza l'*indice di gradimento* è quella mostrata nella Tabella 7.7.

Coefficiente regredito	Variabile	Nome variabile	Codifica	Categoria
-0,6209958	x_2	Recensioni	-1	Disponibili
-0,4160338	x_3	Autore	-1	Straniero
0,2605852	x_4	Novità	1	Assente

Tabella 7.7 *Valori delle variabili esplicative che massimizzano l'indice di gradimento.*

Si evince dunque che la proposta di libro che massimizza l'*indice di gradimento* presenta le seguenti caratteristiche:

- recensioni disponibili;
- autore straniero;
- novità assente.

Per quanto riguarda invece la variabile prezzo, denominata x_1 , essa presenta un coefficiente regredito caratterizzato da un valore negativo; dunque, come ci si aspettava data la natura continua e reale della variabile, all'aumentare del prezzo diminuisce la probabilità di scelta del libro.

Mantenendo fissi i valori delle variabili esplicative x_2 , x_3 ed x_4 alle condizioni ottimali, si lascia variare il prezzo e si opera il confronto tra la probabilità di scelta in funzione del prezzo con la probabilità di scelta delle alternative proposte nel questionario.

Dato che l'alternativa 8 presentava già i valori ottimali per le variabili x_2 , x_3 e x_4 si è proceduto calcolando la probabilità di scelta al variare del prezzo di vendita della alternativa 8, mantenendo i valori delle altre variabili invariati, poiché i loro valori corrispondono alle categorie che massimizzano V_j , come mostrato nella Figura 7.1.

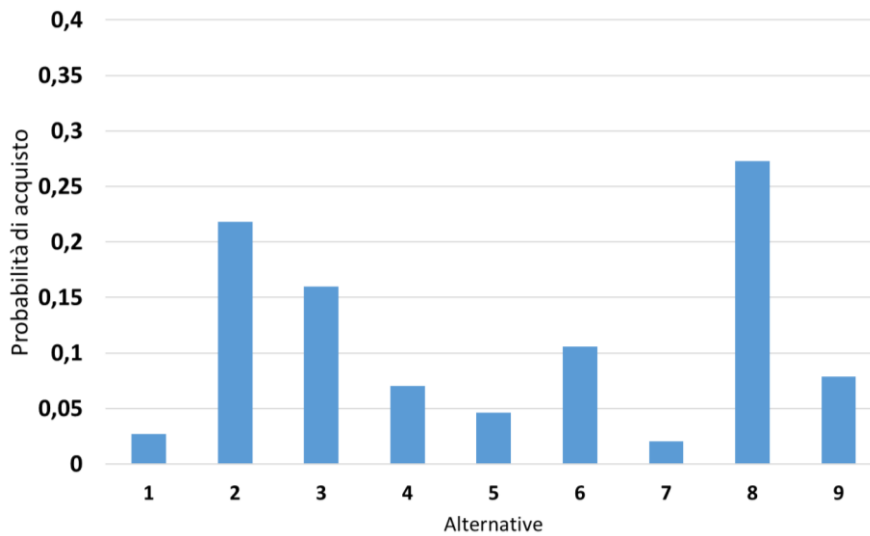


Figura 7.1 Probabilità di acquisto per ogni alternativa.

Sebbene l'alternativa 8 presentasse un livello di prezzo maggiore, pari a 25€, essa si classifica come la migliore tra le alternative, ovvero quella per la quale la probabilità di acquisto è più elevata. Si evidenzia dunque in questa sede che, in corrispondenza di caratteristiche ottimali per quanto riguarda recensioni, autore e novità, la variabile prezzo passa in secondo piano.

Soffermandosi brevemente sulle ragioni per le quali le alternative 1, 4, 5 e 7 hanno una così bassa probabilità di acquisto, si è evinto che ciò è legato alla determinazione delle variabili x_2 , x_3 ed x_4 (Tabella 7.8).

