



UNIVERSITÀ DEGLI STUDI DI PADOVA
FACOLTÀ DI SCIENZE STATISTICHE

TESI DI LAUREA SPECIALISTICA

**TEST DI PERMUTAZIONE SU UNA CLASSE DI MODELLI
MISTURA: APPLICAZIONE AD UN CASO REALE**

**PERMUTATION INFERENCE ON A CLASS OF MIXTURE
MODELS: A REAL CASE APPLICATION**

Relatore: Ch.mo Prof. Fortunato Pesarin

Correlatore: Dott. Stefano Bonnini

Laureanda: Danièle Aurelie Djumboung Lontsi

ANNO ACCADEMICO 2009-2010

Ringraziamenti

Da dove iniziare ... molto probabilmente l'elenco si apre sul Prof. Pesarin; con i suoi insegnamenti ho imparato a affrontare e manipolare i modelli statistici complessi, grazie professore di avermi aperto nuovi orizzonti.

Un grazie, immenso, al Dott. Bonnini, per avermi seguito con molta attenzione nello svolgimento di questa tesi, grazie per il suo sostegno e i suoi consigli che mi hanno aiutato a raggiungere questo obiettivo.

Grazie al Prof Salmaso, al Dott. Corain, alla Dott.ssa Arboretti di aver creduto in me.

Tengo a esprimere tutta la mia profonda gratitudine ai miei genitori, mi hanno insegnato i valori dell'impegno, dell'indipendenza e dell'onestà e che hanno sempre creduto in me. Di sicuro, senza di loro non avrei raggiunto questa meta.

Grazie ai miei fratelli e alle mie sorelle per il loro sostegno lungo questi anni di studi. Un grazie particolare al mio fratellino Prudence Robinson; il suo successo mi ha incredibilmente motivato.

Un grazie infinite a tutte le "statistiche" che mi hanno supportato e sopportato, con le quali ho condiviso più da lontano che da vicino questi ultimi anni di intenso studio. In diversi modi hanno contribuito nel mio percorso formativo, aiutandomi a migliorare sempre di più il mio italiano, a credere in me stessa, suscitando in me nuovi interessi e suggerendomi il modo con cui superare alcuni passi della vita universitaria per raggiungere i miei sogni. Grazie mille ragazze!!!

Grazie alle amiche rimaste in Camerun, nonostante i chilometri di distanza tra di noi e gli anni di separazioni, sano al loro modo essere presente nella mia vita anche extra-universitaria. Per questo ringrazio infinitamente lo sviluppo delle tecnologie e in particolare internet e il web.

Last but not least... il mio tesoro Romeo! Lui ha il merito di avere supportato tutto più di tutti lungo questi anni. Tra i momenti di crisi e quelli relax, tra i periodi d'ansia e di gioia, e di sicuro il futuro ne promette altro ancora.

Sommario

<i>Capitolo 1. Presentazione generale</i>	7
<i>Capitolo 2. Teoria sui CUB</i>	9
2.1 Introduzione	9
2.2 Variabili latenti.....	10
2.3 La nuova classe di modelli	10
2.3.1 Modellazione del feeling: <i>shifted binomial</i>	10
2.3.2 Modellazione della seconda componente: l'incertezza.....	11
2.3.3 Il modello MUB.....	11
2.3.4 Il modello CUB: Covariate Uniform Binomial	12
2.3.5 Estensione del modello CUB.....	15
2.4 Inferenza sul modello	16
2.5 Software per l'inferenza sui modelli CUB.....	18
<i>Capitolo 3. Il modello CUB: test non parametrico sulle covariate</i>	21
3.1 Introduzione	21
3.2 I test di permutazione	21
3.2.1 I test di permutazione: principi e proprietà	22
3.2.2 Definizioni	23
3.3 Test di permutazione univariati.....	25
3.4 Tecniche di permutazione	26
3.4.1 Il metodo bootstrap	26
3.4.2 Il metodo Monte Carlo condizionale	26
3.5 Test di permutazione univariato per il modello CUB	27
3.5.1 Descrizione del test.....	27
3.5.2 La routine R	30
3.5.3 Osservazioni.....	33
3.6 Studio di simulazione	33
3.6.1 Introduzione	33
3.6.2 Studio della potenza.....	33
3.6.3 Risultati delle simulazioni.....	34
3.6.4 Codici R per la simulazione del CUB(0,1).....	38
<i>Capitolo 4. Applicazione del modello CUB ad un indagine di marketing sul vino passito</i> 40	
4.1 Introduzione	40
4.1.1 Passiti Veneti	41
4.1.2 Profilo dei consumatori di vino in Italia	41
4.2 Disegno dell'indagine	42

4.2.1	Modalità di svolgimento delle interviste e raccolta dei dati	42
4.2.2	Definizione del questionario	42
4.3	Analisi descrittiva dei dati.....	43
4.4	Analisi CUB versione parametrica.....	65
4.4.1	Modello saturo CUB(3,3)	66
4.4.2	Modello saturo CUB(0,0)	74
4.5	Analisi CUB versione non parametrica: test di permutazione	79
<i>Capitolo 5. Conclusioni</i>		81
Appendice A: The E-M algorithm		83
Appendice B: Questionario raccolta dati		87
<i>Indici delle figure e delle tabelle</i>		89
Bibliografia		93

Capitolo 1. Presentazione generale

Il lavoro sviluppato in questa tesi intitolata “test di permutazione su una classe di modelli mistura: applicazione ad un caso reale” si colloca all’intersezione tra due tematiche statistiche, tra l’altro l’introduzione di un nuovo modello parametrico (modelli CUB) e l’innovazione di un metodo non parametrico tramite i test di permutazione sullo stesso modello.

Quando le assunzioni sulla distribuzione di provenienza dei dati non sono rispettate, i risultati parametrici sono poco affidabili. Fisher (1935) e Pittman (1937) hanno proposto una nuova classe di test, i test di permutazione. I test di permutazione fanno riferimento a metodi per lo studio di un campione di dati senza ipotesi sulla loro distribuzione, permettono di mettere in rilievo delle caratteristiche sottostanti ai dati, le quali potranno essere tradotte in un modello per una analisi più semplice. Nonostante siano i test non parametrici più anziani, a causa della complessità dei calcoli richiesti, essi non sono molto diffusi. Con l’aumento della potenza dei calcolatori, questi test sono attualmente utilizzati in diverse situazioni e diversi ambiti di studio.

Il principio del test di permutazione è di lavorare a partire dai soli dati osservati, tale da non fare ipotesi sulla distribuzione della variabile misurata.

Tutti i test di permutazione derivano dallo stesso principio, quest’ultimo abbastanza intuitivo: *sotto l’ipotesi nulla, dati i valori del campione, le permutazioni delle osservazioni del campione hanno tutte la stessa probabilità di essere osservate*. È necessario che ci sia scambiabilità sotto l’ipotesi nulla (vedi capitolo 3). Quindi, un test di permutazione consiste in un paragone tra il valore osservato della statistica test con i valori generati permutando i dati. Le probabilità critiche sono calcolate considerando la legge condizionale rispetto al vettore delle statistiche osservate, e questa legge fornisce la distribuzione empirica rispetto alla quale può essere situata la statistica osservata del test. Per problemi statistici di analisi dei dati di sondaggio, l’obiettivo è la valutazione di un insieme di oggetti e le risposte sono variabili numeriche o ordinali, si parla del metodo di valutazione (Rating) e del metodo delle preferenze (Ranking). Nel nostro lavoro, proponiamo una soluzione di permutazione per l’inferenza nei modelli CUB, basata sulla permutazione dei vettori di covariate. Assumiamo di voler testare l’ipotesi nulla che una o più covariate siano inutili nella modellazione della variabile risposta: l’idea di base è che sotto tale ipotesi nulla vi è scambiabilità all’interno del vettore/i delle covariate di interesse, ovvero, non essendo tali covariate utili per spiegare il processo di decisione dei rispondenti, qualsiasi permutazione dei valori osservati nel campione per tali covariate è ugualmente verosimile.

Il presente lavoro di tesi si suddivide in sei capitoli. Il primo capitolo è introduttivo al lavoro; nel Capitolo 2, è riportata una descrizione teorica dei principali strumenti di statistica inferenziale utilizzati per l’analisi dei dati in esame, ossia: i modelli CUB (M.Iannario & D.Piccolo, 2004), che rappresentano un nuovo modello statistico per dati ordinali, e permettono di esplorare la relazione tra i soggetti e gli item, consentendo di sintetizzare la distribuzione delle preferenze (ranking) espresse dai rispondenti, tramite una scala qualitativa di tipo Likert; il suddetto modello è definito come:

$$\Pr(Y = y) = \underbrace{\pi \binom{m-1}{y-1} (1 - \xi)^{y-1} \xi^{m-y}}_{\text{feeling}} + \underbrace{(1 - \pi) \left[\frac{1}{m} \right]}_{\text{Incertezza}}, y = 1, \dots, m$$

Nel Capitolo 3, si introducono i test di permutazione, descrivendo nella prima parte i principi dei test di permutazioni, nella seconda i test di permutazione univariati, e infine il test di permutazione delle covariate del modello CUB di cui la statistica test di combinazione è dato da: ${}_bT^* = T[y, X_L, {}_bX_2^*, W_L, {}_bW_2^*]$ con $b = 1, \dots, B$, e il p-value

$$\hat{\lambda} = \sum_{b=1}^B \frac{I_{[T_{obs}, \infty)}({}_bT^*)}{B} \quad \text{dove } I_{[T_{obs}, \infty)}(t) = 1 \text{ se } t \geq T_{obs} \text{ e pari a 0 altrimenti}$$

Nel Capitolo 4, si trovano i risultati delle simulazioni effettuate per lo studio della potenza del test non parametrico del modello CUB con una covariata per ξ (CUB(0,1)) il Capitolo 5 è dedicato all'applicazione del modello CUB sia in versione parametrica che non. Il suddetto capitolo si apre con una breve descrizione delle principali tipologie di vini passiti presenti sul territorio italiano, dopo si forniscono i dettagli sulle modalità di svolgimento dell'indagine il quale è stato effettuato con modalità C.A.T.I., con i dati a disposizione viene seguita un'analisi descrittiva attraverso le statistiche univariate, riportando le principali tabelle e distribuzioni di frequenza, il capitolo si conclude con l'applicazione dei dati al modello CUB in ambito parametrico e non parametrico. Da quelle analisi risale che:

- i modelli CUB parametrici, rappresentano uno strumento particolarmente efficace nell'analizzare le preferenze dei consumatori registrate attraverso scale Likert. I modelli CUB inoltre, esplicitano in forma diretta il legame tra la probabilità di fare una scelta su una scala discreta (rating, valutazione, voto) e le covariate del soggetto;
- i modelli CUB non parametrici sono abbastanza potenti

Infine, il sesto e ultimo capitolo, fornisce alcune conclusioni generali.

Capitolo 2. Teoria sui CUB

2.1 Introduzione

Non si misura l'intelligenza chiedendo alle persone se si stimano intelligenti, non si misura la soddisfazione dei clienti chiedendo loro se in effetti sono soddisfatti. Nell'attuale mondo competitivo, la soddisfazione è la chiave del successo, bisogna quindi disporre di strumenti che permettono d'identificare obbiettivamente i bisogni dei clienti, di valutare la loro soddisfazione. Di fatto, in ambito industriale, sanitario, universitario e tanti altri, il barometro della soddisfazione è un questionario (ISO 2004), col quale viene chiesto a ciascun soggetto intervistato di fornire giudizi, preferenze, valutazioni. Potremmo dire che esiste un processo psicologico che il soggetto deve mettere in atto quando gli viene richiesto di dare un giudizio - tangibile o no - rispetto a ciascun dei tanti oggetti (*items: servizi, prodotti, ecc.*) in esame, utilizzando una scala ordinale. Partendo da giudizi di preferenza di carattere qualitativo, è possibile effettuare una comparazione sull'intera distribuzione delle valutazioni espresse (*rating*) in un contesto inferenziale, e collegare le valutazioni a caratteristiche dei soggetti. Questo tipo di procedura è stato ampiamente studiato nella teoria delle variabili latenti (Corduas e al 2009).

Esprimere una valutazione è l'atto finale di cause complesse, e le risposte sono affette da considerazioni reali del problema e da alcune incertezze le quali si collegano alle decisioni umane. Perciò, ogni percezione/valutazione diventa la realizzazione di un fenomeno stocastici.

Il processo decisionale del valutatore si manifesta con due componenti:

- La prima componente, riguarda il feeling o gradimento del rispondente, legato alla conoscenza e alla comprensione generale del problema, e all'esperienza personale.
- La seconda componente, generata dall'incertezza intrinseca, la quale ha effetto sulla scelta definitiva. Riteniamo che questo fatto sia dovuto ai seguenti fattori:
 - il tempo necessario per l'elaborazione della risposta,
 - la limitata gamma di informazioni disponibili,
 - la comprensione parziale della domanda,
 - la pigrizia del soggetto; Questo avviene quando il soggetto non è abbastanza interessato a riflettere per rispondere.

In questo capitolo, presenteremo nel paragrafo 1.2 il concetto di variabili latenti, nel paragrafo 1.3 si introdurranno i modelli CUB, nell'1.4 parleremo dell'inferenza sul modello introdotto, e nell'ultimo paragrafo, presenteremo il programma per l'applicazione a questo modello.

2.2 Variabili latenti

Molte variabili (*items/oggetto/servizi*) che sono di interesse per gli esperti di scienze sociali non sono direttamente osservabili (Magidson e Vermunt 2003). Alcuni esempi sono la *preferenza*, l'*attitudine*, le *azioni premeditate*, i *tratti di personalità*. La misura di questi oggetti però può avvenire indirettamente tramite indicatori osservabili. Diverse tecniche di ordinamento sono state sviluppate per le osservazioni derivanti dalle variabili non osservate di interesse. Finché la soddisfazione e la percezione del servizio non sono osservabili, si definisce il loro ruolo in un gruppo di modelli detti di classe latenti. Le valutazioni fornite dal soggetto possono essere interpretati come un miscuglio di *feeling* e di *incertezza*: esse sono componenti latenti, continue, non osservabili e collegabili con le caratteristiche individuali dei soggetti.

Alcune ricerche statistiche hanno come obiettivo la modellazione empirica delle osservazioni campionarie per studiare il meccanismo stocastico di generazione dei dati ordinali. Nell'ambito dell'analisi delle preferenze, il primo strumento statistico utilizzato è l'ANOVA. L'analisi della varianza - metodo robusto se sono rispettate le assunzioni di normalità e homoschedasticità- è utilizzata per studiare gli effetti di ciascuna caratteristica di ogni preferenza. D'altra parte, possiamo citare come tecniche di analisi le componenti principale (PCa), i minimi quadrati parziali (Partial Least Square), scale di preferenze multidimensionali (MultiDimensional scaling PREFERence) (Piccolo D 2007).

2.3 La nuova classe di modelli

La scelta della risposta finale del soggetto intervistato è la combinazione tra ciò che egli sente, il gradimento verso l'oggetto, - *feeling* - e una *incertezza* intrinseca. Il *feeling* dipende dalle motivazioni del soggetto e quindi è la componente principale della scelta, mentre l'*incertezza* dipende dalla circostanza in cui viene effettuata l'intervista ed è la componente legata alla fiducia della scelta. Il *feeling* è una variabile casuale (nel seguito v.c.) continua, resa discreta dal vincolo che la scelta (il giudizio, le preferenze, la valutazione) deve essere espressa mediante m valori ordinali prefissati (Marshack 1960).

2.3.1 Modellazione del feeling: *shifted binomial*

Quando il soggetto sceglie una tra le m valutazioni possibili, si può pensare ad una esclusione delle altre tramite confronti a coppie (D'Elia 2000). In effetti, se chiamiamo y la realizzazione che indica la scelta del soggetto, cioè la valutazione assegnata all'oggetto con $y \in \{1, \dots, m\}$, ciò vuole dire che y è preferibile alle $y - 1$ valutazioni precedenti secondo la scala usata, e non è preferibile rispetto alle restanti $m - y$ valutazioni. Supponiamo $(1 - \xi)$ la probabilità che una data valutazione sia ritenuta inferiore a quella di cui l'oggetto è meritevole (successo) e ξ la probabilità che una data valutazione sia invece inferiore (insuccesso). L'insieme degli $m - 1$ possibili confronti porta alla sequenza *insuccesso/successo* che ha la probabilità di $(1 - \xi)^{y-1} \xi^{m-y}$. Gli argomenti combinatori dimostrano che la probabilità della scelta finale è la legge binomiale traslata (*shifted binomial distribution*). La probabilità di scelta finale ottenuta dagli argomenti combinatori è la seguente:

$$\binom{m-1}{y-1} (1-\xi)^{y-1} \xi^{m-y} \text{ per } y = 1, \dots, m \quad (2.1)$$

Si assume che le variabili casuali che descrivono il confronto siano indipendenti e identicamente distribuite. Riteniamo quindi che la prima componente la quale descrive il feeling, può essere simulata tramite la distribuzione appena descritta.

2.3.2 Modellazione della seconda componente: l'incertezza

Nella teoria delle probabilità e in statistica, la legge uniforme è una famiglia di leggi di probabilità tale che tutti gli intervalli di stessa lunghezza del supporto hanno la stessa probabilità. Nel caso discreto, la legge uniforme assegna ad ognuno degli m valori possibili la stessa probabilità. Nel nostro caso, possiamo dire che l'incertezza è data da una variabile aleatoria uniforme discreta con supporto $\{1, 2, \dots, m\}$ e funzione di probabilità pari a: $P_y(Y) = \frac{1}{m}$. La distribuzione della v.c. **Uniforme discreta** è un elemento essenziale per modellare l'incertezza presente nelle m alternative (con m finito).

Siccome le scelte di ogni valutatore presentano un certo grado di *feeling* e *incertezza*, esse possono essere rappresentate mediante una distribuzione mista. I pesi della distribuzione mista sono π e $(1 - \pi)$ rispettivamente per il feeling e l'incertezza (Corduas e al 2009) in questo modo sono appena stati messi in evidenza le misure delle componenti latenti.

2.3.3 Il modello MUB

Solitamente, per i dati ordinali vengono utilizzate funzioni delle probabilità cumulate di scelta, di recente D'Elia e Piccolo (2005) hanno sviluppato e proposto un modello esplicito per le probabilità di scelta, chiamato **Mixture Uniform Binomial (MUB)**. L'idea del modello MUB parte dalla considerazione che il *feeling* e la componente di *incertezza* sono diverse variabili casuali ponderate che possono essere combinate in un *modello di mistura*.

Il grado di piacevolezza dell'oggetto/prodotto in studio può essere considerato come una distribuzione gaussiana, ma il rating ($y=1, \dots, m$) assegnato sarà analizzato come una v.c. discreta. Per un prefissato m , se :

$$\Pr(Y = y) = \pi \left[\binom{m-1}{y-1} (1-\xi)^{y-1} \xi^{m-y} \right] + (1-\pi) \left[\frac{1}{m} \right] \text{ per } y = 1, \dots, m \quad (2.2)$$

gli autori definiscono $Y \sim \text{MUB}(\pi, \xi)$ dove $\xi \in [0, 1]$ e $\pi \in [0, 1]$ (D'Elia e Piccolo 2005) con spazio parametrico quadrato unitario. Al variare dei parametri π e ξ , cambiano le forme della distribuzione della v.c., ne esce quindi che il MUB è abbastanza flessibile. Alcune sue caratteristiche rilevanti sono le seguenti:

- Ammette una moda $M_o = 1 + [m(1 - \xi)]$ in qualsiasi valore sul supporto $\{1, \dots, m\}$ la quale coincide con la moda della Binomiale traslata;

- Le code hanno pesi funzioni di $\frac{(1-\pi)}{m}$;
- Se il parametro ξ è pari a $\frac{1}{2}$, la distribuzione ottenuta è simmetrica.

Il modello appena descritto si adatta bene a molte situazioni su dati ordinali (graduatorie, preferenze, percezioni, misure di concordanza, ecc.). Per ottenere migliori risultati, è possibile introdurre le caratteristiche personali o covariate del soggetto (età, sesso, situazione familiare, regione di appartenenza, ecc.) in modo da collegarli con la scelta espressa. Queste caratteristiche permettono di rendere le previsioni più accurate, di discriminare le differenti sotto popolazioni. L'introduzione del legame tra covariate del soggetto e i parametri porta alla definizione di un altro modello: il modello CUB (**Covariate Uniform Binomial**) il quale verrà discusso nel successivo paragrafo.

2.3.4 Il modello CUB: Covariate Uniform Binomial

La corrispondenza che si crea tra i dati e la loro struttura probabilistica molto malleabile, è stata generata da un approfondimento del meccanismo psicologico che genera la scelta tra m alternative. Ne è uscita una serie convergente di proposte, attualmente confluite in una classe molto ampia di modelli statistici, definiti dagli autori **CUB**. Risulta possibile studiare i legami, somiglianze, connessioni nonché stabilità rispetto a diversi gruppi di dati, periodi temporali, ambiti territoriali, circostanze diverse, e così via.

Per tenere conto della natura composita del processo di scelta, è stata introdotta una v.c., Y , nella quale le due componenti di *feeling* ed *incertezza* sono adeguatamente ponderate. Indichiamo con y la realizzazione della v.c. discreta Y definita come la mistura di una v.c. Uniforme e di una v.c. Binomiale traslata. Tale v.c. sarà chiamata CUB e l'acronimo deriva dal fatto che introdurremo **Covariate** nella mistura di variabili casuali **U**niforme e **B**inomiale traslata.

Formalmente, per un $m > 3$ fisso e noto, la sua distribuzione di probabilità è:

$$\Pr(Y = y) = \underbrace{\pi \left[\binom{m-1}{y-1} (1-\xi)^{y-1} \xi^{m-y} \right]}_{\text{feeling}} + \underbrace{(1-\pi) \left[\frac{1}{m} \right]}_{\text{Incertezza}} \cdot y = 1, \dots, m \quad (2.3)$$

dove $\xi \in [0,1]$, $\pi \in (0,1]$ (D'Elia e Piccolo 2005) e $m > 3$ è il valore più elevato della scala Likert utilizzata dai soggetti. Lo spazio parametrico è :

$$\Omega(\pi, \xi) = \{(\pi, \xi) : 0 < \pi \leq 1; 0 \leq \xi \leq 1\}.$$

Viene imposto un valore m maggiore di 3, per evitare di avere una variabile *random* degenerare ($m=1$), un modello indeterminato ($m=2$), o un modello saturo ($m=3$).

La prima componente dell'equazione precedente, con peso π , è denominata *feeling* ed indica la convinzione profonda del soggetto che esprime il giudizio e la forza di attrazione/repulsione che egli prova verso l'oggetto della valutazione.

Il parametro ξ è connesso al *grado di feeling* espresso dal rispondente:

- ✓ Se il *rating* verso l'item esprime una valutazione diretta, allora il gradimento cresce con $1 - \xi$.
- ✓ Se il *ranking* dell'oggetto pone al primo posto l'oggetto preferito, allora il gradimento cresce con ξ .

La seconda componente dell'equazione, con peso $1-\pi$, è definita **incertezza**; è l'incertezza che ciascun soggetto apporta alle scelte laddove deve tradurre una propria convinzione in un punteggio.

È la componente personale intrinsecamente correlata ad ogni scelta umana ed è il risultato di alcuni fattori convergenti:

- *Conoscenza/Ignoranza* dei problemi e/o delle caratteristiche degli oggetti
- *Interesse/Coinvolgimento personale* in attività, oggetti, opinioni, etc.
- *Tempo impiegato* per giungere alla valutazione espressa
- *Pigrizia/Apatia* verso il meccanismo di valutazione.

Il parametro π è inversamente connesso al **grado di incertezza** espresso dal rispondente perché la sua incertezza cresce con $1 - \pi$, ovvero:

- ✓ Se $\pi \rightarrow 0$, il rispondente ha una propensione sempre maggiore verso una scelta estrema di totale incertezza.
- ✓ Se $\pi \rightarrow 1$, il rispondente esprime la sua scelta sulla base di una incertezza molto ridotta ma con propensione elevata dettata esclusivamente dal feeling verso l'item/oggetto.

Questa interpretazione consente di rappresentare ogni modello come un punto nello **spazio parametrico** (quadrato unitario) confrontando i modelli nel tempo, nello spazio, per sottogruppi, al variare delle circostanze esterne, e così via.

Se si collocano i parametri stimati di un modello CUB nel corrispondente spazio parametrico, si sintetizzano gli aspetti salienti contenuti nei dati a disposizione; soprattutto, si possono confrontare più modelli in modo agevole ed efficace, poiché:

1. $1 - \pi$ quantifica la *propensione* dei rispondenti a fornire risposte totalmente casuali, più π è spostato a destra del quadrato unitario e più le valutazioni/preferenze espresse riflettono risposte ben definite (l'incertezza è bassa). Supponendo l'equiprobabilità, nelle applicazioni, si utilizzerà $\frac{1-\pi}{m}$
2. In modo analogo, $1 - \xi$ misura la forza del *feeling* espresso dal soggetto a favore di una diretta e positiva valutazione del soggetto, più vicino è ξ al bordo superiore del quadrato unitario e meno l'item è preferito.

In tal modo, l'effetto di covariate mostra come si modifica il comportamento dei rispondenti al variare delle caratteristiche del soggetto.

Disponendo di covariate del soggetto, si definisce un modello CUB in presenza o in assenza delle stesse.

Per un prefissato $m > 3$, la logica in questo caso è simile a quella dei modelli lineari generalizzati (King G et al 2000) e il modello CUB è definito con:

- Una **componente stocastica**: Essa spiega la variabile risposta Y_i come il risultato di un esperimento casuale

$$Pr(Y_i = y | x_i; w_i) = \binom{m-1}{y-1} \xi_i^{m-y} (1 - \xi_i)^{y-1} + (1 - \pi_i) \binom{1}{m}, \quad (2.4)$$

definita per ogni i -esimo soggetto $i = 1, 2, \dots, n$.

- Due **componenti sistematiche** sono funzioni logistiche e sono definite da:

$$\begin{cases} \pi_i = \frac{1}{1+e^{-x_i\beta}} \\ \xi_i = \frac{1}{1+e^{-w_i\gamma}} \end{cases}, \quad (2.5)$$

dove x_i e w_i sono le covariate del soggetto i -esimo, introdotte per migliorare l'interpretazione di π_i e ξ_i , rispettivamente.

Nei capitoli successivi del presente lavoro, la funzione legame di riferimento sarà la funzione logistica.

La media e la varianza di Y sono date dalle relazioni seguenti:

$$E(Y) = \pi (m-1) \left(\frac{1}{2} - \xi \right) + \frac{(m+1)}{2} \quad (2.6)$$

$$Var(Y) = (m-1) \left\{ \pi \xi (1 - \xi) + (1 - \pi) \left[\frac{(m+1)}{12} + \pi (m-1) \left(\frac{1}{2} - \xi \right)^2 \right] \right\} \quad (2.7)$$

Nella Tabella 1 vengono riassunte le notazioni che utilizzeremo nel seguito per fare riferimento ai modelli CUB:

Tabella 1. Notazione CUB(p,q)

Modelli	Covariate	Parametri	Spazio parametrico	Numero di parametri
CUB (0,0)	<i>senza covariate</i>	$\theta = (\pi, \xi)'$	$(0,1] \times (0,1]$	2
CUB (p,0):	<i>con p covariate per π</i>	$\theta = (\beta', \xi)'$	$\mathbb{R}^{p+1} \times (0,1]$	p+2
CUB (0,q)	<i>con q covariate per ξ</i>	$\theta = (\pi, \gamma)'$	$(0,1] \times \mathbb{R}^{q+1}$	q+2
CUB (p,q)	<i>con p covariate per π e q covariate per ξ</i>	$\theta = (\beta', \gamma)'$	\mathbb{R}^{p+q+2}	p+q+2

Evidentemente, il modello CUB(0,0) non è null'altro che il modello MUB.

Attualmente non esiste ancora in letteratura un metodo ufficialmente riconosciuto per la scelta delle covariate. In generale, nella prassi si verifica se una data covariata influisce su *feeling* e *incertezza*, separatamente per i due parametri, e in caso affermativo si inserisce la covariata nel modello. Seguendo una procedura stepwise vengono eseguiti sequenzialmente test di significatività su diverse covariate, considerate una alla volta, che vengono inserite nel modello in caso di significatività del test.

