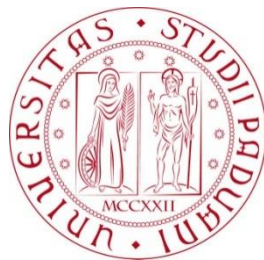


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche Aziendali



**Caso di studio attraverso le *Vignettes*:
scelta di acquisto di uno smartphone**

Relatore Prof. Omar Paccagnella
Dipartimento di Scienze Statistiche

Laureanda: Giulia Andreella
Matricola N. 1093649

Anno Accademico 2016/2017

*Alla parte di me che ha saputo
resistere alle difficoltà e non si è arresa.*

Indice

Introduzione	9
Capitolo 1. Variabili influenzanti	11
1.1 Le variabili scelte	13
1.1.1 Display	13
1.1.2 Memoria interna	13
1.1.3 Prezzo.....	14
1.1.4 Genere del personaggio della <i>vignette</i>	14
1.1.5 Perché non la marca?	15
Capitolo 2. Metodologia	17
2.1 <i>Vignettes Studies</i>.....	17
2.1.1 Cosa sono le <i>Vignettes</i>	18
2.1.2 La selezione di <i>Vignettes</i> per ogni questionario	19
2.2 Modelli	21
2.2.1 <i>Ordered Probit</i>	21
2.2.1.1 Descrizione	21
2.2.1.2 Stima.....	22
2.2.2 Multilivello	23
2.2.2.1 Descrizione	23
2.2.2.2 Il Coefficiente di Correlazione Intraclasse (ICC)	24
2.2.2.3 Notazioni del modello multilivello.....	25
2.2.2.4 Random Intercept Model	26
2.2.2.5 Random Slope Model.....	27
2.2.2.6 Stima.....	28
Capitolo 3. Creazione Questionario	31
3.1 Struttura.....	31
3.2 Somministrazione.....	33

Capitolo 4. Analisi dei dati	35
4.1 Codifica delle variabili	35
4.2 Analisi esplorative	37
4.2.1 Analisi univariate.....	37
4.2.1.1 “Risposta”	37
4.2.1.2 “Scelta”	38
4.2.1.3 “Motivo”	39
4.2.2 Analisi bivariate.....	39
4.2.2.1 “Risposta” vs “Posizione”	39
4.2.2.2 “Risposta” vs “Set”	40
4.2.2.3 “Risposta” vs “Sesso”	41
4.2.2.4 “Risposta” vs “Prezzo”	41
4.2.2.5 “Risposta” vs “Display”	42
4.2.2.6 “Risposta” vs “GB”	43
4.2.2.7 “Risposta” vs “Scelta”	44
4.2.2.8 “Risposta” vs “Motivo”	45
4.2.2.9 “Risposta” vs “Genere”	45
4.2.3 Analisi trivariate.....	46
4.2.3.1 “Risposta” vs “Genere” se “Sesso” =	46
4.2.3.2 “Risposta” vs “Posizione” se “Set”=C	48
4.3 Modello <i>Ordered Probit</i>	49
4.4 Analisi Multilivello	53
Capitolo 5. Conclusioni	57
Appendice	61
A.1 <i>Vignettes</i> del questionario A	61
A.2 <i>Vignettes</i> del questionario B	62
A.3 <i>Vignettes</i> del questionario C	63
A.4 <i>Vignettes</i> del questionario D	65

A.5 <i>Vignettes</i> del questionario E	66
A.6 <i>Vignettes</i> del questionario F	67
A.7 Parte di questionario comune.....	69
Riferimenti Bibliografici	71
Riferimenti Sitografici.....	73
Ringraziamenti	75

Introduzione

I telefoni cellulari al giorno d'oggi hanno ceduto il passo agli smartphone.

I primi permettono di telefonare in mobilità, senza essere ancorati ad una rete fissa, i secondi, il cui nome deriva dall'unione dei due termini inglesi *smart* - 'intelligente' - e *phone* - 'telefono', sono apparecchi che uniscono in sé le funzioni di computer palmare (computer che sta sul 'palmo' di una mano, dotato di uno schermo tattile) e di telefono cellulare, con i quali si può anche navigare in Internet, spedire e-mail e svolgere tutte quelle attività che prima erano demandate ad un computer, ma non solo: ci si orienta finché si viaggia, si può ascoltare musica, vedere film, realizzare foto e/o video, avere un'agenda e gestire dati personali, giocare con videogiochi o, banalmente, guardare l'ora, avere una sveglia o una calcolatrice o bussola o una pila! Inoltre, ed è stato questo a rivoluzionare il mercato dei cellulari, gli smartphone possono essere personalizzati con nuove funzioni e programmi (applicazioni, da cui il termine comune "app") da scaricare dai *marketplace* dedicati.

Attualmente, soprattutto per le persone meno anziane, avere uno smartphone è normale. Ma quali caratteristiche deve avere uno smartphone oggi? Quali sono le variabili che ne influenzano la scelta d'acquisto?

L'obiettivo principale di questa tesi consiste quindi nell'investigare l'importanza, per mezzo della tecnica delle *experimental vignettes*, che un giovane dà alle caratteristiche d'acquisto proposte, qualora dovesse scegliere di acquistare un nuovo smartphone. Da questo obiettivo, in base alle variabili analizzate, si cercherà anche di:

- ❖ esaminare qual è la caratteristica, se ce n'è solo una, che influenza maggiormente la scelta di acquisto;
- ❖ valutare se ci sono o meno differenze a livello di genere rispetto alla valutazione delle *vignettes*;
- ❖ capire l'importanza della marca nel processo di acquisto, poiché, in particolare per chi predilige Apple, la scelta del modello e delle varianti è piuttosto limitata, cosa invece decisamente più ampia per altri marchi leader del mercato.

Per poter capire i parametri di scelta di acquisto, è stato creato un questionario *ad hoc* con la tecnica delle *vignettes* ed è stato successivamente somministrato ad un campione circoscritto ed omogeneo, ovvero a 240 studenti, compresi tra i 19 e i 25

anni, del Dipartimento di Scienze Statistiche dell'Università degli studi di Padova. I dati raccolti per mezzo di tale questionario sono stati quindi analizzati tramite il modello *Ordered Probit* ed il modello multilivello.

La struttura della tesi è sviluppata in 5 capitoli:

- ❖ Il **primo capitolo** giustifica le variabili scelte per la ricerca;
- ❖ Il **secondo capitolo** sviluppa gli aspetti teorici delle *vignettes* e dei modelli statistici utilizzati;
- ❖ Nel **terzo capitolo** viene spiegata nel dettaglio la costruzione del questionario, la scelta del campione la raccolta dei dati;
- ❖ Nel **quarto capitolo** viene presentata l'analisi dei dati;
- ❖ Nel **quinto capitolo** verranno tratte le conclusioni rispetto al lavoro svolto.

Tutte le analisi sono state svolte utilizzando il software statistico STATA.

Capitolo 1.

Variabili influenzanti

Sono moltissime le attività che oggi si possono compiere con uno smartphone. Se fino alla fine degli anni '90 erano più gli utenti che utilizzavano il cellulare a fini lavorativi che per fini personali, dall'esordio di Apple con iPhone del 2007 le cose sono radicalmente cambiate.

Da una ricerca compiuta nel 2006 su 286 impiegati di aziende sparse per il mondo (Sang Hyun Kim, 2008) risulta che l'adozione e l'accettazione di utilizzare *mobile phones* a fini lavorativi dipende dalla facilità d'uso. Una ricerca del 2009 (Verkasalo et al., pubblicata nel 2010) su un campione di 579 utilizzatori Finlandesi di smartphone e non (principalmente utilizzatori di Nokia) evidenzia come l'uso di app base sia accettato se ritenuto utile, mentre l'uso di app avanzate sia meno accettato poiché esiste una "barriera tecnologica" da superare; la possibilità di scegliere cosa installare viene visto come un fattore positivo.

Il primo iPhone aveva dell'incredibile: spariva la tastiera tattile, sostituita da quella presente sul display multi-touch del cellulare, di dimensione 3,5" (pollici)¹.

Worldwide Smartphone Sales to End Users by Vendor (Thousands of Units)										
Vendor	2007 YR	2008 YR	2009 YR	2010 YR	2011 YR	2012 YR	2013 YR	2014 YR	2015 YR	2016 YR
Samsung	2.325	4.748	5.872	25.435	87.126	205.767	299.795	307.597	320.220	306.447
Apple	3.303	11.418	24.890	46.598	89.263	130.133	150.786	191.426	225.851	216.064
Huawei	-	-	33	386	15.636	27.169	46.609	68.081	104.095	132.825
OPPO	-	-	-	-	58	3.072	10.983	20.501	39.489	85.300
Vivo	-	-	-	-	20	2.480	11.163	20.416	35.291	72.409
Others	33.852	27.283	21.320	46.574	96.577	201.029	375.167	585.797	672.492	666.849
Nokia	60.465	60.920	70.875	102.167	84.614	39.263	29.609	15.811	-	-
Research In Motion*	11.768	23.149	34.347	49.652	51.542	34.210	18.606	7.911	4.398	2.109
HTC	3.719	6.536	10.812	24.631	42.997	32.094	21.688	23.109	18.577	10.148
Sharp	6.885	5.234	4.228	3.404	5.057	4.890	3.945	4.091	3.488	3.207
Total	122.316	139.288	172.376	298.847	472.891	680.108	968.351	1.244.740	1.423.900	1.495.358

Note: Nokia, *RIM (BlackBerry since 2013), HTC, and Sharp were the leading vendors in 2007
Source: Gartner (June, 2017)

Tabella 1.1: Evoluzione delle vendite smartphone dal 2007 al 2016 (Fonte: sito internet, per l'indirizzo rifarsi alla nota 2 a piè pagina).

¹ Nel 1994 veniva lanciato Simon Personal Communicator, il primo telefono intelligente; fu prodotto da Mitsubishi per conto di IBM. Simon pesava mezzo chilo e costava 900 dollari. Era un'agenda rivolta al mondo business con una serie di servizi utili come orologio, calendario, calcolatrice, un client di posta e fax e un blocco note per appunti da prendere con il pennino sullo schermo touch screen monocromatico da 4.7". Durò sei mesi poi non fu più prodotto (<http://www.dday.it/redazione/14042/ventanni-di-smartphone-ibm-simon-e-stato-il-primo>). La stessa sorte toccò al primo smartphone capacitivo (con lo schermo sensibile al tocco delle dita) ad opera di LG in collaborazione con Prada, alla fine del 2006.

Come si può osservare dalla *Tabella 1.1*, nel 2007 erano Nokia e BlakBerry i marchi leader, ma l'ingresso di Apple con le sue novità ha cambiato il mercato. L'idea di utilizzare le icone per passare da un servizio ad un altro e la possibilità di personalizzazione del *mobile* in modo semplice, tramite l'acquisto di applicazioni dedicate per dispositivi mobili sui *marketplace*, ha aperto il mercato delle app (creando quella che viene chiamata App's economy)², oltre che incrementare in modo vertiginoso le vendite di cellulari intelligenti.

La possibilità di installare applicazioni nel proprio smartphone ha portato anche il ridimensionamento negli acquisti di PC. Chris Jones, analista di Canals, ha evidenziato come nel 2011 ci sia stato il sorpasso nelle vendite di smartphone (e *pads*) rispetto ai personal computer (PC)³, mentre un'analisi del 2016 di StatCounter, società irlandese specializzata in web analytics, trova che il numero di pagine visitate con i dispositivi mobili ha superato quelle viste con i pc desktop⁴ (*Figura 1.1*).

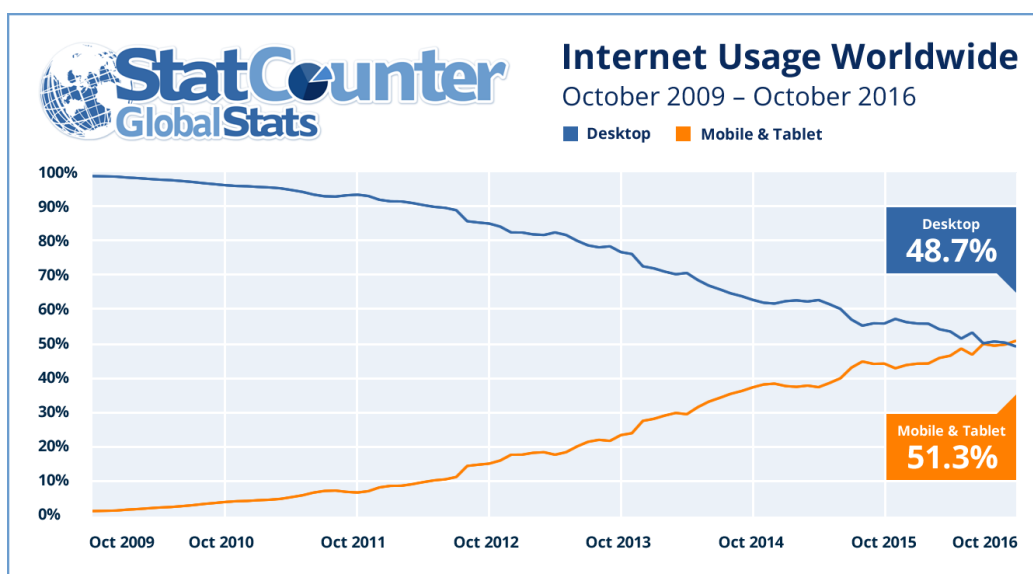


Figura 1.1: Uso di Internet nel mondo con PC e con dispositivi mobili.

² http://www.silicon.it/mobility/i-dieci-anni-delliphone-nascita-e-rivoluzione-del-mercato-smartphone-109885?inf_by=598d6602671db8a6638b48de

³ <https://www.canalys.com/newsroom/smart-phones-overtake-client-pcs-2011>

⁴ <http://gs.statcounter.com/press/mobile-and-tablet-internet-usage-exceeds-desktop-for-first-time-worldwide>

1.1 Le variabili scelte

Considerando le caratteristiche e le potenzialità degli smartphone, le variabili utilizzate nelle *vignettes* del questionario proposto sono:

- ❖ Display;
- ❖ Memoria interna;
- ❖ Prezzo;
- ❖ Genere del personaggio della *vignette*;

1.1.1 Display

L'uso di oggetti multimediali, navigazione, giochi e interazione sui social hanno richiesto la modifica della grandezza degli smartphone: la dimensione del display passa dai 3,5" del 2007 ai 5" ed oltre di oggi. Una ricerca svolta da Siu-Tsen Shen (2015) tra 127 giovani di sei Paesi diversi evidenzia l'importanza della grandezza del display.

Uno studio di Na e Ruifeng (2017) su 842 partecipanti, 464 donne e 378 uomini tra i 20 ed i 51 anni analizza quali possano essere i fattori critici per lo sviluppo della diffusione degli smartphone; è un'analisi multilivello che prende in considerazione nove caratteristiche, tra cui anche la dimensione dello schermo. Anche in questo caso viene evidenziata l'importanza di questa caratteristica.

Per il questionario sono stati utilizzati due valori per il display: inferiore ai 5" e superiore ai 5".

1.1.2 Memoria interna

Non basta avere uno schermo grande, partecipare nei social significa "postare" foto e video. Questo ha un costo in termini di memoria utilizzata, poiché le fotocamere diventano sempre più sofisticate e la qualità delle foto e dei video pesa in termini di byte utilizzati. Anche la possibilità di scaricare app per svolgere le azioni più disparate comporta la necessità di avere una memoria di massa capiente e mano a mano che si va avanti negli anni, la quantità di memoria che le applicazioni richiedono aumenta sempre di più.

Il primo iPhone (2007) aveva 4 GB (gigabyte) e sembrava una caratteristica esagerata in confronto per esempio ai 64 MB (megabyte) dell'HTC Touch o ai 112 MB del Samsung F700, usciti nello stesso anno dell'iPhone 1; i nuovi prodotti spaziano attualmente dai 16 ai 256 GB!

In questo lavoro verranno prese tre grandezze di memoria, che sono quelle più diffuse negli smartphone odierni: 8, 16 e 24 GB.

1.1.3 Prezzo

Una ricerca del 2014 (Daejoong, Heasun e Hyunjoo) relativa ai fattori che influenzano gli studenti universitari (il campione è costituito da 354 studenti di New York, i dati sono stati raccolti all'inizio del 2011) nell'adozione dello smartphone trova che il prezzo sia un fattore discriminante, poiché spesso in questa fascia d'età non c'è autonomia monetaria oppure il reddito ottenuto è basso; in base a ciò le risposte presuppongono un prezzo che sia considerato accettabile da parte degli intervistati. Lo studio si rivolge all'utilità percepita nell'adozione della tecnologia.

Il valore percepito è la valutazione complessiva di un individuo sull'utilità di un prodotto o di un servizio, sulla base delle percezioni di ciò che viene ricevuto e di ciò che viene dato. Il modello di adesione del valore percepito propone che l'adozione di uno smartphone da parte di un individuo sia dovuta ad un compromesso tra i vantaggi percepiti ed i costi percepiti nell'uso del cellulare.

I costi percepiti comprendono non solo il prezzo monetario effettivo di un prodotto, ma anche aspetti non monetari, come lo sforzo e il tempo. I vantaggi percepiti includono l'utilità percepita, definita come il valore totale che un individuo percepisce dall'uso di una nuova tecnologia, tra questi l'accettazione da parte di un gruppo di pari. Uno dei risultati ottenuti dal lavoro è che i costi elevati tendono a vietare l'adozione della tecnologia e che i benefici elevati sono presumibilmente una forte motivazione per l'adozione della tecnologia.

