

UNIVERSITA' DEGLI STUDI DI PADOVA
CORSO DI LAUREA MAGISTRALE IN SCIENZE STATISTICHE



Anchoring Vignettes:
il modello Hopit applicato alla misura
della Customer Satisfaction

Tesi di Laurea in
Scienze Statistiche

Relatore
Prof. Omar Paccagnella

Presentata da
Giovanni Pergola

Anno Accademico 2011/2012

Indice

Elenco delle Figure	iv
Elenco delle Tabelle	vi
Introduzione	1
1 La Customer Satisfaction	3
1.1 Che cos'è	3
1.2 Come si misura	5
1.2.1 Metodi principali	5
1.2.2 Capire gli obiettivi	7
1.2.3 Selezionare gli attributi da valutare	8
1.2.4 Come capire l'importanza degli attributi	9
1.3 I Vantaggi	17
2 Le Anchoring Vignettes	19
2.1 Cosa sono	19
2.2 Assunzioni	20
2.3 Il metodo non parametrico	21
2.4 Il metodo parametrico	22
2.4.1 L'Ordered Probit Model	23
2.4.2 Il modello HOPIT	24
2.5 La stima del modello	27
2.6 Controllare la validità delle assunzioni	27
3 L'indagine	31
3.1 Il LISS panel	32

3.1.1	Il Campione	32
3.2	Il Questionario	33
4	L'analisi	37
4.1	Statistiche Descrittive	38
4.2	Stime dei Modelli	40
4.2.1	Modello Probit Ordinato	40
4.2.2	Modello Hopit	43
4.3	Counterfactuals	47
4.3.1	Benchmark: GruppoAB	48
4.3.2	Benchmark: Lavoro	50
4.3.3	Benchmark: Proprietà	50
	Conclusioni	55
	A Il Questionario	57
	Bibliografia	62
	Ringraziamenti	65

Elenco delle figure

1.1	Confronto tra dealer con più alta soddisfazione della clientela rispetto a quelli con più bassa soddisfazione [Myers, 1999].	18
2.1	Metodo non parametrico per l'utilizzo delle anchoring vignettes	21
3.1	Percentuale di uomini e donne all'interno del campione	34
3.2	Distribuzione dell'età all'interno del campione	35
3.3	Percentuale di intervistati che abita nelle diverse fasce di popolosità (valori per km^2)	35
4.1	Distribuzione dell'età nel sottocampione	39
4.2	Soddisfazione sull'utilizzo dello Smartphone per gli utenti del sottocampione	39
4.3	Soddisfazione sull'utilizzo dello Smartphone divisa per sesso	40
4.4	Probabilità di risposta per le varie categorie con le stime del modello Ordered Probit nel caso di riscontro di problemi o meno	43
4.5	Differenze di proporzioni tra Gruppo A e Gruppo B utilizzando come benchmark se soglie del Gruppo A	49
4.6	Differenze di proporzioni tra Gruppo A e Gruppo B utilizzando come benchmark le soglie del Gruppo B	49
4.7	Differenze di proporzioni tra lavoratori e non utilizzando come benchmark le soglie dei lavoratori	51
4.8	Differenze di proporzioni tra lavoratori e non utilizzando come benchmark le soglie dei non lavoratori	51
4.9	Differenze di proporzioni tra individui proprietari di casa e non utilizzando come benchmark le soglie dei proprietari dell'abitazione dove vivono	52

4.10 Differenze di proporzioni tra individui proprietari di casa e non utilizzando come benchmark le soglie dei non proprietari dell'abitazione dove vivono	53
---	----

Elenco delle tabelle

1.1	Esempio di tabella costruita con il confronto di coppie[Myers, 1999]	12
1.2	Esempio di valori riscaldati con la Thurstone Case V scale [Myers, 1999]	12
4.1	Suddivisione in gruppi del campione	37
4.2	Tipologie di acquisti per rispondenti	38
4.3	Stime del modello Ordered Probit	41
4.4	Probabilità di risposta per le varie categorie con le stime del modello Ordered Probit	42
4.5	Probabilità di risposta per le varie categorie con le stime del modello Ordered Probit di una persona con problemi nell'utilizzo del prodotto	42
4.6	Stime del modello Hopit	43
4.7	Differenze tra le soglie stimate da un modello Probit Ordinato e da un Hopit	46
4.8	Probabilità di risposta per le varie categorie con le stime del modello Ordered Probit confrontate alle stime del modello Hopit, per una persona che non ha riscontrato problemi.	46
4.9	Probabilità di risposta per le varie categorie con le stime del modello Ordered Probit confrontate alle stime del modello Hopit, per una persona che ha riscontrato problemi.	47
4.10	Probabilità di risposta per le varie categorie, di una persona che non vive in una casa di proprietà	47
4.11	Differenze tra le soglie stimate da un modello Hopit, fissando come benchmark il gruppo A e il gruppo B	48
4.12	Differenze tra le soglie stimate da un modello Hopit, fissando come benchmark gli individui lavoratori e non.	50

4.13 Differenze tra le soglie stimate da un modello Hopit, fissando come benchmark gli individui proprietari della casa dove vivono e non. . . . 52

Introduzione

Lo studio della Customer Satisfaction è diventato, negli ultimi tempi, una delle attività più importanti e impegnative per le aziende. Per poterla misurare, ci sono svariate metodologie, ognuna con i suoi pregi e difetti. Uno dei più utilizzati è la compilazione di un questionario, dove il cliente valuta la sua situazione ed esprime il suo stato di soddisfazione, in genere con una scala da 1 a 5. I risultati vengono poi modellati tramite un modello Probit Ordinato. Un difetto di questo approccio è la possibile non compatibilità delle risposte tra persone appartenenti a gruppi o popolazioni differenti, dovuta ad incomprensioni delle domande o ad un utilizzo di scale di risposta differenti. Questa incomparabilità è già conosciuta da tempo in ambiente sociologico e viene definita con DIF, *Differential Item Functioning*.

King *et al.* [2004] hanno introdotto l'approccio delle "Anchoring Vignettes", ovvero delle descrizioni di scenari che dovrebbero aggiungere informazioni, in modo da poter correggere le risposte dal DIF. Questi scenari rappresentano situazioni di vita o di utilizzo di un prodotto (in base all'oggetto dell'indagine) che devono essere valutate dall'intervistato con la stessa scala utilizzata per l'autovalutazione. Il modello che utilizza le informazioni aggiunte dalle vignettes si chiama Hopit ed è un'estensione del modello Probit Ordinato.

Alla base di questo approccio ci sono due assunzioni: *Response Consistency* e *Vignette Equivalence*. La prima prevede che gli intervistati utilizzino la stessa scala di misura sia per valutare la situazione espressa dalla vignettes, sia per la propria. La seconda prevede che le vignettes siano percepite da tutti gli intervistati nello stesso modo. Tuttavia, la validità di queste assunzioni è ancora oggi oggetto di dibattito.

In questo lavoro di tesi si utilizzano dati provenienti dal panel LISS¹, che è parte del programma olandese MESS². Questo panel è utilizzato per molteplici indagi-

¹Longitudinal Internet Studies for the Social Sciences

²Measurement and Experimentation in the Social Sciences

ni, in particolare quella analizzata, denominata VECS³, ha come oggetto la misura della customer satisfaction attraverso l'uso delle "Anchoring Vignettes" ed è stata proposta da Paccagnella Omar, Varriale Roberta e Bassi Francesca.

La tesi è sviluppata su 4 capitoli. Nel Capitolo 1 viene spiegato il concetto di Customer Satisfaction, la sua storia e i metodi per misurarla. Inoltre è presente uno studio che mostra l'importanza di avere e mantenere una clientela soddisfatta.

Nel Capitolo 2 si introducono le Vignettes, le assunzioni necessarie per la validità del modello e i metodi di utilizzo delle informazioni aggiuntive fornite dalle vignettes. Inizialmente viene presentato il metodo non parametrico, che utilizza le risposte date alle vignettes per riscaldare quelle relative all'autovalutazione. Successivamente si introduce il metodo parametrico, concentrandosi prima sul modello Ordered Probit e poi sul modello Hopit. Infine vengono riportati alcuni studi che hanno cercato di dimostrare la validità delle assunzioni alla base dei metodi appena descritti.

Nel Capitolo 3 vengono analizzati l'indagine e il questionario, facendo una breve descrizione del progetto LISS.

Nel Capitolo 4, infine, vengono analizzati i dati raccolti dalla prima rilevazione del progetto VECS nell'indagine LISS, confrontando i risultati ottenuti dal modello Hopit con quelli ottenuti dal modello Ordered Probit, in modo da evidenziare eventuali differenze sostanziali. Nell'ultima sezione vengono esaminati alcuni "Counterfactuals". Questi ultimi vengono proposti studiando come verrebbero modificate le proporzioni delle risposte date da un gruppo di individui, se si fossero utilizzate le soglie appartenenti ad un altro gruppo.

Tutte le analisi sono state effettuate utilizzando il software Stata⁴

³Vignette Evaluation of Customer Satisfaction

⁴©Copyright 1996-2012 StataCorp LP

Capitolo 1

La Customer Satisfaction

1.1 Che cos'è

Customer satisfaction è un termine entrato ormai nel comune linguaggio aziendale. Ma cosa si intende realmente quando si parla di “Customer Satisfaction”? La traduzione letterale italiana risulta “Soddisfazione del Cliente”, un concetto di comprensione apparentemente immediata, ma di non semplice realizzazione (o conseguimento). Il significato di soddisfazione viene spiegato attraverso la teoria della discrepanza, che introduce altri due concetti:

- **Aspettative:** rappresentate dalle attese e dai desideri, che i consumatori ripongono nei beni che hanno intenzione di acquistare o nei servizi di cui vorrebbero usufruire. In qualsiasi acquisto il consumatore pone delle aspettative, anche se non sempre ne è consapevole.
- **Prestazione Percepita:** rappresentata dai benefici e dalle performance percepite dal consumatore all'atto dell'utilizzo del bene o del servizio.

Secondo la teoria della discrepanza, il grado di soddisfazione si quantifica misurando la discrepanza tra le aspettative prima dell'acquisto e la prestazione percepita dopo l'utilizzo. Se quest'ultima risulta superiore rispetto alle aspettative, allora il consumatore si può definire soddisfatto e di conseguenza, probabilmente, fedele alla marca; in caso contrario, il consumatore risulterà insoddisfatto e in aggiunta alla, probabile, mancata fedeltà, potrebbe creare un ulteriore effetto negativo comunicando la propria esperienza ad altri. Nel caso in cui le aspettative raggiungano

esattamente la prestazione percepita, il consumatore risulterà soddisfatto, tuttavia potrebbe non essere fedele alla marca e quindi facilmente attratto da prodotti o servizi concorrenti. [Hayes, 2009]

Ma quando si cominciò a parlare di customer satisfaction? La ricerca della soddisfazione della clientela è un concetto che si è sviluppato solo nel dopoguerra. Un primo “embrione” di customer satisfaction si ebbe negli Stati Uniti negli anni '50 grazie alla General Electric, che propose il “Marketing Concept”, un nuovo modo di vedere il business, mettendo il cliente al centro di tutte le attività e impiegando tutte le risorse per soddisfare i suoi bisogni. Questo primo tentativo non ebbe molto successo, principalmente per tre motivi:

1. La maggior parte delle aziende di allora avevano delegato all'area marketing la ricerca di informazioni sul cliente. In questo modo però essa otteneva maggior potere all'interno dell'azienda, diventando fondamentale per la comprensione dei desideri del consumatore e con la capacità di coordinare le altre funzioni per soddisfare questi desideri. Tuttavia, non sempre esistevano le competenze all'interno dell'area marketing per coordinare le funzioni dell'azienda, portando quindi al fallimento del Marketing Concept.
2. Il marketing era stato studiato ed utilizzato per acquisire nuovi clienti e non per mantenere quelli già esistenti. Solo dopo alcune ricerche effettuate negli anni successivi si scoprì che il costo per ottenere nuovi clienti era cinque volte più elevato rispetto a quello per trattenerli. In questo modo l'impegno dell'azienda non si concentrava sull'obiettivo più profittevole.
3. Proprio perché le aziende non si focalizzavano nel cercare di trattenerne i clienti già acquisiti, non esisteva un modo per misurare la soddisfazione di questi. Diventava quindi impossibile pensare di poterla gestire. Si cominciò a superare questi problemi solo verso gli anni '80, quando nel mercato degli Stati Uniti entrarono aziende straniere, che aumentarono il livello di concorrenza. In questo contesto venne creata una “scorecard” per comprendere il grado di soddisfazione dei bisogni dei consumatori; questo nuovo strumento risultava molto utile, in quanto riusciva a raggiungere ogni angolo dell'organizzazione diffondendo la “voce del consumatore” in modo sistematico e oggettivo. Risultava quindi difficile ignorare o non capire la centralità che doveva la soddisfazione

del consumatore. Le aziende riuscirono così ad aumentare la soddisfazione dei propri clienti e soprattutto a rafforzare la fedeltà di quelli già acquisiti.[Myers, 1999]

1.2 Come si misura

1.2.1 Metodi principali

Esistono molti modi per misurare la customer satisfaction: attraverso la raccolta e lo studio dei reclami, l'analisi della fedeltà della clientela, il confronto della qualità del servizio/prodotto erogato dalla propria azienda con quella dei propri concorrenti, la procedura di autovalutazione, indagini sulla soddisfazione attraverso interviste o compilazione di questionari ed infine attraverso la raccolta e lo studio di commenti dei clienti. Andiamo ora ad analizzarli uno per uno:

- **I reclami:** sono un chiaro indicatore per misurare la soddisfazione della clientela e ogni azienda ha l'obiettivo di minimizzarli. Il problema principale riguardante l'utilizzo dei reclami per verificare la soddisfazione dei consumatori è che se questi risultano pochi o addirittura non sono presenti, non è necessariamente vero che tutti i clienti siano soddisfatti. Questo potrebbe accadere per due motivi:
 - l'impossibilità del cliente insoddisfatto ad esprimere la propria voce a causa dell'assenza di strumenti per raccogliere la testimonianza dei clienti. In questo caso, spesso i reclami vengono fatti al personale di vendita e difficilmente riescono a raggiungere il management e la dirigenza aziendali, rimanendo così inascoltati.
 - la non volontà del cliente insoddisfatto ad esprimere la propria voce: infatti nella maggior parte dei casi, un cliente insoddisfatto, invece di reclamare, preferisce rivolgersi ad un'azienda concorrente.
- **La fedeltà del cliente:** è un'indicatore utile soprattutto nei mercati dove è facile cambiare fornitore. Solo in mercati dinamici, infatti, un cliente fedele risulta sicuramente un cliente soddisfatto. In un mercato chiuso e con

grossi costi legati al cambio di fornitore, può succedere che anche un cliente insoddisfatto sia fedele.