Generalmente, nella pianificazione del questionario, si chiedono covariate che rappresentano caratteristiche personali, demografiche, ambientali, sociali, culturali, etc. che, potrebbero spiegare le risposte ordinali.

2.3.5 Estensione del modello CUB

Oltre al feeling e all'incertezza un altro aspetto che potrebbe contribuire a descrivere la scelta finale del rispondente, è la "pigrizia" dei soggetti e/o il loro grado di interesse/coinvolgimento nei confronti dell'indagine. Molti valutatori infatti spesso hanno propensione a scegliere per tutte o quasi tutte le domande, una categoria standard/semplce tra le m possibilità questo è la scelta "*shelter* (Iannario e Piccolo 2009). Si aggiunge una terza componente per spiegare questo comportamento (*shelter effect*), e si ottiene il modello CUB *esteso* definito per $y = 1, \dots, m$ e $m > 4$ da:

$$\Pr(Y = y|\theta) = \pi_1 \left[\binom{m-1}{y-1} (1-\xi)^{y-1} \xi^{m-y} \right] + \pi_2 \left[\frac{1}{m} \right] + (1 - \pi_1 - \pi_2) D_y^{(c)} \quad (2.9)$$

$\theta = (\pi_1, \pi_2, \xi)'$ parametro esteso, $D_y^{(c)}$ v.c. degenerare con massa di probabilità concentrata a $y = c$:

$$D_y^{(c)} = \begin{cases} 1, & \text{se } y = c; \\ 0, & \text{altrimenti.} \end{cases}$$

Osservazioni:

- a- Se $\pi_1 + \pi_2 = 1$, il modello CUB esteso coincide con il CUB standard;
- b- Il modello esteso è capace di prevedere distribuzioni dove la maggiore parte dei rispondenti hanno scelta concentrata nella categoria intermedia, in queste situazione (rarissime), $\pi_1 = \pi_2 = 0$.

2.4 Inferenza sul modello

Dato un vettore di frequenze osservate $n = (n_1, \dots, n_m)$ per le m modalità ordinali della variabile risposta Y , la frequenza assoluta di $Y = y$ è indicata con n_y ($y = 1, \dots, m$). Posto la distribuzione di probabilità:

$$P_y(\pi, \xi) = P_y(Y = y_i | \beta, \gamma) = \frac{1}{1 + e^{-x_i \beta}} \left[\frac{(m-1)}{(y_i - 1)} \frac{(e^{-w_i \gamma})^{y_i - 1}}{(1 + e^{-w_i \gamma})^{m-1}} - \frac{1}{m} \right] + \frac{1}{m}$$

dove y_i è la risposta dell' i -esimo soggetto, ipotizzando indipendenza delle risposte dei soggetti, la funzione di *log-verosimiglianza* per il modello è pari a:

$$\log L(\theta) = \sum_{i=1}^n \log \left[\frac{1}{1 + e^{-x_i \beta}} \left\{ \left[\frac{(m-1)}{(y_i - 1)} \frac{(e^{-w_i \gamma})^{y_i - 1}}{(1 + e^{-w_i \gamma})^{m-1}} - \frac{1}{m} \right] + \frac{1}{m} \right\} \right]. \quad (2.10)$$

Per ottenere le diverse stime di massima verosimiglianza di π e ξ , si usa l'algoritmo **E-M Expectation Maximization** (D'Elia e Piccolo 2005) il quale assume che la risposta del soggetto segua una distribuzione Binomiale traslata o da un distribuzione Uniforme discreta con probabilità π e $(1 - \pi)$ rispettivamente. Non sapendo da quale componente è stata generata la valutazione fornita, per le stime dei parametri consideriamo i dati come *incompleti*. Vengono introdotte le variabili latenti tale che:

$$\left\{ \begin{array}{ll} Y_{gi} = 1 & \text{Se preferenza proviene dalla distribuzione } g \\ Y_{gi} = 0 & \text{Altrimenti} \end{array} \right.$$

Gli E-M convergono molto lentamente, ma la funzione di *log-verosimiglianza* cresce ad ogni iterazione e assicura la convergenza finale verso il massimo (Piccolo 2006). Per controllare la significatività delle stime ottenute, per costruire regioni di confidenza e per validare il modello stimato, occorre riferirsi alla teoria asintotica degli stimatori di massima verosimiglianza. Nonostante la complessità inferenziale sui vettori di parametri β e γ , l'algoritmo E-M fornisce una soluzione efficace sul piano numerico. La matrice di

varianza e covarianza $V(\theta)$ della stima di massima verosimiglianza $\hat{\theta}$ del parametro θ del modello CUB è basata sull'informazione attesa $i(\theta)$, inversa della matrice Hessiana quando $\theta = \hat{\theta}$ (si veda Pawitan 2001).

La matrice di varianza-covarianza delle stime di θ quando $\theta = \hat{\theta} = (\hat{\pi}, \hat{\xi})'$ è pari a:

$$V(\theta) = [i(\hat{\theta})]^{-1} = - \left(\begin{array}{cc} \frac{\partial^2 l(\theta)}{\partial \pi^2} & \frac{\partial^2 l(\theta)}{\partial \pi \partial \xi} \\ \frac{\partial^2 l(\theta)}{\partial \pi \partial \xi} & \frac{\partial^2 l(\theta)}{\partial \xi^2} \end{array} \right)_{\theta=\hat{\theta}}^{-1} \quad (2.11)$$

Per l'implementazione di questo risultato è disponibile un codice R, il quale serve anche per le stime e l'inferenza dei modelli con e senza covariate, ne discutiamo nel paragrafo successivo. Nei modelli CUB è importante stabilire la rilevanza statistica delle covariate nella spiegazione dei dati. A tal fine, è opportuno confrontare la log-verosimiglianza del modello CUB (0,0) stimato (senza covariate, quindi) con le log-verosimiglianze dei modelli CUB(p,0), CUB(0,q) e CUB(p,q), rispettivamente. Al contrario dei MUB dove la misura utilizzata per effettuare questo confronto è la devianza, per i CUB si utilizza l'*indice di dissimilarità*, questa misura va poi confrontata con il test di χ^2_g .

Da un punto di vista descrittivo, consideriamo l'indice di dissimilarità normalizzato $Diss \in [0,1]$ definita da:

$$Diss = \frac{1}{2} \sum_{y=1}^m \left| \Pr(Y = y | \hat{\theta}) - \frac{n_y}{n} \right|. \quad (2.12)$$

Dove n_y è il vettore delle frequenze assolute, e n quello delle frequenze osservate.

Se la dissimilarità ha un livello superiore a 0.10, il modello non è accettabile. Questa soglia è stata stabilita sulla base di una estesa esperienza (l'applicazione attuale è limitata ai modelli senza covariate).

Tabella 2. confronto tra modelli complessi e semplici e i gradi di libertà associati

Confronto	Test log rapporto di verosimiglianza	Gradi di libertà
CUB(p,0) vs CUB(0,0)	$2(l_{10} - l_{00})$	p
CUB(0,q) vs CUB(0,0)	$2(l_{01} - l_{00})$	q
CUB(p,q) vs CUB(0,0)	$2(l_{11} - l_{00})$	$p+q$

La log - verosimiglianza del modello *saturo* è data da:

$$l_{sat} = -n \log(n) + \sum_{y=1}^m n_y \log(n_y)$$

Una misura della bontà del modello si ottiene definendo uno pseudo R^2 , chiamato ICON (=Information CONtent), che confronta la log – verosimiglianza del modello stimato con quella della v.c. Uniforme discreta. (Iannario e Piccolo, 2008). Esso è definito come:

$$ICON = 1 + \frac{l(\hat{\theta})/n}{\log(m)}$$

Visto quali sono le misure importanti nel modello CUB, è ora di vedere come fare girare l’algoritmo E-M per ottenere le stime sopra descritte. Nel prossimo paragrafo parleremo del programma sviluppato dagli autori con il software R.

2.5 Software per l’inferenza sui modelli CUB

Oltre alle misure menzionate nel paragrafo 1.5, gli autori introducono il BIC (Bayesian Information Criterion, (Schwarz 1978)) il quale è un’approssimazione del calcolo della verosimiglianza dei dati condizionalmente al modello fissato ed è definito come:

$$-2 * \ln(\hat{\theta}) + k * \ln(n) \text{ dove } k \text{ è la dimensione del parametro}$$

e nella versione 2.0 del software sviluppato per fare inferenza sui CUB, si propone un’altra misura testata efficace D^2 .

$$D^2 = \frac{1}{m} \sum_{y=1}^m \left(\frac{\frac{n_y}{n}}{\Pr(Y = y|\hat{\theta})} \right)^2 - 1.$$

Tutte le misure descritte tranne ICON, devono essere minimizzate per avere una buona stima.

Il codice R per implementare il CUB ha la logica delle procedure GAUSS scritte da Piccolo tra 2001 e 2005, alcune convenzioni sono state importate in R. Nell’attuale versione, ci sono delle *micro funzioni* incluse nelle *macro*. Per attivare il programma, basta eseguire la funzione **CUB()** la quale contiene altre funzioni specializzate relative ai diversi modelli CUB(p,q). Le istruzioni per ottenere stime dopo avere importato il data set e fissato il numero di categorie sono qui sotto riportate:

```
source("CUB.R")
CUB(ordinal, X = paicov, W= csicov, shelter = c)
```

Il default CUB(ordinal) corrisponde al CUB senza covariate.

ordinale è matrice di dati da modellare; *paicov* e *csicov* sono le covariate possibili e *c* è la categoria corrispondente all’effetto *shelter*.

Tabella 3. Istruzioni per le stime dei modelli CUB.

<i>Modello</i>	Covariate	Istruzione
CUB (0, 0)		CUB(ordinal)
CUB (p, 0)	paicov	CUB(ordinal, X = paicov)
CUB (0, q)	csicov	CUB(ordinal, W = csicov)
CUB (p, q)	paicov, csicov	CUB(ordinal, X = paicov, W = csicov)
CUB +shelter		CUB(ordinal, shelter=c)

L'output ottenuto varia in funzione della presenza o dell'assenza delle covariate e dalla loro natura dicotomica o no.

Tabella 4. Variabili ottenute dopo l'esecuzione del CUB

<i>Modello</i>	Variabili
CUB (0, 0)	pai, csi, varmat, loglik, diss, n, AICCUB00, BICCUB00
CUB (p, 0)	beta, csi, varmat, loglik, diss, n, AICCUBp0, BICCUBp0
CUB (0, q)	pai, gamma, varmat, loglik, diss, n, AICCUB0q, BICCUB0p
CUB (p, q)	beta, gamma, varmat, loglik, diss, n, AICCUBpq, BICCUBpq
CUB +shelter	pai1, pai2, csi, varmat, loglik, diss

In più delle usuali statistiche (stime, standard errors, p.values) il programma propone alcune informazioni grafiche:

- ✓ CUB (0, 0): il grafico ottenuto è il plot delle frequenze osservate.
- ✓ CUB (1, 0): se e solo se la singola covariata per π è dicotomica e la codifica delle Dummy sia con la parametrizzazione (0, 1), il plot delle stime dei due sottogruppi è automaticamente generato.
- ✓ CUB (1, 0): se e solo se la singola covariata per ξ è dicotomica e la codifica delle Dummy sia con la parametrizzazione (0, 1), il plot delle stime dei due sottogruppi è automaticamente generato.
- ✓ CUB (p, q): il disegno del grafico non avviene in maniera automatica.
- ✓ CUB + shelter: il grafico delle frequenze osservate della distribuzione di dati ordinali e le relative stime di probabilità sono visibili.

Qualora venissero utilizzate più covariate al posto della singola covariata, è importante tener presente che:

- Il comando CUB deve contenere la matrice con le covariate corrispondente (la colonna di 1 per la costante è immessa automaticamente nel programma). Quindi, per stimare sulla variabile ordinale y_i un modello CUB (p,q) con $cov1, \dots, covp$ covariate, per il parametro di incertezza (=pi) e $cov1, \dots, covq$ covariate, per il parametro di feeling (=csi), l'istruzione è la seguente:
- Se si hanno variabili politomiche, occorrono $k-1=2$ variabili dicotomiche, come è standard nei modelli di regressione.

```
with (data= dati, CUB(yi, cbind(cov1,..., covp), cbind(cov1,..., covq)))
```

Per ridefinire i grafici plot e le simulazioni effettive, è possibile sfruttare alcune funzioni *ad hoc* per approfondire i risultati inferenziali, un caso è la possibilità di avere i contours e il profilo della funzione di log – verosimiglianza per entrambi i parametri del modello senza covariate.

```
par(mfrow=c(2,1)); # set 2x1 area del grafico  
  
par(mar=c(4,4,3,1)+0.1); # definire le nuove margini  
  
elleplot(pai,csi,loglik,ordinal); # contours plot  
  
proflik(pai,csi,ordinal); # funzione profilo
```

Un'ultima considerazione è l'esecuzione delle simulazioni, si genera *nsimul* numeri pseudo-casuali dal modello CUB, dando in input *m*, π e ξ . La sua funzione *simcub()* è chiamata dopo avere caricato il CUB.

```
source("CUB.R")
```

Capitolo 3. Il modello CUB: test non parametrico sulle covariate

3.1 Introduzione

L'inferenza sui modelli CUB è stata sviluppata nell'ambito parametrico. In effetti, uno dei problemi inferenziale molto interessanti e più usuale riguarda il test d'ipotesi sulle covariate, in particolare modo, quando l'ipotesi nulla riguarda la significatività di uno o di più coefficienti del modello.

La proposta innovativa che intendiamo presentare in questo capitolo cioè in ambito non parametrico propone test sulle covariate di incertezza e feeling senza rincorrere alla teoria asintotica, utilizzando la stima di massima verosimiglianza del modello CUB, ma sfruttando la distribuzione di permutazione delle statistiche test. Nel primo paragrafo, faremo una breve descrizione di quello che intendiamo per test di permutazione dando i loro principi e le loro proprietà nel paragrafo successivo si introdurranno i test di permutazione univariati, nell'ultimo paragrafo parleremo del test di permutazione univariato per il modello CUB.

3.2 I test di permutazione

In statistica, particolarmente quella inferenziale, gli studiosi sono interessati principalmente al seguente problema: osservando un particolare fenomeno appartenente a un insieme di risultati possibili, quale legge di probabilità e quale famiglia possiamo assegnare a quest'insieme? Si suppone che la legge di probabilità appartenga a una famiglia parametrica di probabilità, cioè è possibile scriverla come: $\{F_\theta, \theta \in \Theta\}$.

il modello statistico è definito con la terna:

- E insieme dei possibili risultati dell'esperimento,
- \mathcal{T} famiglia associata a E ,
- \mathcal{P} famiglia delle leggi di probabilità.

Nell'ambito non parametrico, i problemi sono analoghi però in questo caso, non si suppone che F sia della forma F_θ , e i problemi d'interesse sono principalmente:

- Stimare il funzionale statistico di F ;

Definizione del funzionale statistico: sia F la legge di probabilità di cui il campione indipendente e identicamente distribuito (X_1, X_2, \dots, X_n) è una realizzazione. Un funzionale statistico è una funzione T definita per la legge F e può quindi indicarsi con $T(F)$. Se F è sconosciuta, lo stimatore naturale est $T(F_n)$, con F_n funzione di ripartizione empirica:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}} = \frac{\#\{X_i \leq x\}}{n} \quad e \quad dF_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}(x).$$

- Decidere dell'ipotesi nulla (H_0) e della sua alternativa (H_1) al riguardo di F ,

- Stimare le funzioni legate a F come ad esempio una densità o una funzione di regressione.

Dei problemi appena elencati, consideriamo il secondo, cioè decidere se accettare l'ipotesi nulla o rifiutarla a favore della sua alternativa. Nel seguito parleremo di test statistici.

I test statistici forniscono una regola che permette di decidere se rifiutare o no un'ipotesi in funzione delle osservazioni campionarie. Quando l'ipotesi non è basata sulla conoscenza della distribuzione dei dati e non sono verificate le assunzioni, si utilizzano test non parametrici. Si distinguono i test d'indipendenza (studio della dipendenza che potrebbe esistere in serie di dati come il test sulla correlazione dei ranghi di Spearman) e quelli di omogeneità (confronto tra i parametri di popolazioni appaiate o no per misurare la differenza tra due gruppi come il test di Wilcoxon, test U di Mann-Whitney, e altri, o tra più gruppi come il test di Kruskal-Wallis).

Nelle applicazioni reali, in aree come l'agricoltura, la biologia, le prove cliniche, ingegneria, farmacologia, ecc., ci sono molti problemi multivariati complessi, difficili da risolvere con metodi standard.

L'uso dei metodi di permutazione permette di ottenere sia nei campioni non bilanciati che bilanciati inferenze sui parametri, le quali sono: esatte, non distorte, e incorrelate (caso dell'ANOVA a due vie). È anche possibile fare la differenza tra l'analisi dei dati mancanti completamente a caso (Missing Completely At Random), e quella dei dati mancanti non completamente a caso (Not Missing Completely At Random). I test di permutazione sono tra i test più potenti quando le assunzioni parametriche non sono valide (Good 1994). Ciò nonostante, nei casi complessi, non è sempre ovvio come strutturare la permutazione dei dati per effettuare questi test. Ad esempio, la costruzione di un test di permutazione per le sole covariate di un intero data set.

3.2.1 I test di permutazione: principi e proprietà

Gli approcci parametrici sono molto incerti circa il piano dell'esperimento (ad esempio il campionamento casuale) e il modello della popolazione (ad esempio la distribuzione normale o l'omoschedasticità). Quando non vengono rispettate queste assunzioni, in particolare quando la legge dei dati non si conforma alle esigenze del test, i risultati parametrici sono meno affidabili. I test non parametrici offrono un'alternativa molto importante poiché necessitano di meno assunzioni. Una classe molto importante di tali test sono i *test di permutazione*.

Le opere di Lehmann (1986) e van der (1998), tra gli altri, contengono parti dedicate alle proprietà teoriche dei test di permutazione per le osservazioni univariate. Alcune di queste proprietà sono state approfondite da Strasser e Weber (1999) a osservazioni multivariate. Quando il problema è complesso, in ambito non parametrico, non è necessario specificare o modellare la struttura della relazione di dipendenza per le variabili di interesse.

Il principio dei test di permutazione è il seguente:

Se due esperimenti prendono valori nello stesso spazio campionario, \mathcal{X}^n sotto le distribuzioni rispettive P_1 e P_2 entrambe membri di \mathcal{P} (distribuzione ignota della

popolazione), dato lo stesso dataset \mathbf{x} , allora l'inferenza condizionata a \mathbf{x} , ottenuta utilizzando la stessa statistica test sarà la stessa, supponendo che la scambiabilità per ciascun gruppo sia soddisfatta nell'ipotesi nulla.

L'importanza dei test di permutazione risiede nella loro robustezza e la loro flessibilità. L'idea utilizzando questi metodi è di generare una distribuzione di riferimento a partire dai dati, e ricalcolare la statistica test per ciascuna permutazione dei dati con riferimento a la legge discreta che ne risulta. Alcune proprietà generali dei test di permutazione sono:

- Quando sotto ipotesi nulla la scambiabilità non è soddisfatta, i test di permutazioni diventano una procedura inappropriata e erronea.
- Le regioni di rifiuto –non randomizzate- condizionate sono simili (α -invarianti nel senso di Scheffé (Scheffé 1943)).
- I test di permutazione sono distribution free poiché la statistica test segue una legge che non dipende dalla distribuzione \mathcal{P} .

3.2.2 Definizioni

Definizione 1-

Sia Π l'insieme di permutazioni dei naturali $(1, \dots, n)$. Un insieme di variabile aleatorie X_1, \dots, X_n è scambiabile se per ogni permutazione $\pi = (\pi_1, \dots, \pi_n) \in \Pi$, si ha :

$(X_1, \dots, X_n) \stackrel{d}{=} (X_{\pi_1}, \dots, X_{\pi_n})$, dove $\stackrel{d}{=}$ indica che i due vettori hanno la stessa funzione di ripartizione (l'uguaglianza in legge).

Definizione 2-

Si suppone il campione casuale $x = (x_1, x_2, \dots, x_n)$, realizzazione di una variabile casuale $X \in \mathcal{X}$ di cui le componenti non necessariamente sono indipendenti e identicamente distribuite (i.i.d.). In generale, per ogni problema di verifica d'ipotesi, H_0 assume che x derivi da un'unica popolazione con distribuzione sconosciuta \mathcal{P} . La famiglia \mathcal{P} delle distribuzioni si comporta in maniera non parametrica quando non siamo in grado di trovare il parametro $\theta \in \Theta$ con Θ finito-dimensionale tale che ci sia una relazione tra \mathcal{P} e Θ . La densità congiunta del vettore x , indicata con $f_p^{(n)}(x)$ è invariante rispetto alle permutazioni degli elementi se :

$$f(x) = f(x_1, \dots, x_n) = f(x_{\pi_1}, \dots, x_{\pi_n}).$$

Definizione 3-

Se l'assunto di indipendenza è sostituita dalla scambiabilità (nel contesto di permutazione, si riferisce alla scambiabilità dei dati osservati tra i gruppi) allora:

$$f_p^{(n)} = \prod_{1 \leq i \leq n} f_p(x_i) = f_p^{(n)}(x_1, \dots, x_n) = f(x_{\pi_1}, \dots, x_{\pi_n}).$$

Si ricorda che $f(x_{\pi_1}, \dots, x_{\pi_n})$ sono le densità di permutazione di (x_1, \dots, x_n) , $x_{\pi} \in S_x$. Con S_x insieme delle permutazioni di X_1, \dots, X_n .

Consideriamo il problema di verifica d'ipotesi nulla H_0 di una legge di probabilità di un campione, contro l'alternativa H_1 della stessa legge. Sia $f_0(x)$ la densità di x sotto H_0 , e $f_1(x)$ la densità di x sotto H_1 . Introduciamo il concetto di *test statistico* prima di descrivere un test di permutazione.

Un test non casualizzato ψ è una regola di decisione che prende valore 0 o 1,

$$\Psi = \begin{cases} 1 & \text{si rifiuta } H_0; \\ 0 & \text{si accetta } H_0. \end{cases}$$

In questo caso, applicare un test ψ consiste nel determinare una regione di rifiuto (qui essa corrisponde all'insieme di x tale che $\psi(x) = 1$) tale che, se il campione osservato appartiene a questa regione, si rifiuta l'ipotesi nulla, i criteri della regione di rifiuto di un test sono gli errori di tipo I (rifiutare H_0 quando è vera) e la potenza (probabilità di rifiutare H_0 quando essa è falsa) la regione di rifiuto C , si definisce mediante una variabile aleatoria chiamata statistica test $U(x)$ e si ha: $\psi(x) = 1$ se $x \in C$ (Pace e Salvani 2001).

Definizione 4-

Un test di permutazione con livello di significatività α ha una funzione di ripartizione che si definisce come funzione del vettore delle statistiche di ordine $x_{(j)}$ e dei ranghi $\pi = (\pi_1, \pi_2, \dots, \pi_n)$. quindi $\psi(x) = \tilde{\psi}(x_{(j)}, \pi_1, \pi_2, \dots, \pi_n)$ e è tali che

$$\frac{\sum_{\pi} \tilde{\psi}(x_{(j)}, \pi_1, \pi_2, \dots, \pi_n)}{N} = \alpha$$

con N la cardinalità spaziale delle permutazioni.

Ovvero, $E(\psi(x)) = \alpha$

Se il test non è randomizzato, ossia $\psi(x)$ non prende più valori 1 o 0, e non si può sempre raggiungere α esatto; quindi la precedente definizione si applica ai valori di α sulla forma k/N , $k = 1, 2, \dots, N$.

Per concludere, possiamo segnalare i seguenti punti:

- ❖ I test di permutazione sono procedure condizionali dove il condizionamento si fa rispetto al vettore delle statistiche d'ordine e questo porta a una legge discreta su S_x , spazio delle permutazioni.

- ❖ I test di permutazione sono test liberi di legge poiché la legge della statistica test non dipende della densità di x .
- ❖ I test di permutazione sono sempre esatti o conservatori poiché la probabilità di errore di primo tipo è sotto controllo per tutti i campioni possibili di densità di x

3.3 Test di permutazione univariati

I metodi parametrici utilizzati per la soluzione di problemi di carattere univariato e multivariato hanno, come limitazione, la necessità di dover ricorrere all'introduzione di ipotesi molto restrittive, spesso non facilmente verificabili se non impossibili da giustificare o anche irrealistiche, difficilmente interpretabili e formulate ad hoc per poter fare inferenza. A questo si deve aggiungere che le assunzioni che rendono valida l'applicazione di tali metodi (normalità, omoschedasticità, indipendenza e identica distribuzione della componente stocastica erratica) sono di norma raramente soddisfatte e, quand'anche soddisfatte, i risultati sono spesso ottenuti tramite approssimazione. Sempre più spesso, per problemi multivariati complessi studiati in ambito biomedico, ingegneristico, psicologico, farmacologico, negli esperimenti clinici, nel controllo della qualità, quando non è noto il modello distributivo, non si può invocare la normalità, l'inferenza riguarda variabili di tipo qualitativo, la numerosità del campione è inferiore al numero di variabili o ci sono dati mancanti non a caso, si passa da un approccio parametrico ad uno non parametrico, ovviando così, senza perdita sostanziale di efficienza, alle limitazioni sopra accennate.

Sia X la variabile risposta che assume valori sullo spazio campionario \mathcal{X} univariato, P la distribuzione di probabilità associata ad un esperimento caratterizzato dall'elemento X , si definisce una classe additiva \mathcal{B} -famiglia di Borel- di sottoinsiemi di \mathcal{X} . La distribuzione della popolazione P appartiene alla famiglia non parametrica \mathcal{P} della distribuzione non degenera $(\mathcal{X}, \mathcal{B})$. Generalmente, questi elementi vengono riassunti implicitamente dalla famiglia $(\mathcal{X}, \mathcal{B}, P \in \mathcal{P})$.

Il campione da $(\mathcal{X}, \mathcal{B}, P)$ è un esperimento casuale dato dalla variabile $X^{(n)} = \{X_1, \dots, X_n\}$

Secondo Good (1994), per eseguire un test di permutazione, ci sono quattro step da seguire:

- a- Analisi del problema: definizione dell'ipotesi nulla, dell'ipotesi alternativa, e le assunzioni del test;
- b- Scelta della statistica test appropriata e calcolo del suo valore in funzione del dataset osservato;
- c- Calcolo del valore della statistica test per tutte le permutazioni possibili delle osservazione di partenza, specificando il metodo di permutazione in funzione del problema da studiare;
- d- Decidere: accettare o rifiutare l'ipotesi nulla utilizzando la legge di permutazione della statistica test.

C'è da notare che, se la dimensione del campione non è piccola, un calcolo diretto della legge per enumerazione di tutte le permutazioni diventa difficile dovuto all'elevato numero di permutazioni. Per ovviare questo problema, si utilizzano metodi più efficace

dal punto di vista di calcoli. Il metodo più utilizzato a questo effetto è il metodo *Monte Carlo Condizionale*. Questa tecnica consiste nel selezionare casualmente un campione di dimensione B (B abbastanza grande), tra tutte le possibili permutazioni, e applicare il test sul campione invece di considerare l'intera legge. Molti autori hanno notato (Ernst 2004, Good 1994, Manly 1997) che una B pari a qualche migliaia è sufficiente per avere un stima precisa della probabilità critica (p -value) esatta. Sottolineiamo che tra le prerogative della metodologia di combinazione non parametrica, oltre a fornire un controllo esatto del livello di significatività per l'ipotesi globale vi è anche quella di poter pervenire ad un controllo della significatività sulle singole sub-ipotesi.

3.4 Tecniche di permutazione

3.4.1 Il metodo bootstrap

Si vuole studiare una popolazione \mathcal{X} di distribuzione P di cui si vuole stimare un parametro θ , per fare ciò, si estrae un campione casuale di n individui di questa popolazione e si misura la caratteristica di interesse su ogni individuo del campione. Per stimare la distribuzione del campione $F(\theta)$, esistono due approcci:

- La logica e l'analisi matematica ad esempio il teorema limite centrale;
- L'approccio computazionale o bootstrap che consiste nel determinare questa stima tramite simulazione dei dati.