Uno studio del 2000 (Brucks, Zeithaml e Naylor), relativo all'importanza del prezzo e della marca nell'acquisto di beni durevoli, evidenzia come spesso i consumatori utilizzino queste due dimensioni come "scorciatoie" per dare scontate alcune caratteristiche qualitative (facilità d'uso, durata del prodotto, versatilità, servizio post-vendita, prestigio), senza approfondirne l'effettiva presenza.

Nel lavoro oggetto di tesi si è scelto un prezzo soglia, quale variabile da testare, di 300€, che è il prezzo che separa più o meno i dispositivi di bassa gamma da quelli di alta gamma.

1.1.4 Genere del personaggio della vignette

Per quanto riguarda il genere, ultima caratteristica considerata in questa ricerca, la decisione di inserirla nello studio è dovuta più ad una necessità metodologica che a motivi economici, tecnologici o sociali.

Nello studio di Na e Ruifeng (2017), citato per l'importanza attribuita al display, non vengono evidenziate differenze di genere rispetto l'importanza delle caratteristiche analizzate; nessuna differenza di genere si riscontra nella già citata ricerca di Daejoong, Heasun e Hyunjoo (2014) relativa ai fattori che influenzano la scelta di utilizzare o meno uno smartphone.

Uno studio specifico sull'influenza che età e genere possono avere sulla lealtà alla marca di uno smartphone (Ching-Hsuan et al., 2015), evidenzia che il genere non è una variabile influenzante, mentre lo è l'età.

Differenze di genere vengono ricercate rispetto la scelta della marca al momento del riacquisto di un cellulare (Hew et al., 2016); il risultato evidenzia una diversità rispetto l'utilità percepita e l'uso che viene fatto dello smartphone posseduto in precedenza, non tanto rispetto le sue caratteristiche.

Il metodo delle *anchoring vignettes* (King et al., 2004), la cui logica si ispira a quella delle *experimental vignettes* usate in questa tesi, ha invece evidenziato che la valutazione degli scenari da parte degli intervistati può essere influenzata dal genere della persona descritta (Kapteyn et al., 2007), la stessa evidenza si trova nel metodo *vignettes* base (Wason, Polonsky e Hymann, 2002). Può capitare che le *vignettes* vengano distribuite rispettando il genere dell'intervistato per non generare disturbo, così come i nomi ai personaggi vengano assegnati utilizzando quelli più comuni per la fascia d'età da rappresentare (Jurges e Winter, 2011).

Per il questionario somministrato sono stati scelti quattro nomi, due maschili e due femminili, appartenenti ad un'età più giovanile, ma la cosa verrà approfondita nel capitolo 3.

1.1.5 Perché non la marca?

Tra le variabili che inizialmente si pensava di inserire nelle *vignettes* c'era la marca. Le due più importanti a livello di quantità vendute sono attualmente Apple e Samsung⁵ (Tabella 1.2), ma è stato valutato il fatto che se fosse stata aggiunta anche questa variabile, il numero di combinazioni per le *vignettes* ed i questionari sarebbe levitato. Per esempio, se fossero state scelte 3 marche (Samsung, Apple e Huawei, o "altro"), il numero di *vignettes* sarebbe stato di $2 \times 2 \times 2 \times 3 \times 3 = 72$ invece delle $2 \times 2 \times 2 \times 3 = 24$ prese in considerazione. Inoltre, con tutte queste caratteristiche da considerare, vi era il forte rischio di creare degli scenari irrealistici. Si è pensato allora di aggiungerla comunque, ma esterna alla *vignette* vera e propria. Il motivo per

⁵ 27/04/2017, <http://www.idc.com/getdoc.jsp?containerId=prUS42507917>

il quale si voleva tenere una domanda inerente alla marca è dovuto al fatto che esiste la possibilità che le persone acquistino uno smartphone solo in base a questa.

Vendor	1Q17 Shipment Volume	1Q17 Market Share	1Q16 Shipment Volume	1Q16 Market Share	Year-Over-Year Change
1. Samsung	79.2	22.8%	79.2	23.8%	0.0%
2. Apple	51.6	14.9%	51.2	15.4%	0.8%
3. Huawei	34.2	9.8%	28.1	8.4%	21.7%
4. OPPO	25.6	7.4%	19.7	5.9%	29.8%
5. vivo	18.1	5.2%	14.6	4.4%	23.6%
Others	138.7	39.9%	140.0	42.1%	-1.0%
Total	347.4	100.0%	332.9	100.0%	4.3%
Source: IDC Quarterly Mobile Phone Tracker, April 27, 2017					

Tabella 1.2: Top five venditori di smartphone, volume delle spedizioni (in milioni) in tutto il mondo, quota di mercato (in percentuali), per il primo quadrimestre del 2016 e del 2017 e la crescita dal 1Q 2016 al 1Q 2017.

Capitolo 2.

Metodologia

Il lavoro si è focalizzato su dare una risposta alla domanda “Quali sono i fattori che influiscono sull’acquisto di uno smartphone?”. Per poter studiare questo fenomeno si è deciso di utilizzare l’approccio delle *experimental vignettes*, a cui si è affiancata una parte di questionario più tradizionale nella quale erano presenti delle domande più specifiche per l’intervistato. Questa metodologia prende il nome più generale di *Vignette Study* (Atzmüller e Steiner, 2010).

Per le analisi che verranno presentate nel capitolo 4 sono stati usati 2 tipi di modelli statistici: il modello *Probit* per variabili ordinali e il modello multilivello, o gerarchico, dato che ogni individuo ha fornito risposte a più *vignettes*.

2.1 Vignettes Studies

Le *Vignettes Studies* sono dei questionari che combinano il metodo tradizionale col metodo delle *vignettes* e sono stati introdotti da Atzmüller e Steiner, (2010). Questo nuovo modo di sottoporre un questionario è stato pensato per superare alcune limitazioni del metodo tradizionale come la presenza di multicollinearità delle variabili e la raccolta dei dati in modo passivo (senza uno studio sperimentale o un controllo delle variabili esplicative).

Nei *Vignettes Studies*, oltre alla prima parte, che consiste nel sottoporre agli intervistati una serie di *vignettes*, ce n’è una seconda che comprende domande dirette al rispondente per avere le caratteristiche specifiche individuali (come potrebbero essere il genere, l’età, lo stato civile, ecc...) ed in alcuni casi per un’auto-valutazione, sempre collegata alla storia delle *vignettes*. In questo caso di studio non è presente l’auto-valutazione.

Viene qui analizzato il metodo delle *vignettes*, utilizzato in parte per creare il questionario che verrà poi spiegato ed analizzato nei capitoli 3 e 4.

Per quanto riguarda la consultazione, per tutto il paragrafo 2.1 e sotto paragrafi si fa riferimento, se non viene specificato diversamente, a Wason, Polonsky e Hyman (2002), Atzmüller e Steiner, (2010), Aguinis e Bradley (2014).

2.1.1 Cosa sono le *Vignettes*

Le *vignettes* si possono definire come “breve descrizione di una persona o di una situazione sociale che contiene riferimenti precisi su quali si ritiene siano i fattori più importanti nei processi decisionali o di valutazione degli intervistati” (Alexander e Becker, 1978). Sono quindi delle brevi descrizioni di scenari che vengono osservati da un punto di vista esterno e la percezione oggettiva dei personaggi delle *vignettes* è uguale per tutti i rispondenti; all’intervistato poi viene chiesto un parere riguardo quello che ha letto.

Con le *vignettes* si tengono in considerazione le differenze nell’interpretazione delle domande da parte di intervistati diversi, ovvero si valuta l’eterogeneità nelle risposte.

La *vignette* rappresenta una storia con determinate caratteristiche che la rendono unica rispetto ad ogni altra *vignette*; inoltre alla fine di ognuna viene posta una domanda diretta all’intervistato nella quale viene chiesto, con delle scale di valutazione ordinate, quanto d’accordo è, o quanto si rivede in quella storia.

Lo studio di Aguinis e Bradley pubblicato nel 2014 ripercorre gli articoli che trattano casi di *vignettes*, dal 1994 al 2013, trovando che la tecnica si sia evoluta non solo su carta ma anche con audio e video. Inoltre gli autori sottolineano come le *experimental vignettes* permettano ai ricercatori di includere negli studi solo quei fattori rilevanti da analizzare, escludendone altri che potrebbero creare disturbo.

La tecnica delle *vignettes* è spesso utilizzata per rendere confrontabili gruppi di osservazioni tra loro diversi (individui, Paesi, ecc...).

Lo studio di Wason, Polonsky e Hyman (2002) analizza molta letteratura e sostiene che, in ambito di marketing, gli esperimenti basati sulle *vignettes* siano spesso più adatti rispetto a quelli basati su domande dirette poiché:

- ❖ Forniscono un maggiore realismo offrendo una serie di fattori situazionali o contestuali simili a situazioni reali di decisione;
- ❖ Forniscono aspetti standardizzati a tutti gli intervistati; ciò aumenta l’efficacia del metodo, l’affidabilità della misurazione e la possibilità di poterlo riproporre facilmente;
- ❖ Il metodo permette di concentrare l’attenzione dei rispondenti sulle caratteristiche individuate per poter rispondere alla domanda oggetto di studio;
- ❖ Rafforzano il coinvolgimento dei partecipanti grazie alla drammatizzazione dei problemi presentati.

L'articolo individua anche una serie di limitazioni nell'uso delle *vignettes*:

- ❖ Uno dei problemi di questo metodo è di non riuscire a capire se la percezione dei rispondenti corrisponda a quella voluta dal ricercatore e se sia omogenea all'interno del campione scelto. Può succedere che gli individui possano comprendere le domande che vengono loro poste in maniera molto diversa, ed avendo delle soglie di valutazione dissimili (differenze a livello culturale, sociale, ecc.) l'interpretazione che daranno ad una stessa risposta potrebbe essere diversa;
- ❖ I criteri utilizzati nelle *vignettes* potrebbero mascherare il vero pensiero degli intervistati quando questo non è rappresentato nei *sets* di caratteristiche studiate.

Gli autori suggeriscono le attenzioni da avere nella preparazione delle *vignettes*, tra queste:

- ❖ Assicurarsi che tutte le variabili rilevanti per il caso di studio siano presenti nelle *vignettes*;
- ❖ Utilizzare un numero adeguato di *vignettes*;
- ❖ Decidere se scegliere il campione in base alla ricerca o adattare le *vignettes* al campione;
- ❖ Rendere le *vignettes* credibili;
- ❖ Rendere evidente la variabile che verrà poi utilizzata come variabile risposta.

2.1.2 La selezione di *Vignettes* per ogni questionario

Il metodo delle *vignettes* richiede che "diverse versioni della stessa *vignette* siano allocate in modo casuale a diversi intervistati" (Alexander e Becker 1978).

Ancora prima di assemblare la storia per ogni *vignette*, è necessario decidere quali variabili verranno utilizzate e quanti livelli presenterà ognuna di esse. Si supponga di avere k variabili V_1, \dots, V_K con v_i livelli ($i = 1, \dots, K$), allora il numero totale di *vignettes* (M) si otterrà dalla combinazione di tutti i livelli delle variabili V_k , ovvero $M = v_1 \times v_2 \times \dots \times v_K$.

A scopo esemplificativo, verrà riportato un esempio, proposto da Atzmüller e Steiner, (2010) sulla combinazione di 3 variabili, che verranno chiamate A, B, C ,

ognuna con 2 livelli ($a_0, a_1; b_0, b_1; c_0, c_1$). Guardando la *Tabella 2.1* si nota che risulteranno $2 \times 2 \times 2 = 8$ vignettes.

	$C = c_0$		$C = c_1$	
	$B = b_0$	$B = b_1$	$B = b_0$	$B = b_1$
$A = a_0$	$a_0 b_0 c_0$	$a_0 b_1 c_0$	$a_0 b_0 c_1$	$a_0 b_1 c_1$
$A = a_1$	$a_1 b_0 c_0$	$a_1 b_1 c_0$	$a_1 b_0 c_1$	$a_1 b_1 c_1$

Tabella 2.1: Combinazione di $2 \times 2 \times 2$ livelli di variabili per un totale di 8 vignettes (Fonte: Atzmüller C. & Steiner P., (2010))

Come si può intuire, più livelli ha una variabile e/o più sono le variabili coinvolte, più il numero di combinazioni delle *vignettes* aumenta velocemente. È necessario in questi casi creare dei *sets* di *vignettes*, ovvero delle partizioni, ed in alcuni casi anche tenere solo una parte del numero totale delle *vignettes*.

Per selezionare solo un **sottogruppo** di *vignettes* ci sono 2 metodi: una selezione casuale (*Randomized Vignette Selection*) delle *vignettes*, oppure un “disegno fattoriale frazionato” (*Fractional Factorial Design*), ovvero viene presa una frazione di *vignettes*, scelta in base al numero di variabili selezionate ed ai loro livelli.

Per decidere le **partizioni** bisogna tenere conto delle variabili scelte, del numero di livelli e del numero di *vignettes* che si vogliono mettere nel questionario per rispondente. Il numero di *vignettes* viene equamente suddiviso in s parti (che sta per *set*) così che si abbiano $M_s = \frac{M}{s}$ *sets*. Anche per scegliere i *sets* si utilizza uno dei due metodi proposti nel caso della selezione del sottogruppo di *vignettes*.

Nel caso di studio analizzato verranno create delle partizioni, in tutto 6, ognuna con 4 *vignettes*, ed il modo in cui sono state selezionate è avvenuto cercando di equilibrare di volta in volta le storie perché non risultassero troppo simili tra loro in ogni gruppo di 4 *vignettes* (quindi è stata fatta una prima selezione casuale con aggiustamenti *ad hoc* in seguito).

Il capitolo 3 sarà dedicato alla descrizione di come si è strutturato il questionario realizzato per questo studio.

2.2 Modelli

Si analizzano di seguito i due modelli utilizzati per l'analisi dei dati raccolti.

2.2.1 Ordered Probit

2.2.1.1 Descrizione

Per questo paragrafo si è fatto riferimento ai contenuti di Cappuccio e Orsi, (2005).

I modelli che hanno come variabile dipendente una variabile discreta ordinale, ovvero una variabile che presenta tre o più categorie con un ordinamento logico (crescente o decrescente), vengono solitamente stimati con una generalizzazione di due tipologie di modelli a risposta binaria: il modello di regressione logistica (o Logit), e il modello *Probit*, che è quello che verrà usato per queste analisi.

Nel questionario analizzato la variabile risposta è il grado di accordo con una domanda che presenta 5 categorie, codificate da 1, "Per niente d'accordo", a 5, "Completamente d'accordo"; è quindi ragionevole pensarla come variabile ordinata.

La variabile dipendente osservata y_i si ipotizza sia collegata ad una variabile latente continua y_i^* che dipende linearmente dalle covariate x_i :

$$y_i^* = x_i^T \beta + \varepsilon_i$$

dove β è un vettore di parametri da stimare ed ε_i è un termine d'errore i.i.d. con funzione di ripartizione $F(\cdot)$.

La variabile ordinale osservabile y_i assume invece i suoi valori in base allo schema seguente, che qui a scopo esemplificativo riporta 3 categorie:

$$y_i = \begin{cases} 1 & \text{se } y_i^* \leq \pi_1 \\ 2 & \text{se } \pi_1 < y_i^* \leq \pi_2 \\ 3 & \text{se } y_i^* > \pi_2 \end{cases}$$

Le quantità π_1 e π_2 sono delle soglie che verranno stimate assieme al vettore dei coefficienti β . Solitamente si avranno tante soglie quante sono le modalità della variabile y_i , meno una.

La distribuzione di probabilità della variabile osservabile y_i è la seguente:

$$\begin{cases} P(y_i = 1|x_i) = P(y_i^* \leq \pi_1|x_i) = P(x_i^T \beta + \varepsilon_i \leq \pi_1) = P(\varepsilon_i \leq \pi_1 - x_i^T \beta) = F(\pi_1 - x_i^T \beta) \\ P(y_i = 2|x_i) = P(\pi_1 < y_i^* \leq \pi_2|x_i) = P(\pi_1 - x_i^T \beta < \varepsilon_i \leq \pi_2 - x_i^T \beta) = \\ \quad = P(\varepsilon_i \leq \pi_2 - x_i^T \beta) - P(\varepsilon_i \leq \pi_1 - x_i^T \beta) = F(\pi_2 - x_i^T \beta) - F(\pi_1 - x_i^T \beta) \\ P(y_i = 3|x_i) = P(y_i^* > \pi_2|x_i) = P(\varepsilon_i > \pi_2 - x_i^T \beta) = 1 - F(\pi_2 - x_i^T \beta) \end{cases}$$

Nel caso di un modello *Ordered Probit* si assume $\varepsilon_i \sim N(0, \sigma^2)$ ma per identificare e stimare il modello c'è bisogno che gli errori si distribuiscano come una Normale standard, ovvero che $\sigma^2 = 1$, e definire un valore ad una soglia, per esempio $\pi_1 = 0$. La funzione di ripartizione degli ε_i risulta una normale standard:

$$F(x) = \Phi(x) = \int_{-\infty}^x \phi(t) dt$$

e quindi

$$\begin{cases} P(y_i = 1|x_i) = \Phi(-x_i^T \beta) \\ P(y_i = 2|x_i) = \Phi(\pi_2 - x_i^T \beta) - \Phi(-x_i^T \beta) \\ P(y_i = 3|x_i) = 1 - \Phi(\pi_2 - x_i^T \beta) \end{cases}$$

In questo caso, come detto poco sopra, il modello è identificato. Il caso generale, ove il modello non è identificato è il seguente:

$$y_i^* = \alpha_i + x_i^T \beta + \varepsilon_i \quad \text{con } \varepsilon_i \sim N(0, \sigma^2) \text{ i.i.d.}$$

$$y_i = \begin{cases} 1 \text{ se } y_i^* \leq \pi_1 \\ 2 \text{ se } \pi_1 < y_i^* \leq \pi_2 \\ 3 \text{ se } y_i^* > \pi_2 \end{cases}$$

dove α_i è il vettore di costanti, e x_i vettore di covariate, senza le costanti.