	Categoria ottimale	Alternativa 1	Alternativa 4	Alternativa 5	Alternativa 7
Recensioni	disponibili	non disponibili	disponibili	non disponibili	non disponibili
Autore	straniero	italiano	italiano	italiano	italiano

Novità	assente	presente	presente	assente	presente
---------------	---------	----------	----------	---------	----------

Tabella 7.8 *Confronto tra i valori delle variabili relative alle alternative di scelta peggiori.*

In particolare si è notato come i valori delle variabili esplicative siano opposti alle condizioni ottimali, o perlomeno due su tre dei valori non coincidono con i valori che massimizzano l'*indice di gradimento*. Ciò significa che le variabili esplicative recensioni, autore e novità hanno un peso notevole nella decisione di acquisto.

Conclusioni

Il presente elaborato ha voluto studiare l'andamento delle abitudini di consumo dei lettori, focalizzandosi sul pubblico di lettori di formato cartaceo e sulle loro preferenze sotto vari punti di vista, al fine di identificare delle leve strategiche attraverso le quali piccoli e medi editori possano concentrare miratamente l'offerta da riservare a ciascuno dei segmenti rilevanti di potenziali clienti. A questo scopo è stata condotta una ricerca quantitativa attraverso la somministrazione di un questionario suddiviso in due parti: la prima aveva lo scopo di indagare in maniera generale le abitudini di lettura con domande relative alla frequenza di lettura, alla quantità di libri letti ed acquistati, sui generi più apprezzati e sul formato preferito. La seconda parte del questionario, strutturata secondo la modalità delle alternative raggruppate in cartellini, aveva lo scopo di mettere in luce l'incidenza del prezzo, della disponibilità delle recensioni, della nazionalità dell'autore e della novità del libro sulla scelta di acquisto da parte dei lettori cartacei, che formavano l'87,40% del campione. Il questionario è stato distribuito a 281 individui, suddivisi tra non lettori e lettori, per un totale in quest'ultimo caso di 254 persone, ed i lettori sono stati suddivisi tra digitali e cartacei, i cui studi approfonditi sono stati effettuati su quest'ultima categoria, la quale contava un totale di 222 individui. Le risposte alla prima parte del questionario, considerando il dettaglio dei lettori, hanno evidenziato una forte inclinazione del campione alla lettura, tant'è che la maggior parte di essi si è dichiarata "lettore forte" o "lettore medio", con una frequenza di lettura che supera le 10 volte al mese in quasi il 60% dei casi, ed una percentuale di acquisto del libro che si attesta a circa il 90%. Dalle risposte si attesta che i generi sui quali puntare, poiché rappresentano quelli più apprezzati, sono il genere narrativo drammatico, thriller/crimine/giallo, storico, poetico e fiabesco ed infine il genere avventuroso e d'azione. Per quanto riguarda il dettaglio socio-demografico del campione, si è registrata una prevalenza di femmine, comprese tra i 18 ed i 29 anni, con un titolo di studio almeno pari al diploma di maturità superiore e, in molti casi, alla laurea. La maggior parte del campione risulta studente, residente in una famiglia con 4 componenti totali, tra cui vi sono 2 persone che percepiscono un reddito (si può presumere che si tratti di rispondenti che vivono in casa con genitori, fratelli e sorelle).

È stato eseguito un approfondimento relativo ai canali informativi più frequentati dai lettori, ed è stato accertato che il canale più significativo è rappresentato dai media tradizionali, comprendendo nel novero i giornali, le riviste, la televisione ma anche i social network. Si è

potuto suddividere il campione in due gruppi, il primo comprendente 89 individui che hanno assegnato valutazioni sotto la media a questo canale informativo, e caratterizzato da individui con titolo di studio più elevato e che, in prevalenza, lavorano. Il secondo gruppo formatosi conteneva 165 individui i quali hanno assegnato valutazioni sopra la media per quanto riguarda i media tradizionali utilizzati per informarsi, in cui comparivano in prevalenza individui con titolo di studio inferiore alla laurea e che non hanno ancora completato il percorso di studi.