Esistono due forme di bootstrap; quello parametrico e quello non parametrico. Il secondo caso nasce quando non siamo in grado di attribuire un modello ai dati, si usa una approssimazione empirica F_n della distribuzione F . Dalla distribuzione empirica, assumendo ci sia indipendenza e equiprobabilità tra le osservazioni IID(F_n), si genera b campioni di dimensione n , e si calcola nuovamente i b valori $\theta_1^*, \dots, \theta_b^*$ di θ per i campioni simulati. Si parla in questo caso di *ricampionamento*, in quanto viene ricostruito un insieme di campioni partendo dal campione iniziale.

3.4.2 Il metodo Monte Carlo condizionale

Innanzitutto, notiamo che sotto l'ipotesi nulla, e con l'assunto di scambiabilità, tutti i punti dello spazio $\mathcal{X}_{/x}$ hanno lo stesso grado di interesse. La trattazione matematica in forma chiusa della distribuzione di permutazione k -variata è notoriamente impossibile, sia generalmente per la elevata cardinalità dell'insieme delle possibili permutazioni ma soprattutto per la difficoltà nella caratterizzazione analitica della regione critica del test che si voglia adottare. È invece possibile e computazionalmente appropriata l'applicazione di un metodo di combinazione non parametrica dei test parziali, cui resta associata una regione critica più facilmente caratterizzabile ed il cui valore- p globale viene di fatto stimato tramite un procedimento di simulazione Monte Carlo condizionale consistente nel considerare un campione casuale semplice di B elementi dall'insieme di tutte le possibili permutazioni dei dati. Tra le diverse tecniche Monte Carlo, la più semplice è tramite il campione casuale semplice.

Senza perdita di generalità, si assume che la statistica test T di permutazione univariata è significativa per valori grandi. La distribuzione di permutazione di ogni T in corrispondenza del dataset \mathbf{X} è $F_T(z|\mathbf{X}), \forall z \in \mathfrak{R}^1$.

La procedura CMC per la stima della funzione di distribuzione cumulativa, e il p-value associato λ indotto dalla statistica T è descritto nei seguenti steps:

- Calcolare il valore osservato della statistica T : $T_{obs} = T(\mathbf{X})$
- Considerare la permutazione \mathbf{X}^* casualmente selezionata da $\mathcal{X}_{/x}$, e considerare il corrispondente valore di T , $T^* = T(\mathbf{X}^*)$.
- Ripetere indipendentemente B volte il punto precedente. L'insieme dei risultati CMC $\{T^*, r = 1, \dots, B\}$ è il campione casuale della distribuzione nulla di permutazione univariata di T .
- La stima non distorta e consistente della permutazione di $F(z|\mathbf{X})$, di T dato \mathbf{X} , è data dalla funzione di ripartizione empirica:

$$\hat{F}_B^*(z) = \sum_{r=1}^B \frac{I(T_r^* \leq z)}{B}, \forall z \in \mathfrak{R}^1,$$

dove $I(\cdot) = 1$ se la relazione $()$ è soddisfatta e 0 altrimenti, per il teorema di Glivenko-Cantelli, fornisce una stima fortemente consistente uniformemente della distribuzione $F(z|\mathbf{X})$ di T . inoltre, il p-value $\lambda = \Pr \{T^* \geq T_{obs} | \mathbf{X}\}$ ha stima non distorta e consistente, la sua relazione è data da:

$$\hat{\lambda} = \sum_{r=1}^B \frac{I(T_r^* \geq T_{obs})}{B},$$

- Se per ogni livello di significatività α fissato e positivo, $\hat{\lambda} \leq \alpha$, l'ipotesi nulla viene rifiutata.

Tabella 5. Metodo Monte Carlo Condizionale e la distribuzione di permutazione di T

\mathbf{X}	\mathbf{X}_1^*	...	\mathbf{X}_r^*	...	\mathbf{X}_B^*
\mathbf{T}_0	\mathbf{T}_1^*	...	\mathbf{T}_r^*	...	\mathbf{T}_B^*

3.5 Test di permutazione univariata per il modello CUB

3.5.1 Descrizione del test

I problemi di test per le variabili categoriali ordinali sono di grande interesse in molte discipline applicative, dove un numero limitato di tali variabili è osservato su ogni singola unità statistica. Diverse soluzioni sono state proposte per il caso univariata, la maggior parte di queste sono basate sul criterio del rapporto di verosimiglianza massimo limitato (Pesarin e Salmaso 2006).

In questo paragrafo, supponendo non note le assunzioni parametriche, con il solo principio di scambiabilità delle osservazioni, tratteremo il test di permutazione per il modello CUB descritto nel capitolo precedente. Ricordiamo che i dati sono composti di n soggetti, con le loro caratteristiche personali (covariate), e dalle variabili risposte (*domande*). L'obiettivo è di studiare l'effetto delle covariate sul modello. In effetti, si vuole effettuare un test di permutazione per valutare eventuali differenze significative tra le risposte dei soggetti in funzione delle loro caratteristiche personali. Perciò, mantenendo fisso il nostro dataset, si scambiano tra di loro solo le righe delle covariate testate. Consideriamo il modello $CUB(p_1, q_1)$ con $p_1 \leq p$, $q_1 \leq q$ e $q_1 + p_1 \leq q + p$, le covariate da sottoporre al test sono X_2 per π e W_2 per ξ .

Nel seguito, parleremo di modello ridotto per indicare il modello specificato sotto l'ipotesi nulla, e di modello completo per indicare quello specificato sotto l'ipotesi alternativa, ossia il modello $CUB(p; q)$.

La soluzione di permutazione per testare il sistema di ipotesi sopra formalizzato può essere descritta come segue:

- a- Il modello completo e quello ridotto vengono stimati sui dati originali, e viene calcolato il valore osservato della statistica test t (in principio qualsiasi statistica test per il confronto tra modelli può essere utilizzata);
- b- I vettori di covariate testate sotto l'ipotesi nulla vengono permutati, ottenendo la nuova matrice di covariate con righe $(x^*_{i,p}, w^*_{i,p})$ per $i = 1, \dots, n$;
- c- Il modello completo viene stimato sulla nuova matrice di dati permutati $(y_i, x^*_{iS}, x^*_{i\bar{S}}, w^*_{iS}, w^*_{i\bar{S}})$, con $i = 1, \dots, n$ e, considerando il modello ridotto stimato sui dati originali, si calcola la statistica test T^* ;
- d- Si ripetono gli step b e c B volte, producendo un campione della distribuzione nulla della statistica test;
- e- Si calcola il p-value del test confrontando il valore osservato della statistica con la distribuzione di permutazione sotto l'ipotesi nulla.

Le strutture dei data set permutati e non possono essere rappresentate come segue.

- $T_T'' = \max_i (1 - \lambda_i)$ è analogo ad una forma di statistica di Kolmogorov-Smirnov con aggiustamento secondo Anderson_Darling per variabili discrete.
- $T_L'' = \sum_i \phi^{-1} (1 - \lambda_i)$ dove ϕ rappresenta la funzione di ripartizione della distribuzione normale standardizzata.

Si effettuano B permutazioni delle righe delle covariate testate. Il p-value del test è calcolato con la seguente relazione :

$$\hat{\lambda} = \sum_{b=1}^B \frac{I_{[T_{obs}, \infty)}(bT^*)}{B} \quad \text{dove } I_{[T_{obs}, \infty)}(t) = 1 \text{ se } t \geq T_{obs} \text{ e pari a 0 altrimenti}$$

Sotto l'ipotesi nulla, le covariate non hanno effetto sulle risposte fornite dai soggetti intervistati, questo significa che per i valori di $W_i = w_i$ e $X_i = x_i$ fissati, è possibile ottenere qualsiasi valore di y , mentre sotto l'ipotesi alternativa, almeno una delle covariate ha effetto sulla scelta del soggetto.

Le statistiche utilizzabili per la decisione sono:

- $T_1 = 2[l(\hat{\theta}) - l(\hat{\theta}_0)]$ dove $l(\hat{\theta})$ è il valore della log-verosimiglianza del modello completo (CUB(p,q)), e $l(\hat{\theta}_0)$ è quello del modello ridotto (CUB(p₁,q₁))
- $T_2 = \left[\left| \frac{\hat{\beta}_{p_{j+1}}}{se(\hat{\beta}_{p_{j+1}})} \right|, \dots, \left| \frac{\hat{\beta}_p}{se(\hat{\beta}_p)} \right|, \left| \frac{\hat{\gamma}_{q_{j+1}}}{se(\hat{\gamma}_{q_{j+1}})} \right|, \dots, \left| \frac{\hat{\gamma}_q}{se(\hat{\gamma}_q)} \right| \right] * \mathbf{c}$

Dove $\mathbf{c} \in \mathfrak{R}_+^{p_1+q_1}$ è il vettore dei pesi, è una combinazione diretta, i p-value ottenuti sono parziali, ogni parametro viene testato.

- $T_s = \psi_s[\lambda_1, \lambda_2, \dots, \lambda_{q_2+p_2}; \mathbf{c}]$

è la generalizzazione della statistica test T_2 , con il controllo di molteplicità, in questo caso vengono combinati i k p-value parziali per ottenere il test globale T_{glob} .

dove λ_k è il p-value del test parziale relativo ad una covariata specifica, \mathbf{c} il vettore dei pesi e, ψ_s è una funzione di combinazione appropriata data la metodologia npc test (F. Pesarin 2001), ed è scelta tra le funzioni di Fisher Liptack e Tippett descritte nel capitolo successivo.

3.5.2 La routine R

La routine scritta per R implementa il test di permutazione descritto sopra. E' necessario importare precedentemente il codice per la stima del modello (cubR-perm.txt): tale codice è quasi identico alla routine CUB proposta dagli autori, le sole differenze riguardano l'output delle due routine, in quanto cubR-perm.txt restituisce i valori delle stime di massima verosimiglianza dei parametri e della verosimiglianza del modello stimato (valori che servono per costruire la statistica test ad ogni permutazione). Di seguito il codice commentato, con i riferimenti agli Step della procedura descritti sopra.

```
perm_test_cub = function(y,covariates_pi,covariates_xi,B,null_indexes){
```

Questa funzione esegue la soluzione di permutazione basata sulla permutazione del vettore delle covariate testate, per la verifica d'ipotesi nulla dei coefficienti associate a una o più covariate nel modello CUB.

Arguments:

y: di dimensione n; vettore delle risposte ordinali

covariates_pi: (n x p) vettore delle covariate per \pi;

covariates_xi: (n x q) vettore delle covariate per \xi;

B: numero di permutazione;

null_indexes: (max(p,q) x 2) vettore con la posizione del parametro testato.

Sotto l'ipotesi nulla, stimiamo il modello ridotto sui dati originali, escludendo le covariate testate (Step (a)).

y_observed = y

if(length(dim(covariates_xi))==0){

if(length(dim(covariates_pi))!=0){

if(dim(covariates_pi[,-c(null_indexes[,1])])[2]==0){

CUB_null = CUB(y_observed)

if(dim(covariates_pi[,-c(null_indexes[,1])])[2]!=0){

CUB_null = CUB(y_observed,Y=covariates_pi[,-c(null_indexes[,1])])}

if(length(dim(covariates_pi))==0){

CUB_null = CUB(y_observed)

}

if(length(dim(covariates_xi))!=0){

if(dim(covariates_xi[,-c(null_indexes[,1])])[2]==0){

if(length(dim(covariates_pi))!=0){

if(dim(covariates_pi[,-c(null_indexes[,1])])[2]==0){

CUB_null = CUB(y_observed)

if(dim(covariates_pi[,-c(null_indexes[,1])])[2]!=0){

CUB_null = CUB(y_observed,Y=covariates_pi[,-c(null_indexes[,1])])}

if(length(dim(covariates_pi))==0){

CUB_null = CUB(y_observed)}

if(dim(covariates_xi[,-c(null_indexes[,1])])[2]!=0){

if(length(dim(covariates_pi))!=0){

if(dim(covariates_pi[,-c(null_indexes[,1])])[2]==0){

CUB_null = CUB(y_observed)

if(dim(covariates_pi[,-c(null_indexes[,1])])[2]!=0){

CUB_null = CUB(y_observed,Y=covariates_pi[,-c(null_indexes[,1])],

W=covariates_xi[,-c(null_indexes[,2])])}

if(length(dim(covariates_pi))==0){

CUB_null = CUB(y_observed,W=covariates_xi[,-c(null_indexes[,2])])}

}

Stimiamo il modello completo su:

- i dati originali: all'interno del ciclo for, per $i = 1$ il modello completo viene stimato sui dati originali, in quanto i vettori delle covariate di interesse non sono ancora stati permutati (Step (a));
- i dati permutati: all'interno del ciclo for, per $i \geq 2$ il modello completo viene stimato sui dati permutati, in quanto i vettori delle covariate di interesse vengono ora, per ogni $i = 2, \dots, B + 1$, permutati (Step (b)) e (c)).

```
test_statistic = NULL
full_model_loglik = NULL
null_model_loglik = NULL
for(i in 1:(B+1)){
  if(length(dim(covariates_pi))!=0 && length(dim(covariates_xi))==0){
    CUB_full = CUB(y,Y=covariates_pi)}
  if(length(dim(covariates_pi))==0 && length(dim(covariates_xi))!=0){
    CUB_full = CUB(y,W=covariates_xi)}
  if(length(dim(covariates_pi))!=0 && length(dim(covariates_xi))!=0){
    CUB_full = CUB(y,Y=covariates_pi,W=covariates_xi)}
```

Ad ogni iterazione calcoliamo (per $i = 1$) il valore osservato della statistica test (Step (a)) e (per $i \geq 2$) i valori permutati per costruire la distribuzione di permutazione (Steps (c) e (d)). In questo caso abbiamo considerato il test rapporto di verosimiglianza:

```
full_model_loglik[i] = CUB_full$log_lik
null_model_loglik = CUB_null$log_lik
test_statistic[i] = 2*(full_model_loglik[i]-null_model_loglik)
```

Precisamente qui di seguito vengono, ad ogni iterazione del ciclo for, permutati i vettori delle covariate di interesse, mentre le altre covariate (quelle non testate sotto l'ipotesi nulla) vengono lasciate invariate rispetto alle osservazioni originali:

```
perm_positions = sample(c(1:n),size=n,replace=FALSE)
if(null_indexes[,1]!=0){
  covariates_pi[,c(null_indexes[,1])]<-covariates_pi[perm_positions,
  c(null_indexes[,1])]}
if(null_indexes[,2]!=0){
  covariates_xi[,c(null_indexes[,2])]<-covariates_xi[perm_positions,
  c(null_indexes[,2])]}
print(i)
}
```

Infine (Step (e)) il p-value del test di permutazione viene calcolato confrontando il valore osservato della statistica test con la sua distribuzione nulla di permutazione:

```
p_value = sum(test_statistic[-1]>test_statistic[1])/B
list("P.value"=p_value,"Test.statistic"=test_statistic)
}
```

3.5.3 Osservazioni

- 1- Calcolando la distribuzione della statistica test, solo le righe delle covariate studiate sono permutate e queste permutazioni sono sincronizzate, quindi la dipendenza tra le covariate testate è nonparametricamente presa in considerazione;
- 2- Dopo ogni permutazione, il calcolo della statistica test implica il calcolo della stima di massima verosimiglianza dei parametri del modello completo del dataset permutato;
- 3- Anche se il metodo proposto può essere considerato semi-parametrico, esso può opportunamente essere classificato come un approccio di permutazione poiché le tecniche parametriche prese in considerazione sono solo una via di elaborazione dei dati per il calcolo della vera e propria statistica test. Non è data nessuna assunzione, e non è applicata la teoria asintotica per descrivere esplicitamente la distribuzione della statistica.

3.6 Studio di simulazione

3.6.1 Introduzione

Metodi basati su simulazione sono stati introdotti recentemente nel contesto dell'inferenza statistica per risolvere problemi computazionali che sorgono quando la derivazione di uno stimatore o di un test richiede il calcolo di integrali complicati o senza soluzione analitica, usualmente affrontato con metodi numerici. Tali difficoltà computazionali caratterizzano, per esempio, modelli non lineari o contenenti variabili non osservabili. Mentre negli ultimi anni lo studio e l'applicazione di metodi basati su simulazione per la stima ha ricevuto una attenzione crescente, l'uso dei metodi simulati nel contesto della verifica di ipotesi si presta a nuovi sviluppi sia teorici che applicativi. In questo caso studio, per valutare la consistenza e il comportamento dei test, sono state effettuate simulazione tramite il metodo di Monte Carlo. Lo studio è stato effettuato su una serie di dati con 1000 permutazioni per ogni test effettuato. Il modello CUB è il modello (2.3) presentato nel capitolo 2, e le componenti sistematiche sono quelle descritte in (2.5), e l'ipotesi nulla di interesse è che i parametri delle componenti siano pari a zero, cioè nessun effetto delle covariate sul modello CUB. I dati sono simulati per i modelli CUB(0,1), particolarmente adatto a descrivere molti casi reali.

3.6.2 Studio della potenza

In questa simulazione relativa al modello CUB(0,1), si studia la stima della potenza del:

- test di permutazione basato sulla statistica di log-verosimiglianza (T_1);
- test di permutazione basato sull'errore standard dei coefficienti γ_1 (T_2);
- test parametrico di Wald (W). Nel test di Wald la stima di massima verosimiglianza $\hat{\theta}$ del parametro di interesse θ è confrontata con un valore proposto, θ_0 , sotto l'assunzione che la differenza tra le due avrà una distribuzione che si può approssimare con una normale. Tipicamente, il quadrato delle

differenze viene confrontato con una distribuzione Chi-quadro. Nel caso univariato, la statistica di Wald è :

$$\frac{(\hat{\theta} - \theta_0)^2}{var(\hat{\theta})}$$

La potenza è funzione della differenza tra i valori di ξ_0 e di π sotto H_0 e H_1 . Assumiamo di voler testare l'ipotesi nulla che una o più covariate siano inutili nella modellazione della variabile risposta: l'idea di base è che sotto tale ipotesi nulla vi è scambiabilità all'interno del vettore/i delle covariate di interesse, ovvero, non essendo tali covariate utili per spiegare il processo di decisione dei rispondenti, qualsiasi permutazione dei valori osservati nel campione per tali covariate è ugualmente verosimile. Per i modelli CUB(0,1) e CUB(1,0) con covariata dicotomica D, possiamo scrivere il sistema di verifica d'ipotesi come segue:

per ξ : CUB(0,1)

$$\begin{cases} H_0: \gamma_1 = 0 \equiv H_0: \xi = \xi_{(0)} = \frac{1}{1 + \exp(-\gamma_0)} \equiv \gamma_0 = \log\left(\frac{\xi_{(0)}}{1 - \xi_{(0)}}\right) \\ H_1: \gamma_1 \neq 0 \equiv se D = 0, H_1: \xi = \xi_{(1)} = \frac{1}{1 + \exp(-\gamma_0 - \gamma_1)} \equiv \gamma_1 = \log\left(\frac{\xi_{(1)}(1 - \xi_{(0)})}{\xi_{(0)}(1 - \xi_{(1)})}\right) \end{cases}$$

per π : CUB(1,0)

$$\begin{cases} H_0: \beta_1 = 0 \equiv H_0: \pi = \pi_{(0)} = \frac{1}{1 + \exp(-\beta_0)} \equiv \beta_0 = \log\left(\frac{\pi_{(0)}}{1 - \pi_{(0)}}\right) \\ H_1: \beta_1 \neq 0 \equiv se D = 0, H_1: \pi = \pi_{(1)} = \frac{1}{1 + \exp(-\beta_0 - \beta_1)} \equiv \beta_1 = \log\left(\frac{\pi_{(1)}(1 - \pi_{(0)})}{\pi_{(0)}(1 - \pi_{(1)})}\right) \end{cases}$$

3.6.3 Risultati delle simulazioni

Presentiamo successivamente i risultati delle simulazioni per il test di permutazione del modello su una covariata di ξ (CUB(0,1)) abbiamo scelto questo modello per dare un maggiore interesse al feeling, con CMC = 1000, e B = 1000, la stima del parametro di π è pari a 0.80.

Secondo gli autori del modello CUB, perché il modello funzioni, deve essere verificata la relazione: $\frac{n}{m} \geq 7$. Quindi, per i dati simulati di dimensione $n = 50$, la lunghezza della scala di valutazione m deve essere minore o uguale a 7. Nelle nostre simulazioni, abbiamo scelto in primo luogo $n = 25$, $m = 6$, la covariata di π è nulla, quella di ξ è una variabile dummy. La tabella 6 riassume il comportamento di γ_1 e ξ sotto l'ipotesi nulla e sotto l'ipotesi alternativa essendo che:

- Sotto l'ipotesi nulla siamo in presenza di un modello MUB o CUB(0,0), e il parametro da stimare è γ_0 ;

$$H_0: \xi = \xi_{(0)}$$

- Sotto l'ipotesi alternativa i parametri da stimare sono γ_0 e γ_1 il modello è il CUB(0,1);

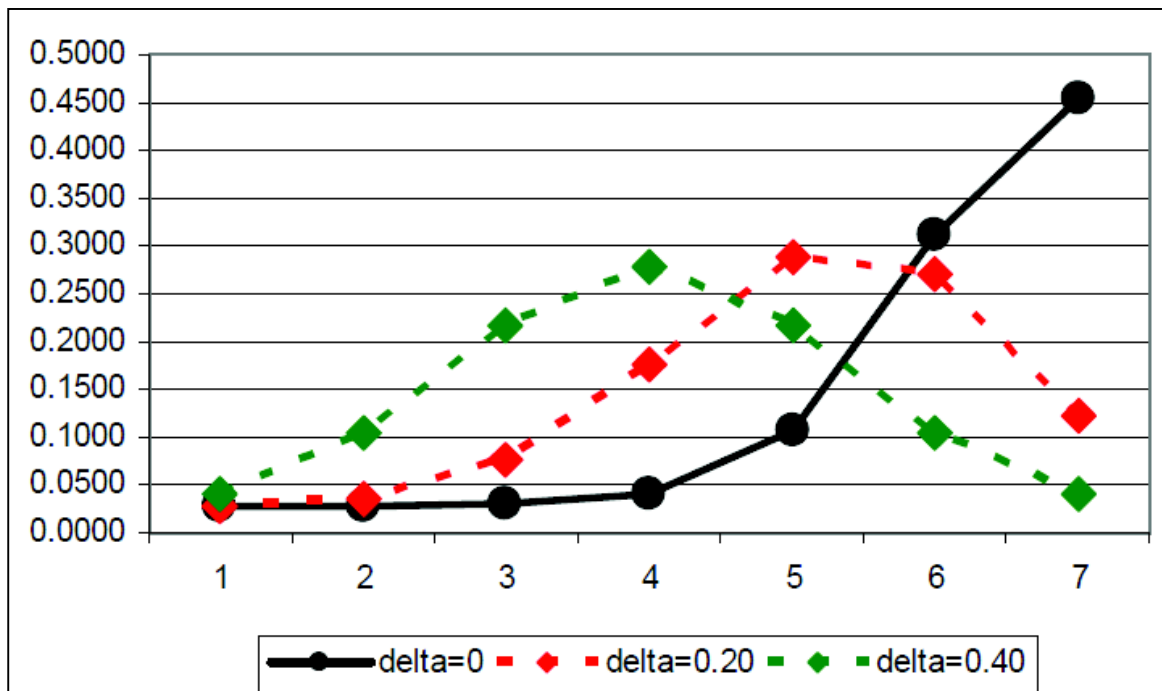
$$H_1: \xi = \xi_{(1)} = \xi_{(0)} + \delta$$

dove δ è una costante nota. Il tasso di rigetto deve essere crescente per i valori crescenti di δ .

Tabella 6: Comportamento di ξ sotto H_0 CUB(0,1)

$\xi_{(0)}$	γ_0	δ	$\xi_{(1)} = \xi_{(0)} + \delta$	γ_1	Ipotesi
0,15	-1.7346011	0.00	0.15	0.00	H_0
0,15	-1.7346011	0.05	0.20	0.35	H_1
0,15	-1.7346011	0.10	0.25	0.64	
0,15	-1.7346011	0.15	0.30	0.89	
0,15	-1.7346011	0.20	0.35	1.12	
0,15	-1.7346011	0.25	0.40	1.33	
0,15	-1.7346011	0.30	0.45	1.53	
0,15	-1.7346011	0.35	0.50	1.73	
0,15	-1.7346011	0.40	0.55	1.94	

Figura 2: distribuzione della variabile risposta al variare di delta per $m = 7$.



Dal grafico 2, vediamo che per il valore delta pari a 0, la curva è monotona crescente, mentre per delta pari a 0.20, l'andamento della curva si vede cambiato dopo una crescita tra la 1 e 5, essa decresce, vale lo stesso per la curva con delta pari a 0.40, però in questo caso, la curva cambia il verso alla modalità 4. Sembra esserci sempre di più un spostamento verso la modalità più piccola al crescere di delta.

La stima della potenza dei test T_1 , T_2 e W , per il livello $\alpha = 0.05$; $m = 7$; $n_1 = n_2 = 25$, $\pi = 0.8$ è riassunta nella tabella 7. Ricordiamo che T_1 è il test di permutazione basato sulla log-verosimiglianza T_2 è il test basato sull'errore di stima, e W è il test di Wald.

Tabella 7: Studio della potenza al variare di delta per $m=7$, $n_1=n_2=25$.

δ	T_1	T_2	W	Ipotesi
0.00	0.048	0.050	0.042	H_0
0.05	0.173	0.178	0.164	H_1
0.10	0.460	0.447	0.449	
0.15	0.725	0.721	0.729	
0.20	0.893	0.877	0.889	
0.25	0.968	0.962	0.963	
0.30	0.985	0.990	0.992	
0.35	0.996	0.993	0.993	
0.40	0.998	1.000	0.998	

Tabella 8: Studio della potenza al variare di delta per $m=6$, $n_1=n_2=25$.

δ	T_1	T_2	W	Ipotesi
0.00	0.061	0.061	0.045	H_0
0.05	0.157	0.136	0.151	H_1
0.10	0.460	0.453	0.401	
0.15	0.754	0.720	0.722	
0.20	0.893	0.877	0.889	
0.25	0.979	0.969	0.968	
0.30	0.992	0.986	0.971	
0.35	0.997	1.000	0.990	
0.40	0.998	1.000	0.994	

Sotto H_1 le potenze sono molto vicine, quindi possiamo dire che le performance di tutti e tre i test sono abbastanza simili, in più si nota che per il valore di delta maggiore di 0.25, la potenza è maggiore del 95%.

Tabella 9: Studio della potenza al variare di delta per $m=5$, $n_1=n_2=20$.

δ	T_1	T_2	W	Ipotesi
0.00	0.052	0.050	0.025	H_0
0.05	0.123	0.122	0.077	H_1
0.10	0.254	0.232	0.129	
0.15	0.453	0.431	0.363	
0.20	0.673	0.651	0.609	
0.25	0.799	0.752	0.714	
0.30	0.885	0.861	0.855	
0.35	0.962	0.895	0.891	
0.40	0.975	0.954	0.947	

Tabella 10: Studio della potenza al variare di delta per $m=4$, $n_1=n_2=15$.