Per l'osservazione i -esima si ha quindi

$$P(y_i = 1|x_i) = P(\alpha + x_i^T \beta + \varepsilon_i \leq \pi_1 | x_i) = \Phi \left[\frac{\pi_1 - \alpha}{\sigma} - x_i^T \left(\frac{\beta}{\sigma} \right) \right]$$

Con questa formulazione non risulta possibile avere quattro stime separate per i parametri $(\alpha, \beta, \pi_1, \sigma)$ poiché combinazioni diverse delle stime di questi parametri possono generare lo stesso valore della log-verosimiglianza. Per avere dei valori diversi e quindi identificare il modello bisogna imporre dei vincoli, come quello riportato prima della varianza unitaria e una soglia pari a zero, oppure per esempio imporre sempre $\sigma^2 = 1$, ponendo a 0, invece di una soglia, la costante α .

2.2.1.2 Stima

La stima dei parametri può essere fatta tramite il metodo della Massima Verosimiglianza. Seguendo il ragionamento iniziato nel paragrafo precedente e quindi mantenendo tre categorie per la variabile risposta, la log-verosimiglianza che risulta è la seguente:

$$l(\beta, \pi) = \sum_{y_i=1} \log \Phi(-x_i^T \beta) + \sum_{y_i=2} \log[\Phi(\pi - x_i^T \beta) - \Phi(-x_i^T \beta)] + \sum_{y_i=3} \log[1 - \Phi(\pi - x_i^T \beta)]$$

dove si ricorda che Φ è la funzione di densità di una Normale $N \sim (0,1)$ e gli y_i delle sommatorie si intendono per $i = 1, \dots, n$, ovvero le somme sono calcolate per tutte le osservazioni, con $y_i = m, m = 1,2,3$.

Essendo questo modello non lineare, le stime dei coefficienti β non indicano gli effetti marginali dei rispettivi regressori x sulle probabilità degli y_i , ma può indicare la direzione, data dal segno del coefficiente, dell'effetto del regressore sulla variabile risposta. Per questi modelli si può quindi intendere che a parità di condizioni, avere una certa caratteristica (nel caso di esplicative categoriali) porta ad una maggiore o minore probabilità di essere d'accordo o meno rispetto ad essere del "gruppo base", ovvero del gruppo preso a riferimento.

Nel caso in cui la variabile risposta abbia un numero di categorie pari a $j = 1,2,3, \dots, J$, semplicemente viene aumentato il numero di soglie π_j in base alle categorie e le soglie sono in ordine crescente $-\infty = \pi_0 < \pi_1 < \pi_2 < \dots < \pi_j = +\infty$.

2.2.2 Multilivello

Per analizzare i dati si è fatto ricorso, oltre ai modelli *Ordered Probit*, ai modelli di regressione multilivello. Per quanto riguarda la consultazione per tutto il paragrafo 2.2.2 e sotto paragrafi si fa riferimento ai volumi di Kreft e De Leeuw (1998), Snijders e Bosker (1999), Hox (2002), Raudenbush e Bryk (2002), Bickel (2007), De Leeuw e Meijer (2008), Goldstein (2010).

2.2.2.1 Descrizione

I modelli multilivello, detti anche gerarchici, a coefficienti casuali, ad effetti casuali o ad effetti misti, possono essere visti come un'estensione dei modelli di regressione lineare, dove presentano, differentemente dagli ultimi, un termine d'errore per ciascun livello. Questi modelli sfruttano il fatto che i dati presentano una struttura a livelli, gerarchica appunto, dove le osservazioni a livello inferiore sono annidate a quelle del livello superiore. Si possono creare più livelli ma a scopo esemplificativo ne verranno tenuti solo 2, poiché così verranno utilizzati per l'analisi dei dati raccolti. Viene riportata la tabella esemplificativa (*Tabella 2.2*) con l'elenco dei termini per descrivere le unità in ognuno dei due livelli:

Unità di 2° livello	Unità di 1° livello
Unità a livello macro	Unità a livello micro
Unità Macro	Unità Micro
Unità Between	Unità Within
Gruppi	Individui
Clusters	Unità elementari
Unità primarie	Unità secondarie

Tabella 2.2: Termini per descrivere le unità nel caso di modello multilivello con due livelli.
Fonte: Snijders T & Bosker R. (1999)

L'appartenenza dell'unità di primo livello ad un gruppo può dipendere da varie caratteristiche, che possono essere individuali, sociali, scandite da regole, casuali. Indipendentemente da queste caratteristiche, i gruppi si differenziano gli uni dagli altri per due motivi principali: innanzitutto i comportamenti individuali vengono influenzati dal gruppo di appartenenza, e in secondo luogo le specifiche del gruppo vengono condizionate dagli individui che ne fanno parte. Lo scopo di questi modelli è studiare se è presente omogeneità tra osservazioni all'interno dello stesso gruppo ed eterogeneità tra un gruppo e l'altro, facendo così decadere l'ipotesi di indipendenza delle osservazioni.

Nel caso di studio qui analizzato si vuole valutare se c'è differenza tra gli individui (unità di secondo livello) che hanno valutato ciascuno 4 *vignettes* (unità di primo livello) con una scala che va da 1 a 5 (variabile risposta), sull'essere in accordo o meno con una determinata scelta effettuata dai personaggi delle *vignettes*.

Nei prossimi sotto paragrafi verranno analizzati nel dettaglio 2 tipi di modello multilivello: *Random Intercept Model* e *Random Slope Model*. Il *Random Slope Model* si differenzia dal primo per l'aggiunta di una variazione casuale tra i gruppi non solo dell'intercetta (come accade nell'*Intercept*) ma anche dei coefficienti di regressione.

Questi due modelli sono stati utilizzati per le analisi del caso di studio.

2.2.2.2 Il Coefficiente di Correlazione Intraclasse (ICC)

Il Coefficiente di Correlazione Intraclasse esprime il grado di correlazione che intercorre tra individui appartenenti allo stesso gruppo ed è generalmente indicato con la lettera greca ρ . Indica quindi se la scelta di studiare i dati con un modello gerarchico fa la differenza o meno (Bickel, 2007).

L'ICC misura la proporzione di variabilità totale (che si può scomporre in varianza tra i gruppi, o varianza *between*, e varianza entro i gruppi, detta anche varianza

within) spiegata dalla variabilità tra i gruppi, ed è una misura di *omogeneità di gruppo* (Kreft, De Leeuw, 1998). Essendo una proporzione, ρ varia tra 0 ed 1 e più si avvicina ad 1, più la struttura gerarchica può avere effetto sulla variabile risposta, mentre più si avvicina allo 0, più è probabile che basti implementare un modello di regressione tradizionale per studiare il fenomeno, poiché il raggruppamento in clusters non determina effetti significativi.

Questo coefficiente indica anche quali variabili di secondo livello servano di più per spiegare la variabilità tra i gruppi: se viene aggiunta una variabile “macro” e questa aiuta a spiegare le differenze tra i gruppi, allora la variabilità between diminuisce, definendo una diminuzione dell’ICC (Kreft e De Leeuw, 1998).

Questo coefficiente è costante per il *Random Intercept Model* mentre non lo è per lo *Slope Model*, ed è il motivo per cui con il secondo modello non risulta appropriato calcolarlo.

2.2.2.3 Notazioni del modello multilivello

Brevemente verranno elencate le notazioni utilizzate nel modello multilivello proposto:

- ❖ $j = 1, 2, \dots, J$ è l’indice delle unità di 2° livello
- ❖ $i = 1, 2, \dots, n_j$ è l’indice delle unità di 1° livello all’interno del gruppo j-esimo
- ❖ Y_{ij} è la variabile risposta di 1° livello (riferita all’unità i-esima del gruppo j-esimo)
- ❖ X_{ij} è la variabile esplicativa di 1° livello (riferita all’unità i-esima del gruppo j-esimo)
- ❖ W_j è la variabile esplicativa al 2° livello (riferita al gruppo j-esimo) e può essere di 2 tipi:
 - ◆ Z_j è una variabile che esprime una caratteristica specifica del gruppo j-esimo
 - ◆ \bar{X}_j è la media di gruppo, ovvero il valore medio del gruppo j-esimo di tutte le caratteristiche X_{ij}
- ❖ ε_{ij} è l’effetto non osservato specifico di 1° livello
- ❖ U_j è l’effetto non osservato specifico di 2° livello

Inoltre per rendere più semplice la notazione, verrà considerata:

- ❖ Una sola variabile risposta Y_{ij}
- ❖ Una sola variabile esplicativa X_{ij}

❖ Una sola variabile esplicativa W_j

Nei prossimi sotto paragrafi verranno analizzati i due principali modelli multilivello: il *Random Intercept Model* e il *Random Slope Model*.

2.2.2.4 Random Intercept Model

Il modello multilivello ad intercetta casuale viene definito come segue:

$$Y_{ij} = \alpha_j + \beta_j X_{ij} + \varepsilon_{ij}$$

con

$$\alpha_j = \gamma_{00} + \gamma_{01} W_j + u_{0j}$$

$$\beta_j = \gamma_{10} + \gamma_{11} W_j$$

Queste 2 equazioni definiscono i parametri del modello. Gli α_j possono essere trattati come realizzazioni di variabili casuali, mentre ε_{ij} e u_{0j} sono dei termini di errore.

Nell'equazione di α_j , il termine $\gamma_{00} + \gamma_{01} W_j$ indica il valore medio dell'intercetta per gruppi con le caratteristiche W_j , mentre il termine di errore u_{0j} indica la deviazione del gruppo j dal valore medio appena menzionato; per l'equazione β_j è l'analogo discorso tranne che per la mancanza del termine di errore.

La forma compatta che si ottiene dalle equazioni precedenti è la seguente:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} W_j + \gamma_{11} W_j X_{ij} + u_{0j} + \varepsilon_{ij}$$

La prima parte dell'equazione, ovvero $\gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} W_j + \gamma_{11} W_j X_{ij}$ è detta parte fissa del modello, mentre $u_{0j} + \varepsilon_{ij}$ è detta parte casuale del modello (contiene gli effetti casuali a tutti i livelli e le interazioni con le variabili osservate).

Viene specificata la distribuzione dei termini d'errore come segue:

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

$$u_{0j} \sim N(0, \sigma_{u_0}^2)$$

$$\text{Cov}(\varepsilon_{ij}, u_{0j}) = 0$$

Dopo aver fatto queste assunzioni, vale la seguente formulazione di varianza e covarianza:

$$Var(Y_{ij}|X_{ij}, W_j) = Var(\varepsilon_{ij}) + Var(u_{0j}) = \sigma_\varepsilon^2 + \sigma_{u_0}^2$$

$$Cov(Y_{ij}, Y_{i'j'}|X_{ij}, W_j) = \begin{cases} \sigma_{u_0}^2 & \text{se } j = j', i \neq i' \\ 0 & \text{se } j \neq j' \end{cases}$$

Questo modello è semplice dal punto di vista interpretativo e presenta omoschedasticità poiché la varianza è composta solo dalle due varianze degli errori; in questo modo si capisce quanta variabilità è dovuta alla componente *between* e quanta a quella *within*.

Il Coefficiente di Correlazione Intraclasse per un modello Random Intercept Model è costante, ed è il seguente:

$$\rho(Y_{ij}, Y_{i'j'}|X_{ij}, X_{i'j'}, W_j, W_{j'}) = \begin{cases} \frac{Cov(Y_{ij}, Y_{i'j'}|X_{ij}, X_{i'j'}, W_j)}{\sqrt{Var(Y_{ij}|\dots) + Var(Y_{i'j'}|\dots)}} & \text{se } j = j', i \neq i' \\ 0 & \text{se } j \neq j' \end{cases} =$$

$$= \begin{cases} \frac{\sigma_{u_0}^2}{\sigma_\varepsilon^2 + \sigma_{u_0}^2} & \text{se } j = j', i \neq i' \\ 0 & \text{se } j \neq j' \end{cases}$$

2.2.2.5 Random Slope Model

Nel modello multilivello a pendenza casuale si ipotizza che i gruppi si differenzino casualmente fra loro non solo rispetto al valore atteso della variabile dipendente, ma anche rispetto alla pendenza di ogni singola retta. Viene quindi definito come segue:

$$Y_{ij} = \alpha_j + \beta_j X_{ij} + \varepsilon_{ij}$$

con

$$\alpha_j = \gamma_{00} + \gamma_{01} W_j + u_{0j}$$

$$\beta_j = \gamma_{10} + \gamma_{11} W_j + u_{1j}$$

Queste 2 equazioni definiscono i parametri del modello ed ogni gruppo ha la propria retta di regressione, quindi si andranno a stimare tante rette quanti saranno i gruppi. I coefficienti α_j e β_j possono essere trattati come variabili dipendenti in una regressione che modella la popolazione dei gruppi, ma sono “regressioni latenti”, perché hanno bisogno dell'errore (che qui sono rispettivamente u_{0j} e u_{1j} e sono errori di secondo livello) per essere osservati (Snijders e Bosker, 1999).

Nell'equazione di α_j , il termine $\gamma_{00} + \gamma_{01}W_j$ indica il valore medio dell'intercetta per gruppi con le caratteristiche W_j , mentre il termine di errore u_{0j} indica la deviazione del gruppo j dal valore medio appena menzionato; per l'equazione β_j è l'analogo discorso.

La forma compatta che si ottiene dalle equazioni precedenti è la seguente:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_j + \gamma_{11}W_jX_{ij} + u_{0j} + u_{1j}X_{ij} + \varepsilon_{ij}$$

La prima parte dell'equazione, ovvero $\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_j + \gamma_{11}W_jX_{ij}$ è detta parte fissa del modello, mentre $u_{0j} + u_{1j}X_{ij} + \varepsilon_{ij}$ è detta parte casuale del modello (contiene gli effetti casuali a tutti i livelli e le interazioni con le variabili osservate).

Nel modello multilivello è presente variabilità non spiegata ad entrambi i livelli, e viene specificata la distribuzione dei termini d'errore come segue:

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

$$U_j = \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N \left(\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}, \begin{bmatrix} \sigma_{u_0}^2 & \sigma_{u_0u_1} \\ \sigma_{u_0u_1} & \sigma_{u_1}^2 \end{bmatrix} \right)$$

$$\text{Cov}(\varepsilon_{ij}, U_j) = 0$$

Il vettore U_j contiene i residui a livello di gruppo, una volta depurato dell'effetto delle variabili osservabili.

Dopo aver fatto queste assunzioni, vale:

$$\text{Var}(Y_{ij}|X_{ij}, W_j) = \sigma_\varepsilon^2 + \sigma_{u_0}^2 + \sigma_{u_1}^2 X_{ij}^2 + 2\sigma_{u_0u_1} X_{ij}$$

$$\text{Cov}(Y_{ij}, Y_{i'j'}|X_{ij}, W_j) = \begin{cases} \sigma_{u_0}^2 + \sigma_{u_1}^2 X_{ij}X_{i'j'} + \sigma_{u_0u_1}(X_{ij} + X_{i'j'}) & \text{se } j = j', i \neq i' \\ 0 & \text{se } j \neq j' \end{cases}$$

Con il *Random Slope Model* non è presente omoschedasticità, infatti la varianza dipende dal valore delle covariate e non è costante. Di conseguenza anche l'ICC non è costante come nel modello precedente, e dipende dalle covariate.

2.2.2.6 Stima

Per questo paragrafo si fa riferimento principalmente a Hedeker e Gibbons (1994), Snijders e Bosker (1999), Pinheiro e Chao (2006).

Se la variabile risposta del caso di studio fosse continua e si assumesse la normalità delle componenti casuali, i due principali metodi per la stima di un modello

multilivello sarebbero quello della Massima Verosimiglianza (Maximum Likelihood: ML) e quello della Massima Verosimiglianza Residuale o Ristretta (REstricted Maximum Likelihood: REML). Le differenze nella stima dei coefficienti della parte fissa del modello sono minime, mentre sono più accentuate nella stima delle componenti di varianza. Una differenza è che il metodo REML, quando stima le componenti di varianza, tiene conto della perdita dei gradi di libertà dovuta alla stima dei coefficienti della parte fissa del modello, mentre la stima di ML non la tiene in considerazione; questo porta le stime ML delle componenti della varianza ad essere distorte verso il basso. Il metodo REML usualmente fornisce delle stime più accurate delle componenti di varianza.

Se con questi 2 metodi di stima non si dovesse raggiungere la convergenza, ci sono alternative attraverso dei metodi iterativi.