- **La qualità del servizio/prodotto erogato:** consiste nel confrontare le proprie performance e livelli di qualità con quelli medi del settore o usando come benchmark i propri concorrenti. Questo è un metodo indiretto di misurazione della qualità, dove non si conosce direttamente dal cliente il suo grado di soddisfazione, ma è possibile intuirlo: se le performance sono migliori dei concorrenti, probabilmente si avranno clienti maggiormente soddisfatti.
- **Autovalutazione aziendale:** con questo metodo l'azienda si immedesima nel cliente, valutando essa stessa quanto potrebbe essere soddisfatto il consumatore. Questo è un metodo molto pericoloso e, soprattutto se le percezioni dell'azienda e quelle dei clienti sono distanti, potrebbe portare a decisioni completamente sbagliate.
- **Mystery Shopper:** più che misurare la soddisfazione, il mystery shopper misura la qualità del servizio, tema molto vicino alla soddisfazione. Il mystery shopper è il ricercatore che diventa consumatore, per sperimentare e valutare la qualità del servizio erogato. Il problema è che la valutazione è molto soggettiva e comunque, il giudizio del ricercatore potrebbe comunque essere distorto dal ruolo che sta esercitando.
- **Indagini sulla soddisfazione:** consistono nell'intervistare i clienti chiedendo direttamente come giudicano la loro soddisfazione. E' forse l'indicatore più adatto, anche se è un metodo piuttosto costoso. Inoltre, se effettuato presso il punto vendita i risultati potrebbero essere distorti a causa della presenza del personale nel luogo dell'intervista o durante la compilazione del questionario.
- **Commenti dei clienti:** si differenziano dai reclami in quanto si offre al consumatore non solo la possibilità di lamentarsi, ma anche di far conoscere la propria opinione con suggerimenti propositivi e commenti sull'operato dell'azienda. Anche con questo metodo, però, esiste la possibilità che solamente i clienti interessati ad un servizio migliore esprimano le proprie opinioni, mentre quelli scontenti preferiscano, nel caso in cui ce ne sia la possibilità, cambiare fornitore. In questo caso, i commenti non rispecchierebbero realmente il grado

di soddisfazione della clientela, risultando così distorti.[Gramma [1993];Myers [1999]]

1.2.2 Capire gli obiettivi

Di tutti questi metodi, questo lavoro si concentrerà principalmente sulle indagini tramite questionario poiché è questo il metodo utilizzato per lo studio. Prima di decidere come costruire questo questionario, bisogna capire quali sono gli obiettivi dell'indagine. Questi possono essere molteplici:

- **Curiosità dell'esecutivo:** Soprattutto nei primi anni di CSM¹ alcune aziende commissionavano delle indagini per avere una rapida occhiata sulla capacità dell'azienda di soddisfare i propri clienti e per conoscere eventuali problemi. Ad esempio una *health maintenance organization* ha chiesto ad un campione di suoi clienti di valutare l'azienda, su circa 20 attributi selezionati dal dipartimento di ricerca. Questi attributi furono selezionati senza prima aver effettuato un'indagine per conoscerne l'importanza relativa.
- **Trend e confronti:** molte aziende hanno superato la fase relativa alla mera curiosità ed hanno cominciato ad effettuare indagini ripetute e costanti nel tempo. Queste indagini poi, vengono utilizzate dal reparto marketing oppure da aziende specializzate nella customer satisfaction.
- **Incentivi ai dipendenti:** In questi casi la soddisfazione viene misurata per uno scopo interno, dare un'incentivo (spesso monetario) ai propri dipendenti, per cercare di aumentare la profittabilità dell'azienda. La misurazione diventa, quindi, sia una forma di controllo, sia una forma di valutazione del dipendente stesso.
- **Internal ownership:** tramite le indagini sulla soddisfazione vengono evidenziati eventuali problemi dell'azienda. Grazie a questi risultati, alcune persone vengono rese responsabili di risolverli in modo permanente. Questo metodo viene chiamato *ownership* dei risultati delle indagini sulla customer satisfaction.

¹customer satisfaction measurement

- **Mission aziendale:** alcune aziende, soprattutto ultimamente, hanno inserito la customer satisfaction come mission aziendale. In questi casi le indagini vengono realizzate proprio per controllare l'andamento della soddisfazione dei clienti e per essere pronti ad agire qualora questa scendesse sotto i termini dettati dalla mission.
- **Obiettivi multipli:** molte aziende hanno più di un obiettivo per il loro programma di *CSM*. In questo modo le indagini possono essere molto flessibili e servire a molteplici funzioni. Il vero problema è che questi obiettivi devono essere specificati in modo chiaro dal management, poiché ciascuno di questi può influenzare la struttura, i contenuti e le procedure del programma. [Myers, 1999]

1.2.3 Selezionare gli attributi da valutare

Le indagini sulla soddisfazione della clientela si basano sulla valutazione di una serie di attributi, unita ad una valutazione generale sul tema chiave dell'indagine. Questo può essere la *soddisfazione* (come nel nostro caso), la *fedeltà* e molti altri. Le aziende più grandi, per selezionare gli attributi da inserire nell'indagine, si affidano a ditte specializzate in indagini sulla customer satisfaction. Quelle più piccole, non potendo permettersi di esternalizzare questa funzione, selezionano loro stesse gli attributi che **pensano** siano importanti per il cliente, applicandoci una qualsiasi scala e conducendo l'indagine via mail o telefono. Ciò è sbagliato poiché la scelta degli attributi è una delle azioni più importanti da fare prima dell'inizio di un'indagine. Tutto ciò che non verrà inserito nell'indagine *fin dall'inizio* non potrà essere estratto in seguito durante l'analisi.

Uno dei metodi più utilizzati per conoscere quali sono gli attributi più importanti sono i **focus group**. Questo metodo ha però alcuni limiti: funziona bene quando i clienti sono concentrati in aree geografiche vicine, in modo che possano trovarsi con facilità e a basso costo. Alcune aziende, soprattutto quando si entra nell'area del BtoB², per ovviare a questo problema, organizzano dei focus group telefonici od in video-conferenza.

²business to business, azienda dedicata al commercio o all'erogazione di servizi per altre aziende e non per utilizzatori finali.

Esistono alcune categorie di prodotti e servizi che hanno dei clienti che non è possibile far riunire in un unico posto, oppure questi prodotti e servizi sono personali e valutabili soggettivamente. In questi casi il focus group non è la soluzione migliore, le aziende quindi preferiscono condurre delle interviste “in profondità” con utilizzatori o potenziali utilizzatori. In questo modo, risulta semplice per un intervistatore preparato, andare in profondità con l’intervista, indagando sui sentimenti e problemi più nascosti delle persone (ovviamente inerenti al prodotto/servizio). E’ molto facile che escano degli attributi importanti da inserire nell’indagine.

Un’altra fonte di attributi sono i clienti passati, coloro che hanno deciso di non acquistare più quel determinato prodotto o di non usufruire più del servizio. Questi in genere non vengono mai presi in considerazione, però potrebbero dare importanti informazioni sul perché hanno abbandonato il prodotto/servizio. Un’intervista profonda con loro può mettere in luce importanti attributi riguardanti la fedeltà e la soddisfazione.

Un’ultimo, ma non per questo meno importante sistema per cercare di capire quali attributi inserire, sono le analisi precedenti. Sono un modo poco costoso di recuperare attributi importanti e permettono anche il confronto con indagini già eseguite. Attenzione però, l’oggetto e l’obiettivo dello studio potrebbero essere cambiati in modo consistente.

1.2.4 Come capire l’importanza degli attributi

Capire l’importanza relativa di un attributo, è sempre stato un processo molto importante e discusso. Sono stati trovati metodi a volte molto semplici ed intuitivi, altri difficili sia nella comprensione, che nel calcolo. I motivi per farlo sono molti e dipendono dalla fase in cui si trova l’indagine: nel *pre-test*, valutare l’importanza relativa serve per poter decidere quali attributi eliminare dall’indagine, in quanto poco o per nulla significativi; si va così ad alleggerire il questionario. Se l’indagine è arrivata alla versione definitiva, la valutazione dell’importanza degli attributi serve per pesare il voto dato agli attributi stessi, in modo da costruire un indice di soddisfazione. Infine, se ci si trova in un programma già avviato, queste valutazioni servono al management per comprendere quali siano gli attributi su cui agire per aumentare la soddisfazione in modo significativo. Esistono molti metodi per misurare l’importanza degli attributi e si dividono in *diretti* ed *indiretti*. I diretti sono metodi

“trasparenti” al consumatore: si chiede direttamente a loro di valutare l’importanza relativa dei singoli attributi. Di solito viene chiesto ad un ristretto gruppo di intervistati ed i risultati sono facilmente comprensibili. Per contro, i metodi indiretti sono molto più “oscuri” agli intervistati. Non viene chiesto di valutare l’importanza di un attributo, e molto spesso l’intervistato non sa nemmeno che lo sta valutando. L’importanza viene ricavata e derivata dalle risposte date tramite analisi statistiche.

I principali metodi diretti sono:

- Estrapolazione delle informazioni (Elicitation)
- Valutazione dell’importanza
- Metodo della somma costante (Constant Sum)
- Paired Comparison
- Determinance
- Laddering

I principali metodi indiretti:

- Analisi di correlazione/regressione
- Conjoint analysis (full profile)
- Conjoint analysis (pairwise trade-offs)
- Differenze degli Estremi
- Livello di Indifferenza

METODI DIRETTI

Estrapolare le informazioni (Elicitation)

Probabilmente è il metodo più semplice tra tutti quelli analizzati. Si basa sul semplice fatto che se si vuole sapere che cosa è importante per una persona, semplicemente basta chiederglielo. Le domande potrebbero essere “Quali sono le caratteristiche più importanti che vuoi in...?” oppure “Quali sono le caratteristiche più importanti che ricerchi quando compri un/una...?”. I vantaggi di questo metodo

sono immediati: è un metodo molto semplice da comprendere sia per gli intervistati che per il management ed è anche molto semplice da eseguire. Purtroppo questo metodo presenta anche molti svantaggi. Innanzitutto alcune persone semplicemente si dimenticano di alcuni importanti attributi quando si valuta la soddisfazione: spesso i consumatori tendono a riportare di più i fattori negativi e non riportano le cose effettivamente positive. Altri consumatori sanno cos'è importante per loro ma non lo vogliono esprimere per paura di essere giudicati. Questo metodo è particolarmente inadatto a valutare gli attributi di quei prodotti o servizi particolarmente complessi, in quanto un consumatore farà sicuramente fatica a ricordare tutti gli aspetti. Infine alcuni consumatori rispondono alle domande citando frasi di spot pubblicitari o claim del prodotto.

Valutazione dell'importanza

Un altro semplice metodo per la valutazione diretta è la “valutazione dell'importanza”, dove ai rispondenti viene chiesto di valutare, con una scala generalmente a 7-10 elementi, l'importanza di alcuni attributi pre-selezionati. Anche qui i vantaggi principali sono la facilità di implementazione e la chiarezza dei risultati. Per quando riguarda gli svantaggi invece, i principali sono la tendenza degli intervistati a sopravvalutare gli attributi, dando sempre voti molto alti e la difficoltà nel trovare il numero giusto di attributi da inserire nel test. Non possono essere troppi, ma se sono troppo pochi si rischia di dimenticare qualche attributo importante.

Metodo della somma costante (Constant Sum)

Uno dei metodi per cercare di superare il problema dei voti troppo alti è stata l'introduzione della somma costante. Ai rispondenti è chiesto di allocare un numero fissato di punti (in genere 100) su degli attributi pre-selezionati. Questo metodo riesce ad evitare che l'intervistato valuti tutti gli attributi con voti alti, si assume che le valutazioni siano più ponderate. Funziona bene solo con pochi attributi (in genere meno di 10) perché se fossero di più, l'intervistato potrebbe pensare di non aver abbastanza punti per valutare tutti gli attributi in modo appropriato. Più il numero totale di punti è elevato, più aumenta la complessità del test, in quanto potrebbe capitare che l'intervistato sbaglia ad allocare i punteggi, non arrivando a dividere tutti i punti o dividendone troppi. Un grosso problema di questo metodo è che l'importanza rilevata è relativa e non assoluta, nel senso che se i punteggi risultano tutti uguali, gli attributi sono indifferenti al rispondente, ma non si conosce

se vengono reputati positivi o negativi.

Confronto di coppie (Paired Comparison)

Nel metodo diretto del confronto di coppie, all'intervistato vengono presentati gli attributi a coppie e viene chiesto di valutare quale ritiene più importante tra i due. Vengono presentate tutte le possibili combinazioni di attributi. Una volta terminato il confronto, viene costruita una tabella (vedi Tab 1.1) dove vengono riportate le proporzioni di individui che hanno valutato gli attributi in colonna migliori degli attributi in riga. I valori vengono poi normalizzati. Una volta riscalati i valori

	A	B	C	D	E
A	.50	.82	.69	.25	.35
B	.18	.50	.27	.07	.15
C	.31	.73	.50	.16	.25
D	.75	.93	.84	.50	.59
E	.65	.85	.75	.41	.50

Tabella 1.1: Esempio di tabella costruita con il confronto di coppie [Myers, 1999]

vengono sommati in colonna e riscalati per eliminare i valori negativi (vedi Tabella 1.2) . Per farlo viene sommato il valore assoluto dell'attributo con valore più basso, si va così a creare una scala di valori di importanza relativa. I vantaggi di questo

	A	B	C	D	E
	-0.09	1.21	0.43	-0.57	-0.92
Valori Riscalati	0.83	2.13	1.35	0.35	0

Tabella 1.2: Esempio di valori riscalati con la Thurstone Case V scale [Myers, 1999]

metodo sono la capacità di trasformare una semplice scelta fra due componenti ad una scala di attributi e di "obbligare" gli intervistati a scegliere una delle due alternative. Gli svantaggi sono la possibilità di utilizzarlo con molti attributi, in quanto le possibili combinazioni aumenterebbero a dismisura all'aumentare delle componenti, la possibilità di calcolare la scala solo a livello aggregato e non per i singoli componenti e un'analisi un po' più complicata rispetto ai metodi precedenti

Determinance

L'idea alla base della determinance è che anche se un attributo è molto importante, esso non è determinante per la scelta di un preciso prodotto o la soddisfazione di un cliente. Questo può avvenire quando l'attributo in questione è una caratteristica ormai ovvia e presente in tutti i dispositivi. Questo metodo quindi si prefigge di misurare non solo l'importanza dell'attributo, ma anche la differenza dell'attributo rispetto ai principali competitors. Ad esempio può essere chiesto: “Quanto importanti sono questi attributi nella scelta di comprare la tua auto?” e “Quanta differenza credi ci sia tra queste auto relativamente gli attributi di cui sopra”. Ovviamente se degli attributi, anche importanti, hanno poca differenza tra un'auto e l'altra, questi saranno di poca utilità (a meno che non siano mancanti, in quel caso avranno una forte influenza sulla soddisfazione, ma in senso negativo). Per costruire poi una scala, le due valutazioni vanno moltiplicate tra loro. I principali vantaggi di questo metodo sono proprio la capacità di individuare quegli attributi che sono importanti per i consumatori ma che non inficiano sulla loro soddisfazione, la possibilità di valutare l'unicità di un attributo rispetto ai concorrenti e ovviamente la capacità di rilevare due informazioni per ogni attributo. Quest'ultimo vantaggio si rivela essere anche uno svantaggio in quanto per avere più informazioni è necessario fare più domande, rendendo così l'intervista più lunga e meno piacevole per gli intervistati.

Scalata

La tecnica chiamata “Scalata” è l'unica di quelle viste finora che è stata sviluppata espressamente per il marketing. E' basata sulla teoria “*Means-End*”, la quale dice che in ogni consumatore esiste una gerarchia di valori (ma anche di attributi) e che questa può essere rivelata tramite un processo di domande conosciuto come “laddering” (scalata). Uno studio di questo tipo comincia con una serie di domande per ogni individuo, come:

1. Si chiede ad un individuo di esprimere gli attributi che lui reputa più importanti da un gruppo di prodotti/servizi/marchi. Questi prodotti devono essere bipolari (ad esempio il gusto di un vino secco-dolce).
2. Si chiede quale polo preferisce
3. Si sceglie un attributo chiave per la scalata
4. Si chiede perché ha scelto quel determinato polo

5. Si chiede perché quel polo è importante, quali sono i benefici per il rispondente?
6. Si continua a chiedere perché ogni scelta è importante, in modo da “scalare” per rivelare come uno specifico prodotto o servizio può soddisfare uno o più dei valori e obiettivi personali del rispondente

Al termine, tutte le risposte vengono messe in input ad un programma che aggrega le varie scalate dei diversi intervistati, per produrre un singolo diagramma che mostra la sequenza e le frequenze relative delle coppie di attributi.

I vantaggi sono che riesce a costruire una struttura di valori, identificando quali sono i più importanti motivatori all’acquisto. E’ anche un metodo molto semplice per chi risponde all’intervista, in quanto si basa sull’interazione intervistatore/intervistato. Gli svantaggi principali sono la lunghezza richiesta dall’intervista e dal bisogno di intervistatori preparati ed esperti.