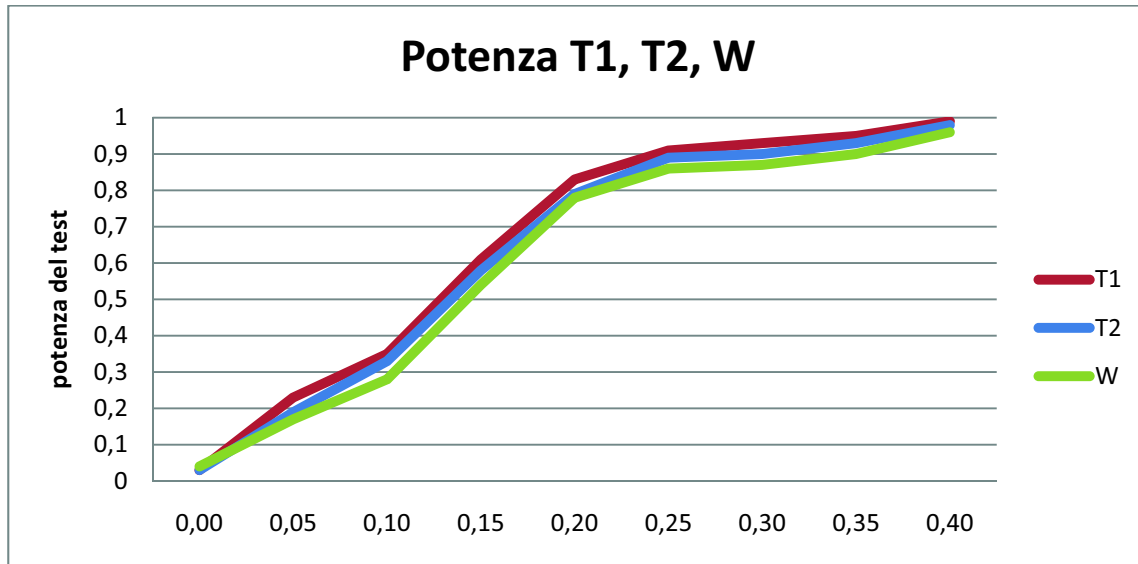
δ	T_1	T_2	W	Ipotesi
0.00	0,032	0,033	0,040	H_0
0.05	0,235	0,191	0,178	H_1
0.10	0,354	0,336	0,281	
0.15	0,619	0,582	0,547	
0.20	0,830	0,799	0,783	
0.25	0,911	0,897	0,869	
0.30	0,933	0,904	0,878	
0.35	0,950	0,938	0,900	
0.40	0,999	0,982	0,967	

Tutti i test simulati; test di permutazione basato sulla log-verosimiglianza, test basato sull'errore, il test di Wald sono abbastanza potenti, il test T_2 converge a 1 e talvolta, il test T_1 è meglio del test W . Le simulazioni mediante il CMC mostra un buon comportamento

della potenza in diverse situazioni. In effetti, per campioni meno numerosi, in questo caso abbiamo simulato il modello con $m = 4$ e $n_1 = n_2 = 15$, $\pi = 0.6$ e $\xi = 0.15$, il risultato è riassunto nella Tabella 10, si nota che è verificata la disuguaglianza;

$$\text{potenza}(T_1) > \text{potenza}(T_2) > \text{potenza}(W)$$

Figura 3: Andamento della potenza per i tre test nel caso $n_1 = n_2 = 15$, $m=4$.



3.6.4 Codici R per la simulazione del CUB(0,1)

```
source("cubR_perm.txt")
source("z_perm_test_functions_cub.txt")
source("combine.txt")
source("t2p.r")
source("FWEminP.r")
```

// con questi comandi, richiamo:

- la funzione CUB,
- La funzione permutazione,
- La funzione FWE (*Familywise Error Rate*) la quale calcola la probabilità di commettere almeno un errore di primo tipo univariato
- La funzione t2p, la quale fornisce una matrice dei p-value parziali dalla matrice relativa alla distribuzione nulla multivariata.
- La funzione combine realizza la combinazione non parametrica di prove dipendenti con alcune funzioni di combinazioni ed alternative limitate.

```
##### simulazioni univariato #####
### CUB(0,1)
Alpha = 0.05
// livello di significatività
n_vector = 50
```

```

// dimensione del campione
m = 6
// scale di valutazioni
par_pi = c(.6)
// stima di pai
xi_0 = 0.15
covariates_pi = NULL
covariates_xi = array(c(rep(0,n_vector/2),rep(1,n_vector/2)),dim=c(n_vector,1))
null_indexes = array(c(0,1),dim=c(1,2))
CMC=500
// numero di simulazione Monte Carlo
B=1000
// numero di simulazione bootstrap
pv=array(0,c(9,CMC,3))
// matrice dei pvalue
nvalid=array(0,c(9,CMC))
power=array(0,c(9,4))
// matrice delle potenze
for (d in 0:8){
// inizializzazione del ciclo
delta=0.05*d
xi_1 = xi_0+delta
par_xi = c(log(xi_0/(1-xi_0)),log((xi_1*(1-xi_0))/(xi_0*(1-xi_1))))
for (s in 1:CMC){
y2 = simulate_values(n_vector,m,par_pi,par_xi,covariates_pi,covariates_xi)
test_perm = perm_test_cub(y2,covariates_pi,covariates_xi,B,null_indexes,m)
test_para = CUB(y2,W=covariates_xi)
pv[d+1,s,1]=test_perm$P.value_1
pv[d+1,s,2]=test_perm$P.value_2
pv[d+1,s,3]=test_para$p_value[3]
nvalid[d+1,s]=test_perm$Perm.valid_1
}
power[d+1,1]=delta
power[d+1,2]=sum(pv[d+1,,1]<alpha)/CMC
power[d+1,3]=sum(pv[d+1,,2]<alpha)/CMC
power[d+1,4]=sum(pv[d+1,,3]<alpha)/CMC
}
}

```

Capitolo 4. Applicazione del modello CUB ad un indagine di marketing sul vino passito

4.1 Introduzione

Nel presente capitolo verrà effettuata una analisi dei dati reali di un indagine sul vino passito, applicando i modelli CUB sia in ambito parametrico che non parametrico. Nella prima parte si parlerà dell'origine dei nostri dati, successivamente di un'analisi descrittiva dei dati, dell'applicazione dei CUB parametrici, e in fine si eseguirà una analisi di tipo non parametrico applicando i test di permutazione.

La legge del 20 febbraio 2006 n.82 “Disposizioni di attuazione della normativa comunitaria concernente l'organizzazione comune di mercato (OCM) del vino”, definisce per la prima volta in un testo ufficiale cosa si deve intendere per vino passito.

Per «vino passito» o «passito» si intende un vino sottoposto ad appassimento, anche parziale, naturale sulla pianta o dopo la raccolta. L'appassimento può essere realizzato mediante uno o più procedimenti e tecniche, anche con l'ausilio di specifiche attrezzature. Nella produzione dei vini passiti non è consentita alcuna pratica di arricchimento del titolo alcolimetrico naturale delle uve prima o dopo l'appassimento. La definizione di vino passito si applica ai vini da uve stramature, nonché ai vini ad indicazione geografica tipica e ai vini di qualità prodotti in regioni determinate (VQPRD), per i quali è prevista tale tipologia nei singoli disciplinari di produzione. I vini passiti possono essere ottenuti da uve di tutte le varietà autorizzate alla produzione di vino, fatte salve eventuali limitazioni presenti nei disciplinari dei vini ad indicazione geografica tipica e a denominazione di origine. La menzione «vino passito liquoroso» o «passito liquoroso» è riservata ai vini liquorosi ad indicazione geografica tipica e a denominazione di origine i cui disciplinari prevedono tale tipologia. La menzione «vino passito» o «passito», ai sensi dell'articolo 12, paragrafo 1, lettera *b*), del regolamento (CE) n. 753/2002 della Commissione, del 29 aprile 2002, e successive modificazioni, sostituisce la denominazione «vino da uve stramature» e può essere accompagnata in etichetta dalla citazione «vendemmia tardiva». Il nome «passito» o «vino passito» può inoltre essere sostituito in etichetta dai tradizionali «Vin santo», «vino santo», «vinsanto» esclusivamente nel caso di VQPRD, i cui disciplinari prevedono tali menzioni (Gazzetta ufficiale n.60 13 marzo 2006).

In Italia li possiamo classificare in questo modo:

- passiti poco o non invecchiati,
- passiti invecchiati a lungo,
- passiti liquorosi, che per il 27% provengono da vitigni rossi e per il 73% da vitigni bianchi.

“Il consumatore ricerca nel vino passito le qualità organolettiche, l'originalità del prodotto, la storia e tradizione, il legame col territorio, e la riscoperta della produzione tipica” (Fregoni, 2006).

4.1.1 Passiti Veneti

Il Veneto è il maggior produttore di vini passiti italiani, conta ben 8 diverse denominazioni. Nella nostra indagine di mercato sono state considerate le seguenti:

- **Torchiato di Fregona:** È un vino da meditazione, limpido e dal colore giallo dorato intenso, con riflessi a volte ambrati. Esso viene prodotto in tre comuni: Fregona, Cappella Maggiore e Sarmede. Il nome Torchiato deriva dalla “Veneto Orientaleazione”, ossia dalla torchiatura.
- **Refrontolo Passito (Marzemino):** Il Marzemino di Refrontolo è un vino diverso da quello prodotto in altre zone, è un vino da dessert, da conversazione, di un bel colore rosso rubino con orli violacei, amabile, sapido, frizzante, ricco di profumi con una nota di mora di rovo e di morasca e può raggiungere i 12-14 gradi. Le origini del Marzemino si perdono nel tempo, sono nebulose e controverse.
- **Fior d’arancio passito dei Colli Euganei:** Da tempi remoti è una tradizione contadina quella di produrre il passito con le uve di moscato fior d’arancio.
- **Torcolato di Breganze:** Il Torcolato, vino bevuto dall'antichità sia dalle famiglie ricche che da quelle più umili, al quale sono stati attribuiti molti riconoscimenti ad iniziare dal 1909 quando, presentato da Arnaldo Carli all'Esposizione di Lonigo, vince la medaglia d'oro. Nel 1980 varca il confine nazionale raggiungendo l'America, nel 1983 conquista un articolo sulla rivista "Gourmet", e nel '93 è presente anche sul "The New York Times".
- **Recioto di Valpolicella:** Viene prodotto in 19 comuni della fascia settentrionale della provincia di Verona, da est a ovest.
- **Recioto di Soave:** Il Recioto di Soave è un vino DOCG la cui produzione è permessa nella provincia di *Verona*, precisamente nei comuni di *Soave*, *Monteforte d'Alpone* e *Montecchia di Crosara*. Il nome deriva da un “Veneto Occidentaleabolo” dialettale veneto: deriva da “*recia*” (che è la parte alta del grappolo di Garganega, quella più ricca di zuccheri e meglio esposta alla luce solare). Nel 1998 è stato il primo vino veneto ad ottenere la denominazione di origine controllata e garantita. Il nome “recioto”, riservato esclusivamente ai vini veronesi ed al Recioto di Gambellara, è protetto dalla Legge n. 46 del 01/03/1975.

4.1.2 Profilo dei consumatori di vino in Italia

Gli italiani che consumano vino sono 24 milioni. Proprio mentre il consumo pro capite scende a 49 litri (50 litri nel 2002 e 104 nel 1975), 16 milioni di italiani bevono quotidianamente vino e nell’88% dei casi viene consumato tutti i giorni. Gli altri 8 milioni di consumatori non sono invece abituali, privilegiano la qualità e bevono vino soprattutto nei momenti conviviali con gli amici, i giorni festivi e i pasti al ristorante. In conclusione gli italiani bevono meno, ma ricercano di più la qualità. Il numero dei consumatori abituali si sta però lentamente contraendo, mentre cresce l’importanza dei

consumatori non abituali che hanno rinunciato a bere quotidianamente vino ma lo consumano in occasioni di incontro con gli amici (58%), nei giorni festivi (48%), nei pasti al ristorante (45%) o in altre ricorrenze (18%). (WineNews; www.winenews.it).

La contrazione dei consumi d'inizio 2009 pare essere più legata alla paura di perdere la patente e pagare multe salate che non alle ricadute della crisi economica iniziata nel 2008. Si rileva dall'indagine effettuata che per gli enoappassionati, è preferibile tagliare altre spese che di rinunciare alle proprie etichette.

Nel presente capitolo, nel primo paragrafo faremo breve descrizione dell'indagine campionaria effettuata nonché la struttura del questionario, nel secondo paragrafo, condurremo una analisi descrittiva dei dati disponibili, nel successivo paragrafo condurremo una analisi parametrica del modello CUB descritto nel capitolo 2, e finalmente nel quinto capitolo applicheremo il test di permutazione descritto nel precedente capitolo.

4.2 Disegno dell'indagine

Al fine di studiare l'andamento della domanda di vini passiti in Veneto è stata fatta un'indagine sul campo, somministrando agli intervistati il questionario riportato in Appendice B.

4.2.1 Modalità di svolgimento delle interviste e raccolta dei dati

La raccolta dati è avvenuta principalmente in modalità CATI (Computer-Assisted Telephone Interviewing), che è una modalità di rilevazione diretta realizzata attraverso interviste telefoniche, dove l'intervistatore legge le domande all'intervistato e registra le risposte su un computer, tramite un apposito software. Produrre interviste avvalendosi di questo sistema permette una documentazione precisa dei dati elementari ed esclude ogni possibile errore sistematico durante il rilevamento dei dati, in quanto il questionario statistico è contenuto nel computer per cui le domande vengono poste esattamente come compaiono sul video e le risposte sono registrate direttamente su un dispositivo di memorizzazione. Il software utilizzato inoltre effettua alcuni controlli di qualità sui dati automaticamente all'immissione degli stessi, cosicché i tempi dell'indagine risultano essere notevolmente accorciati. L'indagine è stata condotta in circa 5 mesi nel periodo compreso tra Aprile 2009 e Agosto 2009. Il numero complessivo di soggetti intervistati è pari a 386.

4.2.2 Definizione del questionario

L'intervista si articola in tre parti principali: una prima parte, nella quale sono rilevati l'età, il sesso e la provincia di residenza dei rispondenti; una seconda parte in cui sono stati rilevate informazioni circa le loro abitudini di acquisto e di consumo di vino passito e le loro preferenze in termini di gusto.

Le risposte alle domande presentate sono state di diverso tipo:

- Chiuse:
 - dicotomiche
 - politomiche
 - scale verbali
- Aperte

Le domande di tipo dicotomico sono quelle che richiedono una sola risposta tra due modalità (SÌ/NO).

Le domande politomiche prevedono una sola risposta tra tre o più opzioni previste dal quesito.

Le scale verbali di tipo Likert sono scale a 5 (1= *mai*, 5= *sempre*) o a 7 (1= *per nulla*, 7= *moltissimo*) modalità che sottendono valori numerici. L'ipotesi che sta dietro a questo tipo di scale è che la distanza tra le modalità sia la stessa e che sia possibile effettuare calcoli numerici. Si tratta quindi di scale di tipo additivo che consentono, se messe insieme in una batteria di domande, di totalizzare punteggi complessivi. L'utilizzo di questo tipo di scale permette di cogliere la multidimensionalità che sta alla base dei fenomeni che si vogliono esplorare. Ogni fenomeno è quindi scomposto in sottodimensioni e ogni singola scala rileva una sottodimensione. Assunto che le risposte siano organizzate in modo da presentare dipendenze positive.

4.3 Analisi descrittiva dei dati

I risultati riportati in questo paragrafo riguardano l'analisi dei dati a raccolti mediante misure, percentuali e di diagrammi a barre.

Le Tabelle 1 e 2 contengono informazioni preliminari sulle caratteristiche delle persone intervistate, le domande a loro fatte riguardano il consumo di vino e in particolare di quello passito.

Come mostrato in Tabella 1, la maggior parte delle persone intervistate risiede in Veneto e ovvero nelle provincie di Belluno, Padova, Rovigo, Treviso, Venezia, Vicenza, Verona. Tra queste, la provincia con il numero più elevato di intervistati è Padova, con una quota del 16% sul totale, seguita dal 15% di Treviso e dal 14% delle altre città. Le persone che risiedono fuori regione, costituiscono l'1% del campione.

Tabella 11. *Distribuzione percentuale geografica degli intervistati*

Provincia di residenza	N	%
Belluno	52	14%
Padova	60	16%
Rovigo	52	14%
Treviso	57	15%
Venezia	52	14%
Vicenza	54	14%
Verona	53	14%
altro	4	1%
Non Risposto	2	-
Totale	386	100%

Analizzando i dati, in base alla fascia di età al sesso degli intervistati (Tabella 12), emerge che il campione è costituito prevalentemente da persone di sesso maschile (61.35%). Da notare anche che, sia nel gruppo dei maschi che in quello delle femmine, la fascia di età predominante è compresa tra i 20 e 29 anni, la percentuale di intervistati con meno di 20 anni e tra 70 e 80 è molto bassa.

Tabella 12. *Distribuzione di frequenze assolute e percentuali di intervistati per età condizionate al sesso*

Età	Sesso		Sesso	
	F	M	F	M
≤20	3	5	2%	2%
20-29	74	150	50%	64%
30-39	12	30	8%	13%
40-49	15	11	10%	5%
50-59	25	20	17%	9%
60-69	14	17	9%	7%
70-80	5	2	3%	1%
Totale	148	235	100%	100%

La prima domanda posta agli intervistati riguarda il loro gradimento del vino. È stato chiesto di esprimere la loro preferenza attraverso un punteggio, con riferimento ad una scala di valutazione da 1 (= *per nulla*) a 7 (= *moltissimo*).

In Tabella 13, sono riportati i risultanti ottenuti. Si nota che, alla maggior parte delle persone piace il vino, infatti le percentuali più alte si riscontrano nei punteggi da 4 a 7, con un picco del 21% per il grado di preferenza pari a 5.

Lo scarto quadratico medio è di 1.8, ciò sta ad indicare una forte eterogeneità delle risposte. Ci aspettiamo inoltre, che la maggior parte dei giudizi si raggruppi nel raggio di 1.8 dalla media, vale a dire tra nell'intervallo $[4.4 \pm 1.8]$ pari a $[2.4 ; 6.2]$. In effetti,

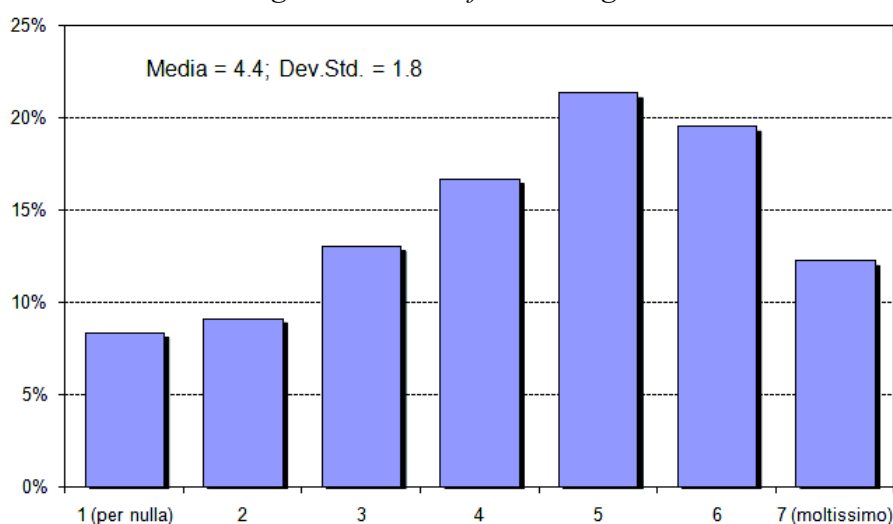
possiamo osservare che le preferenze del 70.2% degli intervistati cadono in questo intervallo.

Tabella 13. *Distribuzione percentuale di intervistati per grado di soddisfazione nel bere vino, data la scala di valutazione da 1 a 7*

Quanto gradisce bere vino?	N	%
1 (per nulla)	32	8%
2	35	9%
3	50	13%
4	64	17%
5	82	21%
6	75	19%
7 (moltissimo)	47	12%
NA	1	-
Totale	386	100%
Media	4,4	
Dev. Std.	1,8	

La Figura 4 mostra graficamente quanto riportato nella tabella precedente. Sull'asse delle ascisse troviamo la scala delle preferenze da 1 a 7 e sulle ordinate la percentuale delle persone che hanno espresso un determinato giudizio. I picchi più alti risultano per i punteggi 5 e 6, quindi valutazioni positive alla domanda sul gradimento del vino. Osservando il grafico si nota, anche, che la distribuzione dei dati ha una asimmetria negativa.

Figura 4. *Percentuale del grado di soddisfazione degli intervistati nel bere vino*



Con la seconda domanda si chiede agli intervistati quanto frequentemente bevano vino durante i pasti, proponendo loro una scala di preferenze che, in questo caso, va da 1 (= *mai*) a 5 (= *regolarmente*). Si nota che, più della metà delle persone ha dato un punteggio

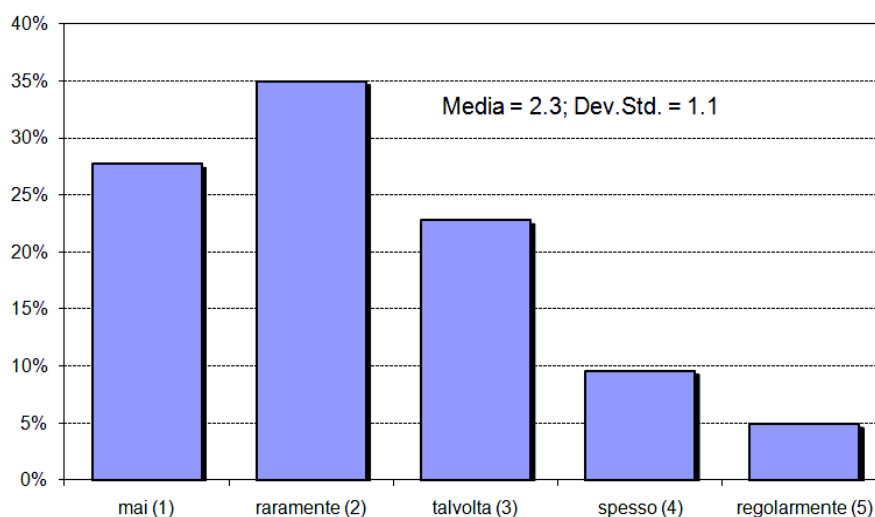
basso, quindi siamo di fronte ad un campione che solitamente non beve vino durante il pranzo o la cena, a casa.

Tabella 14. *Distribuzione percentuale di intervistati per frequenza con cui bevono vino durante i pasti*

In quale occasione beve vino?		
A casa durante i pasti	N	%
mai (1)	107	28%
raramente (2)	135	35%
talvolta (3)	88	23%
spesso (4)	37	10%
regolarmente (5)	19	5%
Totale	386	100%
Media	2,3	
Dev. Std.	1,1	

Lo scarto quadratico medio è pari a 1.1 e la media a 2.3, questo comporta che il 57.7% dei giudizi si distribuiranno in un range di punteggi compresi tra 1.2 e 3.4, nel nostro caso tra 2 e 3. Quanto detto si riscontra nel diagramma a barre in Figura 5, nel quale appare evidente come la percentuale di intervistati che ha espresso giudizi bassi prevale rispetto agli altri. Questo è evidenziato anche dall'asimmetria positiva del grafico.

Figura 5. *Percentuale delle volte che gli intervistati bevono vino durante i pasti*



Un'altra domanda viene fatta relativamente alla frequenza con cui l'intervistato beve vino, ma stavolta si fa riferimento al vino preso al bar/enoteca. La scala di valutazione è sempre compresa tra 1 (= *mai*) e 5 (= *regolarmente*). A differenza di prima, i punteggi dati sono stati leggermente più alti, spostando il picco di preferenza sul valore 3, infatti il 34% degli intervistati dichiara che solo *talvolta* beve del vino al bar. Una percentuale piuttosto alta di persone ha dato ancora punteggio 2, affermando di bere *raramente* del

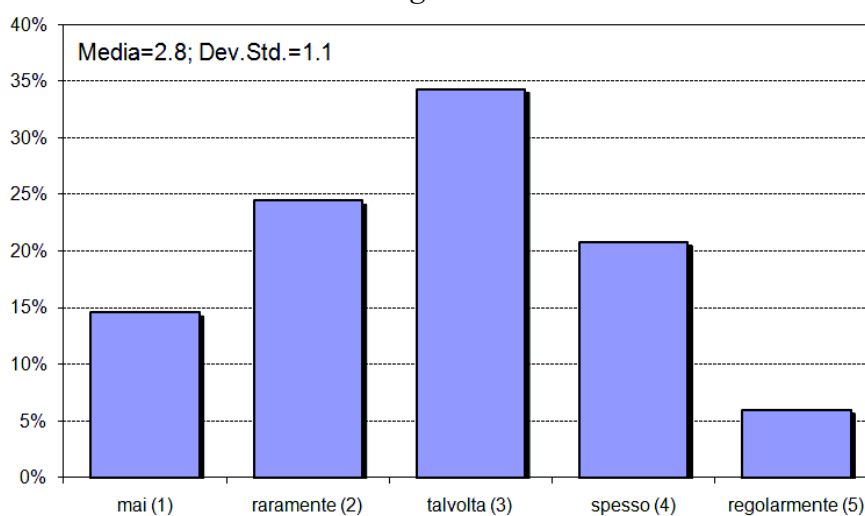
vino fuori casa. Quanto detto si riscontra in media e deviazione standard, che assumono rispettivamente i valori 2.8 e 1.1. c'è da notare che la media risulta più alta rispetto al valore assunto nella domanda precedentemente, mentre la deviazione standard è rimasta invariata.

Tabella 15. *Distribuzione percentuale di intervistati per la frequenza con cui bevono vino al bar/ enoteca*

In quale occasione beve vino?		
Al bar/enoteca	N	%
mai (1)	56	15%
raramente (2)	94	24%
talvolta (3)	132	34%
spesso (4)	80	21%
regolarmente (5)	23	6%
Non Risposto	1	-
Totale	386	100%
Media	2,8	
Dev. Std.	1,1	

La Figura 6 evidenzia che il picco massimo è rappresentato dal 34% degli intervistati che talvolta prende vino al bar/enoteca, mentre il minimo è dato dal 6% del campione che prende *regolarmente* vino al bar. Globalmente si osserva che la distribuzione è leggermente obliqua a sinistra. Analizzando media e deviazione standard vediamo che le risposte del 58.5% dei rispondenti cadono all'interno dell'intervallo di valori compreso tra 1.7 e 3.9, ciò sta a indicare che i nostri intervistati in generale non prendono vino in queste occasioni.

Figura 6. *Percentuale delle volte che gli intervistati bevono vino bar/enoteca*



Un'ultima domanda, riguardante l'occasione in cui l'intervistato beve vino, fa riferimento al pranzo/cena al ristorante/pizzeria. La risposta dominante risulta assegnata al valore 3 (= *talvolta*), la seconda preferenza è stavolta data al punteggio 4 (= *spesso*). Come evidenzia

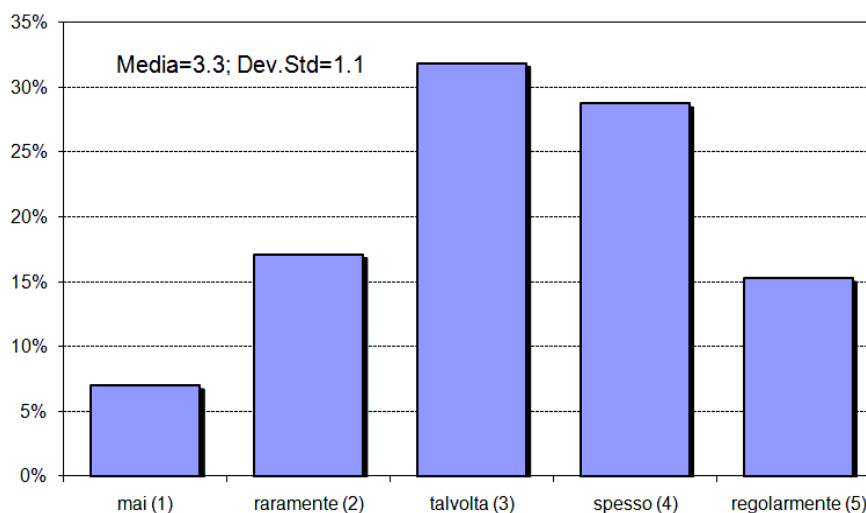
la media, che è pari a 3.3, abbiamo un aumento delle preferenze verso valori più elevati della nostra scala di riferimento, infatti il 60.6% della popolazione ha dato punteggi compresi tra 2.2 e 4.4, o meglio tra 3 e 4.

Tabella 16. *Distribuzione percentuale di intervistati per frequenza con cui bevono vino a pranzo/cena al ristorante/pizzeria*

In quale occasione beve vino?		
A pranzo/cena in ristorante/pizzeria	N	%
mai (1)	27	7%
raramente (2)	66	17%
talvolta (3)	123	32%
spesso (4)	111	29%
regolarmente (5)	59	15%
Totale	386	100%
Media	3,3	
Dev. Std.	1,1	

Come mostra la Figura 7, si può notare come la maggior parte del campione abbia dichiarato che gli capita talvolta o spesso di prendere del vino al ristorante/pizzeria. Mettendo al confronto questo risultato con quello dei grafici in Figura 6 e in Figura 7, potremmo dire che i nostri intervistati generalmente preferiscano bere vino durante i pasti, ma fuori casa.

