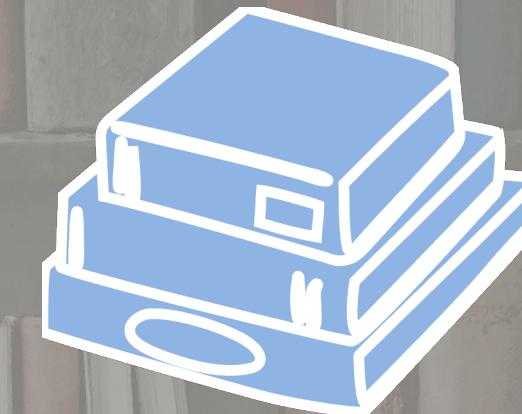


# BOOKS BY MAIL

## LA SCELTA DI CANALE



CLAMSEI, *Modelli di Marketing*  
Anno accademico 2021/2022

Dataset BOOKS\_A

Gaia Bartolini, Alice Biancacci, Roberta Fazio,  
Sara Lelii, Beatrice Nardella, Laura Tirabassi



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

# OVERVIEW e DOMANDA DI RICERCA

**Books by Mail** è un'impresa di direct marketing (divisione club del libro) che offre i propri prodotti tramite **3 canali di vendita**:



Mobile



Internet



Punto vendita

Il dataset fornito dall'impresa contiene le informazioni sugli acquisti effettuati da un campione di 1000 clienti osservati per 14719 occasioni di acquisto.

L'obiettivo conseguito è stato quello di studiare la scelta del canale fornendo conseguentemente iniziative di direct marketing più moderne e innovative in modo tale da: migliorare la qualità dei servizi più deboli, migliorare il rapporto tra impresa e singolo cliente e incentivare gli acquisti.

# Trattamento dei dati e statistiche descrittive

54%  
degli acquisti avviene nei  
**Punti vendita**

scelta	Freq.	Percent	Cum.
mobile	5,882	39.96	39.96
internet	846	5.75	45.71
ptovendita	7,991	54.29	100.00
Total	14,719	100.00	

Età media\*  
**43 anni**

Variable	Obs	Mean	Std. dev.	Min	Max
age	14,719	43.94198	14.62497	21	88

\* i 205 dati anomali sono stati sostituiti con la media

**4.128** uomini\*

**10.591** donne\*

genere	Freq.	Percent	Cum.
0	4,128	28.05	28.05
1	10,591	71.95	100.00
Total	14,719	100.00	

\* Considerando ogni riga come un singolo consumatore

I multicanale sono solo il  
**17%**  
(persone restie a cambiare canale)

. tab multi			
multi	Freq.	Percent	Cum.
0	12,188	82.80	82.80
1	2,531	17.20	100.00
Total	14,719	100.00	

Nel **97%** delle volte  
che si acquista è perché si viene  
reclutati da agenti di  
vendita

. sum emails					
Variable	Obs	Mean	Std. dev.	Min	Max
emails	14,719	.568854	1.39005	.0721874	8.019269
. sum catalogs					
Variable	Obs	Mean	Std. dev.	Min	Max
catalogs	14,719	1.7927	.4043657	1.102285	11.93756
. tab agenti					
agenti	Freq.	Percent	Cum.		
0	380	2.58	2.58		
1	14,339	97.42	100.00		
Total	14,719	100.00			

Abbiamo generato una variabile  
dummy  
**"reso"**  
che vale 1 per chi ha effettuato un  
reso\*

. tab reso			
reso	Freq.	Percent	Cum.
0	14,384	97.72	97.72
1	335	2.28	100.00
Total	14,719	100.00	
. sum resi_euro if reso==1			
Variable	Obs	Mean	Std. dev.
resi_euro	335	20.10648	12.92182
		.99	104.4

\* La standard deviation è alta, c'è molta variabilità: c'è chi manda indietro un alto valore di libri, chi poco

# Multinomial logit

"scelta"

Baseline

Variabile dipendente  
qualitativa multinomiale

Modalità più frequente:

Modalità	Frequenza
Mobile	39,96%
Internet	5,75%
Punto vendita	54,29%

Modello  
Multinomial logit

Punto vendita

Tra le variabili predittive inseriamo "reso" (anziché "resi\_euro") ed eliminiamo "tot" poiché collineare con le tre variabili "Mtot", "PVtot", "Itot".