I METODI INDIRETTI

Analisi di correlazione/regressione In questo metodo, si assume che più è alta la correlazione tra due variabile, ad esempio un attributo e la soddisfazione, più questo attributo è importante per il consumatore. Questa è un’assunzione, perché la correlazione non implica necessariamente causalità, ma è ragionevole pensarlo. Questo tipo di analisi prende in esame le performance, non l’importanza, dei vari attributi. Nelle ricerche di marketing, infatti, agli intervistati è chiesto di valutare la performance di un’azienda, rispetto alle concorrenti, valutando uno o più attributi. Viene anche chiesta una valutazione generale, ad esempio sulla soddisfazione. L’importanza relativa di ogni attributo viene determinata calcolando un coefficiente di correlazione (r) tra le valutazioni sulla performance e la valutazione generale. Questo coefficiente rappresenta quanto della variazione di una variabile può essere spiegato dall’altra, più è alto, più il legame tra le due è forte. I più grandi vantaggi di questo metodo sono la facilità di risposta e di analisi dei risultati, d’altro canto, come già accennato, la correlazione non è la causalità, e quindi i risultati non sono certi, ma solo probabili.

Conjoint Analysis (full profile)

La conjoint analysis, è ad oggi una delle tecniche più utilizzate nella marketing research. In questo contesto andiamo ad analizzarla solo come strumento per la

misura dell'importanza degli attributi. Agli intervistati è chiesto di valutare la preferenza generale su 10 o 20 prodotti *ipotetici*. Questi sono costruiti *ad-hoc* dai ricercatori, utilizzando 4/8 attributi indipendenti divisi a loro volta in 2/5 livelli chiaramente definiti. In questo modo i prodotti da valutare risultano totalmente incorrelati tra i diversi "profili". Utilizzando la classificazione delle preferenze da parte degli intervistati, un computer calcola l'importanza relativa di ogni livello di ogni attributo. Da questa "utilità", viene calcolata l'importanza relativa di ogni attributo sottraendo al valore di utilità maggiore, quello minore. Gli attributo con differenza maggiore, sono quelli più importanti.³ Le utilità vengono poi confrontate tra loro per trovare l'attributo più importante per gli individui. In questo modo la conjoint analysis (full profile) fornisce, non solo l'importanza relativa di ogni attributo, ma anche il valore che ogni livello di ogni attributo ha per i rispondenti. Ciò rende possibile calcolare il prodotto "ideale" fornendo al prodotto gli attributi con maggiore utilità e a questi associare il livello con maggiore utilità. I vantaggi di questo metodo sono la possibilità di poter ottenere l'importanza di ogni attributo per ogni cliente, pur essendo un metodo indiretto. Questo rende possibile segmentare il mercato sulla base di attributi o prodotti che il cliente ritiene più importanti.

Conjoint Analysis (Pairwise Trade-Offs)

Ques'altra forma di conjoint analysis, consiste nel far giudicare agli intervistati, non un prodotto costruito, ma tutte le possibili coppie di livelli di attributi (da qui il nome di Pairwise Trade-Offs). Queste classificazioni diventano poi la variabile dipendente di una regressione multipla che calcola l'importanza relativa di ogni livello di ogni attributo. Questa è la stessa analisi statistica che si effettua nella versione full profile e quindi i risultati sono comparabili. Anche questa modalità ha il grande vantaggio di riuscire a misurare l'importanza degli attributi, tramite un metodo indiretto, per ogni individuo.

Differenze degli estremi

I metodi finora presentati hanno alcuni limiti: l'analisi di correlazione/regressione non riesce a calcolare l'importanza degli attributi per ogni singolo individuo, ma solo in modo aggregato. La conjoint analysis, invece, ha il problema di non riuscire a

³Attenzione però che questo metodo ha dei problemi: se, ad esempio, due rispondenti valutano un profilo esattamente l'opposto uno dall'altro, l'importanza calcolata sarà esattamente la stessa, anche se i rispondenti hanno due preferenze completamente diverse.

calcolare l'importanza per un numero elevato di attributi. A volte capita di dover riuscire a calcolare l'importanza a livello individuale di un numero elevato di attributi, per far ciò si utilizza il metodo delle differenze degli estremi. Agli intervistati viene innanzitutto chiesto di nominare il marchio preferito e quello meno preferito di un particolare prodotto. Viene poi chiesto di valutare un gran numero di attributi usando una scala a 7/10 punti. Poi, per ogni attributo, alla valutazione più elevata, viene sottratta la valutazione più bassa. La grandezza della differenza indica l'importanza relativa di quell'attributo per quel determinato individuo. La logica alla base di questo metodo è che se non ci sono differenze tra la marca preferita e la marca ritenuta meno appetibile, allora quel determinato attributo non è chiave per determinare la scelta di una marca o l'altra, e quindi è meno importante. Il problema di questo metodo è che rispetto ai precedenti ³ è meno statisticamente valido e quindi va utilizzato con cautela.

Livello d'indifferenza

In questo metodo, ad ogni individuo vengono mostrati due profili di uno stesso prodotto. Tutti gli attributi hanno lo stesso livello eccetto due: in uno un profilo ha un livello più alto, risultando più appetibile; nel secondo, un profilo ha un livello, l'altro è vuoto e viene chiesto all'intervistato di inserire un valore per equilibrare le due scelte. Ad esempio: i due profili si riferiscono ad un'auto, tutti gli attributi sono uguali a meno del prezzo, che per una è 12000€, per l'altra 15000€. L'attributo "consumi" ha valore 17km/l per l'auto da 12000, per l'altra il valore è da inserire. Il rispondente valuta che il consumo per rendere la scelta indifferente è di 20km/l. Grazie a questa affermazione possiamo concludere che un aumento di 3km/l vale 3000 per gli intervistati.

I confronti più significativi sono, come nel caso in esempio, con il prezzo, in questo si riesce a quantificare la disponibilità a pagare per un attributo migliore.

QUALE METODO SCEGLIERE?

Esiste un metodo migliore rispetto agli altri? No, la scelta dipende da vari fattori, come obiettivi dell'indagine, numero e tipo di attributi. Sfortunatamente questi metodi forniscono anche risultati differenti, quindi la scelta non è indifferente. Una soluzione sarebbe quella di utilizzare sia metodi diretti che indiretti nello stesso studio, quando questo è possibile. Per farlo bisognerebbe chiedere agli intervistati sia

l'importanza che la performance dei vari attributi. In questo modo, i ricercatori potrebbero calcolare sia la misura dell'importanza, sia la correlazione tra la performance e la misura generale sulla soddisfazione, questo per ogni attributo. Se entrambi i metodi (diretto ed indiretto) sono in accordo, si avrà un buon grado di certezza sull'importanza, o meno, dell'attributo preso in esame.

1.3 I Vantaggi

Quali sono i vantaggi di avere una clientela soddisfatta? I metodi descritti in precedenza sono utili per sapere se i propri clienti risultano soddisfatti, ma non danno nessuna informazione su quanto questa soddisfazione incida sull'andamento aziendale. Per capire questo, la Toyota Motorsales, U.S.A. ha confrontato i risultati finanziari dei suoi dealers, ponendo da una parte quelli con soddisfazione della clientela maggiore, dall'altra quelli con soddisfazione minima. I risultati sono indicativi di quanto sia importante la soddisfazione. Infatti, come si nota in Figura 1.1, i dealer con soddisfazione più alta hanno il 46% in più di profitti netti, 81% di profitti netti come percentuale sulle vendite in più, il 67% di profitti in più per impiegato, il 43% in meno di turnover del personale e il 30% di costi in meno di promozione per veicolo.

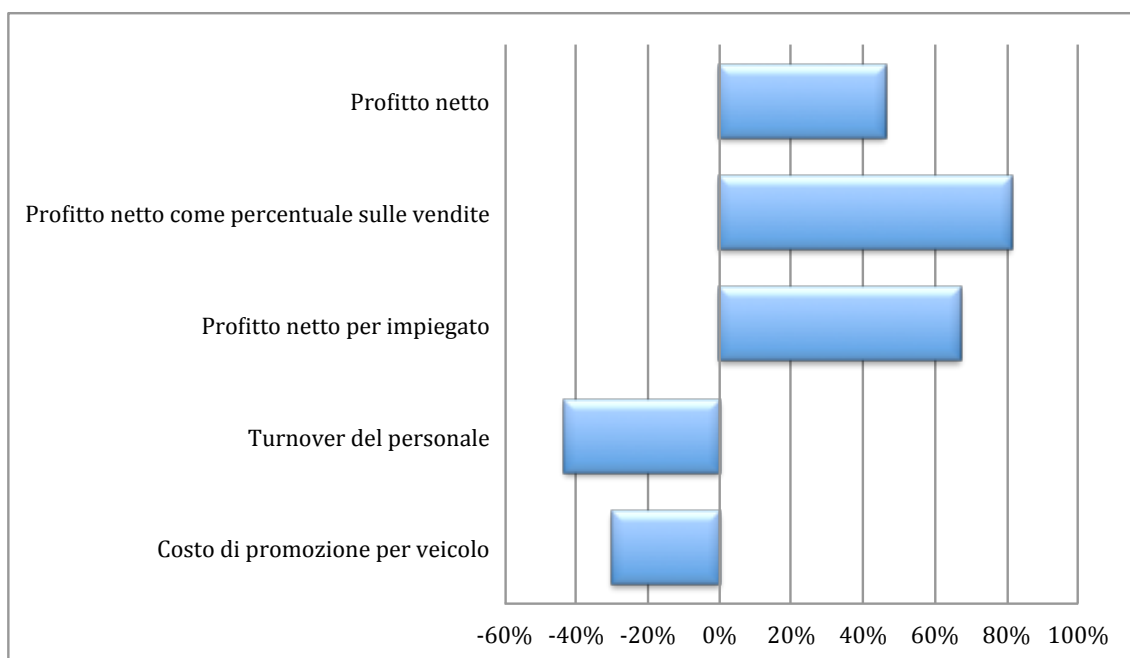


Figura 1.1: Confronto tra dealer con più alta soddisfazione della clientela rispetto a quelli con più bassa soddisfazione [Myers, 1999].

Capitolo 2

Le Anchoring Vignettes

2.1 Cosa sono

Nel primo capitolo è stato spiegato come uno dei modi per poter misurare la soddisfazione della clientela sia di interpellare direttamente il cliente, tramite ad esempio, un questionario. Questo è un ottimo modo se si analizza una popolazione con caratteristiche simili, ma quando l'indagine coinvolge popolazioni differenti le cose si complicano, perché non è più così immediato confrontare i risultati. Frequentemente infatti, vengono misurati concetti che sono definibili solo tramite esempi, come la libertà, la salute, la soddisfazione. Inoltre gli individui, spesso, comprendono le domande poste loro in maniera molto diversa.[Green e Hensher, 2010] Queste differenze interpersonali e interculturali nell'interpretare, capire o utilizzare categorie di risposta per una stessa domanda, sono noti in psicomelia come DIF (Differential Item Functioning)[Paccagnella, 2011]. King *et al.* [2004] proposero l'utilizzo delle anchoring vignettes per superare questo problema. Questo approccio consiste nell'utilizzare una serie di scenari (chiamati vignettes), uguali per tutti gli intervistati, in modo da poter stimare il DIF di ognuno e quindi correggere i risultati dell'indagine [Green e Hensher, 2010].

Nel 2002 in un'indagine dell'Organizzazione Mondiale della Sanità fu chiesto ad un campione di cittadini cinesi e messicani di giudicare la propria “political efficacy¹” in una scala che va da “nessuna” a “illimitata”. I primi risultati furono sorprendenti

¹Per political efficacy si intende la fiducia che un cittadino ha nel suo governo e la credenza che questo ha nel poter capire ed influenzare le decisioni politiche.

poiché il 50% dei messicani rispose “nessuna”, contro il 27% dei cinesi. Questi risultati porterebbero alla conclusione, dopo un’indagine abbastanza approssimativa, che gli intervistati messicani abbiano una minore political efficacy rispetto ai cinesi, e questo risulterebbe alquanto strano. Rifacendo l’analisi ed inserendo le anchoring vignettes, le conclusioni sono state completamente stravolte: il 40% dei cinesi, infatti, reputa la propria condizione peggiore della peggior “vignette”, mentre ciò accade solo nell’11% dei messicani [Chevalier e Fielding, 2011] Questo esempio aiuta a comprendere quanto possano essere importanti le anchoring vignettes per confrontare in modo corretto i risultati di indagini su popolazioni diverse.

2.2 Assunzioni

Per poter utilizzare le anchoring vignettes sono necessarie due assunzioni, senza le quali non è possibile correggere il DIF.

La prima è chiamata *Response Consistency*²: con essa si assume che ogni individuo utilizzi le stesse categorie di risposta sia valutando la propria situazione, sia valutando quella descritta dalla vignette. Questo significa che può esistere del DIF tra gli intervistati, ma non ce ne può essere tra le domande che ogni singolo individuo va a valutare.[King *et al.*, 2004]

La seconda è la *Vignettes Equivalence*³: essa indica che il livello della variabile presentato in ogni vignettes viene percepito alla stessa maniera da tutti gli individui. Ovvero che gli intervistati possono avere delle differenze nella percezione del livello della variabile in ogni scenario, ma che queste devono essere casuali e indipendenti dalle caratteristiche della vignette stessa.[King *et al.*, 2004]

La validità di queste due assunzioni è tuttora oggetto di dibattito. Verranno presentati, più avanti nel capitolo, degli studi in proposito.

Si andrà ora ad analizzare due metodi che utilizzano le anchoring vignettes per correggere il DIF: il primo non parametrico, il secondo parametrico.

²D’ora in avanti abbreviata con RC

³Abbreviata con VE

2.3 Il metodo non parametrico

Il metodo più semplice per utilizzare le anchoring vignettes è quello non parametrico. Questo metodo consiste nello riscalare l'auto-valutazione, con la valutazione data alle vignettes [Chevalier e Fielding, 2011].

Con questo esempio si chiarifica il concetto: sono state esposte 5 vignettes ordinate nel seguente modo A B C D E (dove A indica uno stato migliore di E, ad esempio A: molto soddisfatto, E: per niente soddisfatto). Gli intervistati valuteranno il loro stato (ad esempio la propria soddisfazione riguardante il lavoro) con una scala da 1 a 5. Due intervistati valutano la loro stessa soddisfazione entrambi 3, poi però andando a votare le vignettes, il primo valuta la situazione espressa dalla vignetta D con voto 3, mentre il secondo valuta con voto 3 la situazione espressa in vignetta B. Come si nota in Figura 2.1 questa situazione evidenzia come il primo individuo sarà meno soddisfatto del secondo.

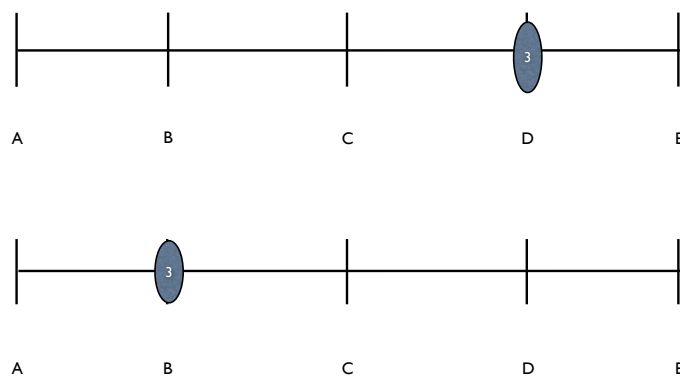


Figura 2.1: Metodo non parametrico per l'utilizzo delle anchoring vignettes

Un sistema ancora più semplice e che evita di dover disegnare grafici, è una

semplice ricodifica delle risposte date:

$$C_i = \begin{cases} 1 & \text{se } y_i < z_{i1} \\ 2 & \text{se } y_i = z_{i1} \\ 3 & \text{se } z_{i1} < y_i < z_{i2} \\ \vdots & \vdots \\ 2J + 1 & \text{se } y_i > z_{iJ} \end{cases}$$

dove y_i è l'auto-valutazione dell' i -esimo intervistato (con $i = 1, \dots, n$) e z_{ij} è la valutazione dell' i -esimo intervistato della situazione presentata nella j -esima vignette. In questo modo, il primo intervistato (che si è valutato 3 e ha valutato 3 la vignette D) avrà una valutazione $C = 4$, mentre il secondo (che si auto-valutava come l'individuo descritto in vignette B) avrà una valutazione $C = 8$, chiaramente più elevata. Questo semplice esempio, come quello precedente su Messico e Cina rivela quanto può essere pericoloso ignorare il DIF quando si analizzano i risultati di indagini con domande di auto-valutazione.