Figura 7. *Percentuale delle volte che gli intervistati bevono vino a pranzo/cena al ristorante*



Una volta raccolte le informazioni relative al vino in generale, agli intervistati si pongono delle domande focalizzate su un particolare vino, quello passito. Per prima cosa si chiede loro se conoscano il vino passito e, come in precedenza, la risposta deve essere data attraverso un punteggio da 1 (= *per nulla*) a 7 (= *moltissimo*). La Tabella 17 mostra che la metà circa del campione conosce poco o per nulla questo vino, mentre per la restante metà solo il 19% può dire di conoscere bene o molto bene il passito. Lo scarto quadratico

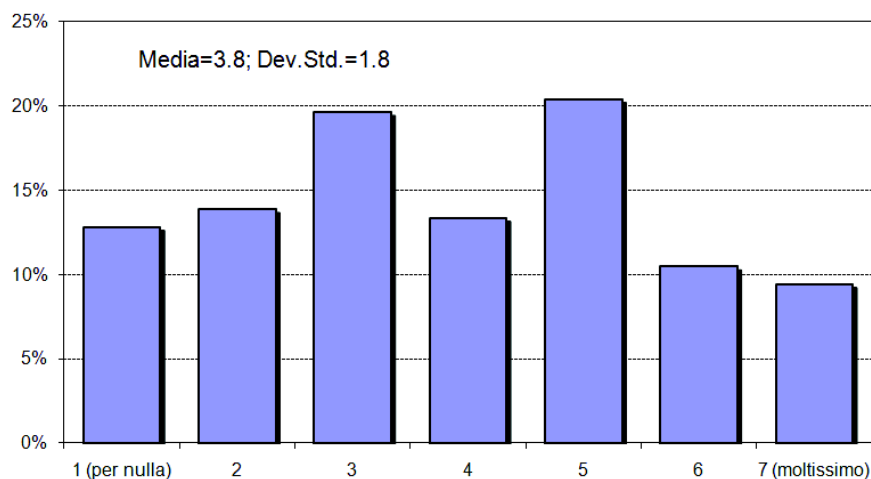
medio è pari a 1.8, quindi, come si vede in Figura 8, si ha una certa eterogeneità nelle risposte, la distribuzione percentuale si aggira tra un 10% e 20%, senza nessun picco evidente.

Tabella 17. *Distribuzione percentuale di intervistati per livello di conoscenza del vino passito dolce, data una scala di giudizio da 1 a 7*

Conosce il vino passito dolce?	N	%
1 (per nulla)	49	13%
2	53	14%
3	75	20%
4	51	13%
5	78	20%
6	40	10%
7 (moltissimo)	36	9%
Non Risposto	4	-
Totale	386	100%
Media	3,8	
Dev. Std.	1,8	

Il diagramma a barre in Figura 8 evidenzia questa eterogeneità delle risposte, con percentuali più elevate per quanto riguarda i punteggi 3 e 5, scelti entrambi da circa un quinto dei rispondenti. Inoltre solamente il 9% del campione dichiara di conoscere molto bene il vino passito.

Figura 8. *Percentuale di intervistati per livello di conoscenza del vino passito*



In Tabella 18, sono presentati alcuni tipi di vino passito del Veneto, ed accanto ad ognuno ritroviamo la percentuale degli intervistati che ha dichiarato di conoscere quel particolare tipo di vino. Va evidenziato che, a differenza delle domande precedenti, l'intervistato stavolta non deve tradurre la propria risposta in un punteggio, sulla base di una scala di

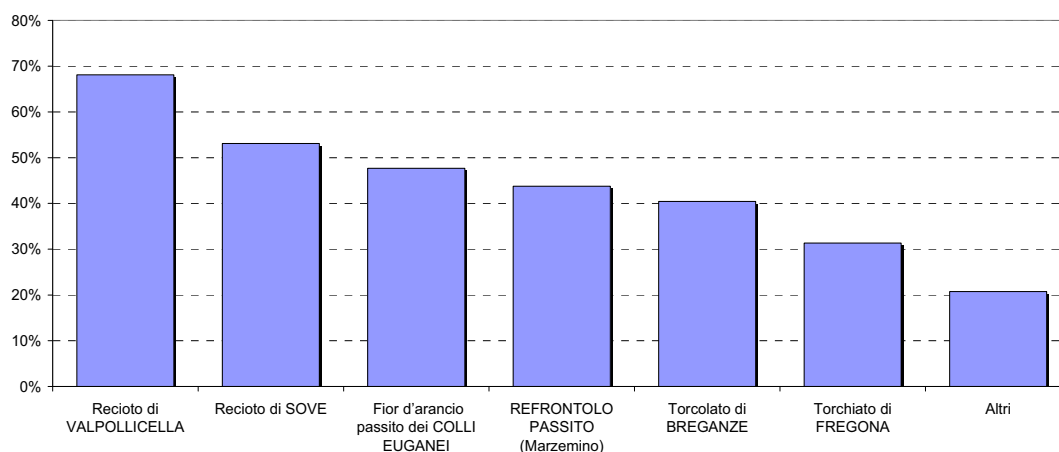
preferenze, ma deve semplicemente fare una scelta tra *conosce/non conosce* (SI/NO) ciascun tipo di vino presentato.

Tabella 18. *Frequenza assoluta e percentuale di intervistati che conoscono i seguenti tipi di passiti del Veneto*

Quali dei seguenti passiti del Veneto conosce?	N	%
Recioto di VALPOLLICELLA	263	68%
Recioto di SOVE	205	53%
Fior d'arancio passito dei COLLI EUGANEI	184	48%
REFRONTOLO PASSITO (Marzemino)	169	44%
Torcolato di BREGANZE	156	40%
Torchiato di FREGONA	121	31%
Altri	80	21%

La Figura 9 dà un'immagine chiara di quanto siano noti i tipi di vino presentati al nostro campione. Il più conosciuto è il Recioto di Valpolicella. Una percentuale del 70% del nostro campione, l'ha bevuto o ne ha sentito parlare. È seguito dal Recioto di SOAVE, dal Fior d'arancio passito dei Colli Euganei, dal Refrontolo Passito (Marzemino) e dal Torcolato di Fregona, che sono all'incirca conosciuti dalla metà del campione. Ben 80 persone hanno dichiarato di conoscere altri tipi di vino passito (etichetta *altri*, sia in Tabella 18 che in Figura 9). Tra questi i più noti sono Amarone, Vin Santo, Pantelleria.

Figura 9. *Percentuale degli intervistati che conoscono alcuni tipi di vino passito del Veneto*



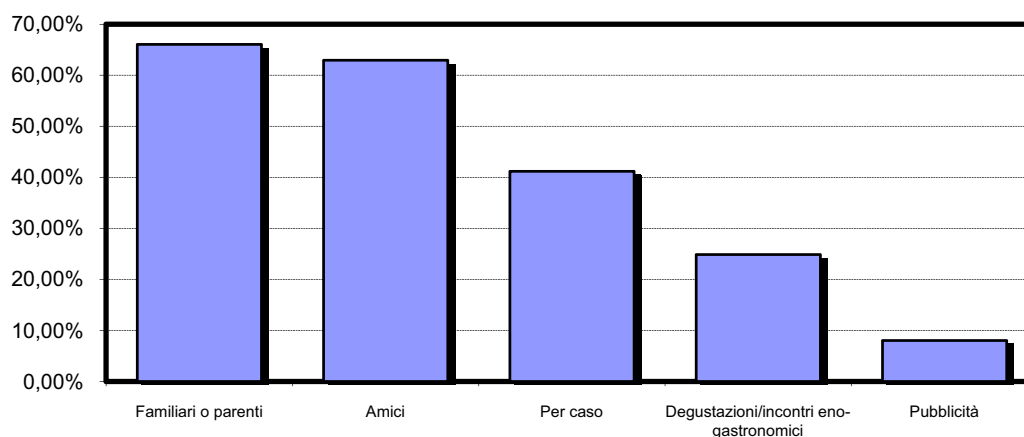
Una volta raccolti i nomi dei vini passiti noti al nostro campione, si è indagato su come ne siano venuti a conoscenza. Anche in questo caso l'intervistato doveva semplicemente rispondere con un SI/NO. In Tabella 19, sono riportate cinque delle principali fonti che portano alla conoscenza del vino.

Tabella 19. *Distribuzione percentuale di intervistati che hanno conosciuto il vino passito attraverso i familiari, amici, per caso, grazie a degustazioni o tramite la pubblicità*

In che modo ha conosciuto il vino passito?	%SI	% NO	% Non Risposto	Totale
Familiari o parenti	66.06%	27.46%	6.48%	100.00%
Amici	62.95%	30.57%	6.48%	100.00%
Per caso	41.19%	50.78%	8.03%	100.00%
Degustazioni/incontri eno-gastronomici	24.87%	67.88%	7.25%	100.00%
Pubblicità	8.03%	86.27%	5.70%	100.00%

Come la tabella precedente, anche la Figura 7, mostra che più del 60% degli intervistati ha conosciuto il vino passito, grazie a familiari e ad amici, il 40% per caso e il 32.9% tramite degustazioni/incontri eno-gastronomici e pubblicità.

Figura 10. *Percentuale di intervistati che hanno conosciuto il vino passito attraverso familiari, amici, per caso, grazie a degustazioni o tramite la pubblicità.*



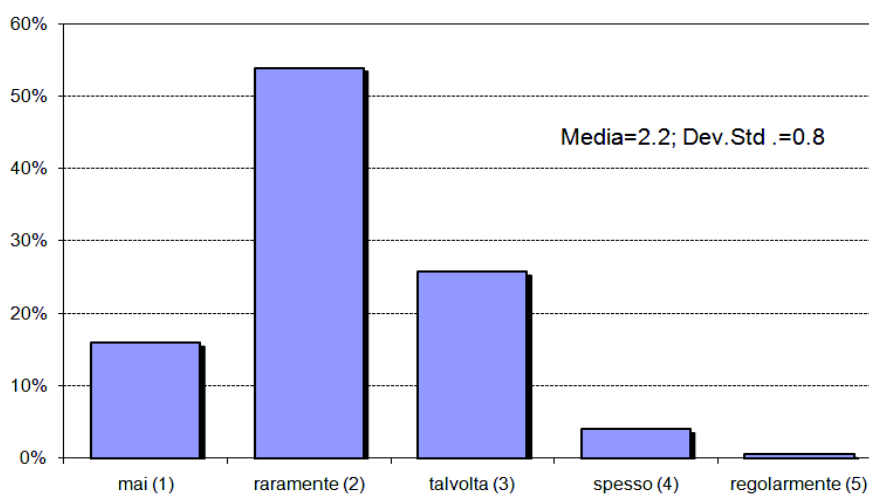
Alla domanda “*Quanto frequentemente beve vino passito?*” più del 50% del campione ha dichiarato di berlo solo *raramente*, giudizio che corrisponde al valore 2 nella scala di preferenze che va da 1 (= *mai*) a 5 (= *regolarmente*). Questo comporta che anche il valore medio sia basso, e precisamente pari a 2.2. Inoltre, va evidenziato, che a questa domanda gli intervistati hanno risposto con una certa omogeneità e ciò si riscontra non solo graficamente, come si può vedere in Figura 11, ma anche dal valore della deviazione standard, che è pari a 0.8.

Tabella 20. *Distribuzione percentuale di intervistati per frequenza con cui bevono vino passito*

Quanto frequentemente beve vino passito?	%SI
mai (1)	15.54%
raramente (2)	52.59%
talvolta (3)	25.13%
spesso (4)	3.89%
regolarmente (5)	0.52%
Non Risposto	2.33%
Totale	100%
Media	2.2
Dev. Std.	0.8

La Figura 11 evidenzia chiaramente che più della metà del campione beve vino passito *raramente*, che una percentuale molto bassa lo beve *spesso* o *regolarmente* e che, in generale, non è una bevanda che viene presa abitualmente.

Figura 11. *Percentuale di intervistati per frequenza con cui bevono vino passito*



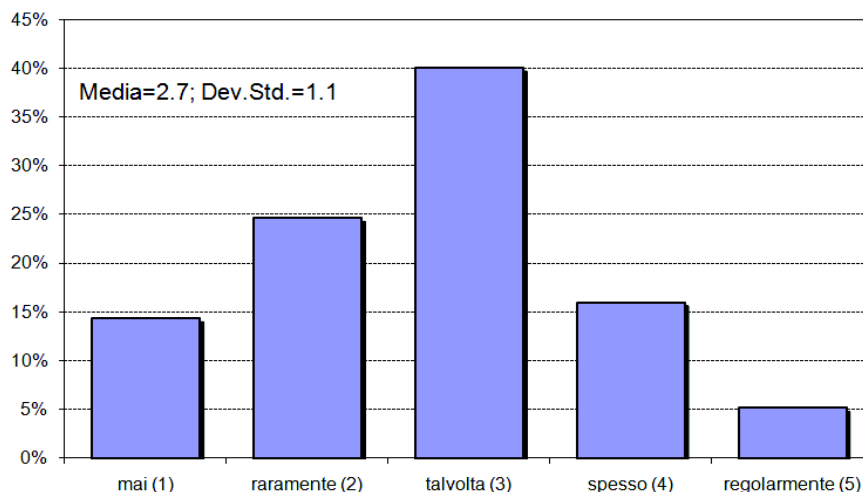
Altra informazione utile è conoscere l'occasione in cui le persone bevono vino passito. Quindi la prima domanda che poniamo al campione è con che frequenza lo beva durante le ricorrenze, come ad esempio Natale, Pasqua, cerimonie in generale. Dalla Tabella 21 emerge che il 40% degli intervistati lo prende *talvolta*, del restante 60% la maggioranza non lo prende quasi mai, mentre una bassa percentuale lo prende spesso. La media delle risposte risulta essere 2.7, molto vicino al valore 3 (= *talvolta*); lo scarto quadratico medio è pari a 1.1, dovuto al picco di preferenze per il punteggio 3.

Tabella 21. Distribuzione percentuale di intervistati per frequenza con cui bevono vino passito durante ricorrenze, come Natale, Pasqua, matrimoni, ecc.

Ricorrenze (Natale, Pasqua, matrimoni)	%SI
mai (1)	13.73%
raramente (2)	23.58%
talvolta (3)	38.34%
spesso (4)	15.28%
regolarmente (5)	4.92%
Non Risposto	4.15%
Totale	100%
Media	2.7
Dev. Std.	1.1

Il diagramma a barre in Figura 12 evidenzia graficamente che le preferenze si concentrano in corrispondenza del valore 3, e che la restante parte del campione ha dato valutazioni basse piuttosto che alte.

Figura 12. Percentuale di intervistati per frequenza con cui bevono vino passito durante ricorrenze, come Natale, Pasqua, matrimoni, ecc.



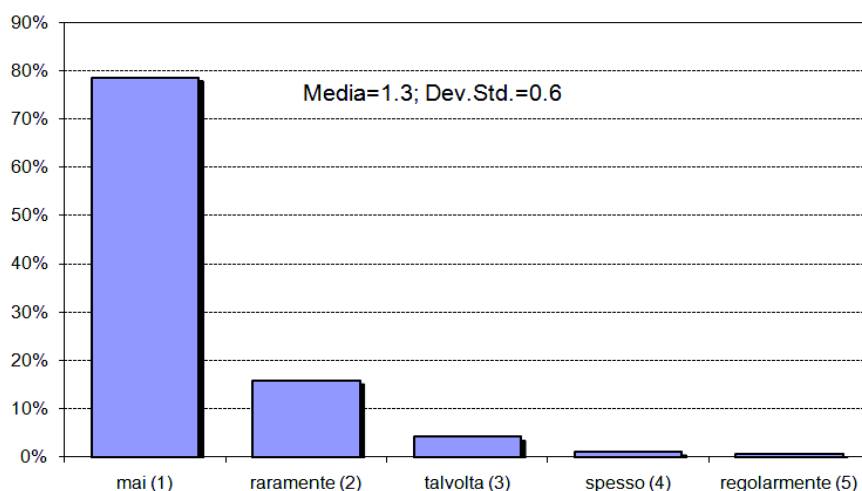
Altra domanda che è stata fatta riguarda la frequenza con cui bevono vino passito da soli, semplicemente quando ne hanno voglia. Come è logico immaginare la maggior parte delle persone, precisamente il 74.61% del campione, dichiara che ciò non accade *mai*, della restante parte degli intervistati il 15.03% lo prende *raramente* da solo e soltanto ad una piccola percentuale di persone capita di berlo ogni volta che lo desidera, 4.92% non risponde alla domanda. Ovviamente questo comporta che anche la media sia spostata verso i valori più bassi della nostra scala di valutazione, in questo caso è pari a 1.3, molto vicina alla scelta *mai*. Data l'omogeneità delle risposte, la deviazione standard risulta essere 0.6.

Tabella 22. *Distribuzione percentuale di intervistati per frequenza con cui bevono vino passito ogni volta lo desiderano, a prescindere dalle occasioni o ricorrenze particolari*

Da solo quando ne ho voglia	Perc.SI
mai (1)	74.61%
raramente (2)	15.03%
talvolta (3)	3.89%
spesso (4)	1.04%
regolarmente (5)	0.52%
Non Risposto	4.92%
Totale	100%
Media	1.3
Dev. Std.	0.6

Il diagramma a barre in Figura 13 evidenzia il picco di preferenze sul valore *mai* e percentuali molto basse per gli altri punteggi, e questo sottolinea ancora una volta che il campione generalmente non beve vino passito.

Figura 13. *Percentuale di intervistati per frequenza con cui bevono vino passito quando lo desiderano*



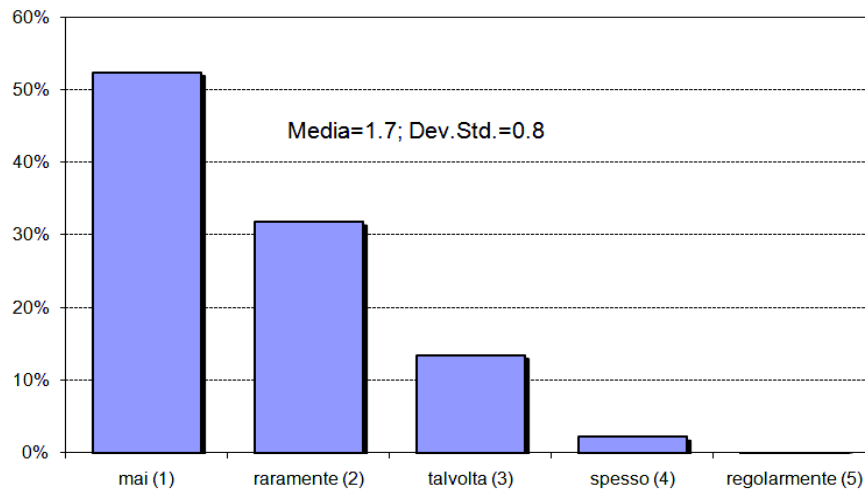
Si è poi chiesto agli intervistati di definire, anche in questo caso, con un punteggio da 1 (= *mai*) a 5 (= *regolarmente*) la frequenza con cui gli capita di prendere del vino passito alla fine dei pasti. Come si vede in Tabella 23, è emerso che a metà del campione non capita *mai*, della restante parte, circa ad un 30% capita *raramente* e ad un 13% *talvolta*. Anche in questo caso i punteggi elevati sono stati pochissimi, quindi anche la media risulta molto bassa, e la deviazione standard, pari a 0.8, evidenzia la tendenza delle risposte verso un'unica preferenza.

Tabella 23. *Distribuzione percentuale di intervistati per frequenza con cui bevono vino passito alla fine dei pasti*

Alla fine dei pasti	Perc.SI
mai (1)	49.48%
raramente (2)	30.05%
talvolta (3)	12.69%
spesso (4)	2.07%
regolarmente (5)	0.26%
Non Risposto	5.44%
Totale	100%
Media	1.7
Dev. Std.	0.8

Quanto detto si riscontra anche in Figura 14, in cui si nota l'andamento decrescente del diagramma a barre man mano che si passa da valori bassi della scala di preferenza a valori alti, con un picco in corrispondenza del punteggio 1 e un minimo per il punteggio 5.

Figura 14. *Percentuale di intervistati per frequenza con cui bevono vino passito alla fine dei pasti*



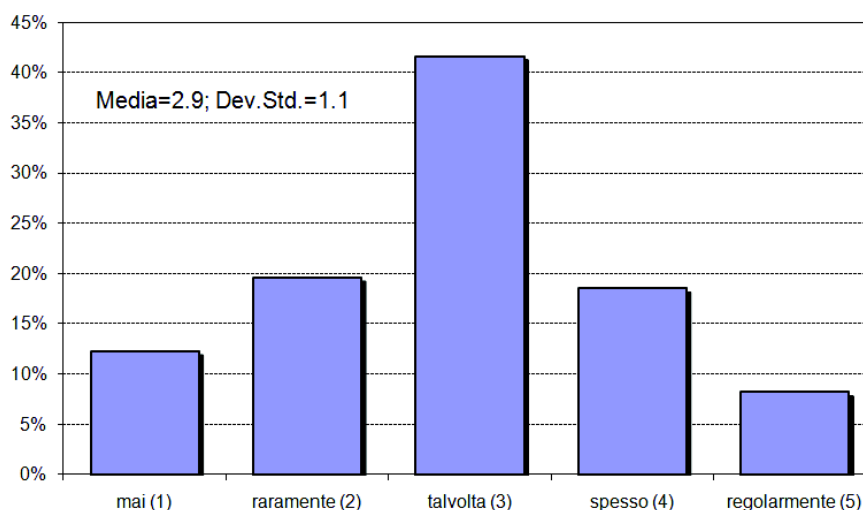
Infine la Tabella 14 riporta la frequenza con cui il campione dichiara di bere vino passito quando gli viene offerto o capita l'occasione. La distribuzione risulta leggermente più eterogenea rispetto a prima, come confermato dallo scarto quadratico medio, pari a 1.1, mentre la media, con valore 2.9, si sposta verso valori più centrali della scala delle preferenze, evidenziando come siano stati dati punteggi più elevati.

Tabella 24. *Distribuzione percentuale di intervistati per frequenza con cui bevono vino passito, perché viene loro offerto o perché capita l'occasione*

	Perc.SI
mai (1)	11.66%
raramente (2)	18.65%
talvolta (3)	39.64%
spesso (4)	17.62%
regolarmente (5)	7.77%
Non Risposto	4.66%
Totale	100%
Media	2.9
Dev. Std.	1.1

Come si può vedere in Figura 15, il diagramma a barre può sembrare quasi perfettamente simmetrico, infatti, media e mediana tendono a coincidere, con valore pari a 2.9 la prima e 3 la seconda. Questo è dovuto solamente ad una coincidenza, resta vero il fatto, però, che la percentuale di persone che ha espresso punteggi positivi (= 4, 5) e negativi (= 1, 2) è quasi la stessa. In generale possiamo dire che le persone intervistate hanno manifestato la tendenza a bere vino passito qualora gli venga offerto o durante ricorrenze particolari.

Figura 15. *Percentuale di intervistati per frequenza con cui bevono vino passito, perché viene loro offerto o perché capita l'occasione*



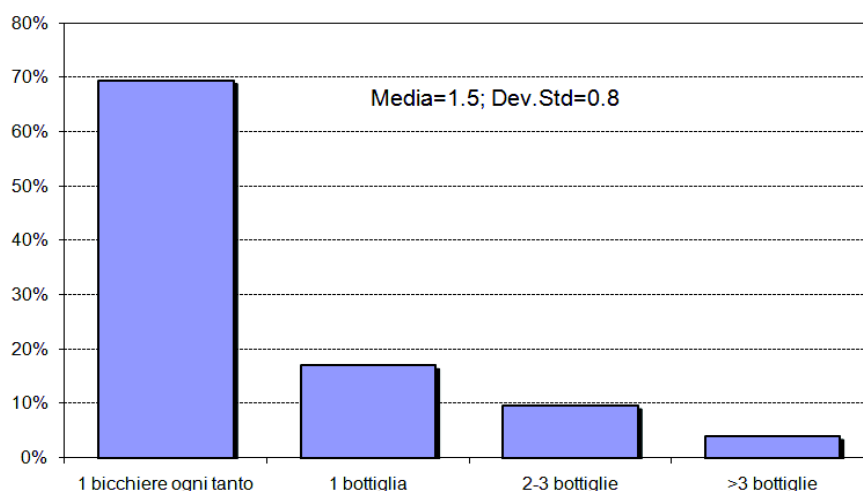
Si è poi chiesto agli intervistati quanto vino passito bevano durante l'anno, presentando quattro diverse opzioni tra cui scegliere, come mostrato in Tabella 25, ed il risultato è stato che circa il 64% ha dichiarato di bere un bicchiere ogni tanto, mentre solo il 15% ne beve una bottiglia e il restante campione beve più di 2 bottiglie l'anno. Questo comporta che la media si stabilizzi verso valori bassi (= *un bicchiere ogni tanto*) della nostra scala. Mentre lo scarto quadratico medio assume un valore pari a 0.8, e ciò dimostra una certa omogeneità nei risultati.

Tabella 25. *Distribuzione percentuale di intervistati per quantità di vino passito bevuto all'anno*

Quanto vino passito beve all'anno?	Perc.SI
1 bicchiere ogni tanto	63.47%
1 bottiglia	15.54%
2-3 bottiglie	8.81%
>3 bottiglie	3.63%
Non Risposto	8.55%
Totale	100%
Media	1.5
Dev. Std.	0.8

Il diagramma a barre in Figura 16 rappresenta perfettamente l'omogeneità delle risposte, con un picco per l'opzione "un bicchiere ogni tanto", e valori molto bassi per le restanti alternative. La percentuale di persone che ha dato punteggi più alti diminuisce man mano che aumenta la quantità di vino bevuta all'anno.

Figura 16. *Percentuale di intervistati per quantità di vino passito bevuto all'anno*



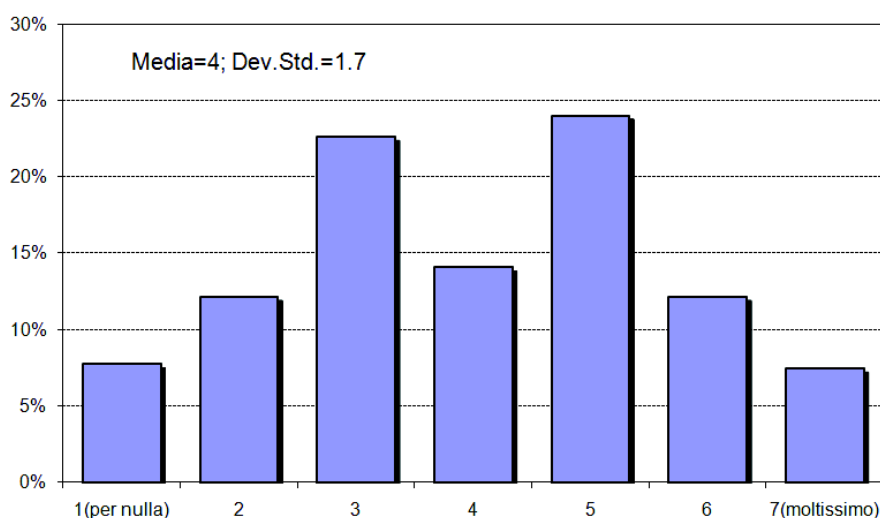
La domanda che viene fatta successivamente riguarda il gradimento del vino passito, gli intervistati hanno espresso la loro preferenza attraverso un punteggio, facendo riferimento a una scala di valutazione che va da 1 (= *per nulla*) a 7 (= *moltissimo*), come fatto per alcune domande precedentemente. Osservando il valore assunto dalla deviazione standard, si nota subito, come le risposte date siano abbastanza eterogenee, analizzando poi i valori percentuali, sempre in Tabella 26, si riscontra quanto detto, ossia che non vi è un punteggio che prevale nettamente sugli altri. Guardando poi la media, che è pari a 4.0, si può concludere che al nostro campione mediamente piace bere vino passito.