Nel caso in cui la variabile risposta sia discreta, i metodi di stima attualmente più usati sono quello della Quadratura Gaussiana (Gauss-Hermite) e della Quadratura Gaussiana Adattiva (AGQ). *“Il metodo Gauss-Hermite viene usato per integrare numericamente la distribuzione degli effetti casuali. Per dati multilivello con 2 livelli, questo modello consente diversi effetti casuali al 2° livello”* (Hedeker e Gibbons, 1994); inoltre ha una buona accuratezza nella stima dei coefficienti della parte fissa del modello mentre sottostima le componenti della varianza. Il metodo AGQ ha una buona accuratezza (e bassa/nulla distorsione) per i coefficienti della parte fissa anche con campioni molto ridotti però sottostima le componenti della varianza anche su campioni di grosse dimensioni. L'AGQ è iterativo e aumentando la complessità computazionale permette di produrre stime sempre meno distorte.

Per analizzare i dati di questo lavoro, sono stati stimati sia un *Random Intercept Model* che un *Random Slope Model* utilizzando la funzione risposta *Probit* per variabili ordinali.

Capitolo 3.

Creazione Questionario

3.1 Struttura

La domanda sulla quale si è incentrata la creazione del questionario è stata: “Quali sono i fattori che influiscono l’acquisto di uno smartphone?”.

Si è deciso di scegliere principalmente quattro variabili per le *vignettes*:

- ❖ Prezzo, diviso in due livelli usando come soglia i 300€ (quindi inferiore ai 300€ e superiore). Si è scelto questo valore perché è quello che attualmente separa gli smartphone di fascia medio-bassa da quelli di fascia medio-alta;
- ❖ Display, anche questo con due livelli, inferiore o superiore ai 5 pollici (5”);
- ❖ Memoria interna, divisa in tre livelli, da 8, 16 o 32 GB;
- ❖ Genere, maschio o femmina.

In calce ad ogni *vignette* agli intervistati veniva posta la domanda su quanto d’accordo fossero sulla scelta dell’acquisto di uno smartphone che presentasse tutte le caratteristiche individuate dalla combinazione in oggetto (1 per niente d’accordo, 5 completamente d’accordo). Con questa domanda si è creata la variabile risposta, categoriale ordinale.

La parte di questionario tradizionale conteneva 3 ulteriori domande. La prima riguardava la marca: non si voleva nominare un brand specifico, ma solo capire se la scelta dipendesse da questo piuttosto che dalle prestazioni o da entrambe.

La seconda domanda è stata posta per comprendere il motivo principale per il quale una persona potrebbe decidere di cambiare il proprio smartphone, dando 6 alternative tra le quali la rottura del prodotto, l’obsolescenza, un’offerta vantaggiosa o l’uscita del nuovo modello. Il questionario si concludeva domandando all’intervistato se fosse maschio o femmina.

Sono stati creati 6 tipi di questionari diversi (dalla A alla F, per distinguerli), ognuno con 4 *vignettes* nelle quali i personaggi erano 2 maschili, Alessandro (a) e Francesco (c) e 2 femminili, Sofia (b) e Martina (d). I nomi sono stati scelti in base alla

“top 10” dei nomi più diffusi in Italia dal 1999 al 2015⁶. Sono stati scelti questi 4 specifici nomi poiché si voleva creare meno “rumore” possibile per il questionario, cercando quindi di non influenzare ulteriormente possibili valutazioni soggettive dovute al collegamento di un nome non comune con persone di una coorte meno recente piuttosto che di un'altra (ne fanno riferimento Jorges H. & Winter J., 2013), mentre mettendo nomi molto frequenti per persone nate dal 1999 ad oggi è possibile che il campione si orienti su coetanei, poiché come verrà spiegato poco più avanti l'età dei rispondenti è giovane. In ogni sottogruppo di questionari è stato fatto ruotare l'ordine dei personaggi, quindi per esempio nel questionario A c'erano 4 varianti, ovvero i 4 personaggi a turno erano i primi con la loro *vignette*, seguiti dai restanti 3.

Per capire meglio la struttura dei questionari si riporta come esempio la *Tabella 3.1*, dove è mostrata la suddivisione dei casi come appena spiegato.

		A	B	C	D	E	F
Alessandro (a)	Genere	M	M	M	M	M	M
	300 €	sup	sup	sup	sup	sup	sup
	5"	sup	sup	sup	inf	inf	inf
	GB	8	16	32	8	16	32
Sofia (b)	Genere	F	F	F	F	F	F
	300 €	sup	sup	sup	sup	sup	sup
	5"	inf	inf	inf	sup	sup	sup
	GB	32	8	16	16	32	8
Francesco (c)	Genere	M	M	M	M	M	M
	300 €	inf	inf	inf	inf	inf	inf
	5"	inf	sup	sup	inf	inf	sup
	GB	16	32	8	32	8	16
Martina (d)	Genere	F	F	F	F	F	F
	300 €	inf	inf	inf	inf	inf	inf
	5"	sup	inf	inf	sup	sup	inf
	GB	8	16	8	32	16	32

Tabella 3.1: Suddivisione delle *vignettes* nei 6 tipi di questionari, dalla A alla F; ogni *vignette* ha una combinazione unica ed è evidenziata a scacchiera per essere distinta dalle altre. Ognuno dei 6 tipi di questionari ha 4 *vignettes* (a, b, c, d).

Di seguito viene spiegato il significato di sup e inf:

-sup: valore superiore rispetto a quanto indicato nell'intestazione di colonna;

-inf: valore inferiore rispetto a quanto indicato nell'intestazione di colonna.

⁶ <https://www.lenius.it/nomi-piu-diffusi-in-italia/>,

<https://www.istat.it/it/prodotti/contenuti-interattivi/calcolatori/nomi>

La rotazione delle 4 *vignettes* all'interno di ognuno dei 6 tipi di questionario è la seguente: a-b-c-d, b-c-d-a, c-d-a-b, d-a-b-c. La scelta di cambiare l'ordine e di non lasciare sempre Alessandro (a) primo e Martina (d) ultima è dovuta al fatto che il rispondente potrebbe leggere con più attenzione la prima, o le prime due *vignettes*, e rispondere in maniera più distratta e frettolosa alle altre. Ruotando le *vignettes* ci si aspetta più omogeneità e meno rumore nei risultati.

I questionari con i diversi *sets* di *vignettes* si possono trovare in Appendice.

3.2 Somministrazione

Creati i questionari, si è scelto di sottoporli a 240 rispondenti, divisi in 120 maschi e 120 femmine, nel modo seguente: ogni tipo di questionario (dalla A alla F) è stato compilato da 40 persone diverse, divise equamente per genere per arrivare ai $40 \times 6 = 240$ questionari totali. Poiché in ogni questionario erano presenti 4 *vignettes*, in totale si ha un numero di *vignettes* compilate pari a $240 \times 4 = 960$.

I dati sono stati raccolti nel mese di Marzo 2017 ed il campione è stato circoscritto agli studenti del Dipartimento di Scienze Statistiche dell'università degli studi di Padova, in modo tale da avere un gruppo di rispondenti che fosse omogeneo al suo interno, con età compresa tra i 19 e i 25 anni.

Capitolo 4.

Analisi dei dati

L'analisi dei dati verrà strutturata in 4 paragrafi, in ognuno dei quali troveranno spazio degli approfondimenti specifici per argomento.

Inizialmente verrà esplicitata la codifica di ogni variabile, per potersi orientare al momento dell'analisi svolta con i modelli statistici (paragrafi 4.3 e 4.4). Prima dei modelli verranno eseguite le analisi esplorative rispetto le variabili usate, analisi univariate, bivariate e trivariate, accompagnate dai rispettivi grafici. Gli ultimi due paragrafi sono dedicati alle stime dei modelli *Ordered Probit* e multilivello.

Sia per il modello *Ordered Probit* che per quello gerarchico sono state svolte due analisi, studiando prima solo le variabili delle *vignettes* (riportate nella tabella col titolo "Solo *Vignettes*") e poi considerando tutte le variabili presenti sul questionario ("Tutte le variabili"). Le tabelle con i risultati delle due analisi verranno affiancate l'una all'altra per poter agevolare il confronto.

4.1 Codifica delle variabili

La variabile scelta come dipendente si chiama "**Risposta**"; assume 5 livelli che indicano quanto d'accordo o meno è il soggetto rispondente con la *vignette* a lui proposta. La variabile è ordinata e va da **1**, "Per niente d'accordo", a **5**, "Completamente d'accordo".

Le prossime 6 variabili sono quelle relative alle *vignettes*.

Le variabili "**Posizione**" e "**Set**" non sono state direttamente inserite nella storia vera e propria della *vignette*, ma rappresentano rispettivamente:

- ❖ La posizione della *vignette* all'interno del questionario, ovvero se era **1°**, **2°**, **3°** o **4°** in ordine di lettura;
- ❖ Quale dei 6 questionari si trattava (se A, B, C, D, E o F, codificati da **1** al **6**).

Si tratta in questo caso di variabili categoriali.

Il genere del personaggio della *vignette* viene codificato con la parola "**Sesso**" ed è una variabile dummy: assume il valore **0** per identificare le persone di genere femminile (Sofia o Martina) ed **1** quelli di genere maschile (Alessandro o Francesco).

Le tre variabili relative alle caratteristiche dello smartphone sono state chiamate “Prezzo”, “Display” e “GB”, quest’ultima indica la memoria interna del dispositivo. “Prezzo” e “Display” sono due variabili dummy, la prima assume valore **0** se la cifra che viene proposta per l’acquisto dello smartphone è inferiore ai 300€, mentre **1** se è superiore; la seconda variabile vale **0** nel caso di display inferiore ai 5” ed **1** per uno schermo più grande della soglia individuata. “GB” è una variabile categoriale, con 3 livelli, ed assume il valore **0** se i GigaBytes sono 8, **1** nel caso di 16 GB e **2** se la memoria è di 32 GB.

Le ultime tre variabili studiate sono state inserite con la modalità del questionario tradizionale, ovvero sono state raccolte per mezzo di domande dirette all’intervistato.

“Scelta” è una variabile categoriale sconnessa con 5 categorie. Rappresenta la scelta che un individuo potrebbe fare al momento dell’acquisto di un nuovo smartphone, ovvero:

- ❖ Valutare solo in base alle prestazioni indipendentemente dalla marca (**1**);
- ❖ Valutare inizialmente le prestazioni ma poi considerare la marca (**2**);
- ❖ Valutare in ugual misura sia le prestazioni che la marca (**3**);
- ❖ Scegliere principalmente in base alla marca ma poi valutare anche le prestazioni (**4**);
- ❖ Scegliere solo in base alla marca, indipendentemente dalle prestazioni (**5**).

“Motivo” ha una categoria in più di “Scelta” e rappresenta il motivo per cui l’intervistato avrebbe scelto, o potrebbe scegliere, di cambiare smartphone:

- ❖ Rottura del proprio smartphone (**1**);
- ❖ Obsolescenza del dispositivo, ovvero se la tecnologia diventa antiquata e sostituita da altre più recenti (**2**);
- ❖ In negozio/on-line viene vista un’offerta vantaggiosa o una promozione su un altro smartphone, anche se quello attualmente in uso fosse ancora funzionante (**3**);
- ❖ Un operatore telefonico (attuale o diverso dal proprio) propone una buona offerta tariffaria comprensiva di smartphone (**4**);
- ❖ Esce il nuovo modello della marca posseduta (**5**);
- ❖ Altro non specificato (**6**).

Per ultima si trova la variabile dicotomica “Genere”, chiamata così per differenziarla da “Sesso”, ed indica il genere dell’intervistato. Assume valore **0** se si tratta di una femmina mentre vale **1** se il rispondente è un maschio.

4.2 Analisi esplorative

In questo paragrafo verranno calcolate alcune semplici statistiche univariate, bivariate e trivariate in merito alle variabili del questionario. Nelle analisi bivariate e trivariate, per ogni tabella a doppia/tripla entrata è stato calcolato l'indice di correlazione chi-quadro di Pearson (χ^2) che verrà riportato nella descrizione di ogni grafico; questa misura porta risultati che potrebbero essere diversi (e in alcuni casi sarà così) rispetto all'analisi multivariata.

4.2.1 Analisi univariate

Verranno svolte analisi univariate solo per le variabili acquisite nella parte del questionario tradizionale ("Scelta" e "Motivo") e per la variabile "Risposta" (che fa parte delle *vignettes*). Tutte le altre variabili sono state costruite *ad hoc* per cui un'analisi di questo tipo non avrebbe alcun senso.

4.2.1.1 "Risposta"

Viene analizzata singolarmente la variabile risposta per avere una panoramica della percentuale di voti che sono stati dati per ogni categoria.

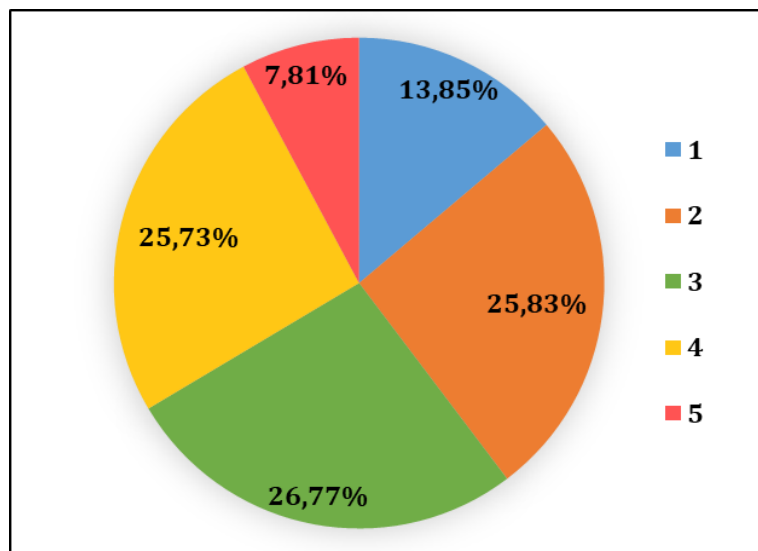


Grafico 4.1: Distribuzione percentuale delle categorie della variabile "Risposta" che va da 1: "Per niente d'accordo", a 5 "Completamente d'accordo".

Osservando il *Grafico 4.1* si nota che la maggior parte delle risposte è stata data ai tre livelli centrali (78,33% totale) contro un 21,66% di risposte tra "completamente in disaccordo" e "completamente d'accordo".

La scelta più gettonata è stata quella centrale che equivale in questo caso alla neutralità della valutazione, ovvero quando non si è né in accordo né in disaccordo. Per questo a volte è preferibile non mettere la scelta centrale: per evitare che l'intervistato faccia "il pigro" e non prenda le parti di una o dell'altra fazione.

In questo questionario sono state presentate le 5 scelte:

Per niente d'accordo				Completamente d'accordo
①	②	③	④	⑤

Sono state date solo la prima e l'ultima "definizione" alle numerazioni, quindi non è certo se i rispondenti che hanno votato 3 intendessero rimanere neutri o se invece l'interpretazione sia stata di "Sufficientemente d'accordo" (o simili).

4.2.1.2 "Scelta"

Viene analizzata la percentuale di risposte che sono state date alle categorie della variabile "Scelta", ovvero si analizza quanto una persona sceglie lo smartphone in base alle prestazioni piuttosto che alla marca.

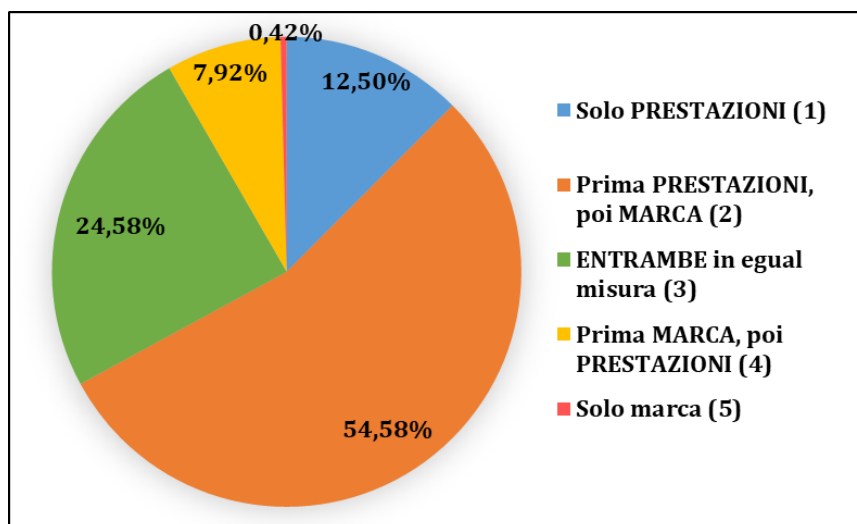


Grafico 4.2: Distribuzione percentuale delle categorie della variabile "Scelta".

Dal *Grafico 4.2* si vede che la maggior parte degli studenti di Scienze Statistiche intervistati caratterizzerebbe la scelta di uno smartphone basandosi primariamente sulle prestazioni, ma tenendo poi anche conto della marca (54,58%). Soltanto uno 0,42% sceglierebbe solo ed esclusivamente in base alla marca..

Un 24,58% pondera la scelta guardando in egual misura sia la marca che le prestazioni, mentre un 12,50% non tiene per niente conto della marca e si focalizza solo sulle caratteristiche prestazionali.