Questo metodo risulta molto semplice da implementare, presenta però due aspetti negativi che rendono preferito, nella letteratura di settore, il metodo parametrico: innanzitutto richiede che le vignettes vengano valutate da ogni individuo, sia nella prima indagine, sia nelle successive, rendendo queste più lunga e, a volte, più costose; in secondo luogo, come molti altri metodi statistici non parametrici, in alcune circostanze, risulta statisticamente non efficiente. Un altro punto a sfavore di questo approccio non parametrico, da non sottovalutare, è che gli individui comprendano tutti allo stesso modo, l'ordinamento delle vignettes. Se, ad esempio, un intervistato reputa la vignette B peggiore della vignette C, sarà impossibile poi, riuscire a correggere il DIF. Il metodo parametrico, spiegato nel successivo paragrafo, supera questi ostacoli.

2.4 Il metodo parametrico

In questa sezione verrà introdotto il modello conosciuto come HOPIT⁴. Può essere visto come un'estensione del modello probit ordinale, con le soglie tra le categorie che vengono specificate individuo per individuo, invece che essere fissate.

⁴hierarchical ordered probit

2.4.1 L'Ordered Probit Model

Prima di spiegare il modello HOPIT è utile conoscere la versione da cui deriva, ovvero l'ordered probit model. Come si può intuire dal nome, questo modello appartiene al gruppo dei modelli a risposta ordinata, che sono un'estensione dei modelli a risposta binaria. Ipotizziamo ad esempio che una variabile y possa assumere tre valori (1, 2, 3) e che sia il risultato osservato di una variabile continua latente y^* :

$$y^* = X\beta + \epsilon$$

$$y = \begin{cases} 1 & \text{se } y^* \leq 0 \\ 2 & \text{se } 0 < y^* \leq \tau \\ 3 & \text{se } y^* > \tau \end{cases}$$

dove ϵ è un termine d'errore i.i.d., simmetrico, con FdR G , X una matrice di variabili esplicative e τ soglie da stimare, uguali per tutti gli individui. Essendo un modello Probit, la Funzione di Ripartizione G è una Normale Standard:

$$G(x) = \Phi(x) = \int_{-\infty}^x \phi(v) dv$$

$$\Downarrow$$

$$P(y = 1|X) = \Phi(-X\beta)$$

$$P(y = 3|X) = 1 - \Phi(\tau - X\beta)$$

$$P(y = 2|X) = \Phi(\tau - X\beta) - \Phi(-X\beta)$$

A differenza dei modelli a risposta binaria, ci sono da stimare oltre al vettore di parametri β anche i parametri τ . Si hanno in genere tante soglie quante sono le modalità della variabile osservata meno una.

Il modello sopra riportato può essere stimato poiché identificato, ma se consideriamo il modello generale,

$$y^* = \alpha + \tilde{X}\beta + \epsilon \quad \text{con } \epsilon \sim (0, \sigma^2)\text{i.i.d.}$$

$$y = \begin{cases} 1 & \text{se } y^* \leq \tau_1 \\ 2 & \text{se } \tau_1 < y^* \leq \tau_2 \\ 3 & \text{se } y^* > \tau_2 \end{cases}$$

con \tilde{X} matrice senza costante, per l' i -esima osservazione vale:

$$P(y_i = 1|\tilde{x}_i) = P(\alpha + \tilde{x}_i' \beta + \epsilon_i \leq \tau_1|\tilde{x}_i) = \Phi \left[\frac{\tau_1 - \alpha}{\sigma} - \tilde{x}_i' \left(\frac{\beta}{\sigma} \right) \right]$$

In questo caso combinazioni diverse dei parametri $(\alpha, \beta, \tau_1, \sigma)$ generano lo stesso valore della log-verosimiglianza, quindi il modello non è identificato.

Per stimare questo modello bisogna quindi inserire dei vincoli:

- Varianza unitaria e prima soglia pari a 0 ($\sigma = 1; \tau_1 = 0$)

o

- Varianza unitaria e costante pari a 0 ($\sigma = 1; \alpha = 0$)

In quest'ultimo caso:

$$y^* = \tilde{X}\beta + \epsilon \quad \text{con } \epsilon \sim (0, 1)\text{i.i.d.}$$

$$y = \begin{cases} 1 & \text{se } y^* \leq \tau_1 \\ 2 & \text{se } \tau_1 < y^* \leq \tau_2 \\ 3 & \text{se } y^* > \tau_2 \end{cases}$$

con \tilde{X} matrice senza costante e:

$$\begin{aligned} P(y = 1|\tilde{X}) &= \Phi(\tau_1 - \tilde{X}\beta) \\ P(y = 2|\tilde{X}) &= \Phi(\tau_2 - \tilde{X}\beta) - \Phi(\tau_1 - \tilde{X}\beta) \\ P(y = 3|\tilde{X}) &= 1 - \Phi(\tau_2 - \tilde{X}\beta) \end{aligned}$$

Il caso generale è con M alternative logicamente ordinate (con M=2 si torna esattamente al caso con le variabili a risposta binaria).

Il problema principale di questo modello è che le soglie τ sono uguali per tutti e questo implica che non si tiene conto del problema del DIF.[Giudici, 2005]

2.4.2 Il modello HOPIT

Il modello HOPIT è stato introdotto proprio per cercare di risolvere questo problema, stimando delle soglie che siano specifiche di ogni persona e non fissate per tutti. Con questo modello è possibile rendere le indagini successive meno dispendiose, sia

in termini temporali che monetari, chiedendo di valutare gli scenari delle vignettes solo in una prima indagine, estendendo poi i risultati ottenuti, nelle successive. [King *et al.*, 2004] Il modello HOPIT è, come detto in precedenza, una versione modificata dell'Ordered Probit Model. Per includere le informazioni derivanti dalle vignettes il modello si dividerà in due componenti (due modelli a risposta ordinata), una riferito alle domande riguardanti l'autovalutazione e una riferito alle vignettes. Prima di esporre queste componenti, è utile introdurre i concetti di livello *attuale*, *percepito* e *riportato* della variabile presa in oggetto. Il primo indica il livello effettivo (di libertà, political efficacy, salute, ecc.) che ogni individuo presenta. Esso è non osservato. Gli individui *percepiscono* questo livello effettivo con un errore, ma correttamente in media. Anche il livello percepito non è osservato. Infine quando questi rispondono ad un questionario, esprimono il proprio livello percepito, ma ogni individuo utilizza delle soglie differenti. Questo livello riportato è osservato, ma non è possibile confrontarlo con altri individui [King *et al.*, 2004]

Prima componente, l'autovalutazione

Si denota il livello *attuale* dell' i -esimo intervistato con μ_i in una scala crescente continua, illimitata e unidimensionale. L'intervistato *percepisce* il livello μ_i solo con un errore (normale standard), come avviene nell'ordered probit model. Il livello percepito dell' i -esimo rispondente risulta quindi:

$$Y_i^* \sim N(\mu_i, 1)$$

Il livello attuale, che varia tra gli individui è una funzione lineare di caratteristiche, espresse dal vettore x_i , ed un effetto casuale normale indipendente da x , chiamato ε_i

$$\mu_i = x_i' \beta + \varepsilon_i, \quad \varepsilon \sim N(0, 1)$$

dove x_i' è un vettore di covariate senza termine costante, β è un vettore di parametri e gli errori ε sono indipendenti dalle covariate. L'assenza della costante e la varianza unitaria di ε sono necessarie all'identificazione. Il livello *riportato* viene rilevato tramite una domanda sull'auto-valutazione avente k ($k = 1, \dots, K$) categorie di risposta. Gli individui, quindi, trasferiscono il livello percepito (continuo) nel livello

rappresentato (discreto) attraverso delle soglie τ :

$$y_i = k \quad \text{se} \quad \tau_i^{k-1} < Y_i^* < \tau_i^k$$

dove $\tau_i^0 = -\infty$, $\tau_i^K = \infty$ e $\tau_i^{k-1} < \tau_i^k$

Le soglie sono così modellate:

$$\begin{aligned} \tau_i^1 &= v_i' \gamma^1 \\ \tau_i^k &= \tau_i^{k-1} + \exp(v_i' \gamma^k), \quad k = 2, \dots, K-1 \end{aligned}$$

con v_i' un vettore di covariate (che potrebbe coincidere con x_i') e γ^k , per $k = 1, \dots, K-1$, vettori di parametri.⁵[King *et al.*, 2004] L'esponenziale nell'equazione delle soglie garantisce che queste siano sempre crescenti.

Se si utilizzasse solo questa componente, esisterebbe però un problema di identificazione in quanto β e γ non sono identificabili separatamente; per farlo c'è bisogno di ulteriore informazione, in questo caso fornita dalle vignettes.

Questa prima componente può essere estesa ad un modello più generale comprendente più domande sull'auto-valutazione, indicate con s ($s = 1, \dots, S$), caso non studiato in questa tesi.

Seconda componente, le vignettes

L'assunzione di *vignettes equivalence*, come riportato in precedenza, indica che il vero livello della variabile presa in oggetto è percepito in modo uguale da tutti i rispondenti, a meno di un errore casuale. Quest'assunzione può essere formalizzata introducendo il parametro θ_j , indipendente dagli intervistati ($j = 1, \dots, J$ indica le vignettes). Il livello percepito della variabile espressa nella j -esima vignette per l'individuo i è, come nella prima componente, non osservato. Esso può essere modellato come

$$Z_{ij}^* = \theta_j + u_{ij} \quad u_{ij} \sim N(0, \sigma_u^2)$$

u_{ij} indipendente da ε e X_i . L' i -esimo individuo manifesta il livello percepito in una categoria Z_{ij} mediante un modello a soglie non fisse, esattamente come nella prima

⁵Esiste una variante, portata da van Soest *et al.* [2011], che introduce un termine di errore u_i all'interno della prima soglia. Questo effetto casuale u_i si assume che sia distribuito normalmente con media 0 e varianza σ_u^2 e indipendente dalle covariate x_i .

componente:

$$Z_{ij} = k \quad \text{se} \quad \tau_{vi}^{k-1} \leq Z_{ij}^* \leq \tau_{vi}^k$$

Con l'assunzione di *response consistency*, le soglie τ^k della seconda componente sono identiche a quella della prima⁶:

$$(\tau_s^k = \tau_v^k)$$

Grazie all'informazione aggiuntiva fornita dalla valutazione delle vignettes, è ora possibile identificare i parametri β e γ . In particolare le soglie sono identificate dalla seconda componente, mentre i β possono essere identificati dalla prima componente.

2.5 La stima del modello

La stima dei parametri è effettuata con il metodo della massima verosimiglianza. E' possibile calcolare due verosimiglianze distinte per le due componenti e, solo in seguito, creare quella congiunta. La verosimiglianza per la prima componente è

$$L_s(\beta, \gamma|y) \propto \prod_{i=1}^n \prod_{k=1}^K [\Phi(\tau_i^k | X_i \beta, 1) - \Phi(\tau_i^{k-1} | X_i \beta, 1)]^{\mathbf{I}(y_i=k)}$$

con $i = (1, \dots, n)$ l'indicatore degli individui, $k = (1, \dots, K)$ l'indicatore delle categorie di risposta e $\mathbf{I}(y_i = k) = 1$ se $y_i = k$, 0 altrimenti.

La verosimiglianza per la seconda componente risulta:

$$L_v(\theta, \gamma|z) \propto \prod_{i=1}^n \prod_{j=1}^J \prod_{k=1}^K [\Phi(\tau_i^k | \theta_j, \sigma_u^2) - \Phi(\tau_i^{k-1} | \theta_j, \sigma_u^2)]^{\mathbf{I}(z_{ij}=k)}$$

con $i = (1, \dots, n)$ l'indicatore degli individui, $k = (1, \dots, K)$ l'indicatore delle categorie di risposta, $j = (1, \dots, J)$ l'indicatore delle vignettes e $\mathbf{I}(z_{ij} = k) = 1$ se $z_{ij} = k$, 0 altrimenti. La verosimiglianza congiunta risulterà quindi:

$$L(\beta, \sigma^2, \theta, \gamma|y, z) = L_s(\beta, \gamma|y)L_v(\theta, \gamma|z)$$

2.6 Controllare la validità delle assunzioni

Si è visto che per poter utilizzare le anchoring vignettes per correggere il DIF sono necessarie le assunzioni RC e VE. Bago d'Uva *et al.* [2009] proposero dei test per

⁶s e v indicano la componente di auto-valutazione e la componente delle vignettes rispettivamente.

cercare di validarle utilizzando alcuni dati dell'indagine ELSA⁷, in particolare quelli della rilevazione del 2006-2007, che presentavano al loro interno le risposte ad alcune vignettes. Per testare la prima, hanno utilizzato informazioni aggiuntive provenienti da un indicatore "oggettivo", in modo che ogni variazione dell'autovalutazione, non condizionata a questo indicatore, può essere attribuita alle differenze di interpretazione. In questo caso, esse sono identificate anche senza assumere l'assunzione RC, quindi essa può essere testata confrontando le soglie per controllare se esse sono uguali nel modello con l'assunzione RC e nel modello con l'indicatore oggettivo.

Per la seconda assunzione Bago d'Uva *et al.* [2009] hanno effettuato un test per valutare la condizione necessaria di assenza di variazione sistematica nelle differenze percepite tra i livelli rappresentati in almeno due vignettes.

I due test vengono applicati a due domini, le funzioni cognitive e la mobilità, che rappresentano le dimensioni mentale e fisica della salute.

I risultati del test rifiutano le ipotesi di validità delle due assunzioni per entrambi i domini. Solamente un test più "debole"⁸ per l'ipotesi di RC, ne accetta la validità per la dimensione mentale. [Bago d'Uva *et al.*, 2009]

Un altro studio condotto da van Soest *et al.* [2011] ha utilizzato un'intervista via web a 3350 studenti dell' University College Dublin per testare la validità dell'assunzione RC. Agli studenti è stato chiesto di valutare il loro "rapporto" con l'alcol, chiedendo, ad esempio, di descrivere le proprie abitudini rispetto al consumo di bevande alcoliche tramite una scala ordinale (mild, moderate, some cause for concern, excessive, extreme). E' stato poi chiesto di valutare, sempre tramite la stessa scala, l'utilizzo di alcol descritto in alcune vignettes loro presentate; infine è stato chiesto di esprimere numericamente l'ammontare del proprio consumo.