Tabella 26. Distribuzione percentuale di intervistati per gradimento del vino passito, data una scala di valutazione da 1 a 7

Quanto gradisce bere vino passito?	%SI
1(per nulla)	7.25%
2	11.40%
3	21.24%
4	13.21%
5	22.54%
6	11.40%
7(moltissimo)	6.99%
Non Risposto	5.96%
Totale	100%
Media	4.0
Dev. Std.	1.7

La Figura 17 mostra una distribuzione abbastanza varia, in cui le uniche risposte che prevalgono sono la 3 e la 5, ma con una percentuale modesta.

Figura 17. Percentuale di intervistati per gradimento del vino passito



Da queste informazioni potremmo dire che al nostro campione mediamente piace bere vino passito. È importante evidenziare che dal confronto di questo diagramma a barre con quello riportato in Figura 7, che riguarda il vino in generale, emerge che l'indice di gradimento è inferiore rispetto al precedente. La stessa media è passata dal valore 4 a 4.4.

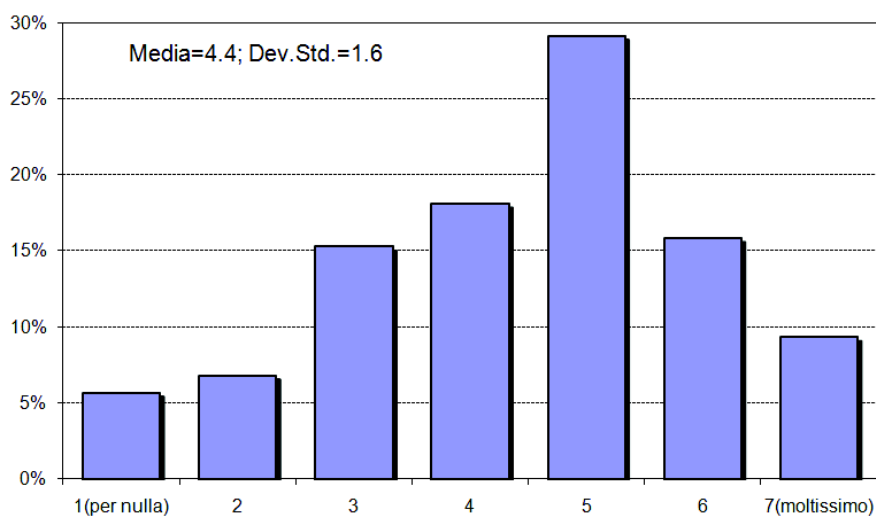
Ci si chiede, ora, quali siano le caratteristiche del vino passito che soddisfano maggiormente il cliente e per prima cosa il nostro campione si trova a giudicarne il profilo aromatico. Dalle informazioni riportate in Tabella 27 emerge che questo aspetto non è gradito allo stesso modo da tutti gli intervistati, portando ad una certa eterogeneità delle risposte, come dimostrato anche dallo scarto quadratico medio. Solo un 9% lo apprezza pienamente, un 30% si ritiene più che soddisfatto e solo una minima parte non lo è per nulla.

Tabella 27. *Distribuzione percentuale di intervistati per il loro grado di soddisfazione riguardo il profilo aromatico del vino passito, data una scala di valutazione da 1 a 7*

il profilo aromatico (aroma e profumo)	%SI
1(per nulla)	5.18%
2	6.22%
3	13.99%
4	16.58%
5	26.68%
6	14.51%
7(moltissimo)	8.55%
Non Risposto	8.29%
Totale	100%
Media	4.4
Dev. Std.	1.6

Il diagramma a barre in Figura 18 evidenzia la diversità delle risposte date, e allo stesso tempo mostra che la maggior parte delle persone ha dato dei punteggi positivi.

Figura 18. *Percentuale di intervistati per il loro grado di soddisfazione riguardo il profilo aromatico del vino passito*



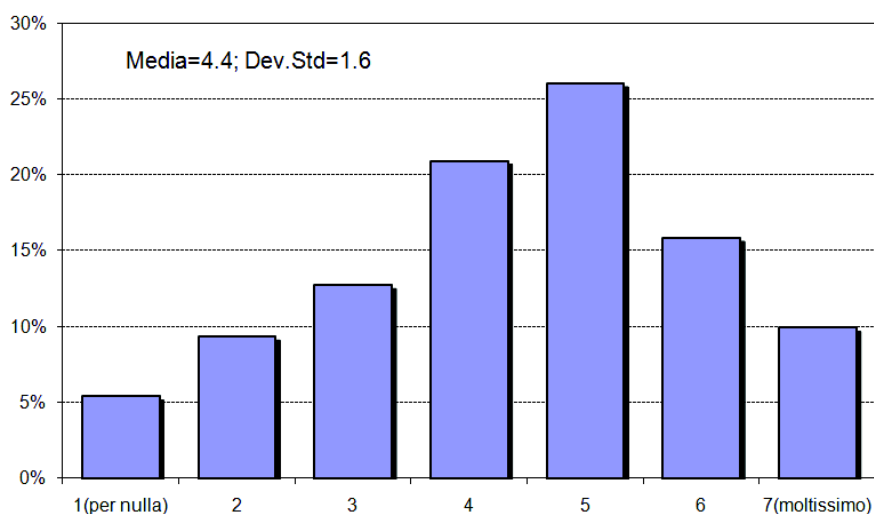
Secondo aspetto del passito che il campione deve giudicare è il suo gusto dolce. Anche in questo caso le risposte sono eterogenee, con un gradimento del 26% del campione sul valore 5. Complessivamente i giudizi sono positivi. Media e deviazione standard coincidono con le precedenti.

Tabella 28. *Distribuzione percentuale di intervistati per grado di soddisfazione riguardo il gusto dolce del vino passito, data una scala di valutazione da 1 a 7*

il gusto dolce	%SI
1(per nulla)	4.92%
2	8.55%
3	11.66%
4	19.17%
5	23.83%
6	14.51%
7(moltissimo)	9.07%
Non Risposto	8.29%
Totale	100%
Media	4.4
Dev. Std.	1.6

Il diagramma a barre in Figura 19 mostra, anche in questo caso, un comportamento non omogeneo delle risposte e la mancanza di un picco evidente di preferenze. La distribuzione risulta asimmetrica a sinistra.

Figura 19. *Percentuale di intervistati per grado di soddisfazione riguardo il gusto dolce del vino passito*



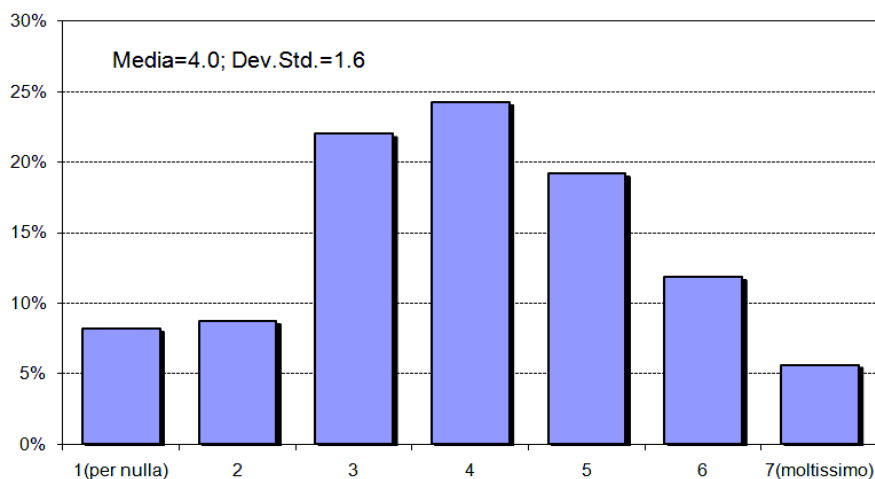
In Tabella 29 troviamo la distribuzione percentuale degli intervistati per grado di soddisfazione relativamente alla gradazione alcolica del vino passito, altro aspetto che il campione deve giudicare, oltre al profilo aromatico e al gusto dolce. Come nei casi precedenti, le risposte risultano eterogenee, con deviazione standard 1.6 e media pari a 4.0.

Tabella 29. *Distribuzione percentuale di intervistati per grado di soddisfazione riguardo la gradazione alcolica del vino passito, data una scala di valutazione da 1 a 7*

la gradazione alcolica	Perc. SI
1(per nulla)	7,51%
2	8,03%
3	20,21%
4	22,28%
5	17,62%
6	10,88%
7(moltissimo)	5,18%
Non Risposto	8,29%
Totale	100%
Media	4,0
Dev. Std.	1,6

La Figura 20 mostra graficamente quanto riportato nella tabella precedente. Sull'asse delle ascisse troviamo la scala delle preferenze da 1 a 7, sulle ordinate la percentuale di persone che ha espresso un determinato giudizio. I picchi più alti si riscontrano per i punteggi 3 e 4, con un 22% di persone che non si sente pienamente soddisfatta della gradazione alcolica di questo vino e un 24% che la ritiene sufficientemente adeguata alle proprie esigenze.

Figura 20. *Percentuale di intervistati per grado di soddisfazione riguardo la gradazione alcolica del vino passito*



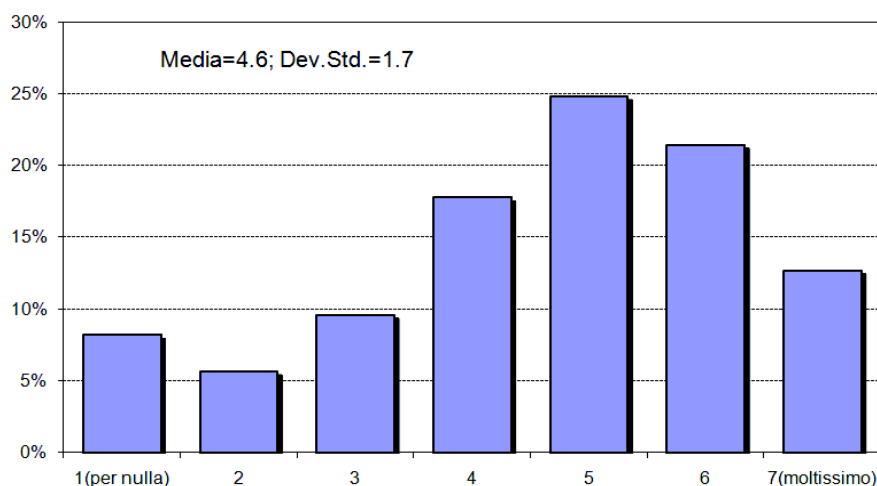
Un'ultima domanda che viene fatta relativamente agli aspetti del vino passito che possono soddisfare o meno il cliente riguarda l'intensità del gusto. A differenza dei giudizi precedenti, gli intervistati hanno espresso valutazioni mediamente più elevate, con una percentuale del 25% per il punteggio 5 e un 21% per il 6. La deviazione standard sottolinea, però, che c'è sempre una certa disomogeneità nelle risposte.

Tabella 30. *Distribuzione percentuale di intervistati per grado di soddisfazione riguardo l'intensità del gusto del vino passito, data una scala di valutazione da 1 a 7*

l'intensità del gusto	Perc. SI
1(per nulla)	7.51%
2	5.18%
3	8.81%
4	16.32%
5	22.80%
6	19.69%
7(moltissimo)	11.66%
Non Risposto	8.03%
Totale	100%
Media	4.6
Dev. Std.	1.7

Il diagramma a barre in Figura 21 mostra chiaramente che a gran parte degli intervistati piace l'intensità del gusto del vino passito, e da un confronto con la Figura 18, la Figura 19 e la Figura 20, emerge che è l'aspetto che più soddisfa il nostro campione.

Figura 21. *Percentuale di intervistati per grado di soddisfazione riguardo l'intensità del gusto del vino passito*



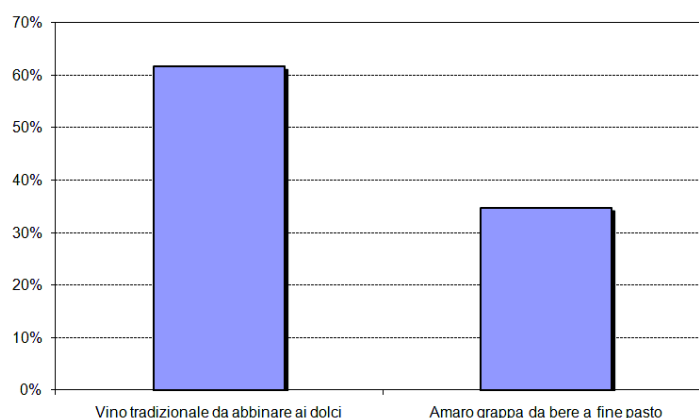
Successivamente si chiede agli intervistati se il vino passito possa essere un'alternativa al tradizionale vino bevuto con i dolci o all'amaro/grappa. Stavolta il campione deve esprimersi con un SI (= è un vino da abbinare ai dolci/ è un'alternativa alla all'amaro, alla grappa)/ NO (= non è un vino da abbinare ai dolci/ non è un'alternativa alla all'amaro, alla grappa). Dalla Tabella 21 emerge che 238 persone su 386, quindi il 63% del campione lo ritiene un vino da abbinare ai dolci e 134 persone su 368 lo sostituirebbero all'amaro/grappa.

Tabella 31. *Distribuzione percentuale degli intervistati che ritengono il vino passito un'alternativa al vino tradizionale da dolce o all'amaro/grappa*

Il vino passito può essere un'alternativa a:	Perc. SI	Perc.NO%	
Vino tradizionale da abbinare ai dolci	63.73%	36.27%	100.00%
Amaro/grappa da bere a fine pasto	38.34%	61.66%	100.00%

Quanto detto si riscontra la Figura 22, dove si vede chiaramente che il vino passito, nella mente del consumatore occupa una posizione di mercato che lo vede molto vicino ai tradizionali vini da dolce, raccogliendo una percentuale di preferenze che supera del 27% la percentuale di coloro che lo definiscono un vino da sostituire alla grappa.

Figura 22. *Percentuale degli intervistati che ritengono il vino passito un'alternativa al vino tradizionale da dolce o all'amaro/grappa*



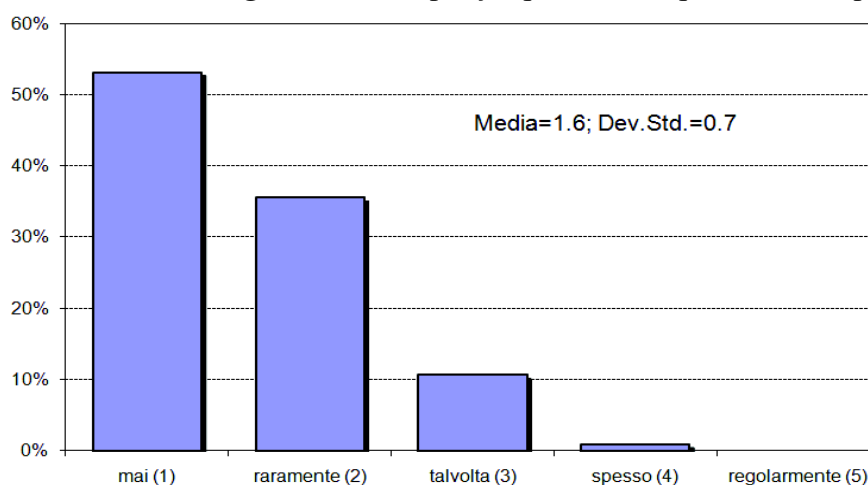
In Tabella 32 troviamo la frequenza con cui gli intervistati acquistano vino passito e il risultato è in accordo con quanto ricavato in Tabelle 10, cioè che la maggioranza del campione non acquista mai questo vino. Va notato anche che nessuna persona ha dichiarato di comprarlo regolarmente e che solo l'1% lo acquista spesso.

Tabella 32. *Distribuzione percentuale degli intervistati per frequenza d'acquisto di vino passito*

Acquista bottiglie di vino passito?	Perc. SI
mai (1)	61.66%
raramente (2)	34.72%
talvolta (3)	3.63%
spesso (4)	61.66%
regolarmente (5)	34.72%
Non Risposto	3.63%
Totale	100%
Media	1.6
Dev. Std.	0.7

Il diagramma a barre in Figura 23 mostra l'andamento decrescente della percentuale delle persone che acquista il passito, passando da una percentuale del 53% per il valore 1 (= *mai*), a una dello 0% per il valore 5 (= *regolarmente*).

Figura 23. *Percentuale degli intervistati per frequenza d'acquisto di vino passito*



Infine è stato chiesto al campione quanto è disposto a pagare per una bottiglia di vino passito da 0.5 litri. Da questa indagine è emerso che il 46% potrebbe spendere tra i 10-15 €, il 38% fino a 10 €, il 12% tra i 15-20 € e quasi nessuno più di 20 €. Il valore della deviazione standard sottolinea una certa omogeneità nelle risposte e la media la tendenza ad acquistare bottiglie che non superino i 15 €.

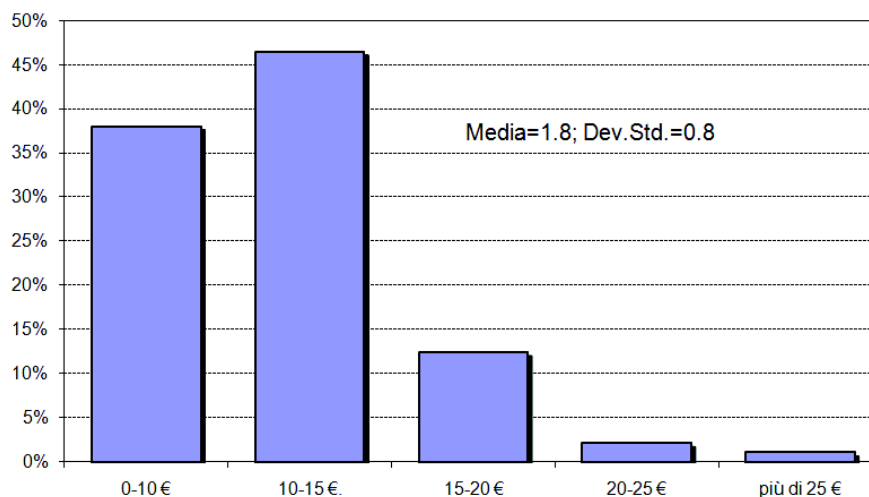
Tabella 33. *Distribuzione percentuale di intervistati per classi di prezzo di vino passito, riferito a una bottiglia da 0,5l*

Quanto è disposto a pagare e per una bottiglia di vino passito? (da 0,5 litri)	Perc. SI
0-10 €	46.89%
10-15 €.	31.35%
15-20 €	9.33%
20-25 €	0.78%
più di 25 €	0.00%
Non Risposto	11.66%
Totale	100%
Media	1.8
Dev. Std.	0.8

Nel diagramma a barre in Figura 24 si ritrova in forma grafica i risultati della tabella precedente. Sull'asse delle ascisse abbiamo le cinque alternative di prezzo poste al cliente e sulle ordinate la percentuale delle persone che ha espresso un preciso giudizio. Come evidenziato prima, si riscontra un picco di preferenze per l'alternativa 10-15 € per

l'acquisto di mezzo litro di vino passito e una percentuale molto bassa di persone che è disposta a spendere più di 20 €.

Figura 24. Percentuale di intervistati per classi di prezzo di vino passito, riferito a una bottiglia da 0,5l



4.4 Analisi CUB versione parametrica

Nel presente paragrafo, applichiamo ai nostri dati i modelli introdotti nel capitolo 2 cioè la loro versione parametrica. Nel Veneto Orientale, abbiamo condotto un'analisi backwise partendo dal modello saturo CUB(3,3), e scegliendo di applicare quelli più semplici solo per le covariate significative per ξ e π .

Le covariate sono state dicotomizzate nel seguente modo:

$$\text{Età:} = \begin{cases} 0 & \text{se } \leq 25 \\ 1 & \text{se } > 25 \end{cases}$$

$$\text{Sesso:} = \begin{cases} 0 & \text{Se femmina} \\ 1 & \text{Se maschio} \end{cases}$$

$$\text{Provincia di residenza:} = \begin{cases} 0 & \text{Se residente nel veneto occidentale} \\ 1 & \text{Se residente nel veneto orientale} \end{cases}$$

I modelli sono stati applicati solamente per le domande ritenute più interessanti, quali: “Gradimento del vino passito” (GRAD_VIN), “Conoscenza del vino” (CONOSCE), “Frequenza con cui beve il vino passito” (FREQ), “Gradimento del vino passito” (GRAD_PAS), “Soddisfazione del profilo aromatico del vino passito” (SODD_1), “Soddisfazione del gusto dolce del vino passito” (SOD_2), “Soddisfazione della gradazione alcolica del vino passito” (SOD_3), “Soddisfazione dell'intensità del gusto del vino passito” (SOD_4), “Frequenza d'acquisto del vino passito” (ACQU), “Quanto è disposto a pagare per una bottiglia di vino passito” (PAGA).

Il livello di significatività α è stato fissato pari a 0.10. Sia l'output del modello saturo che quelli dei modelli ritenuti abbastanza rilevanti sono rapportati in appendice.

4.4.1 Modello saturo CUB(3,3)

Come è già stato detto, la prima analisi viene fatta utilizzando il modello CUB(3,3), ovvero il modello con tutte le covariate sia per il grado de *feeling* che per l'*incertezza*. Dall'output ottenuto, abbiamo dei coefficienti gamma e beta per l'età, il sesso e la provincia di residenza, con queste stime calcoliamo le stime dei parametri ξ e π , ricordando che:

- le componenti sistematiche del modello sono:

$$\begin{cases} \pi_i = \frac{1}{1 + e^{-x_i\beta}} \\ \xi_i = \frac{1}{1 + e^{-w_i\gamma}} \end{cases} \quad \begin{array}{l} \text{dove le } x_i \text{ e } w_i \text{ sono le covariate per } \pi, \\ \xi, \text{rispettivamente} \end{array}$$

- e che:

$$\text{grado di feeling} = 1 - \xi \qquad \text{Grado di incertezza} = \frac{1-\pi}{m}$$

a- *Gradimento del vino*

Per la domanda relativa al gradimento del vino in generale, dal modello CUB (3,3), si hanno due covariate significative (livello di significatività pari a 0.10) come illustrato in Tabella 24 la quale contiene i valori per le stime e i p-value; una per la componente di π in particolare l'età degli intervistati il p-value significativo è pari a 0.053, e per la componente di feeling in particolare il loro sesso il p-value è 0.000. Nel caso della covariata età sull'incertezza, la stima del parametro β_1 ha un valore positivo, quindi al crescere della categoria, cioè passando dai più giovani ai meno giovani, π cresce, e quindi applicando la formula il grado di incertezza diminuisce. Per quanto riguarda il sesso sul feeling, la stima del parametro γ_2 è negativa, cioè al passare dal sesso 0 (femmine) al sesso 1 (maschi), ξ decresce, e di conseguenza, il grado di feeling è più basso per le femmine che per i maschi.

Tabella 34. *Stime dei parametri e p-value*

Covariate	π				ξ			
	Età	Sesso	Residenza		Età	Sesso	Residenza	
Parametri	β_0	β_1	β_2	β_3	γ_0	γ_1	γ_2	γ_3
Stime dei parametri	-1.298	0.982	0.541	-0.590	0.969	0.328	-1.112	-0.187
p-value	0.365	0.053	0.306	0.250	0.102	0.145	0.000	0.347

Tabella 35 a: *Grado di incertezza ricavato dal modello CUB(3,3) gradimento del vino*

età		
	≤ 25	> 25
stima di π	0.214	0.421
grado incertezza	0.112	0.082

Tabella 36 b *Grado di di feeling ricavato dal modello CUB(3,3) gradimento del vino*

Sesso		
	F	M
stima di ξ	0.724	0.464
grado di feeling	0.275	0.535

Da questa prima analisi, possiamo dire che col passare degli anni, le persone diventano più consapevoli di quanto gradiscono il vino, in più si nota che negli uomini un livello di gradimento superiore che per le donne.

b- *Conoscenza del vino passito*

Per quanto riguarda la conoscenza del vino passito, la componente di incertezza risulta non significativa per tutte e tre le covariate, mentre per il feeling sia la residenza che il sesso hanno p-value significativi, rispettivamente 0.035 e 0.000. Nella Tabella 37, sono riassunti i calcoli delle stime di ξ e della componente di feeling. Si nota che le donne sia nel veneto occidentale che in quello orientale hanno un feeling inferiore a quello degli uomini in termini di conoscenza del vino passito, in effetti ha una stima negativa.

Tabella 37. *Stime dei parametri e p-value*

Covariate	π				ξ			
	Età	Sesso	Residenza		Età	Sesso	Residenza	
Parametri	β_0	β_1	β_2	β_3	γ_0	γ_1	γ_2	γ_3
stime dei parametri	-1.213	0.849	-0.589	0.234	1.357	-0.455	-0.890	0.575
p-value	0.470	0.164	0.298	0.699	0.164	0.195	0.000	0.035

Tabella 38 *grado di feeling per la conoscenza del passito*

Sesso		Residenza	
		VENETO OCCIDENTALE	VENETO ORIENTALE
F	Stima di ξ	0.794	0.872
	grado feeling	0.205	0.127
M	Stima di ξ	0.613	0.737
	grado feeling	0.386	0.262

c- *Gradimento del vino passito*

Le covariate significative qui sono:

- Per l'incertezza la residenza con un p-value pari 0.095 e segno del parametro negativo -0.95, ma la differenza è poco rilevante in una zona o nell'altra in termine di incertezza.
- Per il feeling, Il sesso (con p-value pari a 0.0001) e la residenza (con p-value pari 0.05). Come visto in precedenza per il gradimento del vino, passando dalle femmine ai maschi, il feeling cresce, e invece esso diminuisce passando dal veneto occidentale a quello orientale (vedi tabelle 39).

Tabella 39 *Stime dei parametri e p-value*

Covariate	π				ξ			
	Età	Sesso	Residenza		Età	Sesso	Residenza	
Parametri	β_0	β_1	β_2	β_3	γ_0	γ_1	γ_2	γ_3
stime dei parametri	3.198	-0.340	-0.426	-0.950	0.335	0.090	-0.570	0.288
p-value	0.069	0.540	0.449	0.095	0.419	0.551	0.000	0.059

Tabella 40.a *Grado di feeling del gradimento del passito*

Sesso		Residenza	
		VENETO OCCIDENTALE	VENETO ORIENTALE
F	stima di ξ	0.581	0.649
	grado di feeling	0.418	0.350
M	stima di ξ	0.440	0.511
	grado di feeling	0.559	0.488

Tabella 41.b *Grado di incertezza e del gradimento del passito*

	Residenza	
	Veneto Occidentale	Veneto Orientale
stima di π	0.960	0.903
grado di incertezza	0.005	0.0137

d- *Soddisfazione del profilo aromatico del vino passito*

In questo caso, per l'incertezza tutte e tre le covariate sono significative (p-value pari a 0.07 per l'età, 0.06 per il sesso, 0.02 per la residenza). L'incertezza sull'aroma del passito decresce al crescere dell'età, decresce da femmine a maschi, e cresce da ovest a est. Per le covariate del feeling invece, risultano significative sesso, il feeling è maggiore per i maschi a femmine. Quindi sono meno incerti i maschi di età maggiore a 25 anni sul profilo aromatico e hanno un migliore feeling i maschi del veneto orientale.

Tabella 42 *Stime dei parametri e p-value*

Covariate	π				ξ			
	Età	Sesso	Residenza		Età	Sesso	Residenza	
Parametri	β_0	β_1	β_2	β_3	γ_0	γ_1	γ_2	γ_3
stime dei parametri	-0.073	1.153	-1.485	-0.590	0.142	0.039	-0.380	-0.040
p-value	0.964	0.073	0.060	0.027	0.696	0.767	0.005	0.762

Tabella 43.a *Grado di feeling della soddisfazione del profilo aromatico*

	Residenza	
	Femmine	Maschi
stima di ξ	0.535	0.525
grado di feeling	0.464	0.474
stima di ξ	0.440	0.430
grado di feeling	0.559	0.569

Tabella 44.b *Grado di incertezza della soddisfazione del profilo aromatico*

età	sesso	Residenza		
		VENETO OCCIDENTALE	VENETO ORIENTALE	
≤ 25	F	stima di π	0.481	0.174
		grado di incertezza	0.074	0.117
	M	stima di π	0.742	0.396
		grado di incertezza	0.036	0.086
> 25	F	stima di π	0.745	0.400
		grado di incertezza	0.036	0.085
	M	stima di π	0.901	0.674
		grado di incertezza	0.014	0.046

e- *Soddisfazione del gusto dolce del vino passito*

L'età non è significativa né per l'incertezza né per il feeling. per quanto riguarda il sesso, l'incertezza sulla soddisfazione del gusto dolce del passito è inferiore per i maschi del veneto occidentale, mentre il feeling è superiore. Come riportato in Tabella 46.a e 46.b, si nota che l'incertezza è più alta nel veneto orientale mentre il feeling è più basso.