4.2.1.3 “Motivo”

La maggior parte degli studenti di Scienze Statistiche (con un 55,42% dei voti) cambierebbe il telefonino in caso di rottura di quello attualmente posseduto. La seconda motivazione, con una percentuale che è quasi la metà rispetto la prima (29,17%), riguarda l’obsolescenza, quindi la necessità di cambiare smartphone per il progredire della tecnologia e per stare “al passo coi tempi” (*Grafico 4.3*).

Le altre quattro voci sono prese molto meno in considerazione, arrivando complessivamente ad un 15,41%. Solo un 3,33% valuterebbe di cambiare il proprio smartphone per acquistare il nuovo modello della marca già in uso.

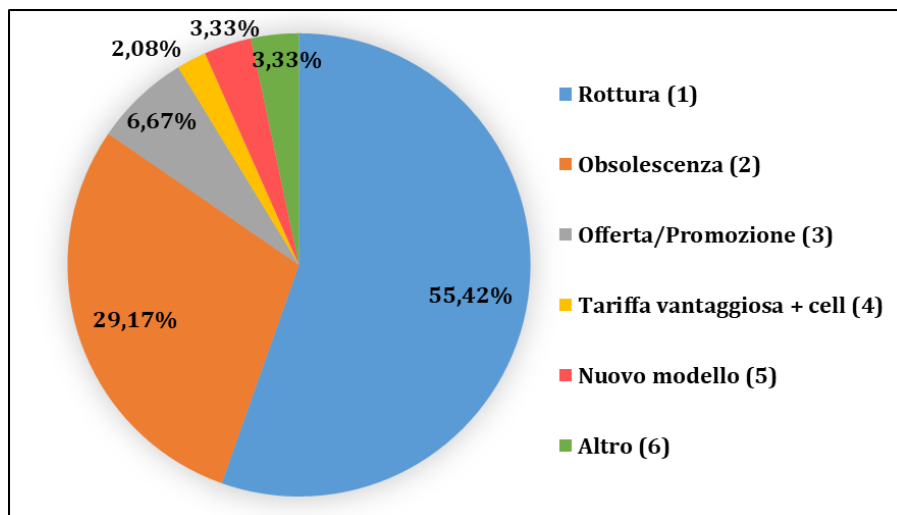


Grafico 4.3: Distribuzione percentuale delle categorie della variabile “Motivo”.

4.2.2 Analisi bivariate

Tutte le analisi di questo paragrafo sono state eseguite mettendo in relazione di volta in volta ogni variabile esplicativa con la variabile risposta.

4.2.2.1 “Risposta” vs “Posizione”

È stato scelto di far ruotare la posizione delle *vignette* all’interno dei questionari per capire se gli intervistati fossero influenzati dall’ordine con cui veniva posta la *vignette*, fosse solo per una questione di attenzione prestata alla lettura.

Il test χ^2 è risultato significativo ad un livello di 1%, a conferma del fatto che c’è correlazione tra le risposte date e la posizione dello scenario presentato. Dal *Grafico*

4.4 si nota come la *vignette* posizionata in testa abbia ricevuto una percentuale di voti “1” (completamente in disaccordo) molto più bassa (13,53%) rispetto all’ultima *vignette* del questionario che ha ricevuto un 33,08% di voti negativi, indipendentemente dal contenuto. Per le risposte che vanno dalla “2” alla “5” non risultano grosse differenze, se non alla risposta “4” della *vignette* posta in seconda posizione, che riceve una percentuale più bassa (20,24%) rispetto la media delle altre 3 (28% circa). Una percentuale simile (19,55%) si trova per la risposta “1” (“Per niente d’accordo”), sempre per la *vignette* in seconda posizione.

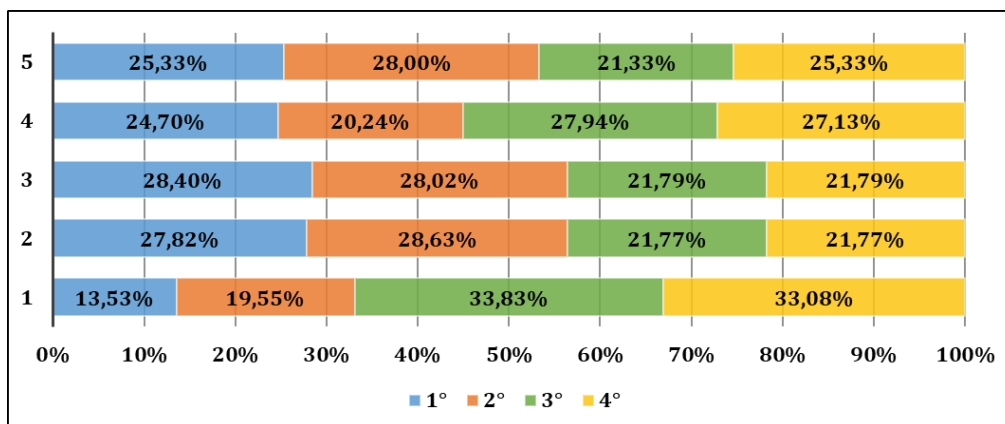


Grafico 4.4: Distribuzione percentuale della variabile "Posizione" in base alla "Risposta". Il test χ^2 è statisticamente significativo ad un livello di 1% (p-value=0.004).

4.2.2.2 “Risposta” vs “Set”

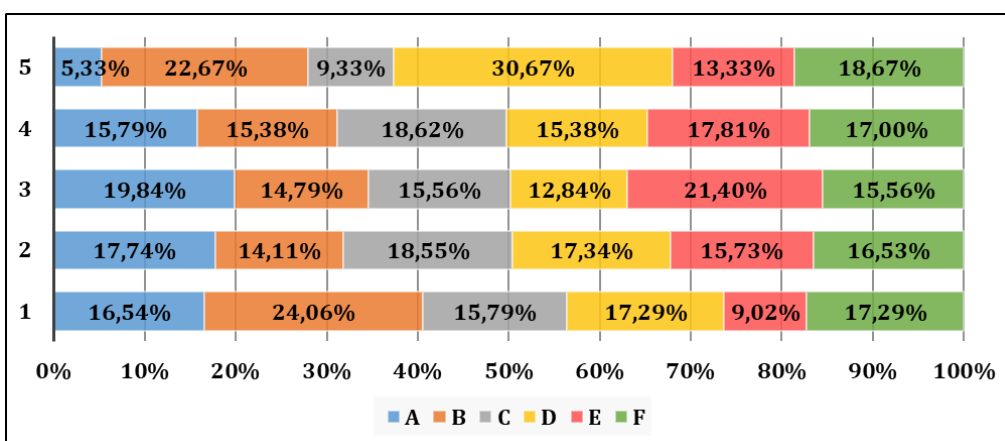


Grafico 4.5: Distribuzione percentuale della variabile "Set" in base alla "Risposta". Il test χ^2 è statisticamente significativo ad un livello di 1% (p-value=0.005).

In questo paragrafo si analizza se può essere influenzante il tipo di questionario (6 in tutto, dalla A alla F) in relazione alla variabile risposta; il test χ^2 risulta

statisticamente significativo ad 1% di livello, indicando così delle differenze tra un tipo di questionario e l'altro, ovviamente prendendo a confronto solo questa variabile con la "Risposta" (è possibile già premettere che invece nelle analisi multivariate le categorie della variabile "Set" non saranno più statisticamente significative). Dal *Grafico 4.5* risulta che la maggior parte dei rispondenti che hanno votato 5 ("completamente d'accordo") sono quelli che hanno ricevuto il questionario D (30,67%) mentre i due questionari con percentuale più bassa riguardo la risposta 5 sono stati la versione A col 5,33% di risposte, e quella C, col 9,33%.

Gli intervistati che hanno votato meno alla risposta 1 sono quelli del questionario E con un 9,08% mentre un buon 24,06% arriva dai rispondenti del questionario C. Per le altre risposte non ci sono differenze importanti da sottolineare.

4.2.2.3 "Risposta" vs "Sesso"

Si vuole controllare se il genere del personaggio che interpreta la *vignette* possa fare la differenza su un campione composto da 240 rispondenti: 120 maschi e 120 femmine. Dal *Grafico 4.6* si nota come ci sia sostanzialmente equilibrio nelle risposte.

Una conferma arriva dall'indice χ^2 che non risulta statisticamente significativo.

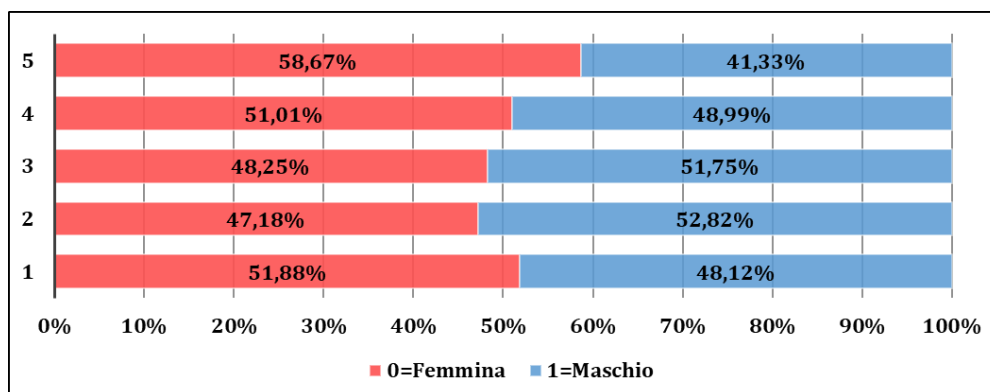


Grafico 4.6: Distribuzione percentuale della variabile "Sesso" (genere del personaggio della *vignette*) in base alla "Risposta". Il test χ^2 non è statisticamente significativo (p-value=0.456).

L'unica variazione che si può evidenziare è quel 58,67% di risposte al livello "5", completo accordo, quando il personaggio della *vignette* è di genere femminile.

4.2.2.4 "Risposta" vs "Prezzo"

Come ci si poteva aspettare, la variabile "Prezzo" influisce molto sulla "Risposta", a confermarlo è la significatività ad 1% di livello del test χ^2 per questa analisi bivariata,

ma questa caratteristica si dimostrerà una importante determinante anche nelle stime dei modelli multivariati.

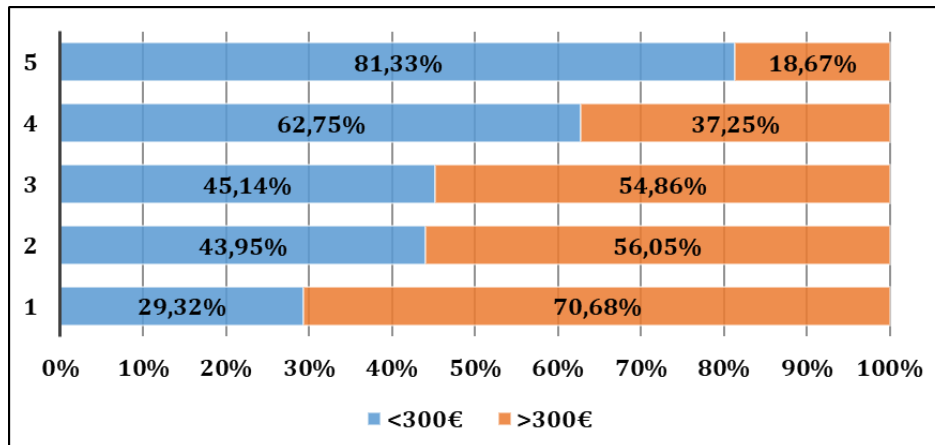


Grafico 4.7: Distribuzione percentuale della variabile "Prezzo" in base alla "Risposta". Il test χ^2 è statisticamente significativo ad un livello di 1% (p-value=0.000).

Questo risultato si ottiene dall'osservazione del *Grafico 4.7* in base al quale questo campione di rispondenti ha un occhio di riguardo rispetto questa variabile. Come già evidenziato nello studio di Daejoong, Heasun e Hyunjoo (2014), a quest'età spesso la disponibilità economica non è elevata e non c'è indipendenza dalla famiglia d'origine.

L'81,33% dei rispondenti è completamente d'accordo con il personaggio della *vignette* quando si tratta di acquistare uno smartphone a meno di 300€, contro il 18,67% dei rispondenti d'accordo con uno superiore a quella cifra. Se si scende di un livello nella risposta (risposta 4), la percentuale di persone in accordo per un prezzo inferiore ai 300€ diventa del 62,75%.

Le risposte "2" e "3" sono piuttosto equilibrate tra superiore ed inferiore ai 300€.

Alla risposta "1" il 70,68% delle persone che hanno votato "Per niente in accordo" avevano nella storia l'acquisto di uno smartphone che costava più di 300€, a conferma che un prezzo basso è preferito dal campione analizzato.

4.2.2.5 "Risposta" vs "Display"

L'analisi tra queste due variabili ha un livello di significatività al 5% quindi da non sottovalutare⁷. Gli studenti che hanno risposto al questionario hanno ritenuto che anche il display potesse essere una delle caratteristiche da tenere presente nel caso di un nuovo acquisto. La grandezza ottimale risulterebbe quindi essere dai 5 pollici in su.

⁷ A conferma di quanto trovato negli studi già citati di Siu-Tsen (2015) e Na e Ruifeng (2017)

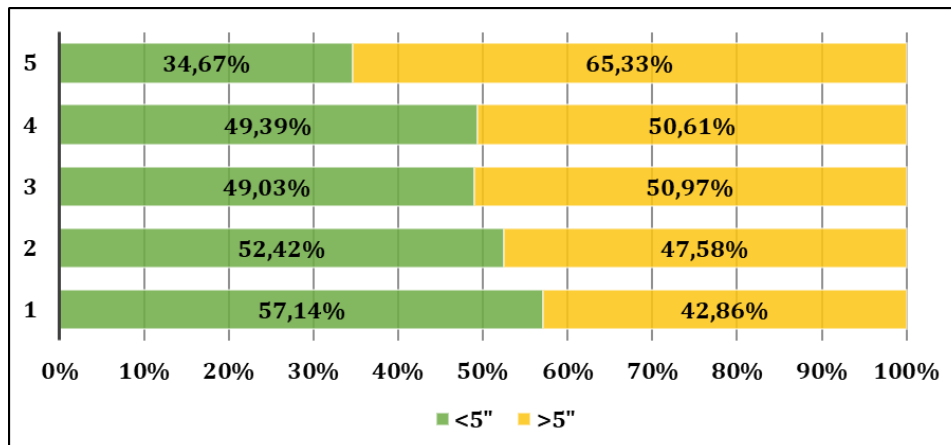


Grafico 4.8: Distribuzione percentuale della variabile "Display" in base alla "Risposta". Il test χ^2 è statisticamente significativo ad un livello di 5% (p-value=0.033).

Alla lettura della *vignette*, quando la valutazione è stata "Completamente d'accordo", un 65,33% ha scelto lo schermo di grandi dimensioni (*Grafico 4.8*). Le risposte centrali non hanno differenze significative mentre il grado di massimo disaccordo (risposta "1") trova un'alta percentuale (57,14%) per il display sotto i 5", a conferma della preferenza attuale del mercato di offrire smartphone di grandi dimensioni.

4.2.2.6 "Risposta" vs "GB"

Come premesso all'inizio di questo lavoro, il rapido avanzamento della tecnologia obbliga ad avere sempre più memoria interna, per supportare l'utilizzo di app e per il salvataggio di dati, sempre più presenti nei nostri cellulari.

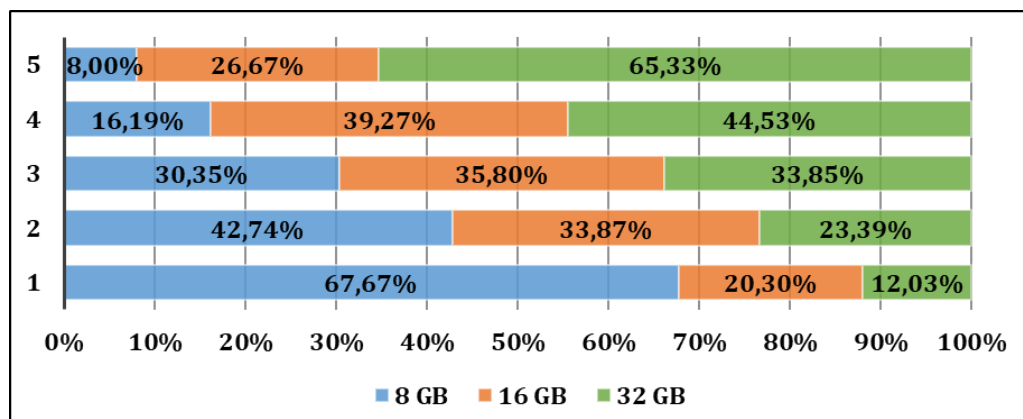


Grafico 4.9: Distribuzione percentuale della variabile "GB" in base alla "Risposta". Il test χ^2 è statisticamente significativo ad un livello di 1% (p-value=0.000).

L'analisi mostra un test con una significatività statistica ad 1% di livello, a conferma della forte rilevanza di questa componente nella valutazione delle *vignettes*. Lo si può notare anche dall'andamento delle barre del *Grafico 4.9* che gradualmente aumentano (o diminuiscono a seconda dell'ordine) man mano che ci si sposta dalla risposta "1" alla "5".

I rispondenti hanno valutato "5" ("Completamente d'accordo") quando si trovavano di fronte ad una *vignette* che presentava uno smartphone con 32GB (65,33%), e solo un 8% ha votato per un nuovo acquisto con 8GB di memoria interna.