Proprio quest'ultima informazione, come nello studio di Bago d'Uva *et al.* [2009], è necessaria per testare la validità dell'assunzione, in quanto si ha bisogno di una

⁷English Longitudinal Study of Ageing

⁸nel test cosiddetto "forte" viene testata l'uguaglianza delle soglie tra il modello con RC e il modello con indicatore oggettivo, nel test "debole" viene testata l'uguaglianza delle *distanze* tra le soglie.

$$\gamma^k - \gamma^{k-1} = \gamma_v^k - \gamma_v^{k-1}$$

con γ_v^k le soglie delle vignettes

misura oggettiva. Questa è misurata anch'essa come un *ordered probit model*:

$$Y_{oi}^* = X_i\beta_o + \xi_{oi}$$

$$Y_{oi} = j \quad \text{se} \quad \tau_o^{j-1} < Y_{oi}^* \leq \tau_o^j, \quad j = 1 \dots 6$$

Con le soglie τ trattate come delle costanti ignote (con $\tau_o^0 = -\infty$ e $\tau_o^6 = \infty$), sulla base dell'assunzione che esse non varino tra gli individui (essendo questa una misura *oggettiva*), Y_{oi} è il livello oggettivo percepito e $j = (1, \dots, J)$ sono le categorie di risposta (nell'esempio $J = 5$). Si assume che ξ_{oi} sia indipendente da X_i , u_i e u_{ij} , ma può correlare con ε_{si} . Grazie a questa informazione aggiuntiva, un test per l'assunzione RC può essere sviluppato comparando un modello "vincolato", avente sia l'indicatore oggettivo che la RC, con uno "libero" (senza RC). E' facilmente dimostrabile che quest'ultimo modello è identificato, in quanto le vignettes sono utilizzate per stimare τ_v^k (stimando γ_v^k) mentre la misura oggettiva può essere usata per stimare $\beta = \beta_s = \beta_o$, così β e γ^1 sono identificate separatamente. Tutte e due le versioni dei modelli possono essere stimate tramite massima verosimiglianza.[van Soest *et al.*, 2011] I risultati dell'indagine, utilizzando l'indicatore AIC⁹ indicano come il modello migliore quello che presenta l'assunzione RC.

Peracchi e Rossetti [2010] proposero un test congiunto per verificare la validità di entrambe le assunzioni, basandosi sul fatto che il modello, sotto entrambe le assunzioni, risulta sovraidentificato. Questo test non ha la necessità di un'informazione oggettiva e può essere applicato a qualsiasi studio presenti almeno una vignette. Per contro, potrebbe rifiutare l'ipotesi di sovraidentificazione anche per motivi differenti dalle due assunzioni, come un'omissione di alcune variabili rilevanti. Sono stati utilizzati dati provenienti dall'indagine SHARE¹⁰.

I risultati del test indicano che nella maggior parte delle situazioni analizzate dall'indagine (dolore, mobilità, problemi a dormire, asma, problemi di concentrazione e depressione) viene rifiutata la sovraidentificazione imposta dalle due assunzioni. Anche se, come detto in precedenza, ciò non esclude la validità delle assunzioni, pone alcuni dubbi sulla validità dell'utilizzo delle vignettes. Peracchi e Rossetti [2010]

Questi esempi mostrano come, non esista un test formale per accertare la validità delle assunzioni e il dibattito sia ancora molto acceso.

⁹Akaike's Information Criterion

¹⁰Survey of Health, Ageing and Retirement in Europe

Capitolo 3

L'indagine

La misura della soddisfazione è sempre stata legata al paradigma della discrepanza, il quale compara le aspettative prima dell'acquisto del prodotto con la performance percepita durante l'utilizzo. Per far questo vengono utilizzate delle misure di scala. Il problema si pone quando si prova a confrontare queste misure tra stati, o gruppi di differente provenienza socio-economica e culturale, in quanto ognuno di questi utilizza una diversa scala per valutare se stesso (il cosiddetto DIF).

L'indagine oggetto di questa tesi, è la prima applicazione delle anchoring vignettes nel campo del marketing e si prefigge di risolvere il problema del DIF collegato alla soddisfazione della clientela.

Gli scopi principali di quest'indagine sono due, uno empirico ed uno metodologico. Il primo consiste nell'arricchire la letteratura di settore, studiando un nuovo metodo di misura della customer satisfaction, che in futuro, se ritenuto soddisfacente, potrà essere implementato. Il secondo è quello di *testare* alcune caratteristiche dell'approccio con le anchoring vignettes:

- valutare le differenze della correzione del DIF cambiando l'ordine delle vignettes, rispetto alla domanda sull'autovalutazione.

Queste valutazioni vengono effettuate sulla base di precedenti studi, effettuati da Hopkins e King [2010]

3.1 Il LISS panel

L'indagine è stata affidata all'istituto di ricerca olandese CentERdata, con sede presso l'università di Tilburg, in modo da poter utilizzare il LISS panel¹, un panel formato da circa 5000 famiglie e 8000 individui olandesi, della durata di 7 anni (2006-2013). Questo panel è la componente principale del MESS², un progetto multidisciplinare e longitudinale che si propone di essere un ambiente di ricerca innovativo permettendo di migliorare la ricerca di molti fenomeni sociali.[Das *et al.*, 2006]

3.1.1 Il Campione

Il panel LISS è costruito per la ricerca scientifica, per questo motivo la qualità e la significatività del campione è essenziale.

Per costruirlo, è stata effettuata un'estrazione casuale semplice di 10150 indirizzi dai registri della popolazione, creati dall'istituto nazionale di statistica olandese³ selezionando in modo casuale il 10% della popolazione residente in modo permanente in Olanda. Per ogni indirizzo possono esserci molteplici persone. In seguito, le persone estratte vengono contattate nei modi tradizionali (lettera, telefono e/o visita a casa) e vengono invitate ad entrare nel panel. Per permettere a tutti i componenti di poter rispondere ai questionari telematici, viene fornita, a coloro che ne necessitano, una connessione a banda larga e un piccolo dispositivo chiamato "simPC". Di questi 10150, a febbraio 2008, sono entrati a far parte del panel 5176. Per ottenere questo campione finale sono stati eliminati gli indirizzi non validi e si è tenuto conto della disponibilità a partecipare al panel.

Nel novembre 2011, il questionario sulla soddisfazione dei consumatori, che verrà analizzato in questa tesi, venne presentato per una prima ondata a 7093 componenti del panel, di cui 5470 lo hanno iniziato e 5449 lo hanno completato (percentuale di risposta del 76,8%).

¹Longitudinal Internet Studies for the Social Sciences

²Measurement and Experimentation in the Social Sciences

³Centraal Bureau voor de Statistiek (CBS), Statistics Nederland

3.2 Il Questionario

Il questionario in esame è un modulo all'interno del LISS Panel, nell'ambito di un progetto portato avanti da Paccagnella Omar, Bassi Francesca e Varriale Roberta. E' stato costruito in modo tale da poter essere completato in pochissimo tempo, questo per ridurre il numero di missing values. Essendo un *panel* le domande vengono ripetute per 3 volte nel giro di un anno, una ogni sei mesi. L'analisi, in questa tesi, è stata però effettuata solo sui dati della prima rilevazione.

Agli intervistati, come prima domanda, è stato chiesto se negli ultimi sei mesi avessero acquistato uno o più dei seguenti prodotti:

- un laptop
- un Apple iPhone
- un Televisore LCD

Coloro i quali avessero risposto positivamente, venivano suddivisi (in modo casuale) in due gruppi, denominati A e B. Nel primo, la domanda di auto-valutazione sulla soddisfazione compariva **prima** delle domande di valutazione sulle vignettes. Nel gruppo B invece, compariva **dopo**.

Un'ulteriore divisione è effettuata all'interno dei due gruppi, A e B, creando così i gruppi A1, A2, B1 e B2. Nei gruppi A1 e B1 sono presenti cinque categorie di risposta (Molto Soddisfatto, Soddisfatto, Né Soddisfatto Né Insoddisfatto, Insoddisfatto, Molto Insoddisfatto). Nei gruppi A2 e B2 invece, la categoria mediana (Né Soddisfatto Né Insoddisfatto) non è presente riducendo il numero di risposte possibili a quattro.

Successivamente alla domanda sugli acquisti, usata per selezionare quali individui proseguiranno l'indagine, sono presenti domande relative all'oggetto acquistato, dove viene chiesto il mese di acquisto del prodotto, se si sono riscontrati problemi nell'utilizzo e/o durante l'acquisto, le aspettative che si riponevano nel prodotto, la qualità di questo e infine, ripensando alle aspettative, se queste sono state soddisfatte.

Da questo punto comincia la suddivisione in gruppi che abbiamo esplicitato in precedenza poiché vengono poste le domande sulla soddisfazione del proprio acquisto

e sulla valutazione delle vignettes. Quest'ultime sono due e rappresentano uno scenario fittizio dove un individuo ha comperato lo stesso prodotto dell'intervistato, e presenta diverse esperienze d'uso.

Le domande si concludono qui, rendendo quindi la compilazione del questionario molto veloce e semplice. In Appendice A è riportato l'intero questionario.

Analizziamo ora alcune statistiche descrittive del campione in modo da conoscerne la composizione. Statistiche più dettagliate verranno analizzate nel capitolo successivo sul campione ridotto, oggetto dell'analisi. Il sesso è quasi equamente distribuito con una leggera maggioranza di donne (53,2%), come si può notare in figura 3.1. L'età ha una media di circa 50 anni e la numerosità delle classi aumenta con l'aumentare dell'età. (Figure 3.2) E' interessante osservare che la maggior parte degli intervistati

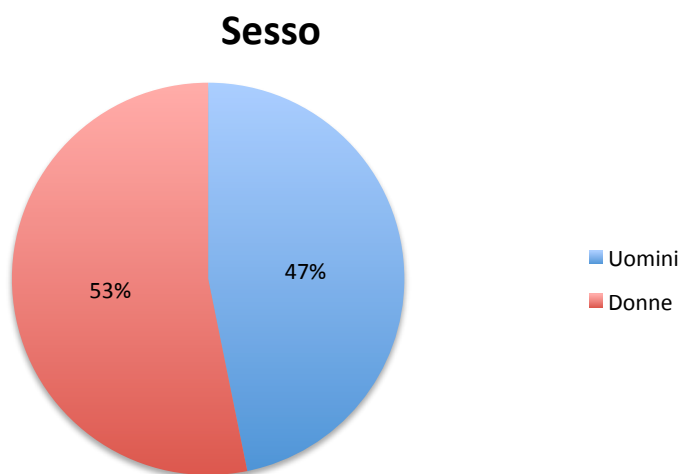


Figura 3.1: Percentuale di uomini e donne all'interno del campione

è sposata (57,8%) e con l'abitazione di proprietà (73,5%). Le località di residenza sono praticamente equidistribuite (a meno dei valori estremi: più di 2500 e meno di 500 abitanti per km^2) come se nota in Fig 3.3. Circa la metà degli intervistati è un lavoratore dipendente (50,8%), il 4,8% è un lavoratore autonomo e i restanti non lavorano; di questi il 47% sono pensionati e 39,3% sono studenti e casalinghe/i.

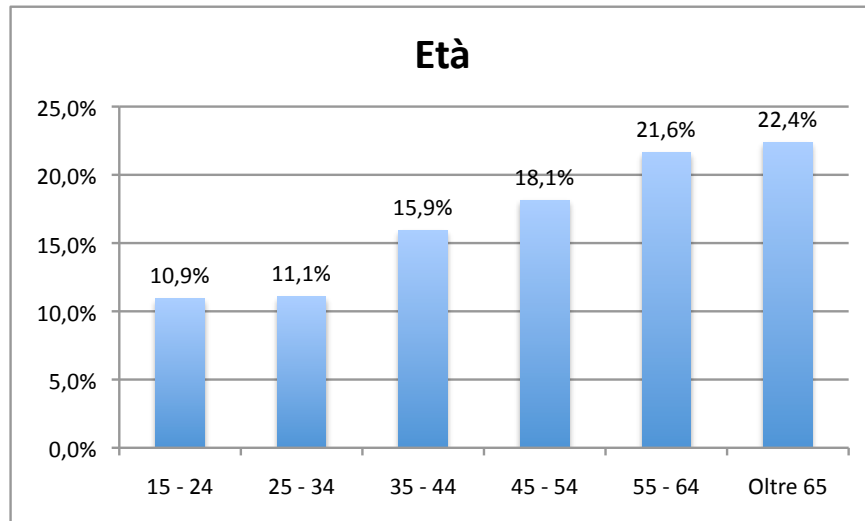


Figura 3.2: Distribuzione dell'età all'interno del campione

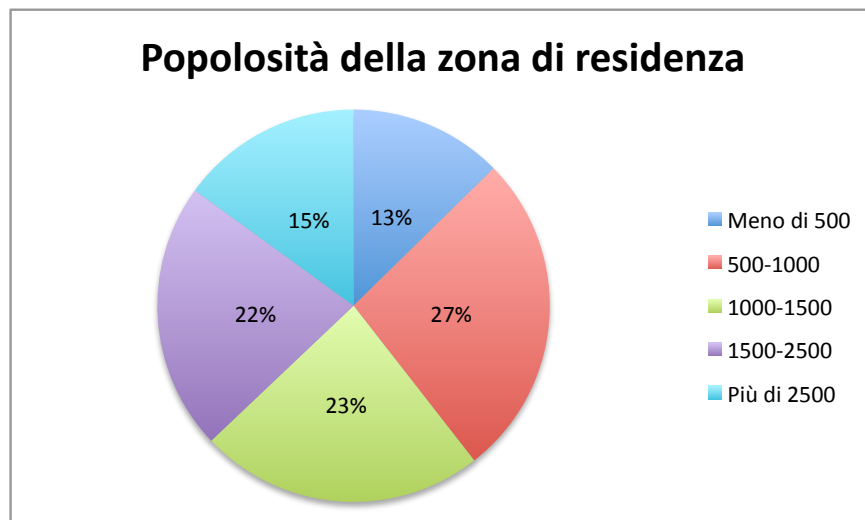


Figura 3.3: Percentuale di intervistati che abita nelle diverse fasce di popolosità (valori per km^2)

Capitolo 4

L'analisi

Come spiegato nel capitolo precedente, questa tesi sfrutterà solamente i dati raccolti per mezzo della prima rilevazione del progetto VECS.

I rispondenti, 5470, vengono casualmente divisi nei gruppi A (2804 persone) e B (2720) e, come esplicito in precedenza, a loro volta nei gruppi 1 e 2, come si vede in Tabella 4.1. A questi viene chiesto se hanno acquistato o meno uno dei tre oggetti

Gruppo	Frequenza	Percentuale
A1	1333	24.37
A2	1369	25.03
B1	1417	25.90
B2	1351	24.70
Totale	5470	100

Tabella 4.1: Suddivisione in gruppi del campione

descritti in precedenza: uno smartphone, un laptop o un TV LCD.

Per semplificare, l'analisi verrà effettuata sui rispondenti che hanno comprato almeno uno smartphone (Tab 4.2) e che sono appartenenti al gruppo 1 (sia A che B).

Il campione risultante è di 270 individui. Per gli intervistati che hanno effettuato più di un acquisto, viene selezionato casualmente il dispositivo su cui fare l'analisi. Dei 270 individui che avevano comprato almeno uno smartphone, quindi, ne rimangono 219, di cui 106 nel gruppo A1 e 113 nel gruppo B1.

Acquisti	Frequenza	Percentuale
Nessun acquisto	4234	77.40
Solo laptop	369	6.75
Solo smartphone	404	7.39
Solo TV	307	5.61
Laptop e Smartphone	71	1.30
Laptop e TV	35	0.64
TV e Smartphone	38	0.69
Tutti e tre	12	0.22
Totale	5470	100

Tabella 4.2: Tipologie di acquisti per rispondenti

4.1 Statistiche Descrittive

Verranno ora analizzate le principali statistiche descrittive del sottocampione preso in esame. Il sesso è quasi equamente distribuito, con una leggerissima maggioranza di uomini (50,7%) mentre l'età (Fig 4.1), contrariamente a quanto avveniva nel campione intero, presenta numerosità decrescenti all'aumentare degli anni degli intervistati. Questo non è insolito, infatti è lecito pensare che siano le persone più giovani ad acquistare e cambiare più spesso smartphone, piuttosto che persone più anziane. La giovane età del campione (media di circa 35 anni contro quella di 50 del campione completo) influisce anche sulle occupazioni, in quanto ben il 31,9% risulta essere ancora studente, mentre il 52,2% lavora. I laureati sono solo il 5%. Le abitazioni di proprietà risultano nel 72,2% del campione, un dato che potrebbe suggerire che il campione è composto da individui mediamente benestanti.