Multinomial logistic regression

Log likelihood = -2391.6137

Number of obs = 14,719  
LR chi2(24) = 20602.38  
Prob > chi2 = 0.0000  
Pseudo R2 = 0.8116

Alto potere  
predittivo: le variabili  
predittive inserite  
spiegano circa l'81%  
della variabilità del  
fenomeno «scelta del  
canale di acquisto».

scelta	RRR	Std. err.	z	P> z	[95% conf. interval]
<b>mobile</b>					
multi	.9774498	.1105162	-0.20	0.840	.7831638 1.219934
emails	1.085265	.0431129	2.06	0.039	1.003971 1.173142
catalogs	.7914966	.1113538	-1.66	0.097	.6007525 1.042804
reso	5.039726	1.756648	4.64	0.000	2.545142 9.979342
qta	.7060334	.0208246	-11.80	0.000	.6663754 .7480516
recency	1.001921	.0006853	2.81	0.005	1.000579 1.003265
Mtot	1.458262	.0309763	17.76	0.000	1.398796 1.520256
PVtot	.7667167	.0141784	-14.36	0.000	.7394251 .7950156
Itot	1.113624	.0314931	3.81	0.000	1.053578 1.177092
agenti	1.061305	.2600376	0.24	0.808	.6565706 1.715532
age	1.006499	.0036959	1.76	0.078	.9992808 1.013769
genere	1.218609	.1470948	1.64	0.101	.961874 1.543869
_cons	.8437144	.3921012	-0.37	0.715	.3393241 2.097859
<b>internet</b>					
multi	1.801928	.3164482	3.35	0.001	1.277181 2.542275
emails	1.53505	.0655162	10.04	0.000	1.411865 1.668983
catalogs	.8945191	.18196	-0.55	0.584	.6004006 1.332717
reso	4.539061	2.12394	3.23	0.001	1.814126 11.35702
qta	.7392251	.0331862	-6.73	0.000	.6769608 .8072161
recency	.9970242	.0009526	-3.12	0.002	.9951589 .9988929
Mtot	1.203466	.0365477	6.10	0.000	1.133924 1.277273
PVtot	.7552824	.0238176	-8.90	0.000	.710014 .8034369
Itot	1.415049	.0474531	10.35	0.000	1.325033 1.51118
agenti	2.615939	.9384225	2.68	0.007	1.295012 5.284225
age	1.009568	.0055194	1.74	0.082	.9988078 1.020444
genere	1.276732	.2073395	1.50	0.132	.9286791 1.755228
_cons	.0435908	.0304784	-4.48	0.000	.0110724 .1716123
ptovendita	(base outcome)				

Poniamo l'attenzione sulle sole variabili predittive significative al 5% e osserviamo il corrispondente *relative risk ratio*.

- All'aumentare del numero medio di e-mails con iniziative di direct marketing, la probabilità di scegliere mobile anziché punto vendita aumenta di circa 1,08 volte cioè aumenta circa dell'8%.

$$\rightarrow (1,08 - 1) \cdot 100$$

- All'aumentare del numero medio di e-mails con iniziative di direct marketing, la probabilità di scegliere internet anziché punto vendita aumenta di circa 1,53 volte cioè aumenta circa del 53%.

$$\rightarrow (1,53 - 1) \cdot 100$$

- Se il "reclutamento" è stato effettuato tramite un agente di vendita, la probabilità di scegliere internet anziché punto vendita aumenta di circa 2,62 volte cioè aumenta circa del 162%.

$$\rightarrow (2,62 - 1) \cdot 100$$

# Valutazione del modello: Lift chart

Suddividendo **in decili**, per ogni scelta, prima il campione di calibrazione poi il campione di validazione e mettendoli a confronto, si conclude che il modello predittivo usato è efficace: andamento decrescente in entrambi i campioni e valori del LIFT superiori a 1 per i primi decili

**Scelta MOBILE**

decili_out	mean	LIFT
1	0,9973	2,4858
2	1,0000	2,4925
3	0,9810	2,4451
4	0,7717	1,9236
5	0,2364	0,5893
6	0,0245	0,0610
7	0,0027	0,0068
8	0	0
9	0	0
10	0	0
total	0,4012	