Mano a mano che diminuisce il grado di accordo, la barra verde (memoria di 32GB) diminuisce a scapito di quella azzurra (8GB) che cresce velocemente (arrivando al 67,67%) per quelli che non sono "per niente d'accordo", lasciando un 12,03% alla memoria da 32GB.

La memoria intermedia da 16GB (barra arancione) rimane pressochè uguale per tutte le categorie di risposta.

4.2.2.7 "Risposta" vs "Scelta"

Questa comparazione non risulta significativa e come già visto per l'analisi univariata sulla variabile "Scelta", le percentuali del *Grafico 4.10* confermano i risultati che si erano avuti dal *Grafico 4.2*: la voce "Solo marca" è pressoché inesistente mentre le barre che colpiscono al primo sguardo sono quelle arancioni, che si riferiscono alla scelta dello smartphone guardandone inizialmente le prestazioni, ma tenendo poi in considerazione anche la marca.

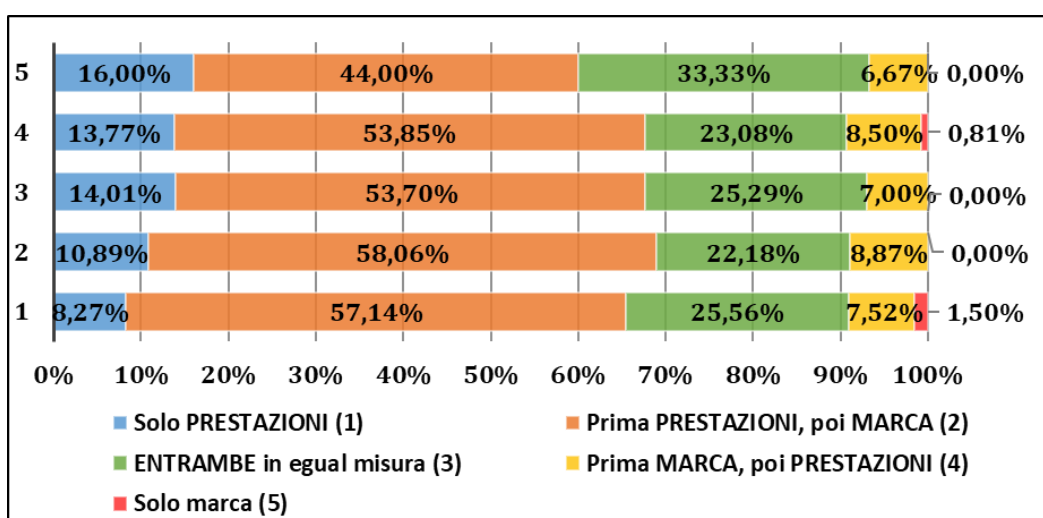


Grafico 4.10: Distribuzione percentuale della variabile "Scelta" in base alla "Risposta". Il test χ^2 non è statisticamente significativo (p-value=0.357).

4.2.2.8 “Risposta” vs “Motivo”

Anche questa analisi dal test χ^2 non emerge una significatività statistica e come per il *Grafico 4.10*, anche questo (*Grafico 4.11*) è in linea col rispettivo grafico univariato relativo alla variabile “Motivo” (*Grafico 4.3*).

Le principali motivazioni che spingono all’acquisto di un nuovo smartphone sono la rottura del precedente e l’obsolescenza. Le altre scelte proposte non hanno avuto gradimento importante.

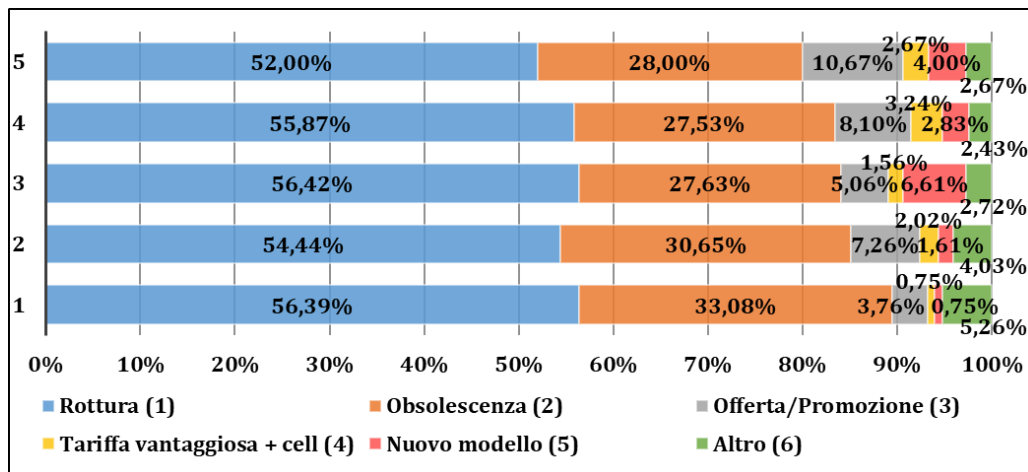


Grafico 4.11: Distribuzione percentuale della variabile "Motivo" in base alla "Risposta". Il test χ^2 non è statisticamente significativo (p-value=0.150).

4.2.2.9 “Risposta” vs “Genere”

“Scelta” e “Motivo”, variabili esterne alle *vignettes*, non sono risultate significative. Lo stesso non si può dire per “Genere” (si riferisce al genere dell’intervistato), altra variabile che non faceva parte dello scenario descritto, ma che risulta di rilievo.

Ciò potrebbe far pensare ad una correlazione tra la scelta della risposta ed il genere di chi la valutava.

Dal *Grafico 4.12* si può notare un’alternanza di risultati tra le risposte: i rispondenti di genere maschile hanno la maggioranza nelle risposte “1”, “3” e “5”, mentre la componente femminile nelle altre due, con delle percentuali che per le risposte dalla “2” alla “5” rimangono circa tra il 45% e il 55%: la risposta “5” evidenzia la più alta maggioranza di un certo genere (in questo caso maschile, con il 64,66%).

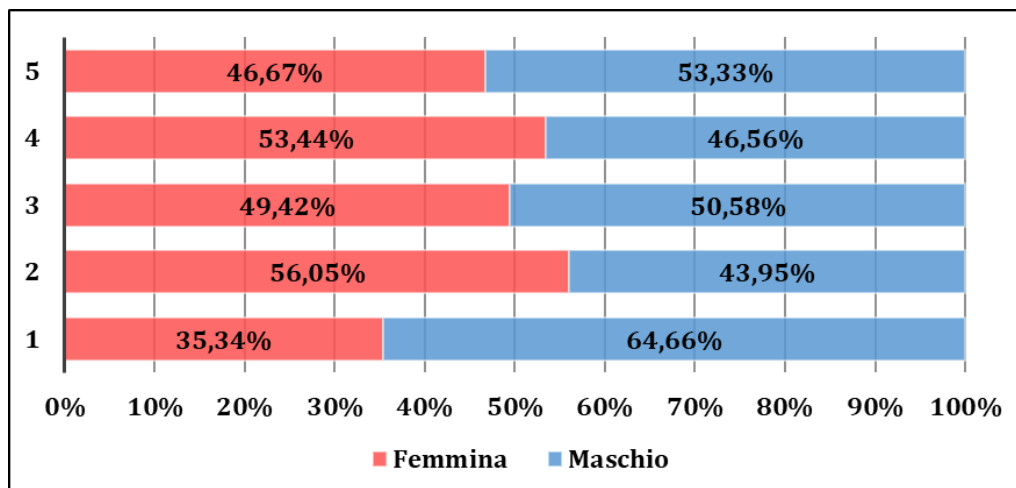


Grafico 4.12: Distribuzione percentuale della variabile "Genere" in base alla "Risposta". Il χ^2 è significativo all'1% Il test χ^2 è statisticamente significativo ad un livello di 1% (p-value=0.002).

4.2.3 Analisi trivariate

I due successivi sotto paragrafi fanno riferimento a:

- ❖ Un'analisi tra la variabile "Risposta" e la variabile "Genere", qualora "Sesso", ovvero il genere del personaggio della *vignette*, assuma valore "femmina" o "maschio";
- ❖ Un'analisi tra "Risposta" e "Posizione", nel caso dei diversi "Set", ovvero dei diversi tipi di questionari.

L'analisi al sotto paragrafo 4.2.3.1 è stata eseguita per capire se le persone siano influenzabili dal fatto che un'asserzione sia formulata da un uomo piuttosto che da una donna. Quella successiva serve a capire se ci siano differenze sostanziali nelle risposte in base alla posizione occupata dalla *vignette* presentata; lo studio è stato svolto prendendo in considerazione ognuno dei 6 questionari creati.

4.2.3.1 "Risposta" vs "Genere" se "Sesso" = ...

...**FEMMINA**: nel caso in cui il personaggio della *vignette* sia di genere femminile, dal test χ^2 non appare una correlazione statisticamente significativa. Se però si osservano i risultati del *Grafico 4.13* si potrebbe giungere a conclusioni diverse. In particolare, ci si riferisce alla risposta "1", dove il 62,32% dei rispondenti è "Per niente d'accordo" ad un acquisto proposto da parte di una donna.

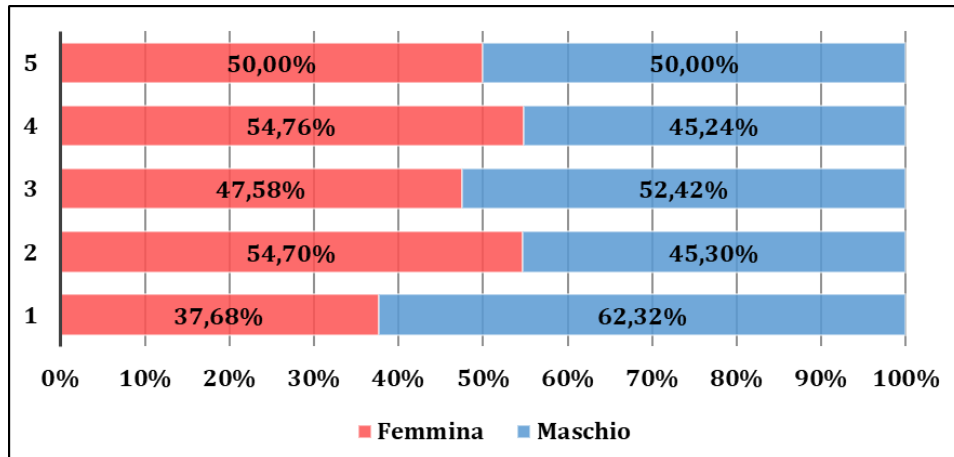


Grafico 4.13: Distribuzione percentuale della variabile "Genere" in base alla "Risposta" se "Sesso" = Femmina. Il test χ^2 non è statisticamente significativo (p-value=0.155).

...**MASCHIO:** rispetto al caso precedente, per il *Grafico 4.14* le considerazioni da fare sono diverse perché il livello di significatività risulta essere del 5%. Dalle analisi multivariate invece non risulterà significativa questa relazione, quindi probabilmente per i personaggi delle *vignettes* di genere maschile ci saranno relazioni tra l'opinione e le altre caratteristiche che vengono confuse dalla caratteristica del genere dell'intervistato.

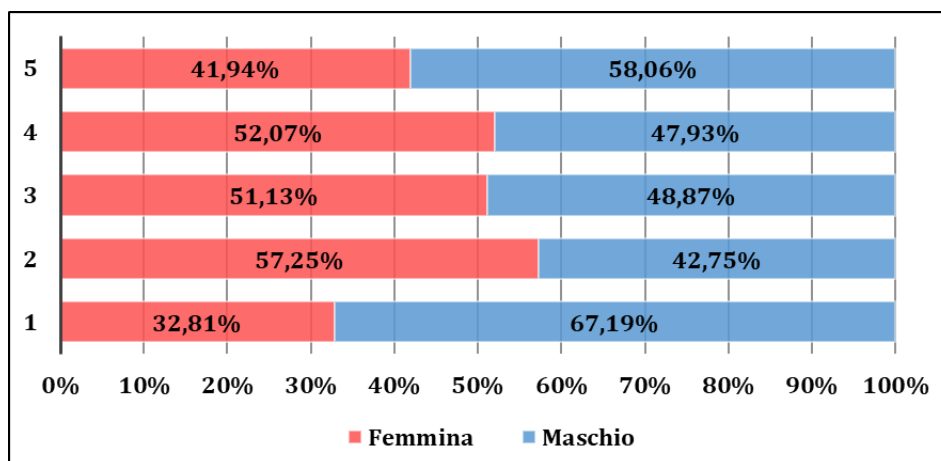


Grafico 4.14: Distribuzione percentuale della variabile "Genere" in base alla "Risposta" se "Sesso"=Maschio. Il test χ^2 è statisticamente significativo ad un livello di 5% (p-value=0.022).

Senza soffermarsi sulle risposte "3" e "4" che sono abbastanza bilanciate, già alla risposta "2" si può notare una maggioranza di voti femminili (57,25%) verso le *vignettes* col personaggio di genere maschile, mentre le risposte "1" e "5" hanno una predominanza di voti maschili. Alla risposta "5" hanno votato un 58,06% di uomini alle *vignettes* con protagonisti Alessandro o Francesco, contro il 41,94% di voti

femminili. Alla risposta “1” – pieno disaccordo - c’è il picco di voti maschili (67,19%) alle *vignettes* con personaggi dello stesso loro genere; solo un 32,81% delle rispondenti donne non era d’accordo con l’acquisto da parte di uno dei due personaggi maschili.

4.2.3.2 “Risposta” vs “Posizione” se “Set”=C

In questa analisi trivariata, per quasi tutti i *sets* l’indice χ^2 di Pearson non è risultato significativo; l’unico caso dove si è avuto un p-value significativo, al 10% (0,062), è stato per il questionario C che sarà quindi l’unico preso in considerazione in questo studio (per il *set* di *vignettes* presente in quel questionario si faccia riferimento all’Appendice A.3).

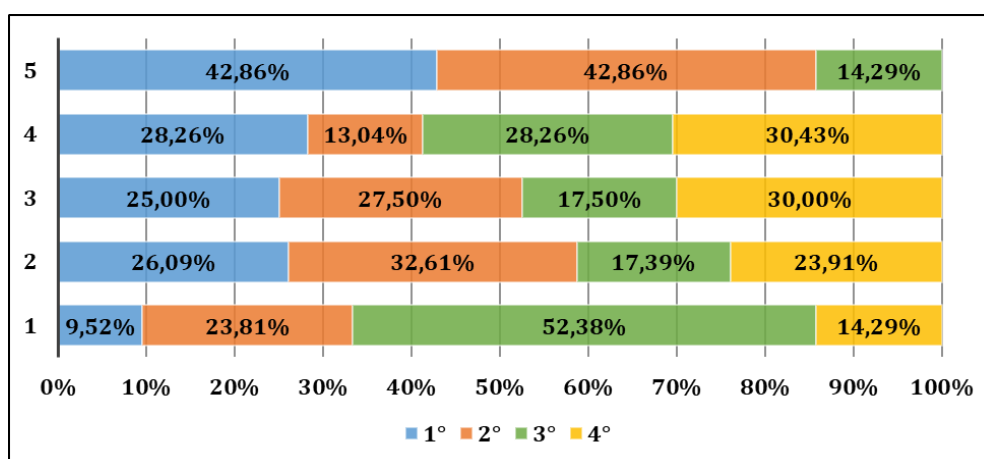


Grafico 4.15: Distribuzione percentuale della variabile "Posizione" in base alla "Risposta" se "Set"=C. Il test χ^2 è statisticamente significativo ad un livello di 10% (p-value=0.062).

La prima cosa che si può notare dal *Grafico 4.15* è l’assenza di risposte “5” alla *vignette* posta in 4° posizione, mentre la maggior parte dei voti (a pari merito) è andata alle due *vignettes* posizionate all’inizio, entrambe con un 42,86% di voti “Completamente d’accordo”.

Per chi invece ha risposto “Per niente d’accordo” la maggior parte dei voti sono andati alla *vignette* in terza posizione, col 52,38%; pochissimi non erano d’accordo con la *vignette* posizionata per prima (9,52%) e solo un po' più alta la preferenza per quella proposta in ultima posizione (14,29%).

Alla risposta “4” si evidenzia quel 13,04% di *vignette* in seconda posizione, mentre per le risposte “2” e “3” sembrano esserci percentuali simili per posizione.

Nell’analisi multivariata tutte le categorie della variabile “Set” non risulteranno significative, mentre per “Posizione” lo risulteranno con un p-value del 10% la 3° e la

4° posizione (la 4° solo per il multilivello), che già dal *Grafico 4.15* mostrano avere andamenti “particolari”.

4.3 Modello *Ordered Probit*

Le categorie di riferimento delle variabili esplicative che sono state inserite nei relativi modelli sono le seguenti:

- ❖ Posizione: 1°;
- ❖ Set: Questionario A;
- ❖ Sesso: Femmina;
- ❖ Prezzo: Inferiore ai 300€;
- ❖ Display: Inferiore a 5”;
- ❖ GB: 8;
- ❖ Scelta: Solo prestazioni;
- ❖ Motivo: Rottura dello smartphone;
- ❖ Genere: Femmina;

Queste modalità fanno riferimento al “gruppo base” delle unità statistiche che presentano le caratteristiche in elenco, rispetto le quali verranno effettuati i confronti.

I modelli che sono stati implementati presentano un buon adattamento ai dati e le stime sono state riportate nella *Tabella 4.1*; non verranno invece riportate le soglie poiché per le analisi svolte non erano di particolare interesse.