Dal grafico sulla soddisfazione (Fig 4.2) è evidente l'asimmetria delle risposte, quasi totalmente positive, dato riscontrabile anche dalla percentuale di persone che hanno effettivamente riscontrato un problema con il loro acquisto, solo il 12,3%. È interessante notare come questa soddisfazione si distribuisca diversamente se divisa per sesso, Fig 4.3, in quanto gli intervistati di sesso maschile sono più propensi ad utilizzare le categorie estreme, rispetto alle donne.

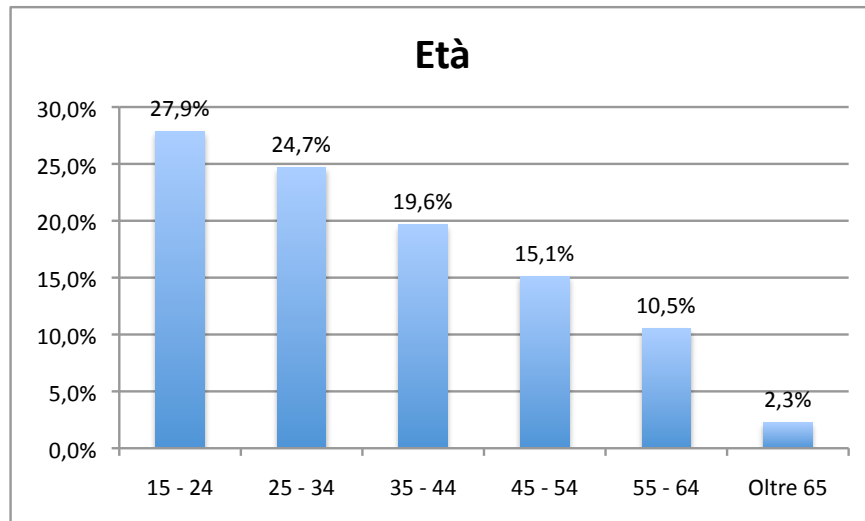


Figura 4.1: Distribuzione dell'età nel sottocampione

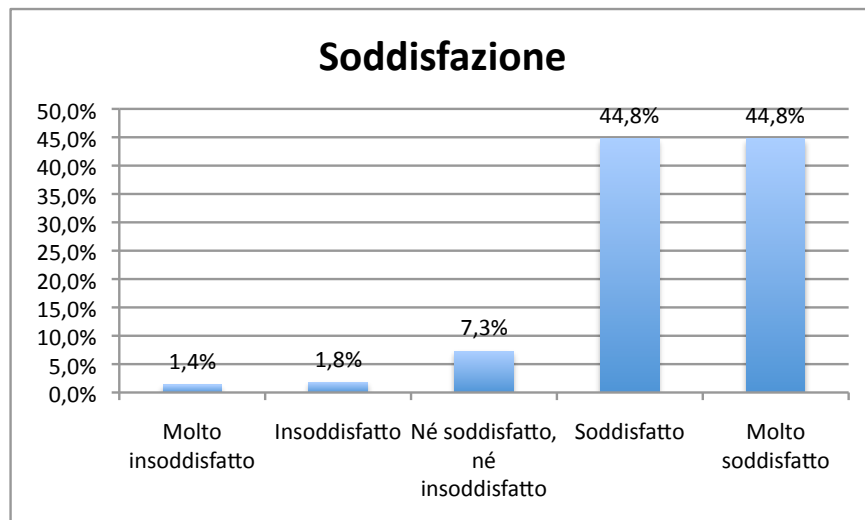


Figura 4.2: Soddisfazione sull'utilizzo dello Smartphone per gli utenti del sottocampione

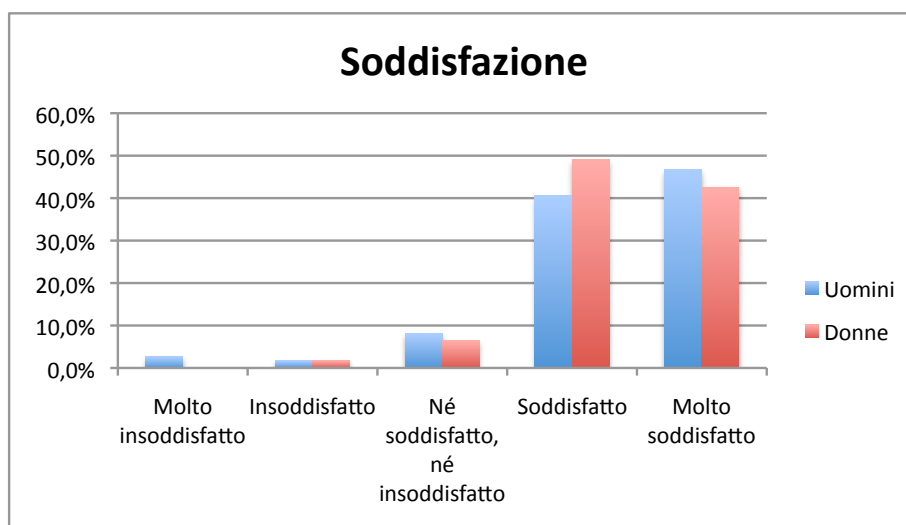


Figura 4.3: Soddisfazione sull'utilizzo dello Smartphone divisa per sesso

4.2 Stime dei Modelli

In questa sezione verranno analizzate prima le stime del modello Probit Ordinato ed in seguito le stime del modello Hopit.

Le variabili di interesse inserite nel modello sono l'età, centrata sui 45 anni (*Età*), il sesso (*sessu*), la presenza di laurea (*laurea*), una variabile dicotomica per indicare se l'intervistato lavora (*lavoro*), una per indicare se la casa dove vive è di proprietà (*proprietà*), una che indica se ha avuto problemi o meno durante l'utilizzo del prodotto (*problemi*) ed infine una dummy che indica se l'intervistato appartiene al gruppo A o B (*gruppoAB*).

4.2.1 Modello Probit Ordinato

Come si è visto nel Capitolo 2, il modello Hopit è un'estensione del modello Probit Ordinato, in quanto il primo stima delle soglie differenti per ogni individuo, mentre il secondo le fissa uguali per tutti.

Le stime del modello Probit Ordinato relative alla soddisfazione sull'utilizzo dello smartphone sono visibili in tabella 4.3. Si nota subito come le variabili Lavoro, Proprietà e Problemi siano fortemente significative e come le prime due, *ceteris paribus*, aumentino la probabilità di essere soddisfatti, mentre la variabile Problemi

Soddisfazione	Coefficiente	s.e.	p-value ^a
Età	0,1089	0,059	0,065*
GruppoAB	0,0739	0,1577	0,639
Sesso	0,0041	0,1594	0,980
Problemi	0,6911	0,2297	0,003***
Laurea	-0,0592	0,3631	0,871
Lavoro	-0,4737	0,1737	0,006***
Proprietà	-0,4977	0,1779	0,005***
Soglia 1	-0,7926	0,2469	
Soglia 2	0,7092	0,2487	
Soglia 3	1,3907	0,2777	
Soglia 4	1,7785	0,3183	

Tabella 4.3: Stime del modello Ordered Probit

^a* < 0,1; ** < 0,05; *** < 0,01

la diminuisca¹. Anche l'età ha un effetto negativo sulla probabilità di essere molto soddisfatti (ma al 10% di livello di significatività), indicando che una persona con più di 45² anni potrebbe essere più esigente e, di conseguenza, meno soddisfatto rispetto alle persone più giovani.

Verranno ora calcolate le probabilità nei vari livelli di soddisfazione per una persona tipo, ad esempio un uomo, appartenente al gruppo A, con un'età di 45 anni, che non ha avuto nessun tipo di problema, non laureato, ma con un lavoro e che vive in una casa di proprietà (Tab 4.4). Una persona con queste caratteristiche ha chiaramente una probabilità molto elevata di essere Soddisfatto e Molto Soddisfatto. Vengono ora stimate le stesse probabilità, riferite ad una stessa persona, ma che ha riscontrato almeno un problema (Tab 4.5).

Come si nota anche in Figura 4.4, l'aver riscontrato problemi nell'utilizzo del dispositivo cambia in modo significativo le probabilità di risposta comportando

¹Il dataset indica come 1 la categoria "molto soddisfatto" e con 5 la categoria "molto insoddisfatto", quindi un segno positivo nel parametro indica che ha un effetto negativo sulla soddisfazione

²L'età è centrata sui 45 anni

Categorie	Probabilità
Molto Soddisfatto	56,9%
Soddisfatto	38,4%
Né soddisfatto, Né insoddisfatto	3,8%
Insoddisfatto	0,6%
Molto Insoddisfatto	0,3%

Tabella 4.4: Probabilità di risposta per le varie categorie con le stime del modello Ordered Probit

Categorie	Probabilità
Molto Soddisfatto	30,3%
Soddisfatto	53,5%
Né soddisfatto, Né insoddisfatto	11,4%
Insoddisfatto	2,8%
Molto Insoddisfatto	2%

Tabella 4.5: Probabilità di risposta per le varie categorie con le stime del modello Ordered Probit di una persona con problemi nell'utilizzo del prodotto

una forte diminuzione della probabilità di essere molto soddisfatto e aumentando le categorie peggiori.

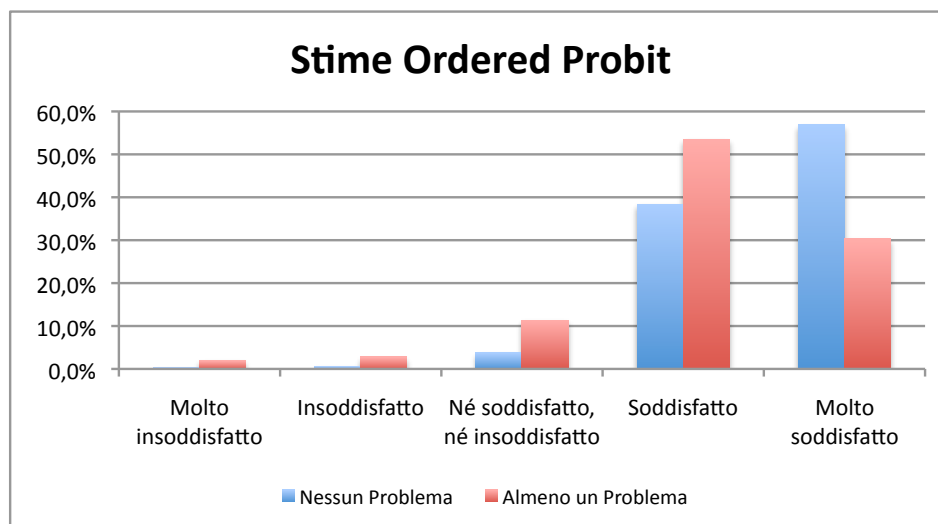


Figura 4.4: Probabilità di risposta per le varie categorie con le stime del modello Ordered Probit nel caso di riscontro di problemi o meno

4.2.2 Modello Hopit

Viene ora introdotta l'informazione fornita dalle vignettes che, attraverso il modello Hopit, permetterà di stimare delle soglie differenti per ogni individuo (Tab 4.6).

Tabella 4.6: Stime del modello Hopit

Soddisfazione	Coefficiente	s.e.	p-value
Età	0,0455	0,0696	0,513
GruppoAB	-0,1109	0,1858	0,550
Sesso	-0,1182	0,1875	0,528
Problemi	0,8186	0,2695	0,002***
Laurea	-0,2794	0,4502	0,535
Lavoro	-0,3	0,2034	0,140
Proprietà	-0,6177	0,2078	0,003***
θ_1	-0,5894	0,2857	0,039**
θ_2	0,6477	0,2819	0,022**

Continua nella pagina successiva —

— Continua dalla pagina precedente

Soglia 1	Coefficiente	s.e.	p-value
Età	-0,066	0,0452	0,146
GruppoAB	-0,2142	0,1174	0,068*
Sesso	-0,1219	0,1184	0,303
Problemi	0,1136	0,1712	0,507
Laurea	-0,2958	0,2908	0,309
Lavoro	0,2645	0,1306	0,043**
Proprietà	-0,2539	0,131	0,052*
Costante	-0,734	0,2556	0,004***
Soglia 2	Coefficiente	s.e.	p-value
Età	-0,005	0,0325	0,877
GruppoAB	0,1428	0,0889	0,108
Sesso	0,0619	0,0907	0,494
Problemi	0,015	0,1324	0,910
Laurea	0,2416	0,1991	0,225
Lavoro	-0,1717	0,0966	0,075*
Proprietà	0,2124	0,1063	0,046**
Costante	0,2317	0,1426	0,104
Soglia 3	Coefficiente	s.e.	p-value
Età	0,0544	0,0626	0,385
GruppoAB	-0,0382	0,1736	0,826
Sesso	-0,1184	0,1794	0,509
Problemi	-0,2728	0,2864	0,341
Laurea	-0,6188	0,5463	0,257
Lavoro	-0,1029	0,1926	0,593
Proprietà	0,3582	0,2201	0,104
Costante	-0,2107	0,2437	0,387
Soglia 4	Coefficiente	s.e.	p-value
Età	0,0167	0,2107	0,937
GruppoAB	-1,0038	0,5887	0,088*
Sesso	-0,8898	0,5707	0,119

Continua nella pagina successiva —

— Continua dalla pagina precedente

Problemi	1,3456	0,8012	0,093*
Laurea	-0,2175	0,7661	0,776
Lavoro	0,8706	0,7082	0,219
Proprietà	0,1898	0,6258	0,762
Costante	-0,3906	0,6512	0,549

Da queste stime si nota come le variabili “Proprietà ” e “Problemi” rimangano significative per la soddisfazione. La prima con un effetto che aumenta la probabilità di essere soddisfatti mentre la seconda con un effetto negativo sulla soddisfazione. Analizzando le stime delle soglie si possono notare dei parametri significativi per le soglie 1, 2 e 4, mentre nessun parametro risulta significativo per la soglia 3. Ciò avvalorata l’ipotesi che ci siano differenze tra gli individui nella percezione delle soglie, e che quindi il modello Ordered Probit potrebbe non essere adatto a stimare questo scenario. Ad esempio la variabile Proprietà nella soglia 1 è significativa con segno negativo, mentre nella soglia 2 lo è con segno positivo; ciò implica che gli intervistati che vivono in una casa di proprietà avranno le soglie che delimitano la probabilità di appartenere alla categoria “soddisfatto” più allargate (probabilità maggiore) rispetto agli intervistati che vivono in affitto.

E’ interessante notare come il parametro che indica l’appartenenza al gruppo (A o B) sia significativo per la prima e l’ultima soglia. Questo indica che, a parità di altre condizioni, la posizione delle vignettes, prima o dopo la domanda sulla soddisfazione, modifica le soglie degli individui. Questo risultato potrebbe rappresentare un’evidenza contro l’assunzione di *Response Consistency*. Questa assunzione implica, infatti, che gli intervistati utilizzino le stesse soglie per rispondere sia alla domanda sull’autovalutazione, sia a quelle sugli scenari rappresentati dalle vignettes. Se le soglie differiscono, a parità di altre condizioni, solamente per la posizione delle vignettes, significa che queste potrebbero influenzare le risposte degli intervistati. Ciò porterebbe al non poter più confrontare i risultati degli intervistati appartenenti ai due gruppi. Tuttavia questo risultato potrebbe essere dovuto alla mancata valutazione di alcune variabili che rendevano gli individui, appartenenti ai due gruppi, differenti. Una più approfondita analisi su questo problema sarebbe quindi necessaria per valutare ulteriormente la validità dell’assunzione di RC.

Le differenze rispetto al modello Ordered Probit, si notano subito dalle soglie. In

tabella 4.7, si vede come le soglie del modello Hopit, relative all'intervistato "tipo" rappresentato in precedenza, differiscono in maniera sostanziale da quelle stimate dal modello Probit Ordinato.