Tabella 45 *Stime dei parametri e p-value*

Covariate	π				ξ			
	Età	Sesso	Residenza		Età	Sesso	Residenza	Età
Parametri	β_0	β_1	β_2	β_3	γ_0	γ_1	γ_2	γ_3
stime dei parametri	1.415	0.593	1.045	-2.030	-0.620	0.337	-0.448	0.276
p-value	0.394	0.292	0.086	0.001	0.099	0.014	0.003	0.065

Tabella 46. *Grado dell'incertezza della soddisfazione del gusto dolce*

Sesso		Residenza	
		VENETO OCCIDENTALE	VENETO ORIENTALE
F	Stima di π	0.803	0.349
	grado di incertezza	0.028	0.092
M	Stima di π	0.920	0.604
	grado di incertezza	0.011	0.056

Tabella 47. *Grado del feeling della soddisfazione del gusto dolce*

Sesso		Residenza	
		VENETO OCCIDENTALE	VENETO ORIENTALE
F	stima di ξ	0.803	0.349
	grado di feeling	0.028	0.092
M	stima di ξ	0.920	0.604
	grado di feeling	0.011	0.056

f- *Soddisfazione della gradazione alcolica del vino passito*

Nella tabella 48, si nota subito che l'età non influisce per niente su incertezza e feeling per quanto riguarda la gradazione alcolica, mentre il sesso influisce solo sul feeling, il quale cresce da femmine a maschi, e infine la residenza influisce sia sull'incertezza che sul feeling. Nel primo caso, cresce quando si passa dal veneto occidentale a quello orientale, mentre il feeling ha un effetto contrario cioè decresce.

Tabella 48 *Stime dei parametri e p-value*

Covariate	π				ξ			
	Età	Sesso	Residenza		Età	Sesso	Residenza	
Parametri	β_0	β_1	β_2	β_3	γ_0	γ_1	γ_2	γ_3
stime dei parametri	1.988	0.523	0.524	-1.60	0.138	0.087	-0.434	0.293
p-value	0.268	0.373	0.381	0.021	0.697	0.480	0.001	0.023

Tabella 49. *Grado di incertezza della soddisfazione della gradazione alcolica*

	Residenza	
	VENETO OCCIDENTALE	VENETO ORIENTALE
Stima di π	0.878	0.593
grado di incertezza	0.017	0.058

Tabella 50. *Grado di feeling della soddisfazione della gradazione alcolica*

	Residenza	
	VENETO OCCIDENTALE	VENETO ORIENTALE
stima di ξ	0.532	0.603
grado di feeling	0.467	0.396
stima di ξ	0.425	0.497
grado di feeling	0.574	0.502

g- *Soddisfazione dell'intensità del gusto del vino passito*

Per questa risposta risulta significativo solamente la covariata sesso, la quale ha effetto sull'incertezza, che cala nei maschi rispetto alle femmine.

Tabella 51.a *Stima dei parametri e p-value*

Covariate	π				ξ			
	Età	Sesso	Residenza		Età	Sesso	Residenza	Età
Parametri	β_0	β_1	β_2	β_3	γ_0	γ_1	γ_2	γ_3
stime dei parametri	-1.975	0.177	1.768	-0.400	-0.732	-0.084	-0.110	0.180
p-value	0.1647	0.732	0.000	0.438	0.165	0.575	0.627	0.224

Tabella 51.b. *Grado di feeling della soddisfazione dell'intensità del gusto del vino*

	Sesso	
	F	M
stima di π	0.112	0.447
grado di incertezza	0.126	0.078

h- *Frequenza d'acquisto del vino passito*

I meno giovani hanno un feeling più elevato rispetto ai più giovani, mentre le donne hanno un feeling inferiore a quello dei maschi, quindi gli uomini più vecchi rispondono

alla frequenza di acquisto di vino passito con un migliore feeling, si nota che la zona di residenza è non significativa per incertezza e feeling (Tabella 53).

Tabella 52 *Stima dei parametri e p-value*

Covariate	π				ξ			
	Età	Sesso	Residenza		Età	Sesso	Residenza	
Parametri	β_0	β_1	β_2	β_3	γ_0	γ_1	γ_2	
stime dei parametri	-21.601	10.473	-1.145	14.628	3.686	-0.845	-0.324	-0.104
p-value	0.969	0.917	0.462	0.978	0.000	0.000	0.072	0.587

Tabella 53 *Grado di incertezza e di feeling della frequenza di acquisto*

		età	
Sesso		≤ 25	> 25
F	stima di ξ	0.975	0.944
	grado di feeling	0.024	0.033
M	stima di ξ	0.966	0.925
	grado di feeling	0.033	0.074

i- *Quanto è disposto a pagare per una bottiglia di vino passito”*

Tabella 54. *Stime parametri e p-value*

Covariate	π				ξ			
	Età	Sesso	Residenza		Età	Sesso	Residenza	Età
Parametri	β_0	β_1	β_2	β_3	γ_0	γ_1	γ_2	γ_3
stime dei parametri	-13.351	11.491	1.054	1.890	2.334	-0.287	-0.315	0.04
p-value	0.907	0.920	0.269	0.109	0.000	0.043	0.031	0.777

Tabella 55. *Grado di feeling della cifra disposta a pagare per una bottiglia*

		età	
Sesso		≤ 25	> 25
F	stima di ξ	0.911	0.885
	grado di feeling	0.088	0.117
M	stima di ξ	0.882	0.850
	grado di feeling	0.117	0.149

Tabella 56. *Grado di incertezza della cifra disposta a pagare per una bottiglia*

Provincia di residenza		
	Veneto Occidentale	Veneto Orientale
stima di π	0.000	0.000
grado di incertezza	0.199	0.199

Anche qui, il comportamento delle covariate è simile al precedente caso cioè, il feeling cresce con l'età ed è maggiore per gli uomini, mentre per l'incertezza c'è significatività

solamente per la regione di residenza e a qualche migliaia, il grado di incertezza è uguale nelle due regioni.

j- *Frequenza con cui beve il vino passito*

Per questa risposta, si ha la significatività solo per il feeling e in particolare sulle covariate età. Quando si tratta della frequenza con la quale uno beve il passito, i meno giovani hanno un grado feeling basso.

Tabella 57. *Stima dei parametri e p-value*

Covariate	π				ξ			
	Età	Sesso	Residenza		Età	Sesso	Residenza	Età
Parametri	β_0	β_1	β_2	β_3	γ_0	γ_1	γ_2	γ_3
stime dei parametri	56.653	-10.945	-10.88	0.084	1.669	-0.322	0.084	0.084
p-value	0.999	0.999	0.999	0.999	0.000	0.004	0.016	0.457

Tabella 58. *Grado di incertezza e di feeling della frequenza con cui si beve*

	età	
	≤ 25	> 25
stima di ξ	0.840	0.792
grado di feeling	0.159	0.199

Alla fine di questa analisi del modello CUB(3,3) si può concludere che gli uomini hanno in generale una incertezza bassa rispetto alle donne, come le persone con età maggiore di 25 anni rispetto ai più giovani. Per la zona di residenza invece, quando le domande riguardano la soddisfazione dell'intensità del gusto, la frequenza di acquisto, la frequenza con cui si beve, quanto si è disposto a pagare per una bottiglia, e il gradimento del vino, non c'è nessun effetto della covariata, quindi è possibile dire che la regione di residenza influisce solamente sulla conoscenza del passito, il suo gradimento, la soddisfazione del gusto, la gradazione alcolica, e l'aroma.

4.4.2 Modello saturo CUB(0,0)

Un'altra analisi è stata fatta sul modello CUB senza covariate ovvero il MUB. I risultati ottenuti sono stati riportati nella Tabella 49.

Tabella 59. *stime dei parametri, errori standard, log-verosimiglianza, I.C. e valore del chi quadro di π e ξ per il modello CUB(0,0)*

Domanda	$\hat{\pi}$	$\hat{\xi}$	log-verosimiglianza ($\hat{\pi}, \hat{\xi}$)	devianza	dissimilarità	I.C per il grado d'incertezza	p-value chi2
Gradimento del vino	0.397	0.332	-730.611	0.602	0.014	[0.068;0.103]	0.961
Conoscenza del passito	0.270	0.542	-743.396	12.335	0.077	[0.085;0.123]	0.014
Gradimento del passito	0.549	0.494	-718.34	6.142	0.051	[0.046;0.081]	0.008
Soddisfazione del profilo aromatico del passito	0.683	0.379	-688.65	20.964	0.10	[0.030;0.059]	0.000
Soddisfazione del gusto dolce del passito	0.642	0.381	-696.56	12.619	0.0837	[0.035;0.066]	0.012
Soddisfazione della gradazione alcolica del passito	0.677	0.500	-693.505	10.489	0.066	[0.030;0.0611]	0.032
Soddisfazione dell'intensità del gusto del passito	0.608	0.317	-696.222	9.801	0.071	[0.041; 0.070]	0.038
Frequenza d'acquisto del passito	0.980	0.831	-363.0137	3.389	0.035	[-0.006;0.016]	0.061
Frequenza con cui beve il passito	1.000	0.702	-451.099	38.31	0.136	[-0.011;0.012]	0.000
Quanto è disposto a pagare per una bottiglia di passito	0.969	0.804	-431.841	10.029	0.076	[-0.001;0.013]	0.006

Come sottolineato nel secondo capitolo del nostro lavoro, basandosi sulle applicazioni passate, gli autori definiscono un modello è valido quello con un indice di dissimilarità minore di 0.10. Nel nostro caso studio, quando non è presa in considerazione nessuna covariata, il modello con variabile di interesse il gradimento del vino e quello con interesse la frequenza con cui un soggetto beve il passito non risultano validi avendo rispettivamente indici di dissimilarità pari a 0.014 e 0.136.

C'è da notare sempre in base ai risultati della nostra tabella 49 che, tranne nel modello con variabile risposta la conoscenza del passito, la stima di ξ risulta inferiore rispetto a quella di π , quindi si ha che il feeling è abbastanza elevato rispetto all'incertezza, addirittura nel modello sulla frequenza con cui si il vino il grado di incertezza è pari a 0 (ricordiamo che il grado di incertezza è dato dalla relazione $\frac{1-\hat{\pi}}{m}$). Queste considerazioni sono rafforzate dagli intervalli di confidenza $\left(I.C = \left(\frac{1-\hat{\pi}}{m} \pm \frac{es(\hat{\pi})}{m} \right) \right)$ del gado di incertezza con livello di significatività pari a 0.05 forniti dall'output, in effetti, gli intervalli di confidenza sono centrati e vicini a 0. Calcolando i gradi di incertezza e di

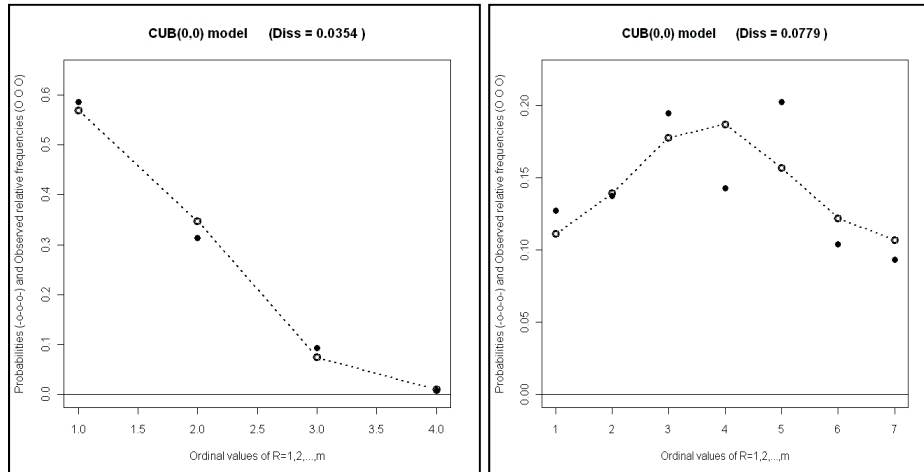
feeling con i valori riportati nella Tabella 50, possiamo notare che gli intervistati hanno presentato un *grado di incertezza* ridotta per tutte le domande, mentre hanno avuto *grado di feeling* abbastanza elevato (maggiore di 0.5) per le domande relative al “*Gradimento del vino*”, “*Gradimento del vino passito*”, “*Soddisfazione del profilo aromatico del vino passito*”, “*Soddisfazione del gusto dolce del vino passito*”, “*Soddisfazione dell'intensità del gusto del vino passito*”.

Tabella 60. *Grado di incertezza e di feeling ricavato dal modello CUB(0,0)*

	grado di incertezza	grado di feeling
Gradimento del vino	0.076	0.676
Conoscenza del vino passito	0.087	0.491
Gradimento del vino passito	0.065	0.508
Soddisfazione del profilo aromatico del vino passito	0.054	0.606
Soddisfazione del gusto dolce del vino passito	0.061	0.610
Soddisfazione della gradazione alcolica del vino passito	0.057	0.494
Soddisfazione dell'intensità del gusto del vino passito	0.066	0.685
Frequenza d'acquisto del vino passito	0.000	0.179
Frequenza con cui beve il vino passito	0.000	0.317
Quanto è disposto a pagare per una bottiglia di vino passito	0.006	0.201

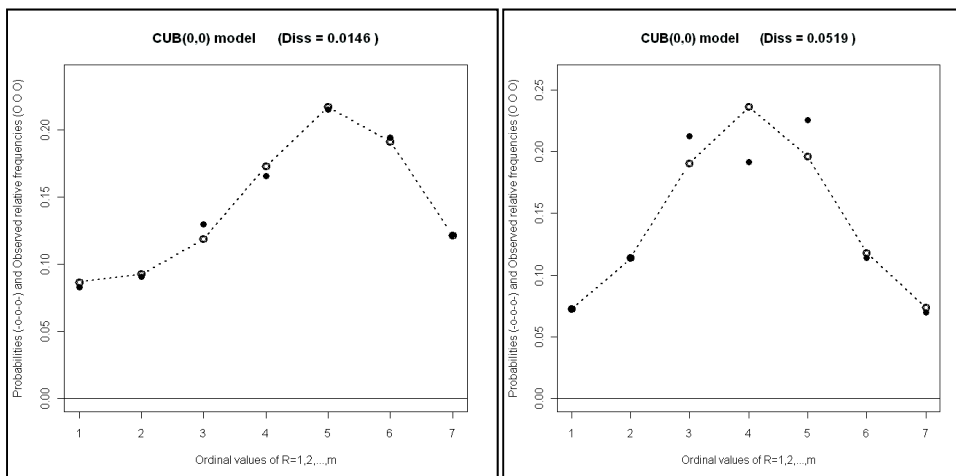
Attraverso il Modello CUB(0,0) possiamo avere anche un riscontro grafico della distribuzione reale (formata da pallini neri) e della distribuzione di probabilità stimata dal modello (realizzata dall'unione di pallini bianchi), come si vede in Figura 25. Come si può facilmente notare, le distribuzioni reali implementate dal modello, forniscono gli stessi risultati dei diagrammi a barre riportati nel paragrafo 3 del presente capitolo.

Figura 25 *Andamento della distribuzione reale e della probabilità stimata dal modello CUB(0,0) senza covariate*



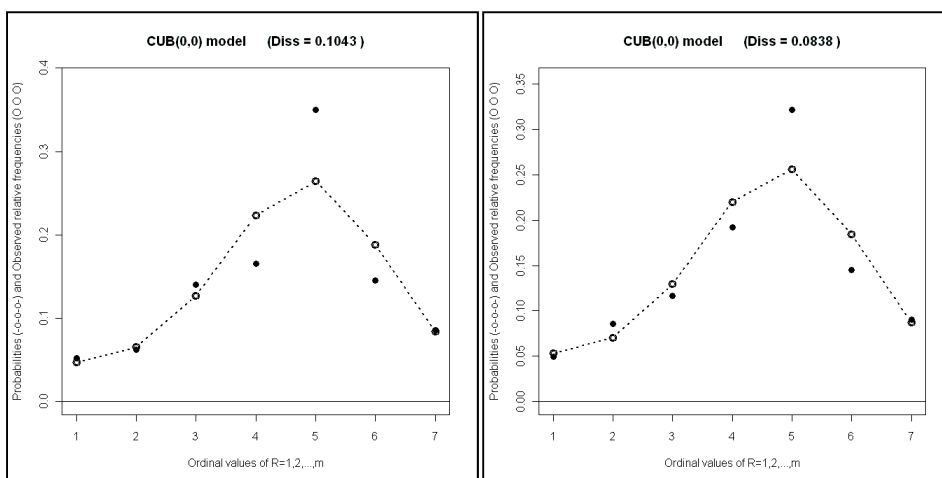
Frequenza d'acquisto del passito

conoscenza del passito



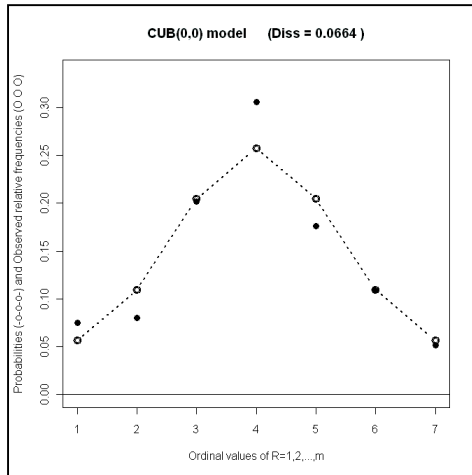
Gradimento del vino

Gradimento del vino passito

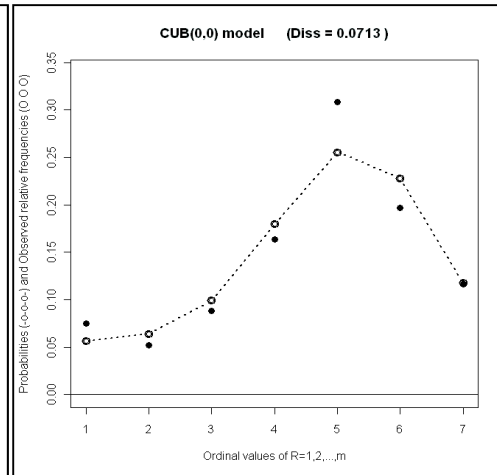


Soddisfazione del profilo aromatico del passito

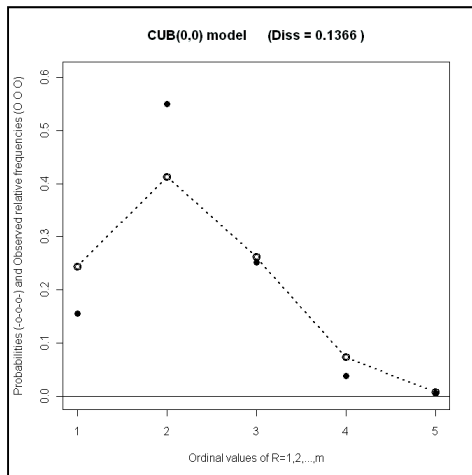
Soddisfazione del gusto dolce del passito



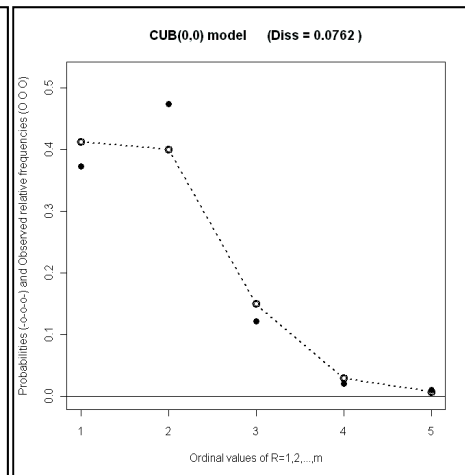
Soddisfazione gradazione alcolica del passito



Soddisfazione intensità del gusto del passito



Frequenza con cui si beve il p passito



Quanto è disposto a pagare per una bottiglia di passito

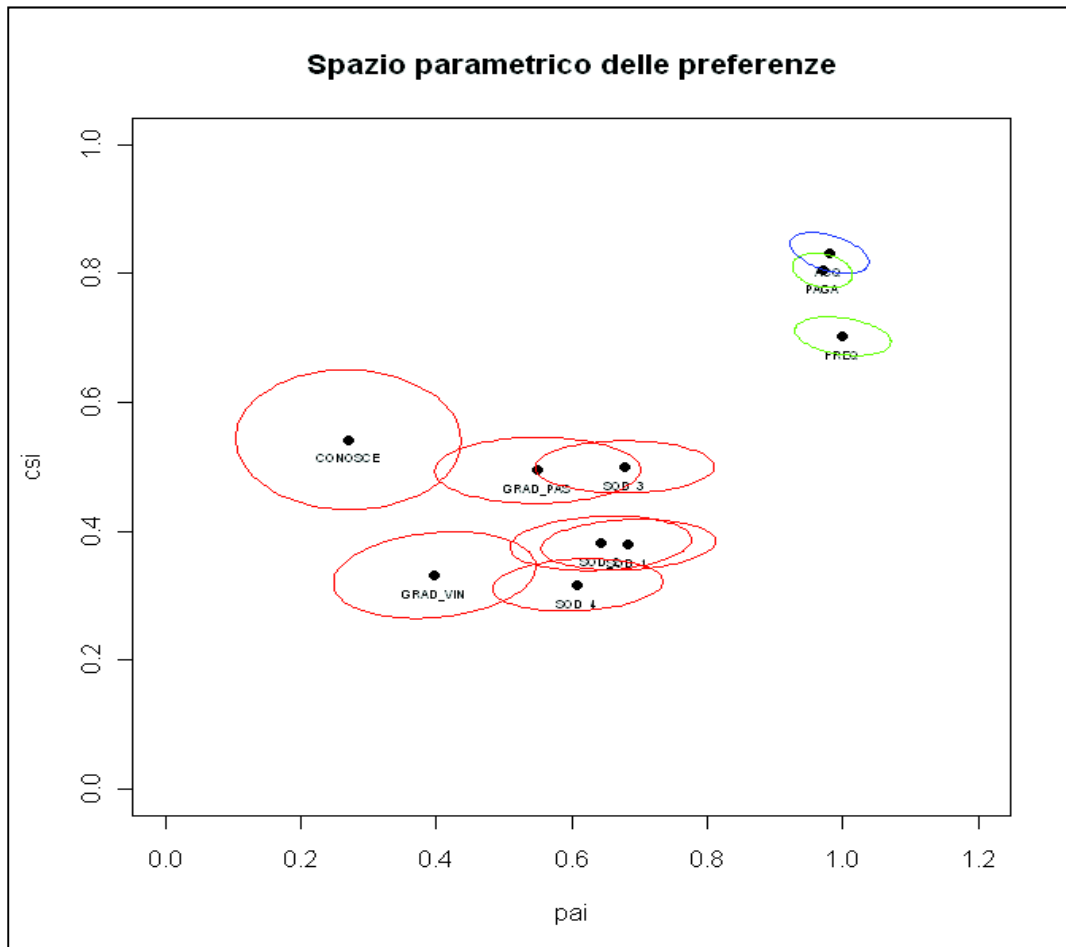
Tale modello ci permette anche di costruire uno spazio parametrico dei valori stimati per π e ζ , creando una regione di confidenza del 95% attorno a questi punti, come mostrato in Figura 24.

In questa immagine vediamo rappresentate in rosso le regioni di confidenza che si riferiscono a quelle domande la cui scala di valutazione va da 1 a 7, in blu le domande con scala da 1 a 5 e infine in verde quelle domande con scala da 1 a 4.

Poiché $1 - \pi$ quantifica la *propensione* dei rispondenti a fornire risposte totalmente casuali, più π è spostato a destra del quadrato unitario e più le valutazioni/preferenze espresse riflettono risposte ben definite (l'incertezza è bassa).

In modo analogo, poiché $1 - \zeta$ misura la forza del *feeling* espressa dal soggetto a favore di una diretta e positiva valutazione del soggetto, più vicino è ζ al bordo superiore del quadrato unitario e meno l'item è preferito.

Figura 26. Regione di confidenza del 95% per π e ζ modello CUB(0,0).



4.5 Analisi CUB versione non parametrica: test di permutazione

Allo scopo di vedere il comportamento dei nostri dati quando non sono valide le assunzioni parametriche e per completare la nostra analisi, abbiamo applicato i test di permutazione ai modelli CUB il cui concetto è stato descritto nel capitolo 3 e le applicazioni nel capitolo 4, e i risultati vengono riportati in questo paragrafo.

Come è stato fatto nel paragrafo precedentemente, con l'utilizzo dei modelli CUB(3,3), anche in questo caso, verranno considerate le covariate *età*, *sex* e *provincia* del campione, per età e sesso, sono considerate persone con età minore o uguale a 25 e maggiore di 25, e per la provincia di residenza verranno analizzati i seguenti raggruppamenti: Belluno- Verona-Vicenza, Padova-Rovigo-Venezia-Treviso.

Naturalmente, le variabili sottoposte allo studio sono sempre: “*Gradimento del vino passito*” (GRAD_VIN), “*Conoscenza del vino passito*” (CONOSCE), “*Frequenza con cui beve il vino passito*” (FREQ), “*Gradimento del vino passito*” (GRAD_PAS), “*Soddisfazione del profilo aromatico del vino passito*” (SODD_1), “*Soddisfazione del gusto dolce del vino passito*” (SOD_2), “*Soddisfazione della gradazione alcolica del vino passito*” (SOD_3), “*Soddisfazione dell'intensità del gusto del vino passito*” (SOD_4), “*Frequenza d'acquisto del vino passito*” (ACQU), “*Quanto è disposto a pagare per una bottiglia di vino passito*” (PAGA).

Come descritto nel capitolo 3, sono state effettuate le permutazioni sull'intero dataset applicando il modello CUB(p,q). Per lo studio della performance dei test di permutazione sui coefficienti (la statistica test è T_2), abbiamo applicato 5 modelli CUB(3,3). Il livello di significatività α è pari a 0.05, il valore di m, cambia in funzione della variabile sulla quale si effettua il test, $n_1 = n_2 = 193$.

Nell'ambito di test di permutazione, non siamo interessati ai valori dei parametri di π e di ξ , ma sottolineiamo tuttavia che essi sono identici a quelli della versione parametrica. Perciò, abbiamo ritenuto opportuno riportare solo i p-value di π e ξ per ogni covariata e per ogni variabile risposta. I risultati sono stati riportati nella tabella 51.

Nella tabella, sono stati evidenziati con colore distintivo i p-value con valori minore del livello prefissato $\alpha = 0.05$, si può infatti notare che:

- Per il gradimento del vino in generale, il p-value è significativo per l'incertezza della covariata età (pari a 0.026), per il feeling della covariata sesso (pari a 0.000). cioè l'età influisce sull'incertezza del gradimento del vino mentre il sesso influisce sul feeling.
- Per quanto riguarda la conoscenza del vino passito, abbiamo la significatività per l'incertezza dell'età (pari a 0.023), il feeling del genere (pari a 0.003) e il feeling della residenza (pari a 0.006). In questo caso, l'età influisce sull'incertezza della conoscenza del vino passito, il sesso e la residenza influiscono sul suo feeling.

- Le permutazioni sul modello con gradimento del vino passito come variabile risposta forniscono un output tale che il sesso e la residenza hanno un p-value significativo al livello di feeling.
- La soddisfazione del profilo aromatico del vino passito ha come p-value significativi quello dell'incertezza del genere e della regione di residenza nonché quella del feeling del genere.
- Per la variabile soddisfazione del gusto dolce, l'unico p-value significativo è quello dell'incertezza della covariata residenza.
- La soddisfazione della gradazione alcolica ha il feeling delle covariate sesso e residenza significative contro l'ipotesi nulla.
- Un dubbio significativo sull'intensità del gusto del passito è presente sia nel sesso che nella regione in cui risiede il soggetto intervistato.
- La frequenza di acquisto ha il feeling significativo per l'età
- La frequenza con la quale uno beve il vino ha anch'essa p-value significativo per l'età,
- E infine per il quanto riguarda il prezzo che uno è disposto a pagare per una bottiglietta di vino passito, il feeling è significativo per il sesso.