Le analisi realizzate sulle due tipologie di modelli (*Ordered Probit* e multilivello) si riferiscono alle stesse variabili esplicative (rispettivamente nei casi “Solo *Vignettes*” e “Tutte le variabili”) e riportano stime puntuali e p-value molto simili. Per questo motivo il comportamento delle variabili esplicative verrà descritto solo in questo paragrafo analizzando la *Tabella 4.1*. Per i modelli multilivello verranno in particolar modo commentati i valori degli ICC ed alcuni p-value che differiscono da quelli stimati nei modelli non gerarchici.

	Solo Vignettes		Tutte le variabili	
Risposta	Stima	P-value	Stima	P-value
Posizione				
2	-.0883	0.358	-.0888	0.356
3	-.1740	0.072	-.1778	0.066
4	-.1481	0.126	-.1520	0.116
Set				
2	-.1090	0.369	-.0995	0.414
3	.0898	0.445	.1060	0.381
4	-.0487	0.687	-.0477	0.699
5	.0810	0.497	.0402	0.740
6	-.1389	0.248	-.1830	0.136
Sesso	-.0711	0.299	-.0708	0.301
Prezzo	-.6485	0.000	-.6556	0.000
Display	.2218	0.001	.2240	0.001
GB				
1	.7674	0.000	.7789	0.000
2	1.1709	0.000	1.1877	0.000
Scelta				
2			-.2886	0.009
3			-.2120	0.085
4			-.3091	0.058
5			-.5926	0.294
Motivo				
2			-.0226	0.781
3			.2291	0.112
4			.3747	0.125
5			.4002	0.048
6			-.2654	0.180
Genere			-.1423	0.040
Log Likelihood	-1331.3798		-1319.6982	

Tabella 4.1: Stime e p-value per i modelli *Ordered Probit* calcolati con le sole variabili delle vignettes (“Solo Vignettes”) e con “Tutte le variabili”.

Una prima considerazione da fare osservando la *Tabella 4.1* (lo stesso vale nel metodo multilivello descritto dalla *Tabella 4.3*) è che le stime comuni dei modelli “Solo Vignettes” e “Tutte le variabili” sono uguali a meno di lievi differenze a livello di terza e quarta cifra decimale. I commenti che verranno espressi riguardo i vari effetti delle variabili varranno sia per l’uno che per l’altro modello.

Molte variabili, tra cui “Posizione” (tranne la categoria “3”), “Set”, “Sesso” ed alcune categorie di “Scelta” e “Motivo”, non presentano stime statisticamente significative per cui, non influenzando la variabile “Risposta”, non saranno analizzate.

A parità di altre condizioni, un prezzo superiore ai 300€ porta ad una maggiore probabilità di non essere d’accordo sull’acquisto dello smartphone, rispetto ad un

prezzo basso, mentre se il display supera i 5" la probabilità di essere favorevoli ad acquistare il dispositivo è più alta rispetto a quella di avere un display più piccolo.

Per la dimensione della memoria del telefono, maggiori sono i GigaBytes proposti, maggiore è la probabilità che l'intervistato sia d'accordo con l'acquisto del telefono.

La stima del genere dell'intervistato, con una significatività del 5%, indica che i maschi, rispetto alle femmine, hanno minore probabilità di concordare con la scelta di acquisto di uno smartphone, a parità di altre condizioni.

Rispetto la variabile "Posizione", è risultata significativa (p-value al 10%) solo la 3° categoria: ciò denota che la probabilità dell'intervistato di essere d'accordo è più bassa se le *vignettes* sono poste in 3° posizione rispetto a quelle poste all'inizio. Nell'analisi multilivello è significativa però anche la 4° posizione.

Quando si tratta di scegliere lo smartphone, se si guardano le prestazioni come prima opzione e poi si valuta la marca, la probabilità di essere d'accordo sull'acquisto cala rispetto a scegliere lo smartphone solo in base alle prestazioni; con un p-value del 10% la probabilità di concordare sull'acquisto cala anche se si tiene conto di marca e prestazioni allo stesso modo, o se si comincia a dare più importanza alla marca rispetto alle prestazioni; entrambi sono da riferirsi rispetto a chi sceglie lo smartphone unicamente in base alle prestazioni.

Per la variabile "Motivo" si riesce a fare un confronto significativo (al 5%) solo tra la rottura dello smartphone e l'uscita del nuovo modello: a parità di altre condizioni, l'uscita del nuovo modello porta ad una maggiore probabilità dell'intervistato di essere d'accordo con l'acquisto del telefono con le caratteristiche presentate nella vignette rispetto al caso in cui avvenga la rottura del proprio smartphone.

La *Tabella 4.2*, che mostra l'analisi fatta con un modello *Multinomial Probit* per variabili non ordinali, è stata inserita per mostrare semplicemente che se si fosse scelto di codificare la variabile risposta come categoriale sconnessa, le relazioni con le variabili esplicative di fatto non sarebbero cambiate: osservando le stime si nota che mantengono tutte gli stessi segni e che i p-value continuano ad indicare, chi più chi meno ma sempre nel *range*, gli stessi livelli di significatività del modello *Ordered Probit* della *Tabella 4.1*.

Risposta 2	Stima	P-value		Risposta 3	Stima	P-value
Costante	.8551	0.004		Costante	.7481	0.011
Posizione				Posizione		
2	-.2525	0.290		2	-.2796	0.244
3	-.8449	0.000		3	-.8709	0.000
4	-.8243	0.000		4	-.8466	0.000
Set				Set		
2	-.4662	0.090		2	-.5930	0.030
3	-.0042	0.987		3	-.2236	0.399
4	-.0279	0.920		4	-.4554	0.107
5	.1652	0.580		5	.2519	0.392
6	-.1668	0.539		6	-.4385	0.106
Sesso	.1839	0.909		Sesso	-.0060	0.970
Prezzo	-.3812	0.021		Prezzo	-.4414	0.007
Display	.1647	0.315		Display	.2990	0.068
GB				GB		
1	.6915	0.000		1	.9920	0.000
2	.7060	0.001		2	1.3062	0.000
Risposta 4	Stima	P-value		Risposta 5	Stima	P-value
Costante	.3228	0.292		Costante	-.9090	0.030
Posizione				Posizione		
2	-.4394	0.082		2	-.2623	0.378
3	-.6072	0.011		3	-.8768	0.003
4	-.6093	0.011		4	-.7407	0.012
Set				Set		
2	-.6078	0.037		2	.0449	0.911
3	.1956	0.473		3	.3708	0.354
4	-.4159	0.165		4	.1979	0.628
5	.2021	0.507		5	.3917	0.364
6	-.4090	0.152		6	.1482	0.699
Sesso	-.1091	0.514		Sesso	-.4022	0.064
Prezzo	-1.1012	0.000		Prezzo	-1.4841	0.000
Display	.3265	0.054		Display	.7077	0.001
GB				GB		
1	1.6790	0.000		1	1.5269	0.000
2	2.1433	0.000		2	2.4370	0.000

Tabella 4.2: Stime e p-value per il modello *Multinomial Probit* con variabile risposta non ordinata, calcolato con le sole variabili delle *vignettes*.
La stima della log-verosimiglianza in questo modello vale -1331.3798.

4.4 Analisi Multilivello

Le modalità di riferimento delle variabili esplicative che sono state inserite nei relativi modelli gerarchici e facenti riferimento al “gruppo base” sono le stesse di quelle considerate per i modelli *Ordered Probit*.

Il modello multilivello è un metodo per tenere conto delle caratteristiche non osservabili degli individui (che sono appunto le unità di secondo livello di questa analisi). Lo scopo di questa analisi è individuare se ci siano differenze di opinioni tra gli individui del questionario, che sono comunque simili tra loro in termini di età e background di istruzione.

Per i modelli multilivello che sono stati implementati, come per gli *Ordered Probit*, si presenta un buon adattamento ai dati. Analogamente al paragrafo precedente non verranno riportate le stime delle soglie.

Della *Tabella 4.3*, come già anticipato, verrà analizzata la stima del valore degli ICC, nonché le variabili e le categorie che differiscono in significatività rispetto alla stima del modello *Ordered Probit*.

Mentre nel modello non gerarchico la posizione “4” era risultata non statisticamente significativa, nella stima del modello multilivello essa evidenzia un p-value di poco inferiore a 10% di livello: la probabilità dell’intervistato di non essere d’accordo è più alta se la *vignette* era posta in 4° posizione rispetto a quella posta in 1°.

La categoria di “Scelta” numero 2, ovvero valutare le prestazioni ma poi tenere conto anche della marca, da un livello di significatività più basso di 1% passa ad uno di 5% nei modelli multilivello.

Per quanto riguarda “Genere”, la variabile continua a risultare significativa, ma non più al 5% bensì al 10% di livello. Lo stesso vale per il motivo addotto per cambio dello smartphone: l’uscita del nuovo modello.

I valori dei due ICC non sono molto alti, ad indicare che le variabili introdotte riescono a spiegare una certa variabilità tra i gruppi, che in questo caso risultano essere gli individui; c’è da sottolineare che questi indici non sono comunque da trascurare poiché si passa da un’ICC pari a 13,61%, ad un ICC che scende di soli 2,38 punti percentuali fino a 11,23%, a fronte dell’introduzione di un buon numero di variabili di secondo livello (colonna di destra della *Tabella 4.3*). Inoltre, considerando

che per ogni gruppo (l'individuo) si hanno solo quattro osservazioni (le *vignettes* presenti nel questionario), i valori ICC indicano che esiste un'eterogeneità di una certa importanza tra gli studenti del dipartimento di Scienze Statistiche.

Risposta	Solo <i>Vignettes</i>		Tutte le variabili	
	Stima	P-value	Stima	P-value
Posizione				
2	-.0950	0.327	-.0944	0.330
3	-.1896	0.052	-.1905	0.051
4	-.1612	0.098	-.1626	0.095
Set				
2	-.1167	0.440	-.1052	0.473
3	.0987	0.505	.1144	0.435
4	-.0540	0.720	-.0509	0.732
5	.0885	0.554	.0433	0.767
6	-.1521	0.312	-.1959	0.186
Sesso	-.0770	0.264	-.0759	0.271
Prezzo	-.6972	0.000	-.6960	0.000
Display	.2379	0.001	.2374	0.001
GB				
1	.8281	0.000	.8287	0.000
2	1.2599	0.000	1.2607	0.000
Scelta				
2			-.3043	0.022
3			-.2231	0.134
4			-.3285	0.096
5			-.6142	0.362
Motivo				
2			-.0234	0.812
3			.2433	0.163
4			.3972	0.180
5			.4260	0.083
6			-.2948	0.220
Genere			-.1517	0.071
Varianza di 2° livello	.1575		.1265	
Log Likelihood	-1322.1725		-1313.3491	
ICC	13.61%		11.23%	

Tabella 4.3: Stime e p-value per i modelli multilivello calcolati con le sole variabili delle *vignettes* ("Solo *Vignettes*") e con "Tutte le variabili".

Nell'implementazione del metodo *Random Slope Model*, è stato introdotto un termine d'errore di 2° livello che è stato fatto interagire con la variabile di 1° livello "Sesso". Questa variabile non era risultata statisticamente significativa quando inserita come effetto diretto sulla variabile risposta. Facendola interagire col termine di errore di 2° livello si voleva valutare se gli intervistati inconsciamente o

indirettamente, giudicassero in modo diverso una stessa *vignette* interpretata da un uomo piuttosto che da una donna. È questo il motivo per cui è stata eseguita anche un'analisi trivariata. Nel confronto tra *Random Intercept Model* e *Random Slope Model* utilizzando l'LR test, questo è risultato non significativo, a confermare che lo *Slope* non serve perché non aggiunge ulteriori informazioni rispetto al *Random Intercept Model*; si sono considerate allora solo le analisi dell'*Intercept Model*.

Capitolo 5.

Conclusioni

Questo lavoro di tesi si è basato sulla costruzione, somministrazione e poi analisi di un questionario che ha come metodo quello delle *Vignettes Studies*, ovvero l'unione di due tecniche: quella delle *experimental vignettes* e quella del questionario tradizionale.

La domanda su cui è stata focalizzata l'attenzione, e sulla quale è stato poi creato il questionario, è "Quali sono i fattori che influiscono l'acquisto di uno smartphone?".

Il campione a cui è stato somministrato tale questionario ha caratteristiche simili in termini di età (19-25 anni) e background di istruzione (Dipartimento di Scienze Statistiche dell'Università degli Studi di Padova), inoltre è stato deciso di consegnarlo in ugual numero a maschi e femmine (120 e 120), per un totale di 240 intervistati.

Gli obiettivi posti all'inizio dello studio erano principalmente tre:

- ❖ Individuare una o più caratteristiche influenzanti la scelta di acquisto dello smartphone;
- ❖ Capire se esistono differenze di genere nella valutazione delle *vignettes*, sia rispetto al rispondente che rispetto al personaggio della storia;
- ❖ Analizzare l'importanza della marca nel processo di scelta d'acquisto del bene.

A queste si sono aggiunte considerazioni rispetto:

- ❖ L'influenza che può avere la posizione della *vignette* sulle risposte dell'intervistato;
- ❖ Il motivo per il quale un individuo potrebbe decidere di cambiare smartphone.

Utilizzando i dati raccolti dal questionario, sono state svolte inizialmente delle analisi esplorative. In seguito sono state implementate due tipologie di modelli: il modello *Ordered Probit* ed il modello gerarchico *Random Intercept Model*. Si è anche verificato che il *Random Slope Model* non avrebbe apportato nessuna valutazione aggiuntiva rispetto a quello adottato.

Si è notato che variabili risultate significative nelle analisi bivariate, come "Posizione" e "Set", non lo siano, in parte, più state nell'analisi multivariata. Viceversa variabili come "Scelta" e "Motivo", che non erano significative se analizzate

singolarmente con la risposta, lo sono poi diventate (per alcune categorie) con i modelli stimati.

Le tre caratteristiche, “Prezzo”, “Display” e “GB”, hanno dato conferma della loro influenza nell’acquisto degli smartphone, portando alle seguenti conclusioni:

- ❖ “Prezzo”: la probabilità che uno studente di Scienze Statistiche sia d’accordo con l’acquisto è più alta se il prezzo dello smartphone risulta essere inferiore ai 300€ piuttosto che più alto, a conferma del fatto che gli studenti generalmente non hanno autonomia monetaria oppure il reddito disponibile è basso (Daejoong, Heasun e Hyunjoo, 2014). L’81,33% degli studenti sono stati d’accordo con il personaggio della *vignette* se questo valutava l’acquisto di uno smartphone che costasse meno di 300€, mentre le valutazioni “Per niente d’accordo” sono state date per la maggior parte (70,68%) ad acquisti sopra i 300€;
- ❖ “Display”: quando la grandezza dello schermo supera i 5”, la probabilità di essere favorevoli all’acquisto è maggiore, rispetto ai display più piccoli. A rispondere “Completamente d’accordo” sull’acquisto di uno smartphone superiore ai 5” è stato un 65,33% degli intervistati. Questo risultato sottolinea come il fattore sia molto importante nella scelta di acquisto, un esito confermato anche da altri studi (Siu-Tsen Shen, 2015, Na e Ruifeng, 2017);
- ❖ “GB”: altro fattore che ha ricevuto esiti rilevanti, poiché con lo sviluppo della tecnologia c’è sempre maggiore richiesta di memoria interna per archiviare i dati. Dall’analisi è emerso che più è grande la memoria proposta e più aumenta la probabilità che i rispondenti siano d’accordo con la scelta di acquisto dello smartphone. Un 65,33% degli intervistati era “Completamente d’accordo” con l’acquisto dello smartphone se questo aveva 32GB, mentre un 67,67% degli individui era “Per niente d’accordo” se la memoria era da 8GB.

La stima della variabile relativa al genere dell’intervistato (“Genere”) esprime una probabilità minore negli uomini rispetto alle donne di concordare sull’acquisto di un dispositivo; nell’analisi bivariata era risultata una percentuale più alta di uomini (64,66%) alla risposta “Per niente d’accordo” rispetto alle donne (35,34%).

Gli studenti di Scienze Statistiche non danno molta importanza alla marca (uno 0,42% ha votato di scegliere lo smartphone solo ed esclusivamente in base a quella e un 7,92% oltre alla marca poi valuta le prestazioni) e il 54,58% sceglie lo smartphone

principalmente in base alle prestazioni valutando solo in un secondo momento la marca; se si aggiunge anche la quota di quelli che guardano solo le prestazioni, si arriva al 67,08% della popolazione di riferimento. Un 24,58% guarda in ugual misura le due componenti. Nell'analisi bivariata questa variabile è risultata non significativa, ma ha poi assunto importanza per le categorie "2" ("prima prestazioni e poi marca"), "3" ("entrambe in ugual misura") e "4" ("prima marca e poi prestazioni") nell'analisi multivariata.

Nelle successive analisi multivariate mediante un modello *Ordered Probit* gli individui che guardano solo le prestazioni vengono presi come "gruppo base", rispetto ai quali proporre certi confronti. A parità di altre condizioni chi valuta inizialmente le prestazioni e poi la marca ha una probabilità più bassa di essere d'accordo sull'acquisto; lo stesso vale per chi tiene in considerazione entrambe le caratteristiche e per chi valuta prima marca e poi prestazioni. La marca qui ha un ruolo poco importante poiché si preferiscono le caratteristiche tecniche all'immagine del brand.