Soglie	Modello OProbit	Modello Hopit Senza Problemi	Modello Hopit Con Problemi
Soglia 1	-0,7926	-0,8453	-0.7317
Soglia 2	0,7092	0,5517	0.6864
Soglia 3	1,3907	1,4806	1.3935
Soglia 4	1,7785	2,2831	4,4756

Tabella 4.7: Differenze tra le soglie stimate da un modello Probit Ordinato e da un Hopit

Categorie	Modello OProbit Senza Problemi	Modello Hopit Senza Problemi
Molto Soddisfatto	56,9%	57,6%
Soddisfatto	38,4%	36,8%
Né soddisfatto, Né insoddisfatto	3,8%	5%
Insoddisfatto	0,6%	0,5%
Molto Insoddisfatto	0,3%	0%

Tabella 4.8: Probabilità di risposta per le varie categorie con le stime del modello Ordered Probit confrontate alle stime del modello Hopit, per una persona che non ha riscontrato problemi.

Analizzando le stime presenti in tabella 4.8 e 4.9 si nota che non vi sono differenze sostanziali tra i due modelli. Tuttavia il modello Hopit riesce comunque a cogliere degli effetti che alcune variabili hanno nel modificare le soglie percepite dagli intervistati. Confrontando i modelli anche riguardo la variabile Proprietà (Tab 4.10), si nota una grande differenza tra i due modelli: nel modello Ordered Probit, gli individui che vivono in una casa in affitto sono per la maggior parte "soddisfatti", invece nel modello Hopit, sono per la maggior parte "Molto Soddisfatti". Inoltre, come accennato in precedenza, si nota come gli intervistati che vivono in affitto, ab-

Categorie	Modello OProbit	Modello Hopit
	Con Problemi	Con Problemi
Molto Soddisfatto	30,3%	30,4%
Soddisfatto	53,5%	51,3%
Né soddisfatto, Né insoddisfatto	11,4%	12,9%
Insoddisfatto	2,8%	5,4%
Molto Insoddisfatto	2%	0%

Tabella 4.9: Probabilità di risposta per le varie categorie con le stime del modello Ordered Probit confrontate alle stime del modello Hopit, per una persona che ha riscontrato problemi.

biano una probabilità maggiore ad essere “soddisfatti”, rispetto a quelli che vivono in un’abitazione di proprietà (Tab 4.8).

Categorie	OProbit	Hopit
Molto Soddisfatto	37,3%	43,1%
Soddisfatto	50,7%	39,9%
Né soddisfatto, Né insoddisfatto	8,8%	11,5%
Insoddisfatto	1,9%	4,3%
Molto Insoddisfatto	1,2%	1,2%

Tabella 4.10: Probabilità di risposta per le varie categorie, di una persona che non vive in una casa di proprietà

4.3 Counterfactuals

Il termine *Counterfactuals*, che potrebbe essere tradotto in italiano con “Controfattuale”, indica una situazione che non si è verificata, ma che sarebbe interessante da studiare. Nel contesto di questa tesi, i Counterfactuals indicano degli studi su come le probabilità di risposta possono venire modificate dall’utilizzo di soglie differenti fra individui. Ad esempio ci si potrebbe chiedere: Come cambierebbero le risposte di un individuo “tipo” appartenente al gruppo B, se fossero utilizzate solo le soglie dei rispondenti al gruppo A?

Per far ciò è necessario costruire un benchmark *fissando* la variabile (ad esempio GruppoAB) uguale per tutti, indipendentemente dal gruppo di appartenenza.

Verranno studiati tre counterfactuals:

- Come si modificano le risposte di un individuo appartenente al gruppo B, se si utilizzano le soglie dei rispondenti al gruppo A? E viceversa
- Come si modificano le risposte di un individuo lavoratore, se si utilizzano le soglie dei rispondenti non lavoratori? E viceversa
- Come si modificano le risposte di un individuo che vive in una casa di proprietà, se si utilizzano le soglie dei rispondenti che vivono in case in affitto? E viceversa

4.3.1 Benchmark: GruppoAB

In tabella 4.11 compaiono le soglie utilizzando come benchmark i due gruppi, A e B. Come si era notato in precedenza in tabella 4.6 le differenze significative sono

Soglie	Gruppo A	Gruppo B
Soglia 1	-0,8453	-1.0595
Soglia 2	0,5517	0.5519
Soglia 3	1,4806	1,4459
Soglia 4	2,2831	1,74

Tabella 4.11: Differenze tra le soglie stimate da un modello Hopit, fissando come benchmark il gruppo A e il gruppo B

nella soglia 1 e nella 4. Nelle figure 4.5 e 4.6 si nota come utilizzare le soglie degli appartenenti al gruppo A, fa aumentare la proporzione di intervistati del gruppo B che avrebbero risposto “Molto Soddisfatto”. Per contro utilizzando le soglie del gruppo B diminuirebbe la proporzione di individui del gruppo A che risponderebbero con la categoria di soddisfazione maggiore. Questo significa che, a parità di vero livello di soddisfazione, gli intervistati appartenenti al gruppo A sono più propensi ad essere “Molto Soddisfatti”, rispetto a quelli del gruppo B. Ciò, come analizzato in precedenza potrebbe essere un'evidenza contro la validità dell'assunzione di *Response Consistency*.

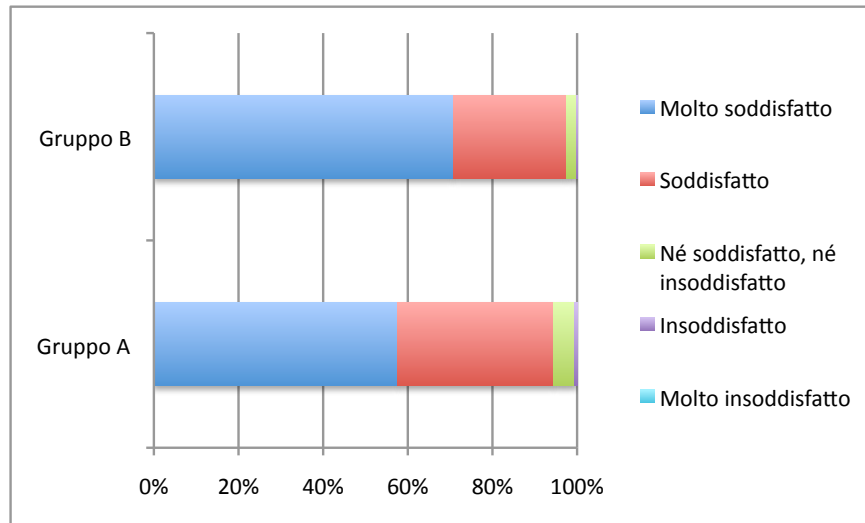


Figura 4.5: Differenze di proporzioni tra Gruppo A e Gruppo B utilizzando come benchmark se soglie del Gruppo A

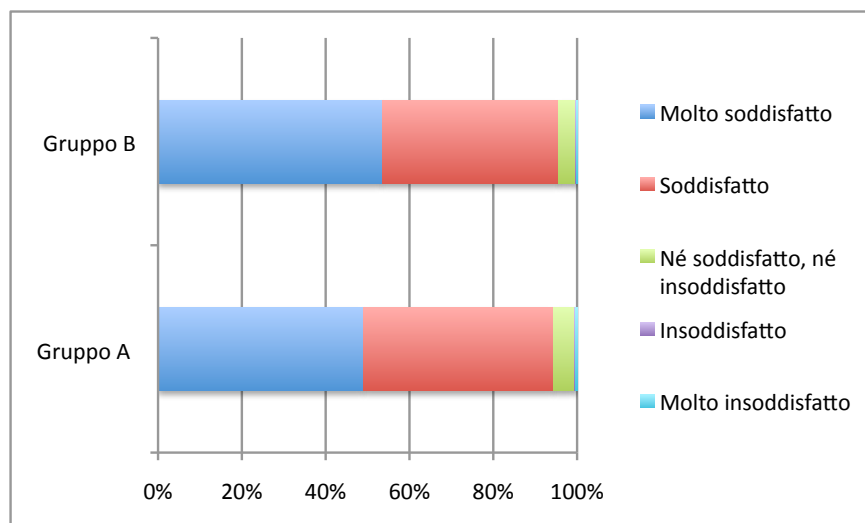


Figura 4.6: Differenze di proporzioni tra Gruppo A e Gruppo B utilizzando come benchmark le soglie del Gruppo B

4.3.2 Benchmark: Lavoro

Risulta anche interessante analizzare come vengono modificate le proporzioni di individui lavoratori che rispondono con la categoria “Molto Soddisfatto”, utilizzando le soglie di un non lavoratore e viceversa. Le soglie utilizzando come benchmark le due tipologie di intervistati sono presenti in tabella 4.12. Come si nota nelle figure 4.7 e 4.8, le proporzioni per la risposta “Molto Soddisfatto” sono più elevate utilizzando come soglie quelle degli intervistati lavoratori. Ciò significa che, a parità di vero livello di soddisfazione, gli intervistati lavoratori hanno la tendenza ad indicare una soddisfazione superiore, rispetto agli individui non lavoratori. Questa situazione potrebbe essere dovuta al fatto che gli individui non lavoratori sono in gran parte studenti; questi potrebbero essere molto più informati rispetto agli adulti sulle novità e le tendenze del settore, quindi pur essendo soddisfatti del loro acquisto, indicano una soddisfazione inferiore poiché conoscono quali caratteristiche aggiuntive il prodotto potrebbe implementare.

Soglie	Lavoratore	Non Lavoratore
Soglia 1	-0,8453	-1,1098
Soglia 2	0,5517	0,5488
Soglia 3	1,4806	1,5783
Soglia 4	2,2831	1,914

Tabella 4.12: Differenze tra le soglie stimate da un modello Hopit, fissando come benchmark gli individui lavoratori e non.

4.3.3 Benchmark: Proprietà

Un'ultima interessante analisi che si può effettuare è utilizzare come benchmark la proprietà della casa dove si vive. Questa variabile era, infatti, significativa nelle prime due soglie.

Le soglie stimate per i due gruppi (Proprietari e non) sono presenti in tabella 4.13. Dai grafici 4.9 e 4.10 si nota che le proporzioni di individui che hanno indicato “Molto Soddisfatto” sono più elevate utilizzando le soglie degli individui non proprietari della casa dove vivono. Questo significa che, a parità del vero livello di soddisfazione gli

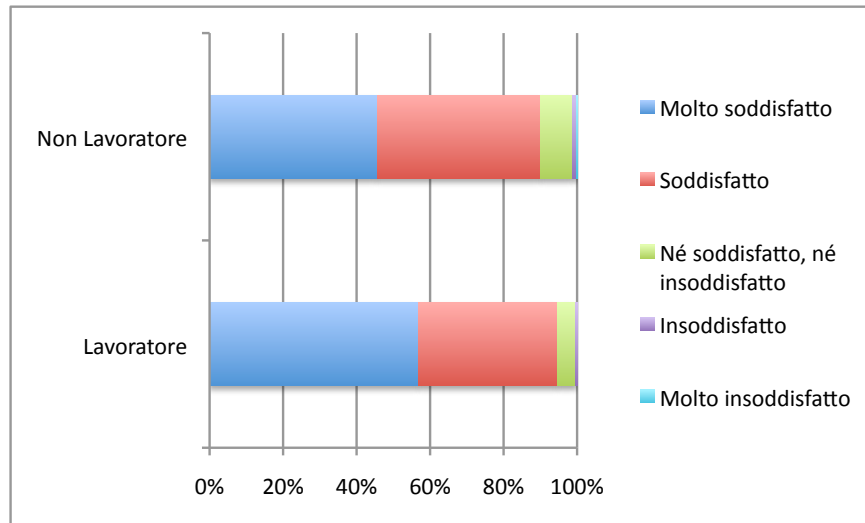


Figura 4.7: Differenze di proporzioni tra lavoratori e non utilizzando come benchmark le soglie dei lavoratori

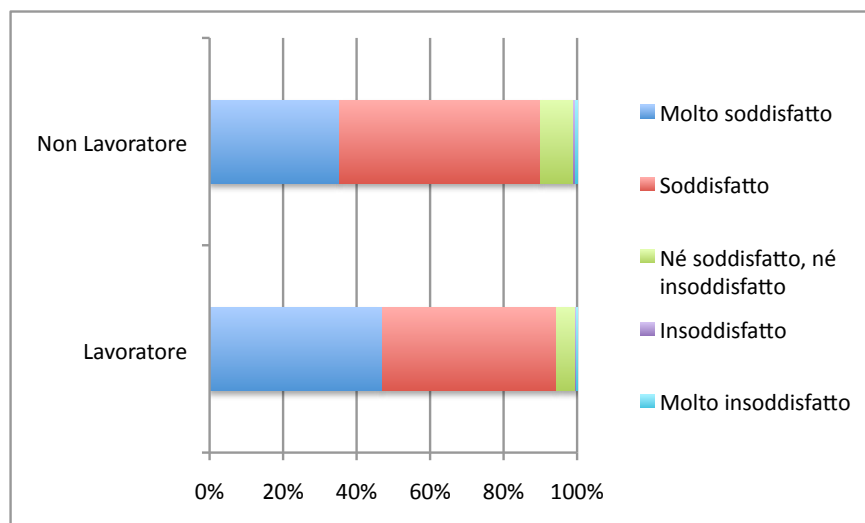


Figura 4.8: Differenze di proporzioni tra lavoratori e non utilizzando come benchmark le soglie dei non lavoratori

Soglie	Casa di Proprietà	Casa non di Proprietà
Soglia 1	-0,8453	-0.5914
Soglia 2	0,5517	0,5384
Soglia 3	1,4806	1,1876
Soglia 4	2,2831	1,8514

Tabella 4.13: Differenze tra le soglie stimate da un modello Hopit, fissando come benchmark gli individui proprietari della casa dove vivono e non.

individui proprietari di casa saranno più propensi a dichiarare di essere “Soddisfatti”, piuttosto che “Molto Soddisfatti” rispetto ai non proprietari.

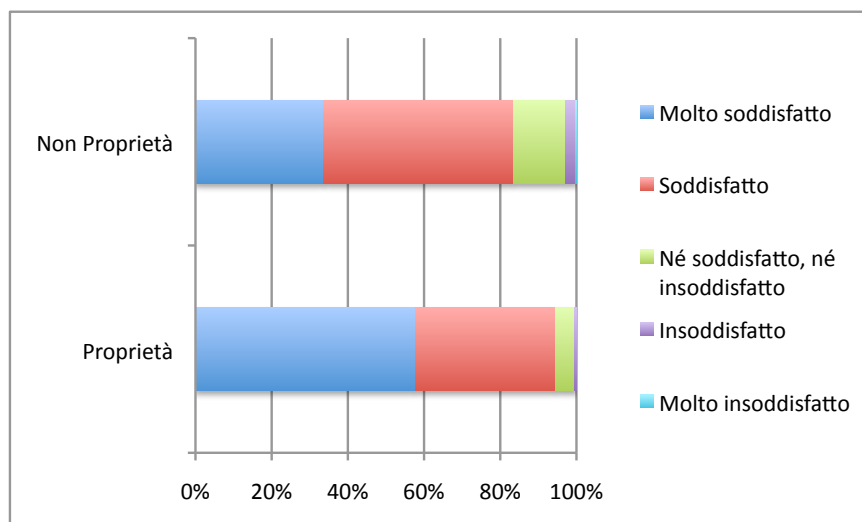


Figura 4.9: Differenze di proporzioni tra individui proprietari di casa e non utilizzando come benchmark le soglie dei proprietari dell'abitazione dove vivono

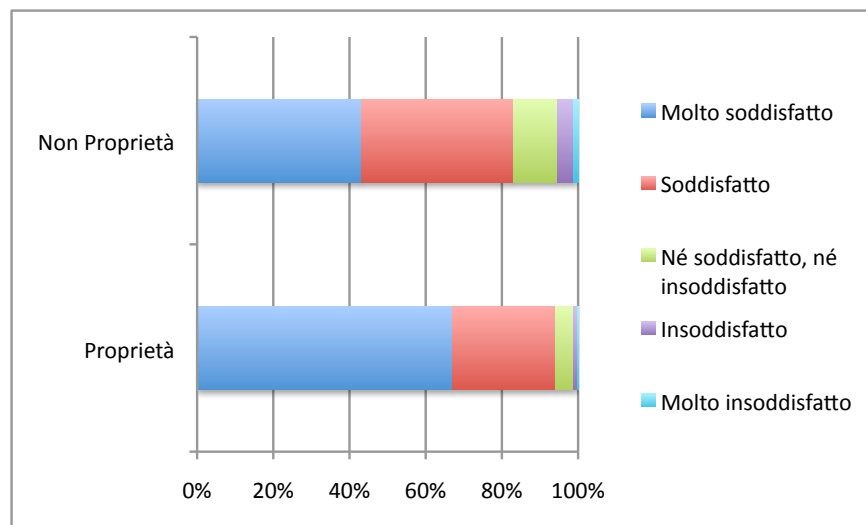


Figura 4.10: Differenze di proporzioni tra individui proprietari di casa e non utilizzando come benchmark le soglie dei non proprietari dell'abitazione dove vivono

Conclusioni

Gli obiettivi di questa tesi erano due: uno empirico ed uno metodologico. Il primo si proponeva di arricchire la letteratura di settore, introducendo uno studio che, in futuro, potrebbe tramutarsi in un nuovo metodo per la misura della Customer Satisfaction. Il secondo si proponeva, invece, di studiare se il cambiamento della posizione delle vignettes rispetto alla domanda sull'autovalutazione avesse qualche effetto sulla correzione del DIF.