2,48 volte più  
probabile rispetto  
alla media

**Scelta INTERNET**

decili_out	mean	LIFT
1	0,5041	9,4619
2	0,0190	0,3570
3	0,0054	0,1020
4	0,0054	0,1020
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
total	0,0533	

9,46 volte più  
probabile rispetto  
alla media

**Scelta PUNTO VENDITA**

decili_out	mean	LIFT
1	1	1,8331
2	1	1,8331
3	1	1,8331
4	0,9973	1,8281
5	0,9674	1,7733
6	0,4701	0,8617
7	0,0109	0,0199
8	0,0109	0,0199
9	0	0
10	0	0
total	0,5455	

1,83 volte più  
probabile rispetto  
alla media

# Test di Hausman e McFadden

Il Multinomial logit si regge sull'assunzione **dell'indipendenza delle alternative irrilevanti (IIA)**, ossia l'assunzione dell'indipendenza tra le componenti di errore che implica la conoscenza del valore degli errori per ogni cliente, alternativa di scelta e occasione. Nella realtà però gli errori possono avere un'alta correlazione:

Test of H0: Difference in coefficients not systematic

$$\begin{aligned} \text{chi2(13)} &= (\mathbf{b}-\mathbf{B})' [(\mathbf{V}_\mathbf{b}-\mathbf{V}_\mathbf{B})^{-1}] (\mathbf{b}-\mathbf{B}) \\ &= \mathbf{61.33} \end{aligned}$$