Ulteriori risultati che sono emersi dall'analisi riguardano la posizione delle *vignettes*. Rispetto ad una *vignette* posta all'inizio, a parità di altre condizioni, l'intervistato ha una probabilità maggiore di non essere d'accordo con l'acquisto qualora la *vignette* sia posizionata al 3° o 4° posto. Tra i rispondenti che avevano votato "Per niente d'accordo", il 33,83% aveva dato quel giudizio alla 3° *vignette* del questionario, un 33,08% lo aveva dato a quella posta in ultima posizione.

Non è risultato significativo il tipo di questionario ("Set" A-B-C-D-E-F), contro una prima analisi esplorativa che dava al "C" una correlazione tra "Posizione" e "Risposta".

Rispetto i motivi per cui si potrebbe scegliere di cambiare lo smartphone, il 55,42% dei rispondenti preferisce aspettarne la rottura, un 29,17% lo cambierebbe perché obsoleto mentre soltanto un 3,33% lo sostituirebbe all'uscita del nuovo modello. Le stime di un modello multivariato portano invece a concludere che seguire il trend del mercato ("uscita di un nuovo modello") è l'unico risultato statisticamente significativo a riguardo dell'aspetto del motivo di un eventuale cambio dello smartphone: se confrontato con chi cambierebbe lo smartphone solo dopo la rottura di quello in uso, la probabilità dell'intervistato di essere d'accordo con l'acquisto dello smartphone, nel caso in cui ci sia l'uscita del nuovo modello, è maggiore.

Il genere del personaggio descritto nello scenario ("Sesso") non ha evidenziato significatività in nessuna analisi, non influisce quindi nella valutazione degli studenti sul grado di accordo per l'acquisto dello smartphone descritto in ogni *vignette*. Questo

contrariamente a quanto emerso in altri studi (Kapteyn et al., 2007, Wason, Polonsky e Hymann, 2002).

Col modello multilivello si può ricavare una ulteriore considerazione riguardante la differenza tra i "gruppi" (che per questa analisi sono gli individui), aspetto che le stime del modello *Probit* non potevano far emergere: l'indice ICC ha fatto emergere un'eterogeneità di una certa importanza tra gli studenti del dipartimento di Scienze Statistiche, pur essendo un campione omogeneo per età e cultura scolastica.

Per approfondire lo studio delle variabili che hanno dato un risultato significativo si potrebbe ampliare la rosa di alternative da proporre. Ad esempio un numero maggiore di categorie per la variabile "Prezzo", proponendo diverse fasce piuttosto che due sole alternative. La soglia ipotizzata in questo lavoro è stata scelta in modo arbitrario e potrebbe aver generato percezioni diverse rispetto al suo significato, come ad esempio l'indicazione della qualità del prodotto da acquistare (Brucks, Zeithaml e Naylor, 2000).

Un ulteriore studio potrebbe prevedere un diverso modo di considerare la marca. Innanzitutto inserendola internamente alla *vignette*, questa volta specificandola (es. Samsung, Apple, Huawei, altro); in alternativa si potrebbe studiare il sistema operativo, chiedendo la preferenza tra Android ed iOS; questa ricerca non darebbe la stessa informazione di quella rispetto alla marca, ma potrebbe essere una valida alternativa.

Incrementare il numero di variabili o di categorie comporta un aumento considerevole di combinazioni di *vignettes*, una maggiore complessità computazionale e difficoltà nella creazione del questionario. In questo caso la complessità si potrebbe gestire creando solo dei sottogruppi di *vignettes*, tenendo quelle la cui storia abbia coerenza rispetto all'insieme delle nuove caratteristiche scelte per lo studio.

Come tutte le indagini, la scelta del campione influenza anche i risultati; il fatto di aver scelto un campione omogeneo per età e background culturale ha portato a dei risultati che probabilmente sarebbero stati diversi con la scelta di un altro gruppo di intervistati. Uno sviluppo potrebbe essere proprio quello di rendere più eterogeneo il campione ed analizzare i nuovi risultati che potrebbero derivarne.

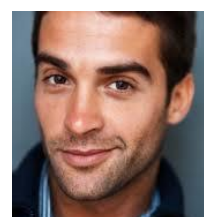
Appendice

In seguito vengono proposti i 6 *sets* di *vignettes* (senza mettere per ognuno anche l'ordine diverso dei singoli *sets*), alla fine dei quali è stata aggiunta la seconda parte del questionario, uguale per tutti i rispondenti.

A.1 *Vignettes* del questionario A

1. Alessandro è orientato su un modello di smartphone con le seguenti caratteristiche:

- Prezzo superiore a 300€;
- Display superiore ai 5”;
- Memoria interna di 8 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

2. Sofia è orientata su un modello di smartphone con le seguenti caratteristiche:

- Prezzo superiore a 300€;
- Display inferiore ai 5”;
- Memoria interna di 32 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

3. Francesco è orientato su un modello di smartphone con le seguenti caratteristiche:

- Prezzo inferiore a 300€;
- Display inferiore ai 5”;
- Memoria interna di 16 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

4. Martina è orientata su un modello di smartphone con le seguenti caratteristiche:

- Prezzo inferiore a 300€;
- Display superiore ai 5";
- Memoria interna di 8 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

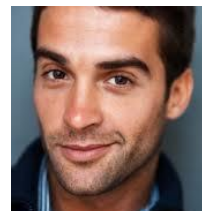
④

⑤

A.2 Vignettes del questionario B

1. Alessandro è orientato su un modello di smartphone con le seguenti caratteristiche:

- Prezzo superiore a 300€;
- Display superiore ai 5";
- Memoria interna di 16 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

2. Sofia è orientata su un modello di smartphone con le seguenti caratteristiche:

- Prezzo superiore a 300€;
- Display inferiore ai 5";
- Memoria interna di 8 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

3. Francesco è orientato su un modello di smartphone con le seguenti caratteristiche:

- Prezzo inferiore a 300€;
- Display superiore ai 5";
- Memoria interna di 32 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

4. Martina è orientata su un modello di smartphone con le seguenti caratteristiche:

- Prezzo inferiore a 300€;
- Display inferiore ai 5";
- Memoria interna di 16 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

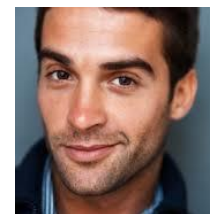
④

⑤

A.3 Vignettes del questionario C

1. Alessandro è orientato su un modello di smartphone con le seguenti caratteristiche:

- Prezzo superiore a 300€;
- Display superiore ai 5";
- Memoria interna di 32 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

2. Sofia è orientata su un modello di smartphone con le seguenti caratteristiche:

- Prezzo superiore a 300€;
- Display inferiore ai 5”;
- Memoria interna di 16 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

3. Francesco è orientato su un modello di smartphone con le seguenti caratteristiche:

- Prezzo inferiore a 300€;
- Display superiore ai 5”;
- Memoria interna di 8 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

4. Martina è orientata su un modello di smartphone con le seguenti caratteristiche:

- Prezzo inferiore a 300€;
- Display inferiore ai 5”;
- Memoria interna di 8 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

A.4 Vignettes del questionario D

1. Alessandro è orientato su un modello di smartphone con le seguenti caratteristiche:

- Prezzo superiore a 300€;
- Display inferiore ai 5”;
- Memoria interna di 8 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

2. Sofia è orientata su un modello di smartphone con le seguenti caratteristiche:

- Prezzo superiore a 300€;
- Display superiore ai 5”;
- Memoria interna di 16 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

3. Francesco è orientato su un modello di smartphone con le seguenti caratteristiche:

- Prezzo inferiore a 300€;
- Display inferiore ai 5”;
- Memoria interna di 32 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

4. Martina è orientata su un modello di smartphone con le seguenti caratteristiche:

- Prezzo inferiore a 300€;
- Display superiore ai 5”;
- Memoria interna di 32 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

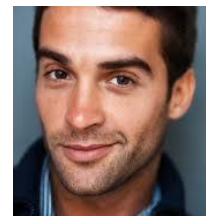
④

⑤

A.5 Vignettes del questionario E

1. Alessandro è orientato su un modello di smartphone con le seguenti caratteristiche:

- Prezzo superiore a 300€;
- Display inferiore ai 5”;
- Memoria interna di 16 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

2. Sofia è orientata su un modello di smartphone con le seguenti caratteristiche:

- Prezzo superiore a 300€;
- Display superiore ai 5”;
- Memoria interna di 32 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

3. Francesco è orientato su un modello di smartphone con le seguenti caratteristiche:

- Prezzo inferiore a 300€;
- Display inferiore ai 5”;
- Memoria interna di 8 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

4. Martina è orientata su un modello di smartphone con le seguenti caratteristiche:

- Prezzo inferiore a 300€;
- Display superiore ai 5”;
- Memoria interna di 16 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

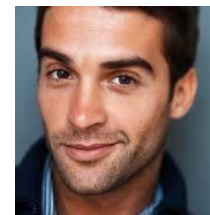
④

⑤

A.6 *Vignettes* del questionario F

1. Alessandro è orientato su un modello di smartphone con le seguenti caratteristiche:

- Prezzo superiore a 300€;
- Display inferiore ai 5”;
- Memoria interna di 32 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

2. Sofia è orientata su un modello di smartphone con le seguenti caratteristiche:

- Prezzo superiore a 300€;
- Display superiore ai 5”;
- Memoria interna di 8 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

3. Francesco è orientato su un modello di smartphone con le seguenti caratteristiche:

- Prezzo inferiore a 300€;
- Display superiore ai 5”;
- Memoria interna di 16 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

4. Martina è orientata su un modello di smartphone con le seguenti caratteristiche:

- Prezzo inferiore a 300€;
- Display inferiore ai 5”;
- Memoria interna di 32 GB.



In una scala da 1 a 5 (dove 1 non sei per niente d'accordo e 5 sei completamente d'accordo), quanto condividi la scelta di questo acquisto?

Per niente d'accordo

①

②

③

Completamente d'accordo

④

⑤

A.7 Parte di questionario comune

Quale delle seguenti affermazioni caratterizzerebbe la scelta del tuo smartphone qualora volessi acquistarne uno?

(Segnare una sola risposta tra quelle proposte)

- Solo le prestazioni, indipendentemente dalla marca
- Le prestazioni come prima scelta, e poi valuti la marca
- Prestazioni e marca in egual misura
- La marca come prima scelta, e poi valuti le prestazioni
- Solo la marca, indipendentemente dalle prestazioni

Qual è il motivo per cui hai cambiato o cambieresti smartphone?

(Segnare una sola risposta tra quelle proposte)

- Rottura del tuo smartphone
- Obsolescenza (tecnologia sorpassata) del tuo smartphone
- Offerta vantaggiosa/Promozione su un altro smartphone (anche se il tuo attuale fosse funzionante) visto in negozio/internet
- Operatore telefonico propone una buona offerta tariffaria comprensiva di smartphone
- Uscita del nuovo modello (es: hai il modello #5 ed esce il #6)
- Altro

Genere dell'intervistato



Riferimenti Bibliografici

- ❖ AGUINIS H. e BRADLEY K. J., “*Best practice recommendations for designing and implementing experimental vignette methodology studies*”, *Organizational Research Methods*, 17 (4), pp. 351-371, 2014.
- ❖ ALEXANDER C., BECKER H., “*The Use of Vignettes in Survey Research*”, *Public Opinion Quarterly*, 42 (1), pp 93-104, 1978.
- ❖ ATZMÜLLER C., STEINER P., “*Experimental Vignette Studies in Survey Research, Methodology*”, 6 (3), pp. 128-138, 2010.
- ❖ BICKEL R., “*Multilevel Analysis for Applied Research. It’s Just Regression!*”, Guilford Press, New York – London, 2007.
- ❖ BRUCKS M., ZEITHAML V., NAYLOR G., “*Price and brand name as indicators of quality dimensions for consumer durables*”, *Journal of the Academy of Marketing Science*, 28 (3), pp. 359-374, 2000.
- ❖ CAPPUCCIO N., ORSI R., “*Econometria*”, il Mulino, 2005.
- ❖ CHING-HSUAN Y., YI-SHUN W., KAILI Y., “*Predicting smartphone brand loyalty: Consumer value and consumer-brand identification perspectives*”, *International Journal of Information Management*, 36 (3), pp. 245–257, 2016.
- ❖ DAEJOONG K., HEASUN C., HYNJOO L., “*Determining the factors that influence college students’ adoption of smartphones*”, *Journal of the Association for Information Science and Technology*, 65 (3), pp. 578-588, 2014.
- ❖ DE LEEUW J., MEIJER E., “*Handbook of Multilevel Analysis*”, Springer, 2008.
- ❖ GOLDSTEIN H., “*Multilevel statistical Models*”, 4th Edition, Wiley, . 2011.
- ❖ HEDEKER D., GIBBONS R., “*A Random-Effects Ordinal Regression Model for Multilevel Analysis*”, *International Biometric Society*, 50 (4), 1994.
- ❖ HEW J., BADARUDDIN M., KRISHNA MOORTHY M., “*Crafting a smartphone repurchase decision making process: Do brand attachment and gender matter?*”, *Telematics and Informatics*, 34 (4), pp. 34–56, 2016.
- ❖ HOLLAND P., WAINER H., “*Differential Item Functioning, Hillsdale*”, NY: Lawrence Erlbaum, 1993.
- ❖ HOX J., “*Multilevel Analysis Techniques and Applications*”, Lea, London, 2002.
- ❖ JURGES H., WINTER J., “*Are Anchoring Vignettes Ratings Sensitive To Vignette Age And Sex?*”, *Health Economics* 22: 1-13, 2013.

- ❖ KAPTEYN A., SMITH J.P., VAN SOEST A., *“Vignettes and self-reports of work disability in the United States and the Netherlands”*, American Economic Review, 97, pp. 461–473, 2007.
- ❖ KING G., MURRAY C., SALOMON J. TANDON A., *“Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research”*, American Political Science Review, 98 (1), pp. 191-207, 2004.
- ❖ KREFT I., DE LEEUW J., *“Introducing Multilevel modeling”*, Sage Publications, London, 1998.
- ❖ NA LIU, RUIFENG YU, *“Identifying design feature factors critical to acceptance and usage behavior of smartphones”*, Computers in Human Behavior, 70, pp. 131-142, 2016.
- ❖ PACCAGNELLA O., Materiale didattico del corso *“Strumenti Statistici per l’Analisi dei Dati Aziendali”*, Università degli studi di Padova, Dipartimento di Scienze Statistiche, 2016.
- ❖ PINHEIRO J., CHAO E., *“Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models”*, Journal of Computational and Graphical Statistics, 15 (1), pp. 58-81, 2006.
- ❖ RAUDENBUSH S., BRYK A., *“Hierarchical Linear Models. Applications and Data Analysis Methods”*, 2nd Edition, Sage Publications, London, 2002.
- ❖ SANG HYUN KIM, *“Moderating effects of Job Relevance and Experience on mobile wireless technology acceptance: Adoption of a smartphone by individuals”*, Information & Management, 45 (6), pp. 387-393, 2008.
- ❖ SIU-TSEN SHEN, *“The Digital Generation: Comparing and Contrasting Smartphone Use in the Digital Age”*, Journal of Internet Technology, 16 (1), pp. 121-127, 2015.
- ❖ SNIJDERS T., BOSKER R. , *“Multilevel Analysis. An introduction to Basic and Advanced Multilevel Modelling”*, Sage Publications, London, 1999.
- ❖ VERKASALO H., LÓPEZ-NICOLÁS C., MOLINA-CASTILLO F., BOUWMAN H. , *“Analysis of users and non-users of smartphone applications”*, Telematics and Informatics, 27 (3), pp. 242–255, 2010.
- ❖ WASON K., POLONSKY M. e HYMAN M., *“Designing Vignette Studies in Marketing”*, Australasian Marketing Journal, 10 (3), pp. 41-58, 2002.

Riferimenti Sitografici

- ❖ <http://www.idc.com/getdoc.jsp?containerId=prUS42507917>
- ❖ <https://www.lenius.it/nomi-piu-diffusi-in-italia/>
- ❖ <https://www.istat.it/it/prodotti/contenuti-interattivi/calco-latori/nomi>
- ❖ <http://www.dday.it/redazione/14042/ventanni-di-smartphone-ibm-simon-e-stato-il-primo>
- ❖ http://www.silicon.it/mobility/i-dieci-anni-delliphone-nascita-e-rivoluzione-del-mercato-smartphone-109885?inf_by=598d6602671db8a6638b48de
- ❖ <https://www.canalys.com/newsroom/smart-phones-overtake-client-pcs-2011>
- ❖ <http://gs.statcounter.com/press/mobile-and-tablet-internet-usage-exceeds-desktop-for-first-time-worldwide>

Ringraziamenti

Un ringraziamento particolare lo dedico al Professor Omar Paccagnella, per la sua grande disponibilità e l'enorme aiuto datomi, anche durante il suo periodo di vacanze.

Un grazie immenso alla mia famiglia, che mi ha sostenuta dall'inizio alla fine di questo percorso. Grazie per tutte le opportunità che mi avete dato Maurizio, Donatella, Alberto!

Ringrazio Stefano per essere sempre stato al mio fianco ed avermi supportata quando ne avevo bisogno e sopportata nei momenti di crisi!

Grazie a tutte quelle persone che nel loro piccolo mi hanno dato uno sprone ed un incitamento a non arrendermi.