I dati che sono stati utilizzati provengono dallo studio denominato VECS (Vignette Evaluation of Customer Satisfaction), proposto da Paccagnella Omar, Bassi Francesca e Varriale Roberta, all'interno del Panel LISS, parte centrale del programma MESS condotto in Olanda da CenERdata, con la collaborazione dell'istituto di Statistica olandese.

Per quanto riguarda l'obiettivo empirico, il modello Hopit è adatto quando si ha la necessità di dover confrontare i risultati di un'indagine tra popolazioni e/o gruppi differenti (ad esempio in un'indagine svolta a livello di Unione Europea). Rispetto al modello Probit Ordinato, il modello Hopit sfrutta le informazioni aggiuntive fornite dalle vignettes per stimare l'impatto delle variabili inserite nel modello, modificando le soglie per ogni individuo. Questo risulta evidente quando si analizzano le probabilità di risposta per un individuo "tipo"; ad esempio, se costui non è proprietario della casa dove vive, secondo il modello Probit Ordinato avrà una probabilità superiore al 50% di essere "Soddisfatto", secondo il modello Hopit, invece, tale probabilità scende al 40%. Sarà interessante proseguire questo studio nelle successive ondate dell'indagine per capire se i risultati risulteranno simili.

Analizzando l'obiettivo metodologico, si può affermare che sicuramente la posizione delle vignettes incide sulla correzione del DIF. Gli intervistati sono stati divisi in due gruppi, A e B. Nel primo la domanda sulla soddisfazione viene prima delle vi-

gnettes, nel secondo avviene il contrario. Si è visto, anche tramite i counterfactuals, che gli intervistati appartenenti al gruppo B hanno, a parità di vero ed ignoto livello di soddisfazione la tendenza a indicare una soddisfazione più elevata rispetto agli appartenenti al gruppo A. Ciò comporta la possibilità che l'assunzione di *Response Consistency* non risulti valida. Questa assunzione prevede la corrispondenza delle soglie utilizzate per l'autovalutazione e quelle utilizzate per valutare le vignettes. Se le soglie vengono modificate, ceteris paribus, a causa del diverso ordinamento delle domande, una possibile conclusione è che le vignettes modifichino le percezioni degli intervistati, modificandone le soglie. Ulteriori analisi sono comunque necessarie per approfondire questo argomento.

Appendice A

Il Questionario

WAVE 1 QUESTIONNAIRE

```
| Q1: In the last 6 months, have you experienced one or more of the following?
|   [NOTE: CODE ALL THAT APPLY. In case of more purchases of the same good in the last 6
|   months, please refer to the most recent purchase]
|   1) I bought a laptop.
|   2) I bought an Apple I-phone.
|   3) I bought a LCD TV.
|   4) None of these
|
|
| | IF Q1== 4
| | Q2: Do you intend to buy a laptop or an Apple I-phone or a LCD TV within the next months?
| |   1) Yes
| |   2) No
| | ENDIF
|
|
| | IF Q1 <> 4
| |   [note: ONLY ONE LOOP IS PLANNED. In case of more than 1 purchase,
| |   only one good (laptop or Apple I-phone or LCD TV) is randomly selected among
| |   the choices in Q1]
| |
| |
| | Q3: In which month did you buy your [laptop/Apple I-phone/LCD TV]?
| |   1) November 2011
| |   2) October 2011
| |   3) September 2011
| |   4) August 2011
| |   5) July 2011
| |   6) June 2011
| |   7) May 2011
| |
| |
```

CAPITOLO A. II Questionario

| | Q4: Did you experience one of the following after the purchase of your [laptop / Apple I-phone /
| | LCD TV]? [NOTE: CODE ALL THAT APPLY]

- | | 1) A manufacturing defect.
- | | 2) A delay in the delivery.
- | | 3) A good with different features with respect to the purchase order.
- | | 4) A price different with respect to the purchase order.
- | | 5) None of these.

| | Q5: Before the purchase of your [laptop/Apple I-phone/LCD TV] you might have known
| | something about the product. Now think back and remember your expectations of the overall
| | quality of the [laptop/Apple I-phone/LCD TV]. Please give me a rating on a 10 point
| | scale on which "1" means your expectations were "not very high" and "10" means your
| | expectations were "very high". How would you rate your expectations of the overall quality
| | of the [laptop/Apple I-phone/LCD TV]?

| | [RECORD RATING 1-10]

| | Q6: Please consider all your experiences to date with your [laptop/Apple I-phone/LCD TV].
| | Using a 10 point scale on which "1" means "not very high" and "10" means "very high", how
| | would you rate the overall quality of your [laptop/Apple I-phone/LCD TV]?

| | [RECORD RATING 1-10]

| | Q7: Considering all of your expectations, to what extent has your [laptop/Apple I-phone/LCD TV]
| | fallen short of or exceeded your expectations? Using a 10 point scale on which "1" now means
| | "falls short of your expectations" and "10" means "exceeds your expectations", to what extent
| | has your [laptop/Apple I-phone/LCD TV] fallen short of or exceeded your expectations?

| | [RECORD RATING 1-10]

| | Q8: Would you suggest the purchase of your [laptop/Apple I-phone/LCD TV] to relatives or
| | friends?

- | | 1) Yes
- | | 2) No

| | [The sample is randomly divided in 2 groups { called A and B}]

| | [GROUP A]

| | Now, we would like to ask one question about how satisfied you are with some aspects of your
| | purchase.

| | [Group A sample is randomly divided in 2 other groups { called 1 and 2}]

| | [GROUP A1]

| |

| | Q9: How satisfied are you with your [laptop/Apple I-phone/LCD TV]?

| | 1) Very satisfied

| | 2) Satisfied

| | 3) Neither satisfied, nor dissatisfied

| | 4) Dissatisfied

| | 5) Very dissatisfied

| |

| |

| | We will now give you some examples of persons who experienced the purchase of some goods. We
 | | would like to know how you evaluate the given aspects of their purchases. Please assume that the
 | | persons have the same age and background that you have.
 | | Please choose one of the five answers for every question.

| | IF Q1== 2 [an Apple I-phone]

| | |

| | | Q10_2: Mark needs a mobile phone for his work. He went to a specialized shop and bought the last
 | | | version of the Apple I-phone because its features are suitable for his work. He had to wait
 | | | for four days in order to receive his I-phone. Reading the user guide, he was able to learn
 | | | its main features in a couple of days. He has never experienced any manufacturing defects.
 | | | How satisfied is Mark with his Apple I-phone?

| | | 1) Very satisfied

| | | 2) Satisfied

| | | 3) Neither satisfied, nor dissatisfied

| | | 4) Dissatisfied

| | | 5) Very dissatisfied

| | |

| | |

| | | Q11_2: Anne works part-time, makes various sports and has a lot of friends. She thinks an Apple
 | | | I-phone can meet her needs. In a shopping centre she immediately bought what she liked.
 | | | Unfortunately, after two months she experienced a problem in the phone book. She came
 | | | back to the shopping centre and the Apple I-phone was withdrawn for the assistance. After
 | | | ten days the phone was delivered to Anne and the problem was solved. She has not
 | | | experienced any other manufacturing defects. How satisfied is Anne with her Apple
 | | | I-phone?

| | | 1) Very satisfied

| | | 2) Satisfied

| | | 3) Neither satisfied, nor dissatisfied

| | | 4) Dissatisfied

| | | 5) Very dissatisfied

| |

| | [GROUP A2]

| |

| | Q9: How satisfied are you with your [laptop/Apple I-phone/LCD TV]?

| | 1) Very satisfied

| | 2) Satisfied

| | 3) Dissatisfied

| | 4) Very dissatisfied

| |

| |

| | We will now give you some examples of persons who experienced the purchase of some goods. We

CAPITOLO A. Il Questionario

| | would like to know how you evaluate the given aspects of their purchases. Please assume that the
| | persons have the same age and background that you have.

| | Please choose one of the four answers for every question.

| | | IF Q1== 2 [an Apple I-phone]

| | |

| | | Q10_2: Mark needs a mobile phone for his work. He went to a specialized shop and bought the last
| | | version of the Apple I-phone because its features are suitable for his work. He had to wait
| | | for four days in order to receive his I-phone. Reading the user guide, he was able to learn
| | | its main features in a couple of days. He has never experienced any manufacturing defects.
| | | How satisfied is Mark with his Apple I-phone?

| | | 1) Very satisfied

| | | 2) Satisfied

| | | 3) Dissatisfied

| | | 4) Very dissatisfied

| | |

| | |

| | | Q11_2: Anne works part-time, makes various sports and has a lot of friends. She thinks an Apple
| | | I-phone can meet her needs. In a shopping centre she immediately bought what she liked.
| | | Unfortunately, after two months she experienced a problem in the phone book. She came
| | | back to the shopping centre and the Apple I-phone was withdrawn for the assistance. After
| | | ten days the phone was delivered to Anne and the problem was solved. She has not
| | | experienced any other manufacturing defects. How satisfied is Anne with her Apple
| | | I-phone?

| | | 1) Very satisfied

| | | 2) Satisfied

| | | 3) Dissatisfied

| | | 4) Very dissatisfied

| | |

[GROUP B]

| | |

| | | [Group B sample is randomly divided in 2 other groups { called 1 and 2}]

| | |

| | |

[GROUP B1]

| | |

| | Now we will give you some examples of persons who experienced the purchase of some goods. We
| | would like to know how you evaluate the given aspects of their purchases. Please assume that the
| | persons have the same age and background that you have.

| | Please choose one of the five answers for every question.

| | | IF Q1== 2 [an Apple I-phone]

| | |

| | | Q10_2: Mark needs a mobile phone for his work. He went to a specialized shop and bought the last
| | | version of the Apple I-phone because its features are suitable for his work. He had to wait
| | | for four days in order to receive his I-phone. Reading the user guide, he was able to learn
| | | its main features in a couple of days. He has never experienced any manufacturing defects.
| | | How satisfied is Mark with his Apple I-phone?

| | | 1) Very satisfied

| | | 2) Satisfied

| | | 3) Neither satisfied, nor dissatisfied

| | | 4) Dissatisfied

| | | 5) Very dissatisfied

| | |

| | |
| | | Q11_2: Anne works part-time, makes various sports and has a lot of friends. She thinks an Apple
| | | I-phone can meet her needs. In a shopping centre she immediately bought what she liked.
| | | Unfortunately, after two months she experienced a problem in the phone book. She came
| | | back to the shopping centre and the Apple I-phone was withdrawn for the assistance. After
| | | ten days the phone was delivered to Anne and the problem was solved. She has not
| | | experienced any other manufacturing defects. How satisfied is Anne with her Apple
| | | I-phone?
| | | 1) Very satisfied
| | | 2) Satisfied
| | | 3) Neither satisfied, nor dissatisfied
| | | 4) Dissatisfied
| | | 5) Very dissatisfied

| | | [GROUP B2]

| | |
| | | Now we will give you some examples of persons who experienced the purchase of some goods. We
| | | would like to know how you evaluate the given aspects of their purchases. Please assume that the
| | | persons have the same age and background that you have.
| | | Please choose one of the four answers for every question
| | | IF Q1== 2 [an Apple I-phone]

| | | Q10_2: Mark needs a mobile phone for his work. He went to a specialized shop and bought the last
| | | version of the Apple I-phone because its features are suitable for his work. He had to wait
| | | for four days in order to receive his I-phone. Reading the user guide, he was able to learn
| | | its main features in a couple of days. He has never experienced any manufacturing defects.
| | | How satisfied is Mark with his Apple I-phone?
| | | 1) Very satisfied
| | | 2) Satisfied
| | | 3) Dissatisfied
| | | 4) Very dissatisfied

| | | Q11_2: Anne works part-time, makes various sports and has a lot of friends. She thinks an Apple
| | | I-phone can meet her needs. In a shopping centre she immediately bought what she liked.
| | | Unfortunately, after two months she experienced a problem in the phone book. She came
| | | back to the shopping centre and the Apple I-phone was withdrawn for the assistance. After
| | | ten days the phone was delivered to Anne and the problem was solved. She has not
| | | experienced any other manufacturing defects. How satisfied is Anne with her Apple
| | | I-phone?
| | | 1) Very satisfied
| | | 2) Satisfied
| | | 3) Dissatisfied
| | | 4) Very dissatisfied

Bibliografia

- Bago d'Uva T.; Lindeboom M.; O'Donnell O.; van Doorslaer E. (2009). Slipping anchor? testing the vignettes approach to identification and correction of reporting heterogeneity.
- Chevalier A.; Fielding A. (2011). An introduction to anchoring vignette. *Journal of the Royal Statistical Society*, **174**, 569–574.
- Das M.; Kepteyn A.; van Soest A. (2006). An advanced multi-disciplinary facility for measurement and experimentation in the social sciences (mess).
- Giudici P. (2005). *Data Mining. Metodi statistici per le applicazioni aziendali*. McGraw-Hill.
- Gramma (1993). *Customer Satisfaction. Misurare e Gestire la Soddisfazione del Cliente*. SEDI.
- Green W. H.; Hensher D. A. (2010). *Modeling Ordered Choices*. Cambridge University Press.
- Hayes B. (2009). *Misurare la Soddisfazione dei Clienti*. FrancoAngeli.
- Hopkins D.; King G. (2010). Designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly*.
- King G.; Murray C. J. L.; Solomon J. A.; Tandon A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, **98**, 191–207.
- Myers J. (1999). *Measuring Customer Satisfaction: hot buttons and other measurement issues*. American Marketing Association.

Paccagnella O. (2011). Anchoring vignettes with sample selection due to non-response. *Journal of the Royal Statistical Society*, **174**, 665–687.

Peracchi F.; Rossetti C. (2010). The heterogeneous thresholds ordered response model: Identification and inference.

van Soest A.; Delaney L.; Harmon C.; Kapteyn A.; Smith J. P. (2011). Validating the use of anchoring vignette for the correction of response scale differences in subjective questions. *Journal of the Royal Statistical Society*, **174**, 575–595.

Ringraziamenti

In this paper use is made of data of the LISS (Longitudinal Internet Studies for the Social sciences) panel administered by CentERdata (Tilburg University, The Netherlands).

Un ringraziamento particolare al Professor Paccagnella per il tempo, le spiegazioni e le tante correzioni di stile.

Grazie ai miei genitori per le opportunità che ho ricevuto.

Un grazie a tutti i miei amici che mi hanno sempre tirato su di morale, anche nei momenti più difficili.

In particolare volevo ringraziare Gege per essere la persona che mi capisce sempre al volo, senza bisogno di parlare.

Infine un grazie alla Marti che è sempre presente.