Tabella 61. *P-value dei test di permutazione sulle covariate.*

Variabile risposta	Età		Sesso		Residenza	
	π	ξ	π	ξ	π	ξ
gradimento del vino	0.026	0.212	0.199	0.000	0.738	0.930
Conoscenza del passito	0.023	0.524	0.069	0.003	0.511	0.006
Gradimento del vino passito	0.515	0.512	0.743	0.000	0.782	0.000
Soddisfazione profilo aromatico	0.233	0.724	0.030	0.027	0.023	0.030
Soddisfazione gusto dolce	0.336	0.023	0.070	0.076	0.043	0.113
Soddisfazione gradazione alcolica	0.425	0.910	0.625	0.017	0.574	0.009
Soddisfazione intensità del gusto	0.897	0.844	0.000	0.743	0.000	0.762
Frequenza di acquisto	0.953	0.003	0.554	0.212	0.159	0.857
Frequenza con cui si beve il passito	0.671	0.006	0.833	0.075	0.601	0.016
quanto si è disponibile a pagare per una bottiglietta	0.873	0.066	0.551	0.047	0.774	0.388

In definitiva, emerge da quest'analisi delle covariate tramite i test di permutazione gli stessi risultati ottenuti in ambito parametrico, cioè in tutte le variabili risposte da noi esaminate, almeno un parametro tra γ_i e β_i ($i = 1, 2, 3$) è diverso da zero e siamo a favore dell'ipotesi alternativa cioè almeno una covariata tra sesso, età e regione di residenza ha un effetto sulla scelta finale che un soggetto intervistato fa sulle scale di valutazioni.

Capitolo 5. Conclusioni

In questo lavoro di tesi, l'obiettivo è stato di studiare il test di permutazione del modello Covariate Uniform Binomial e di fare un confronto tra quest'ultimo e la versione parametrica. Dopo aver descritto il modello sulle due versioni e fatto una serie di simulazioni, abbiamo applicato il modello su un insieme di dati destinato ad un'indagine di mercato.

Il vantaggio di questa applicazione risulta nel fatto che non solo siamo riusciti a paragonare le due versioni del modello, ma abbiamo anche avuto informazioni sull'andamento del profilo del consumatore target del vino passito nel veneto.

Applicando il modello CUB(3,3) sui test di permutazione, l'output fornito da R ha a qualche differenza gli stessi risultati forniti dal modello parametrico. In effetti, oltre alle stime di ξ e di π , alle log-verosimiglianze, alla devianza, l'indice di dissimilarità, e le altre grandezze, ci sono i p-value dei diversi test (T_1 , T_2 , W).

I vantaggi dei test di permutazione sulle covariate del modello CUB sono i seguenti:

- I test di permutazione sono "distribution free", non necessitano che siano note le assunzioni della distribuzione della statistica test,
- Questi test non richiedono l'applicazione delle teorie asintotiche e possono essere applicate anche per campioni piccoli.
- Sotto l'ipotesi nulla, il tasso di rifiuto rispetta il livello di significatività prefissato.
- La stima della potenza del test non è piccola, e a volte è migliore di quella del test di Wald.
- Per i piccoli campioni, la stima della potenza del test segue la disuguaglianza: potenza(T_1) > potenza (T_2) > potenza (W).

Nel caso dello studio condotto su un campione eterogeneo di persone, provenienti da diverse aree geografiche del Veneto, dopo l'applicazione della versione parametrica del modello CUB, si può affermare, con un ridotto margine di errore, che:

- gli intervistati maschi, gradiscono meglio il passito rispetto alle donne, mentre i residenti del veneto occidentale hanno un feeling più basso di quelli della parte orientale. bevono più frequentemente passito durante l'anno, rispetto alle femmine;
- i maschi con età maggiore di 25 anni sono anche maggiormente disposti a pagare di più una bottiglia di passito. Fra tutti, però, le persone diventano più consapevoli di quanto gradiscono il vino sono quelli con più di 45 anni, e tra questi i maschi, ovviamente.

Tutto ciò si potrebbe motivare col fatto che il campione del veneto orientale e particolarmente i maschi, conoscono maggiormente il passito rispetto agli altri intervistati, e questo è dimostrato dalle risposte date quando gli è stato chiesto di dare un punteggio sulla propria conoscenza sul passito.

Passando ad analizzare gli aspetti relativi al gusto che possono piacere di più, si nota che:

- il *profilo aromatico* soddisfa maggiormente il gruppo del veneto orientale. Tra questi la fascia d'età è un altro fattore discriminante del campione in quanto sono meno incerti i maschi di età maggiore a 25 anni, e anche il sesso lo è, in quanto i maschi manifestano preferenze sempre maggiori rispetto le femmine;
- il *gusto dolce* soddisfa leggermente di più il campione veneto orientale, di età inferiore ai 25 anni, mentre valori molto bassi sono stati dati dai residenti nelle province del veneto ovest.
- la *gradazione alcolica* soddisfa i residenti del veneto occidentale
- infine, l'*intensità del gusto*, in caso, punteggi alti provengono anche dal campione femminile.
 - A questo punto si potrebbe definire un consumatore target di passito, ossia un maschio di età superiore ai 25 anni e residente nelle province del veneto orientale, che solitamente beve passito a pranzo/cena, al ristorante/pizzeria e che di questo vino apprezza maggiormente il profilo aromatico e la sua intensità del gusto.

Per quanto riguarda l'applicazione ai dati della versione non parametrica del modello CUB(3,3), siamo a favore dell'ipotesi alternativa, quindi c'è un effetto di almeno una covariata tra età, sesso e residenza sulle valutazioni delle domande di interesse. I risultati ottenuti sono simili a quelli ottenuti dal CUB(3,3) versione parametrica.

Altre estensioni del test di permutazione sono:

- Una possibile applicazione del test di permutazione sui modelli CUB riguarda il modello multivariato, in questo caso, verranno valutate gli effetti delle covariate sulle valutazioni fornite dai rispondenti su più prodotti considerati simultaneamente;
- Applicare il test di permutazione su un oggetto considerando più aspetti delle variabili risposte, con la possibilità di combinare per dominio

Appendix A: The E-M algorithm

Let $y = (y_1, y_2, \dots, y_n)'$ be the sample of ordinal data generated by the survey where n respondents are asked to choose an integer in the support $\{1, 2, \dots, m\}$, for a given $m > 4$. We supposed the location $c \in \{1, 2, \dots, m\}$ of the *shelter choice* is know.

For the extended CUB model, the log – likelihood function $\ell(\theta)$ for the sample y , with $\theta = (\pi_1, \pi_2, \xi)$ is

$$\begin{aligned}\ell(\theta) &= \sum_{i=1}^n \log[\Pr(Y = y_i | \theta)] \\ &= \sum_{i=1}^n \log[\pi_1 b_{y_i}(\xi) + \pi_2 U_{y_i} + (1 - \pi_1 - \pi_2) D_y^{(c)}],\end{aligned}$$

Where the components of the mixture are specified for $i = 1, 2, \dots, n$, by:

$$P_g(y_i; \theta_g) = \begin{cases} b_{y_i}(\xi) = \binom{m-1}{y_i-1} (1-y)^{y_i-1} \xi^{m-y_i}, & g = 1; \\ U_{y_i} = \frac{1}{m}, & g = 2; \\ D_{y_i}^{(c)}, & g = 3. \end{cases}$$

We introduce the unobservable vector $z = (z_1, z_2, \dots, z_n)'$ where $z_i = (z_{1i}, z_{2i}, z_{3i})'$ is a three – dimensional vector such that, for $g = 1, 2, 3$:

$$\begin{cases} z_{gi} = 1, & \text{if the } i\text{-th subject belongs to the } g \text{ group;} \\ z_{gi} = 0, & \text{otherwise.} \end{cases}$$

Simplifying the notation, we let;

$$\pi_g = \begin{cases} \pi_1, & g = 1; \\ \pi_2, & g = 2; \\ 1 - \pi_1 - \pi_2, & g = 3. \end{cases} \quad \theta_g = \begin{cases} \theta_1 = (\pi_1, \xi)', & g = 1; \\ \theta_2 = \pi_2, & g = 2; \\ \theta_3 = 1 - \pi_1 - \pi_2, & g = 3. \end{cases}$$

Then, the likelihood function of the complete-data vector (y', z') is given by:

$$L_c(\theta) = \prod_{g=1}^3 \prod_{i=1}^n [\pi_g p_g(y_i; \theta_g)]^{z_{gi}},$$

and the complete-data log – likelihood function is:

$$\ell_c(\theta) = \sum_{g=1}^3 \sum_{i=1}^n z_{gi} [\log(\pi_g) + \log(p_g(y_i; \theta_g))].$$

The $(k + 1)$ iteration of the EM algorithm consists of the following steps:

- **E-step:**

The conditional expectation of the indicator random variable z_{gi} , given the observed sample y , is:

$$\mathbb{E}(z_{gi} | y, \theta^{(k)}) = \Pr(z_{gi} = 1 | y, \theta^{(k)}) = \frac{\pi_g^{(k)} p_g(y; \theta_g^{(k)})}{\sum_{j=1}^3 \pi_j^{(k)} p_j(y; \theta_j^{(k)})} = \tau_g(y_i, \theta^{(k)}),$$

For any $g = 1, 2, 3$. This quantity is the posterior probability that the i -th subject of the sample with the observed y_i belongs to the g -th component of the mixture.

Then the expected log – likelihood of complete-data vector is obtained as:

$$\mathbb{E}(\ell_c(\theta)) = \sum_{g=1}^3 \sum_{i=1}^n \tau_g(y_i, \theta^{(k)}) \left[\log(\pi_g^{(k)}) + \log(p_g(y; \theta_g^{(k)})) \right].$$

- **M-step:**

At the $(k + 1)$ -th iteration, the function $Q(\theta^{(k)}) = \mathbb{E}(\ell_c(\theta))$ has to be maximized with respect to the parameters (π_1, π_2) and ξ . If the parameters of the components are specified, this quantity may be expressed as follows:

$$\begin{aligned} Q(\theta^{(k)}) &= \sum_{i=1}^n \left[\tau_1(y_i, \theta_1^{(k)}) \log(\pi_1^{(k)}) + \tau_2(y_i, \theta_2^{(k)}) \log(\pi_2^{(k)}) \right. \\ &\quad \left. + \tau_3(y_i, \theta_3^{(k)}) \log(\pi_3^{(k)}) \right] + \sum_{g=1}^3 \sum_{i=1}^n \left[\tau_g(y_i, \theta_g^{(k)}) \log(p_g(y; \theta_g^{(k)})) \right] \\ &= S_1 \log(\pi_1^{(k)}) + S_2 \log(\pi_2^{(k)}) + (n - S_1 - S_2) \log(1 - \pi_1^{(k)} - \pi_2^{(k)}) \\ &\quad + Q^* \end{aligned}$$

Where Q^* is independent from $\pi_g^{(k)}$ parameters, and

$$S_g = \tau_g(y_i, \theta^{(k)}), \quad g = 1, 2, S_3 = n - S_1 - S_2.$$

Then, by solving the system: $\frac{\partial Q(\theta^{(k)})}{\partial \pi_g} = 0$, for $g = 1, 2$, we get:

$$\pi_1^{(k+1)} = \frac{S_1}{n} = \frac{1}{n} \sum_{i=1}^n \tau_1(y_i, \theta^{(k)}); \quad \pi_2^{(k+1)} = \frac{S_2}{n} = \frac{1}{n} \sum_{i=1}^n \tau_2(y_i, \theta^{(k)}).$$

Instead, the estimate of ξ , for a given k , is obtained from:

$$\sum_{i=1}^n \tau_1(y_i, \theta_1^{(k)}) \frac{\partial \log(p_1(y_i; \xi))}{\partial \xi} = 0.$$

A simple algebra produces the solution:

$$\xi^{(k+1)} = \frac{m - \bar{Y}_n(p)}{m - 1}; \quad \bar{Y}_n(p) = \frac{\sum_{i=1}^n y_i \tau_1(y_i, \theta_1^{(k)})}{\sum_{i=1}^n \tau_1(y_i, \theta_1^{(k)})}.$$

Here, $\bar{Y}_n(p)$ is the average of the observed sampled values weighted with the posterior probability that y_i is a realization of the first component of the mixture (that is a shifted Binomial distribution).

Then, E- and M- steps are repeated with new parameters $(\pi_1^{(k+1)}, \pi_2^{(k+1)}, \xi^{(k+1)})'$ until a convergence criterion is satisfied. For instance, this could be given by:

$$|\ell(\theta^{(k+1)}) - \ell(\theta^{(k)})| < \varepsilon, \text{ for a small } \varepsilon > 0.$$

Notice that as far as ML estimation is concerned, sample data y_i is equivalently represented by the vector of absolute frequencies n_y , for the computational efficiency, it is therefore convenient to use in previous steps the log – likelihood function for grouped data. To this end, we will compute:

$$S_g = \sum_{y=1}^m n_y \tau_g(y_i, \theta^{(k)}), \quad g = 1, 2.$$

The step-by-step formulation of the EM algorithm may be easily programmed in formal languages (such as GAUSS[®], MATLAB[®] or R).

Appendice B: Questionario raccolta dati

▪ Et : _____ ▪ Sesso: M F ▪ Provincia di residenza: _____

▪ Quanto gradisce bere vino?

per nulla	1	2	3	4	5	6	7	moltissimo
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

(indichi una risposta da 1 a 7)

▪ In quale occasione beve vino?

mai	raramente	talvolta	spesso	regolarmente
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

– A casa durante i pasti
– Al bar/enoteca
– A pranzo/cena al ristorante/trattoria

▪ Conosce il vino passito dolce?

per nulla	1	2	3	4	5	6	7	molto bene
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

(indichi una risposta da 1 a 7)

▪ Quali dei seguenti passiti del Veneto conosce:

NO	SI		NO	SI
<input type="radio"/>	<input type="radio"/>	– <i>Torcolato</i> di FREGONA	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	– REFRONTOLO PASSITO (Marzemino)	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	– Fior d’arancio passito dei COLLI EUGANEI	<input type="radio"/>	<input type="radio"/>

▪ Conosce altri tipi di passito?

NO	SI	Se SI, quali? 1 _____
<input type="radio"/>	<input type="radio"/>	2 _____ 3 _____

▪ In che modo ha conosciuto il vino passito:

NO	SI		NO	SI
<input type="radio"/>	<input type="radio"/>	– amici	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	– familiari o parenti	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>			

– pubblicit 
– degustazioni/incontri enogastronomici
– per caso

▪ Quanto frequentemente beve vino passito?

mai	raramente	talvolta	spesso	sempre
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<ul style="list-style-type: none"> ▪ In quale occasione beve vino passito? – ricorrenze (Natale, Pasqua, matrimoni) – da solo quando ne ho voglia – alla fine dei pasti – se me lo offrono o capita l'occasione 	mai	raramente	talvolta	spesso	sempre
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<ul style="list-style-type: none"> ▪ Quanto vino passito beve all'anno? 	1 bicchiere ogni tanto	1 bottiglia	2-3 bottiglie	>3 bottiglie
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<ul style="list-style-type: none"> ▪ Quanto gradisce bere vino passito? (indichi una risposta da 1 a 7) 	per nulla	1	2	3	4	5	6	7	moltissimo
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<ul style="list-style-type: none"> ▪ Del vino passito, quanto la soddisfano – il profilo aromatico (aroma e profumo) – il gusto dolce – la gradazione alcolica – l'intensità del gusto 	per nulla	1	2	3	4	5	6	7	moltissimo	
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

<ul style="list-style-type: none"> ▪ Ritieni che il vino passito possa essere una valida alternativa a: – vino tradizionale da abbinare ai dolci – amaro grappa da bere a fine pasto 	NO	SI
	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>

<ul style="list-style-type: none"> ▪ Acquista bottiglie di vino passito? 	mai	raramente	talvolta	spesso	sempre
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<ul style="list-style-type: none"> ▪ Quanto è disposto a pagare per una bottiglia di vino passito? (da 0.5 litri) 	0-10 €	10-15 €	15-20 €	20-25 €	più di 25 €
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Indici delle figure e delle tabelle

Figura 1: <i>Matrice di dati e matrice dei dati permutati.</i>	29
Figura 2: <i>distribuzione della variabile risposta al variare di delta per $m = 7$.</i>	36
Figura 3: <i>Andamento della potenza per i tre test nel caso $n_1 = n_2 = 15, m=4$.</i>	38
Figura 4. <i>Percentuale del grado di soddisfazione degli intervistati nel bere vino</i>	45
Figura 5. <i>Percentuale delle volte che gli intervistati bevono vino durante i pasti</i>	46
Figura 6. <i>Percentuale delle volte che gli intervistati bevono vino bar/enoteca</i>	47
Figura 7. <i>Percentuale delle volte che gli intervistati bevono vino a pranzo/cena al ristorante</i>	48
Figura 8. <i>Percentuale di intervistati per livello di conoscenza del vino passito</i>	49
Figura 9. <i>Percentuale degli intervistati che conoscono alcuni tipi di vino passito del Veneto</i>	50
Figura 10. <i>Percentuale di intervistati che hanno conosciuto il vino passito attraverso familiari, amici, per caso, grazie a degustazioni o tramite la pubblicità.</i>	51
Figura 11. <i>Percentuale di intervistati per frequenza con cui bevono vino passito</i>	52
Figura 12. <i>Percentuale di intervistati per frequenza con cui bevono vino passito durante ricorrenze, come Natale, Pasqua, matrimoni, ecc.</i>	53
Figura 13. <i>Percentuale di intervistati per frequenza con cui bevono vino passito quando lo desiderano</i>	54
Figura 14. <i>Percentuale di intervistati per frequenza con cui bevono vino passito alla fine dei pasti</i>	55
Figura 15. <i>Percentuale di intervistati per frequenza con cui bevono vino passito, perché viene loro offerto o perché capita l'occasione</i>	56
Figura 16. <i>Percentuale di intervistati per quantità di vino passito bevuto all'anno</i>	57
Figura 17. <i>Percentuale di intervistati per gradimento del vino passito</i>	58
Figura 18. <i>Percentuale di intervistati per il loro grado di soddisfazione riguardo il profilo aromatico del vino passito</i>	59
Figura 19. <i>Percentuale di intervistati per grado di soddisfazione riguardo il gusto dolce del vino passito</i>	60
Figura 20. <i>Percentuale di intervistati per grado di soddisfazione riguardo la gradazione alcolica del vino passito</i>	61
Figura 21. <i>Percentuale di intervistati per grado di soddisfazione riguardo l'intensità del gusto del vino passito</i>	62
Figura 22. <i>Percentuale degli intervistati che ritengono il vino passito un'alternativa al vino tradizionale da dolce o all'amaro/grappa</i>	63
Figura 23. <i>Percentuale degli intervistati per frequenza d'acquisto di vino passito</i>	64
Figura 24. <i>Percentuale di intervistati per classi di prezzo di vino passito, riferito a una bottiglia da 0,5l</i>	65
Figura 25 <i>Andamento della distribuzione reale e della probabilità stimata dal modello CUB(0,0) senza covariate</i>	76
Figura 26. <i>Regione di confidenza del 95% per π e ξ modello CUB(0,0).</i>	78

Tabella 1: Notazione CUB(p,q)	15
Tabella 2 confronto tra modelli complessi e semplici e i gradi di libertà associati	17
Tabella 3. Istruzioni per le stime dei modelli CUB.	19
Tabella 4. Variabili ottenute dopo l'esecuzione del CUB	19
<i>Tabella 5: Metodo Monte Carlo Condizionale e la distribuzione di permutazione di T...</i>	<i>27</i>
Tabella 6: Comportamento di ξ sotto H_0 CUB(0,1).....	35
Tabella 7: Studio della potenza al variare di delta per $m=7$, $n_1=n_2=25$	36
Tabella 8: Studio della potenza al variare di delta per $m=6$, $n_1=n_2=25$	37
Tabella 9: Studio della potenza al variare di delta per $m=5$, $n_1=n_2=20$	37
Tabella 10: Studio della potenza al variare di delta per $m=4$, $n_1=n_2=15$	37
Tabella 11. <i>Distribuzione percentuale geografica degli intervistati</i>	44
Tabella 12. <i>Distribuzione di frequenze assolute e percentuali di intervistati per età condizionate al sesso</i>	44
Tabella 13. <i>Distribuzione percentuale di intervistati per grado di soddisfazione nel bere vino, data la scala di valutazione da 1 a 7</i>	45
Tabella 14. <i>Distribuzione percentuale di intervistati per frequenza con cui bevono vino durante i pasti</i>	46
Tabella 15. <i>Distribuzione percentuale di intervistati per la frequenza con cui bevono vino al bar/ enoteca</i>	47
Tabella 16. <i>Distribuzione percentuale di intervistati per frequenza con cui bevono vino a pranzo/cena al ristorante/pizzeria</i>	48
Tabella 17. <i>Distribuzione percentuale di intervistati per livello di conoscenza del vino passito dolce, data una scala di giudizio da 1 a 7</i>	49
Tabella 18. <i>Frequenza assoluta e percentuale di intervistati che conoscono i seguenti tipi di passiti del Veneto</i>	50
Tabella 19. <i>Distribuzione percentuale di intervistati che hanno conosciuto il vino passito attraverso i familiari, amici, per caso, grazie a degustazioni o tramite la pubblicità</i> 51	
Tabella 20. <i>Distribuzione percentuale di intervistati per frequenza con cui bevono vino passito</i> 52	
Tabella 21. <i>Distribuzione percentuale di intervistati per frequenza con cui bevono vino passito durante ricorrenze, come Natale, Pasqua, matrimoni, ecc.</i>	53
Tabella 22. <i>Distribuzione percentuale di intervistati per frequenza con cui bevono vino passito ogni volta lo desiderano, a prescindere dalle occasioni o ricorrenze particolari</i> 54	
Tabella 23. <i>Distribuzione percentuale di intervistati per frequenza con cui bevono vino passito alla fine dei pasti</i>	55
Tabella 24. <i>Distribuzione percentuale di intervistati per frequenza con cui bevono vino passito, perché viene loro offerto o perché capita l'occasione</i>	56
Tabella 25. <i>Distribuzione percentuale di intervistati per quantità di vino passito bevuto all'anno</i> 57	
Tabella 26. <i>Distribuzione percentuale di intervistati per gradimento del vino passito, data una scala di valutazione da 1 a 7</i>	58
Tabella 27. <i>Distribuzione percentuale di intervistati per il loro grado di soddisfazione riguardo il profilo aromatico del vino passito, data una scala di valutazione da 1 a 7</i> ..59	
Tabella 28. <i>Distribuzione percentuale di intervistati per grado di soddisfazione riguardo il gusto dolce del vino passito, data una scala di valutazione da 1 a 7</i>	60
Tabella 29. <i>Distribuzione percentuale di intervistati per grado di soddisfazione riguardo la gradazione alcolica del vino passito, data una scala di valutazione da 1 a 7</i>	

Tabella 30.	<i>Distribuzione percentuale di intervistati per grado di soddisfazione riguardo l'intensità del gusto del vino passito, data una scala di valutazione da 1 a 7 ...</i>	62
Tabella 31.	<i>Distribuzione percentuale degli intervistati che ritengono il vino passito un'alternativa al vino tradizionale da dolce o all'amaro/grappa</i>	63
Tabella 32.	<i>Distribuzione percentuale degli intervistati per frequenza d'acquisto di vino passito</i>	63
Tabella 33.	<i>Distribuzione percentuale di intervistati per classi di prezzo di vino passito, riferito a una bottiglia da 0,5l</i>	64
Tabella 34.	<i>Stime dei parametri e p-value</i>	67
Tabella 35 a:	<i>Grado di incertezza ricavato dal modello CUB(3,3) gradimento del vino..</i>	67
Tabella 36 b	<i>Grado di feeling ricavato dal modello CUB(3,3) gradimento del vino....</i>	67
Tabella 37.	<i>Stime dei parametri e p-value</i>	67
Tabella 38	<i>grado di feeling per la conoscenza del passito</i>	68
Tabella 39	<i>Stime dei parametri e p-value</i>	68
Tabella 40.a	<i>Grado di feeling del gradimento del passito.....</i>	68
Tabella 41	<i>Grado di incertezza e del gradimento del passito.....</i>	68
Tabella 42	<i>Stime dei parametri e p-value</i>	69
Tabella 43.	<i>Grado di feeling della soddisfazione del profilo aromatico</i>	69
Tabella 44	<i>Grado di incertezza della soddisfazione del profilo aromatico</i>	69
Tabella 45	<i>Stime dei parametri e p-value</i>	70
Tabella 46.	<i>Grado dell'incertezza della soddisfazione del gusto dolce.....</i>	70
Tabella 47.	<i>Grado del feeling della soddisfazione del gusto dolce.....</i>	70
Tabella 48	<i>Stime dei parametri e p-value</i>	71
Tabella 49.	<i>Grado di incertezza della soddisfazione della gradazione alcolica</i>	71
Tabella 50.	<i>Grado di feeling della soddisfazione della gradazione alcolica</i>	71
Tabella 51.a	<i>Stima dei parametri e p-value</i>	71
Tabella 52	<i>Stima dei parametri e p-value.....</i>	72
Tabella 53	<i>Grado di incertezza e di feeling della frequenza di acquisto.....</i>	72
Tabella 54.	<i>Stime parametri e p-value</i>	72
Tabella 55.	<i>Grado di feeling della cifra disposta a pagare per una bottiglia.....</i>	72
Tabella 56.	<i>Grado di incertezza della cifra disposta a pagare per una bottiglia.....</i>	72
Tabella 57.	<i>Stima dei parametri e p-value.....</i>	73
Tabella 58.	<i>Grado di incertezza e di feeling della frequenza con cui si beve.....</i>	73
Tabella 59.	<i>stime dei parametri, errori standard, log-verosimiglianza, I.C. e valore del chi quadro di π e ξ per il modello CUB(0,0)</i>	74
Tabella 60.	<i>Grado di incertezza e di feeling ricavato dal modello CUB(0,0)</i>	75
Tabella 61.	<i>P-value dei test di permutazione sulle covariate.</i>	80

Bibliografia

- Gazzetta ufficiale n.60 13 marzo 2006.
- Corduas, M, e al. «A class of statistical models for evaluating services and performances.» 2009.
- Good, P. *Permutation tests: a practical guide to resampling methods for testing hypotheses*. New York: Spinger Verlag, 1994.
- ISO. «10002:2004 Gestione per la qualità - Soddisfazione del cliente - Linee guida per il trattamento dei reclami nelle organizzazioni.» 2004.
- Magidson, J, e K Vermunt. «Latent Variables.» *Encyclopedia of Social Science Research Methods*, 2003.
- Marshall. *Random Utility Models*. 1960.
- Pace, L, e A Salvan. *Introduzione alla statistica II. Inferenza, verosimiglianza, Modelli*. Padova: Cedam, 2001.
- Pesarin, F, e L Salmaso. «Permutation Tests for Univariate and Multivariate Ordered Categorical Data.» *Australian Journal Of Statistics vol 35*, 2006: 315-324.
- Pesarin, Fortunato. *Multivariate Permutation Tests With Application in Biostatistic*. John Wiley & Sons, LTD, 2001.
- Piccolo D, D'Elia A. «A new approach for modelling consumers' preferences.» *Food and Preference*, 2007: 19: 247-259.
- Piccolo, D. «Observed information matrix for MUB models.» *Quaderni di Statistica 8*, 2006: 33 - 78.
- Scheffé, H. «Statistical inference on non-parametric case.» In *Annals of Mathematical Statistics*, 14, 1340-1342. 1943.
- Schwarz, G. «Estimating the dimension of a model.» *Annals of Statistics 6*, 1978: 461-464.
- Strasser, Helmut, e Christian Weber. «On Asymptotic Theory of Permutation Statistics.» *Mathematics Methods of Statistics*, 1999: 220-250.
- van der, Vaart. «Asymptotic Statistics.» Cambridge: University press, 1998.
- WineNews. www.winenews.it (consultato il giorno luglio 21, 2010).