Prob > chi2 = **0.0000**

p-value < 0.05

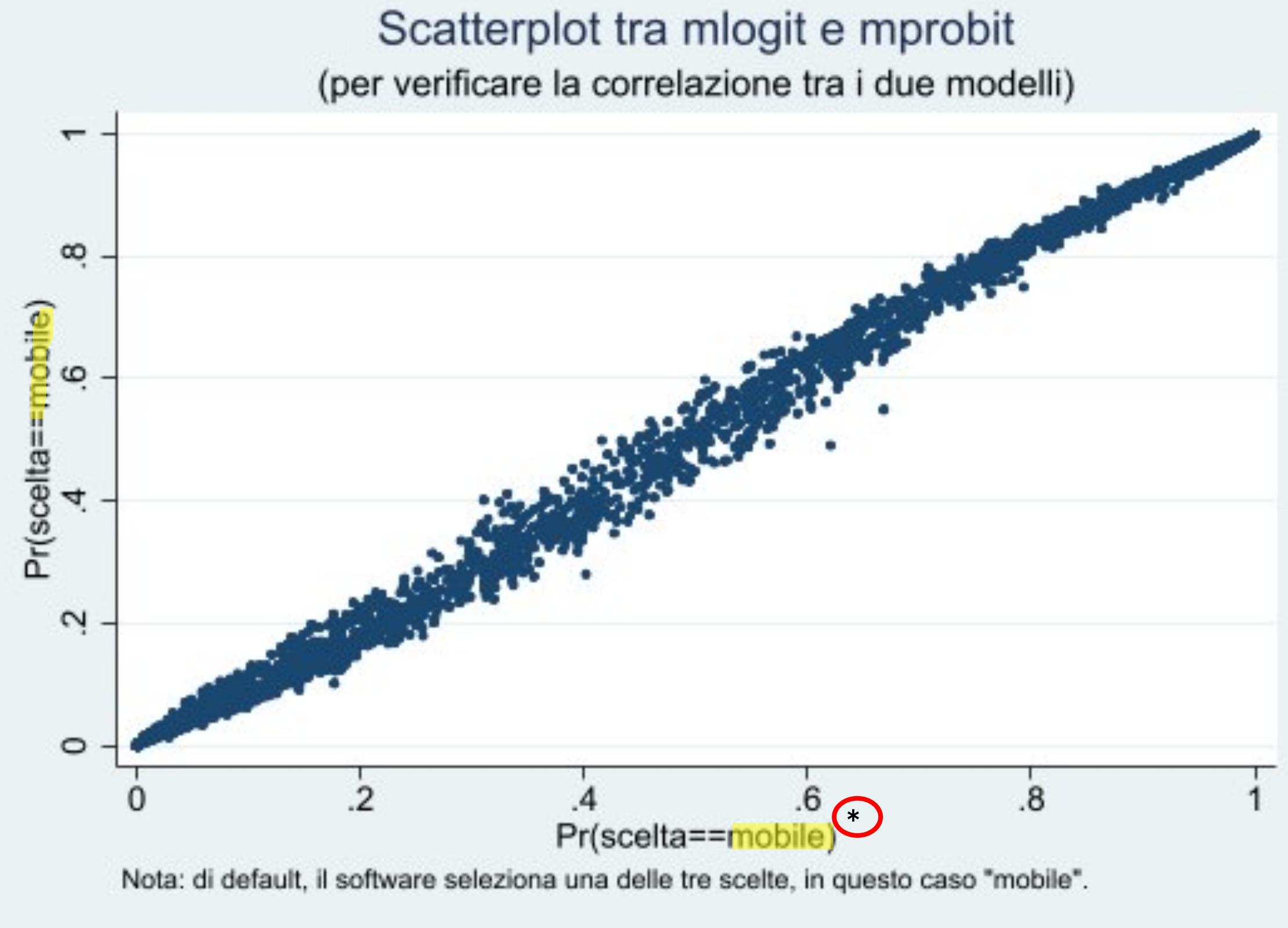


si rifiuta l'ipotesi nulla, dunque si riscontrano differenze sistematiche nei coefficienti: l'esclusione di una delle scelte (esclusione di "mobile") viola la condizione IIA

Dal momento che non vi sono variabili riferite ad una caratteristica specifica di uno dei tre canali, abbiamo optato per la costruzione di un **Multinomial probit** in cui la condizione IIA è più verosimile

# Verifica della correlazione tra i due modelli

Prima di stimare il multinomial probit verifichiamo che i risultati dei due modelli siano tra loro correlati.



\* Scatterplot simili risultano per gli altri due canali, ossia internet e punto vendita

# Multinomial probit

Confronto tra mobile e punto vendita:

scelta	Coefficient	Std. err.	z	P> z	[95% conf. interval]
mobile					
multi	-.0249045	.0762116	-0.33	0.744	-.1742765 .1244675
emails	.0487728	.0272715	1.79	0.074	-.0046783 .1022239
catalogs	-.1484873	.0956964	-1.55	0.121	-.3360487 .0390741
reso	1.06551	.2332775	4.57	0.000	.6082948 1.522726
qta	-.2209947	.0201822	-10.95	0.000	-.2605511 -.1814383
recency	.0013371	.0004608	2.90	0.004	.000434 .0022402
Mtot	.272933	.0142444	19.16	0.000	.2450145 .3008515
PVtot	-.1987083	.0126147	-15.75	0.000	-.2234327 -.1739838
Itot	.069892	.01899	3.68	0.000	.0326722 .1071118
agenti	.0152495	.1767996	0.09	0.931	-.3312715 .3617704
age	.004211	.0025009	1.68	0.092	-.0006907 .0091128
genere	.1046953	.0813129	1.29	0.198	-.0546751 .2640658
_cons	-.1187335	.3218353	-0.37	0.712	-.7495191 .512052

Notiamo che le variabili indipendenti multi, emails, catalogs, agenti, age e genere non sono significative in quanto hanno un  $p - value > 0,05$ .

Tutte le altre, invece, risultano significative al 5% quindi, possiamo ad esempio affermare che:

- se il cliente effettua un reso, a parità di tutte le altre variabili, la probabilità di acquistare tramite mobile rispetto a punto vendita aumenta;
- al crescere della quantità acquistata, a parità di tutto il resto, diminuisce la probabilità di scegliere mobile rispetto a punto vendita.

# Confronto tra internet e punto vendita:

scelta	Coefficient	Std. err.	z	P> z	[95% conf. interval]
<b>internet</b>					
multi	.2176838	.1122183	1.94	0.052	-.0022599 .4376276
emails	.2837353	.0281654	10.07	0.000	.2285321 .3389384
catalogs	-.013803	.1342555	-0.10	0.918	-.2769389 .2493328
reso	.9393753	.3017737	3.11	0.002	.3479097 1.530841
qta	-.1967947	.0304633	-6.46	0.000	-.2565018 -.1370877
recency	-.0018857	.0006338	-2.98	0.003	-.0031279 -.0006435
Mtot	.1319176	.0197324	6.69	0.000	.0932428 .1705924
PVtot	-.1753006	.0202439	-8.66	0.000	-.2149779 -.1356233
Itot	.2670725	.0223234	11.96	0.000	.2233194 .3108256
agenti	.5930386	.245375	2.42	0.016	.1121125 1.073965
age	.0047101	.0036122	1.30	0.192	-.0023697 .0117898
genere	.1233711	.107738	1.15	0.252	-.0877914 .3345337
_cons	-2.150214	.46458	-4.63	0.000	-3.060774 -1.239654