L'87% circa dei lettori rispondenti hanno dichiarato di preferire il formato cartaceo, ed a questi individui è stato chiesto di compilare anche la seconda parte del questionario, relativa alle preferenze tra combinazioni di alternative variamente declinate, formate da combinazioni tra le variabili prezzo, recensioni, autore e novità. L'analisi in questione ha messo in luce che:

- l'aumento di prezzo causa una diminuzione dell'utilità e, dunque, della probabilità di acquisto del libro;
- l'assenza di recensioni causa una diminuzione dell'utilità e, dunque, della probabilità d'acquisto del libro;
- la nazionalità italiana dell'autore diminuisce l'utilità, e quindi la probabilità di acquisto del libro, rispetto alla nazionalità straniera;
- l'assenza del parametro della novità del libro aumenta l'utilità e, dunque, la probabilità di acquisto del libro.

Sulla base dei dati raccolti, esplicitati in maniera esaustiva in questo elaborato, le piccole e medie case editrici possono adoperarsi per sviluppare un piano di marketing e di comunicazione che sia il più possibile mirato, tenendo conto delle caratteristiche relative alle variabili analizzate e della caratterizzazione sociale e demografica del campione.

Si specifica che i risultati dell'indagine sono legati alle caratteristiche del campione preso in esame: non è possibile dunque fare considerazioni generali ed universalmente valide basandosi solo sui risultati della presente indagine.

Bibliografia e Sitografia

- [1] Accademia della Crusca, Claudio Marazzini (2014), *L'editoria italiana nell'era digitale – Tradizione e attualità*.
- [2] Anderson, D. R. (2008), *Model Based Inference in the Life Sciences*, Springer.
- [3] Associazione italiana Editori. Ufficio studi, Istituto poligrafico e Zecca dello Stato (2002), *L'editoria libraria in Italia*.
- [4] Bailey, Ken (1994), *Numerical Taxonomy and Cluster Analysis. Typologies and Taxonomies*.
- [5] Ben-Akiva and Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*.
- [6] Carla Bodo, Emilio Cabasino, Federica Pintaldi (2009), *L'occupazione culturale in Italia*, FrancoAngeli Editore.
- [7] Cattell, R. B. (1943), *The description of personality: Basic traits resolved into clusters*. Journal of Abnormal and Social Psychology.
- [8] Diego di Dio (2016), *Il Mercato dell'Editoria*, Primiceri Editore.
- [9] Giraud, C. (2015), *Introduction to High-Dimensional Statistics*, CRC Press.
- [10] Istat (2016), *Elaborazione e analisi di indicatori sulle istituzioni e le attività culturali. Biblioteche pubbliche statali*.
- [11] Istat (2016), *Multiscopo sulle famiglie: aspetti della vita quotidiana. Internet: accesso e tipo di utilizzo*.
- [12] Istat (2016), *Multiscopo sulle famiglie: aspetti della vita quotidiana. Quotidiani e libri*.
- [13] Kenneth Train (2009), *Discrete Choice Models with Simulation*
- [14] Konishi, S.; Kitagawa, G. (2008), *Information Criteria and Statistical Modeling*, Springer.
- [15] Koppelman and Bhat (2006), *A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models*.
- [16] McFadden, D. (1973) *Conditional logit analysis of qualitative choice behavior*.
- [17] McFadden, D. (1974) *The Measurement of Urban Travel Demand*, Journal of Public Economics, 3.
- [18] McQuarrie, A. D. R.; Tsai, C.-L. (1998), *Regression and Time Series Model Selection*, World Scientific.

- [19] Nicolai Rubinstein, Giovanni Ciappelli (2004), *Editoria Italiana Online*.
- [20] Train, K. (2004) *Discrete Choice Modelling, with Simulations*, Cambridge, University Press.
- [21] Vincenzo Gervasi, Mirko Tavosanis, M. Teresa Lupia (2011), *Editoria digitale*. UTET Università.
- [22] Zani S. e Cerioli A. (2007). *Analisi dei dati e data mining per le decisioni aziendali*, Milano: Giuffrè.
- [23] Zarembka P. ed., *Frontiers in Econometrics*, New-York: Academic Press.

Sitografia

- https://www.istat.it/it/files/2017/12/Report_EditoriaLettura.pdf
- <https://www.istat.it/storage/rapporti-tematici/conoscenza2018/Rapportoconoscenza2018.pdf>
- <https://www.wired.it/play/libri/2018/01/25/libri-crescono-piano-questione-politica/>
- <https://www.ilpost.it/2018/01/25/dati-mercato-libri-italia-2017-aie/>
- https://www.corriere.it/cultura/17_dicembre_27/libri-dati-istat-lettura-italia-2016-683b8b12-eb2b-11e7-ae4-bc31285a7c38.shtml
- <https://www.illibraio.it/italia-mercato-del-libro-2017-726016/>
- http://www.corriere.it/cultura/18_gennaio_25/libro-editoria-dati-vendite-amazon-digitale-ebook-69ed5ab4-01c2-11e8-9ff2-341a2fe0297c.shtml
- http://www.ansa.it/sito/notizie/cultura/libri/2018/01/25/aie-mercato-libro-2017-cresce-5.8_ef3dc46e-e6a8-44b5-b966-299e131ba52b.html
- <https://www.bauer.uh.edu/rsusmel/phd/ec1-18.pdf>
- https://beta.vu.nl/nl/Images/werkstuk-wittink_tcm235-237206.pdf
- http://tfresource.org/Category:Choice_models
- <https://imai.princeton.edu/teaching/files/discrete.pdf>
- <https://infoscience.epfl.ch/record/138827/files/mbiNestle09.pdf>
- <https://cran.r-project.org/web/packages/mlogit/vignettes/mlogit.pdf>
- <https://www.rdocumentation.org/packages/mlogit/versions/0.1-0/topics/mlogit>
- <https://digitalreport.wearesocial.com>

Ringraziamenti

Ringrazio con tutto il cuore tutte le persone che mi sono state vicine in questo percorso, grazie alle quali sono riuscita a superare ogni difficoltà mi si parasse davanti.

Il primo ringraziamento è per la professoressa Procidano, per le competenze e la passione che le sono proprie, nonché per gli strumenti messi a mia disposizione, che mi hanno permesso di portare a termine nel modo migliore il mio lavoro.

Ringrazio tutti i miei amici tra cui Pier, che mi ha dato la forza per andare avanti anche nei momenti di sconforto, motivandomi e ripetendomi di credere in me, dandomi ogni giorno di più la forza di camminare a testa alta.

Ringrazio le mie amiche Elisa, Irene e Camilla per avermi sostenuta nel mio percorso e per essere dei pilastri portanti della mia vita, per farmi ridere ogni giorno e per avermi dato coraggio nelle avversità: come le stelle, che nella notte rappresentano un punto di riferimento. Per me non smetterete mai di brillare.

Alla mia famiglia va il più grande riconoscimento: al nonno Gino, alla nonna Mina, allo zio Giuliano, alla zia Cristina, che mi hanno dato la forza di resistere e rappresentano la più grande certezza della mia vita. Siete la cosa più preziosa che ho.

Un grazie speciale lo dedico a mamma e papà, che mi hanno sempre supportata e coccolata, ai quali sono legata da un affetto ed un amore incommensurabili. Il vostro calore mi accompagna ogni giorno, rendendo la mia vita meravigliosa. Vi amo.

Grazie anche a mia cugina Elo, amica e compagna di vita, per me più preziosa di una sorella. Grazie perché mi fai ridere anche quando non ne ho voglia, grazie perché quando ridi tu, emani una luce che avvolge e riscalda. Ti voglio un mondo di bene.

Dulcis in fundo, il ringraziamento più grande di tutti lo dedico al mio fidanzato Marco, la persona più unica e speciale che conosca, il mio punto di riferimento, la mia roccia, la mia ancora di salvezza. Voglio ringraziarti per tutto: per la tua forza, per l'aiuto nei momenti di difficoltà, per l'amore che mi dai, per avermi sopportato nei momenti più bui, per il tuo carattere fiero ed ottimista che mi fa sentire sempre al sicuro.

La mia vita è meravigliosa da quando ci sei tu, e tutte le parole che conosco non basterebbero per spiegarti quanto ti amo. Non smettere mai di guardarmi anche quando mi sento invisibile,

non smettere mai di stringermi a te quando mi sento insicura, non smettere mai di sorridermi quando mi sento smarrita. Il tuo sorriso è la mia pace.

Un enorme e sincero grazie a tutti, avervi incontrati è stata la più grande fortuna che mi sia mai capitata.

Ed io non vi ringrazierò mai abbastanza.

Sara