Le variabili indipendenti catalogs, age e genere hanno un  $p - value > 0,05$ , dunque non sono significative.

Tutte le altre sono **significative** e in particolare osserviamo che:

- all'aumentare del numero medio di emails che il cliente riceve, a parità di tutto il resto, la probabilità che faccia i suoi acquisti su internet piuttosto che nel punto vendita aumenta;
- se il cliente viene reclutato da degli agenti, a parità del resto, la probabilità che acquisti su internet piuttosto che nel punto vendita subisce un incremento.

# Implicazioni strategiche



Problema: il canale di vendita **internet** risulta  
il più debole



Obiettivo: incentivare i consumatori  
**all'acquisto online** e **acquisire quote di mercato**

Tra le iniziative di direct marketing che Books by Mail mette in atto (catalogo, email e agenti) risultano significative rispetto alla scelta di canale solo le **e-mail** e il **“reclutamento”** da parte degli agenti.

- **CATALOGHI:** 

Dunque consigliamo di evitare l'investimento su cataloghi cartacei poiché quest'ultimi non incentivano l'acquisto tramite il canale internet. Piuttosto proponiamo l'invio di **cataloghi digitali** che siano più sostenibili ed ecologici.



## → E-mail



01

### Personalizzare i contenuti pubblicitari

- Quiz sui libri preferiti dai clienti
- Contenuti riferiti al gusto letterario degli acquirenti

02

### Creare contenuti che spingano all'azione sul sito web

- Invio di coupon con sconti utilizzabili per gli acquisti via internet

03

### Aumentare il numero di e-mail inviate ai clienti

! L'effetto delle e-mail sulla scelta del canale internet non è esponenziale

## → Punto vendita



01

### Incentivare chi acquista nel punto vendita a compiere acquisti online

- Consegnare nel punto vendita dei buoni sconto da utilizzare al prossimo acquisto online

01

## Sistemi di CRM e target marketing mirato

- Fornire promo ad hoc per l'acquisto online ai clienti che di solito acquistano in negozio

01

## Ridurre i resi sugli acquisti via internet

- Buzz marketing
- Mettere a disposizione l'anteprima di alcuni libri sulla pagina Web

02

## Eventi in collaborazione con importanti case editrici

- Organizzare presentazioni di libri e firmacopie da parte degli autori
- Offrire scontistiche per l'acquisto online a chi partecipa a questi eventi

Agenti ←



02

## Digital marketing e collaborazioni

- Co-branding nella forma del marketing congiunto sui Social media
- Fornire scontistiche ai clienti delle imprese collaboratrici se acquistano da Books by mail



Sito Web ←  
e Social Media



# Limiti della metodologia

## Dati Panel

Abbiamo trattato  
ogni riga come  
un singolo  
consumatore

## Dati anomali in “age”

Abbiamo rilevato  
205 dati anomali in  
“age” e li abbiamo  
trattati come dati  
mancanti,  
sostituendoli con la  
media

## Utilizzo Multinomial probit

Il modello impiegato  
non permette di  
cogliere l'intensità  
dell'impatto dei  
predittori sulla  
variabile dipendente

# CONCLUSIONI

## La scelta di canale

Dall'analisi prima descrittiva poi predittiva del dataset, abbiamo individuato il canale più utilizzato (punto vendita), il più debole (internet), abbiamo riscontrato una certa avversione alla multicanalità e abbiamo captato le iniziative promozionali poco significative per la scelta di canale.

Di conseguenza abbiamo profilato una serie di proposte di direct marketing ma anche di digital marketing (servizio che l'impresa potrebbe esternalizzare) che mirano sia ad obiettivi di quota di mercato che al potenziamento del canale e-commerce.





**GRAZIE PER L'ATTENZIONE**