



**UNIVERSITÀ
DI TORINO**

Università degli Studi di Torino

Corso di Laurea

Economia e Statistica per le Organizzazioni

Titolo

Fattori determinanti nelle aspettative d'inflazione delle
famiglie italiane

Relatrice

Cinzia Carota

Candidato

Ferrero Andrea

Matricola 949959

Anno Accademico 2022/2023

Abstract

Il lavoro consiste in un'analisi statistica sul sondaggio condotto da Banca d'Italia con lo scopo di individuare i fattori che influenzano maggiormente le aspettative d'inflazione delle famiglie italiane. Riprendendo l'analisi svolta da Conrad et al. (2022) sui dati tedeschi, ci si è concentrati particolarmente sulle determinanti nella scelta dei canali informativi in materia economico-finanziaria e sul loro ruolo nella formazione delle aspettative. Si è anche indagata la presenza di effetti significativi in merito alle esperienze, soprattutto durante gli anni impressionabili (tra i 18 e i 25 anni), e alle impressioni sull'inflazione nel periodo precedente al sondaggio.

Indice

Introduzione	7
1 Analisi descrittive	9
1.1 Dati utilizzati	9
1.2 Caratteristiche socio-economiche	9
2 Determinanti dei canali informativi	13
2.1 Si informa?	13
2.1.1 Modello di regressione logistica	14
2.1.2 Diagnostica	17
2.1.3 Analisi	18
2.2 Scelta del canale informativo	22
2.2.1 Modello di regressione multinomiale logistica	22
2.2.2 Test del rapporto di verosimiglianza	23
2.2.3 Analisi	24
3 Determinanti delle aspettative di inflazione	35
3.1 Modello a logit cumulati	37
3.2 Modello logit a categorie adiacenti	38
3.3 Ipotesi di Odds Proporzionali e modelli logit ordinali generalizzati	38
3.4 Analisi con regressione logistica	39
3.5 Analisi con modelli ordinali	41
3.6 Il ruolo delle esperienze	47
Conclusione	51

Introduzione

L'inflazione è un fenomeno economico che consiste in un aumento generale dei prezzi. Individuarne le cause spesso non è facile, ma da tempo gli economisti e le banche centrali sono al corrente del ruolo che svolgono le aspettative degli agenti economici nel determinare l'andamento dell'inflazione. Se le aspettative cominciano a muoversi a favore di un aumento generalizzato dei prezzi, si rischia che i prezzi seguano l'andamento "predetto" dagli agenti economici.

Nel lavoro che segue si è cercato di individuare quali fossero i fattori socio-economici che maggiormente influenzano le aspettative delle famiglie italiane riguardo ai prezzi. I dati utilizzati sono quelli pubblicati dal sondaggio della Banca d'Italia, denominato Indagine Straordinaria sulle Famiglie italiane, e l'analisi ha avuto come punto di riferimento il lavoro svolto da Conrad et al. (2022) sui dati delle famiglie tedesche.

Inizialmente ci si è concentrati sulle determinanti nella scelta dei mezzi informativi in materia economica. Ciò è stato motivato dall'importanza che il tipo di canale informativo ha nell'influenzare le aspettative di inflazione, come riscontrato anche da Conrad et al. (2022). Si riscontra che l'età, il titolo di studio e il reddito, così come l'essere disoccupati e la provenienza geografica, sono fattori socio-economici importanti in questo ambito. In particolare, sono essenziali sia nel determinare se l'individuo si informa in generale, sia nel determinare che tipo di mezzo utilizza. Il risultato più rilevante per le banche centrali e le istituzioni è che chi è disoccupato è più probabile che non si informi con media istituzionali.

Successivamente ci si è concentrati sul tema principale, che è quello delle aspettative. Le analisi indicano che l'età e il reddito sono gli unici fattori socio-economici rilevanti in merito. A questi si aggiunge come fattore importante l'utilizzo di media tradizionali e di quelli istituzionali. Si è rilevato che persone che utilizzano questi media per informarsi sono più propense ad aspettarsi inflazione nel futuro. Un altro elemento essenziale è la percezione dell'aumento dei prezzi. Più gli individui ritengono che i prezzi siano aumentati e più si aspettano che i prezzi aumentino nel futuro.

L'ultimo effetto significativo riscontrato è aver avuto esperienza di inflazione eleva-

ta. In Italia nel corso degli anni '70 e '80 si sono registrati livelli particolarmente alti e sostenuti di inflazione, anche oltre al 20%. Questo sembra aver influenzato le famiglie italiane, poiché coloro che hanno vissuto quel periodo sono più propensi ad aspettarsi inflazione. Si è analizzato anche il ruolo di aver vissuto tale periodo in anni considerati dalla psicologia particolarmente impressionabili, tra i 18 e i 25 anni. Per quanto l'effetto sia risultato significativo, non si sono riscontrate particolari differenze nel livello di inflazione attesa, suggerendo probabilmente che il ruolo dell'impressionabilità in questo contesto sia marginale.

La metodologia utilizzata verrà analizzata dettagliatamente in seguito, e comprende modelli lineari generalizzati per variabili dicotomiche, politomiche nominali e politomiche ordinali.

Ringrazio la mia relatrice, la professoressa Cinzia Carota, per la dedizione con la quale ha supervisionato il lavoro, e i miei cari, per il supporto che mi hanno sempre dimostrato.

Capitolo 1

Analisi descrittive

1.1 Dati utilizzati

Il lavoro svolto utilizza i dati raccolti da Banca d'Italia (2021) durante le 6 edizioni delle indagini straordinarie sulle famiglie italiane.

In particolare, il lavoro si concentra sulla edizione svoltasi tra la fine di agosto e l'inizio di settembre 2021. Lo scopo del sondaggio era intervistare un campione rappresentativo della popolazione italiana al fine di raccogliere informazioni sulla situazione economica del paese.

L'indagine è stata condotta a distanza su un campione totale di 2063 famiglie, e le osservazioni ottenute sono da attribuirsi al capofamiglia. Da questo campione sono state ignorate osservazioni ove la risposta a domande di interesse, come le aspettative sui prezzi, era del tipo "Non so", in quanto ritenute di scarso rilievo. Il campione finale ammonta a 1917 individui.

Per ogni rispondente sono state registrate molteplici caratteristiche socioeconomiche e in particolare le risposte a diverse domande legate alle aspettative sull'andamento dei prezzi e alla situazione economico-patrimoniale.

1.2 Caratteristiche socio-economiche

Età

Come si può vedere nel primo panel della figura 1.1 il campione si distribuisce maggiormente tra i 40 e i 70 anni, con un'età mediana di 55 anni (linea blu tratteggiata). Il campione è infatti stato estratto al fine di rappresentare al meglio la popolazione italia-

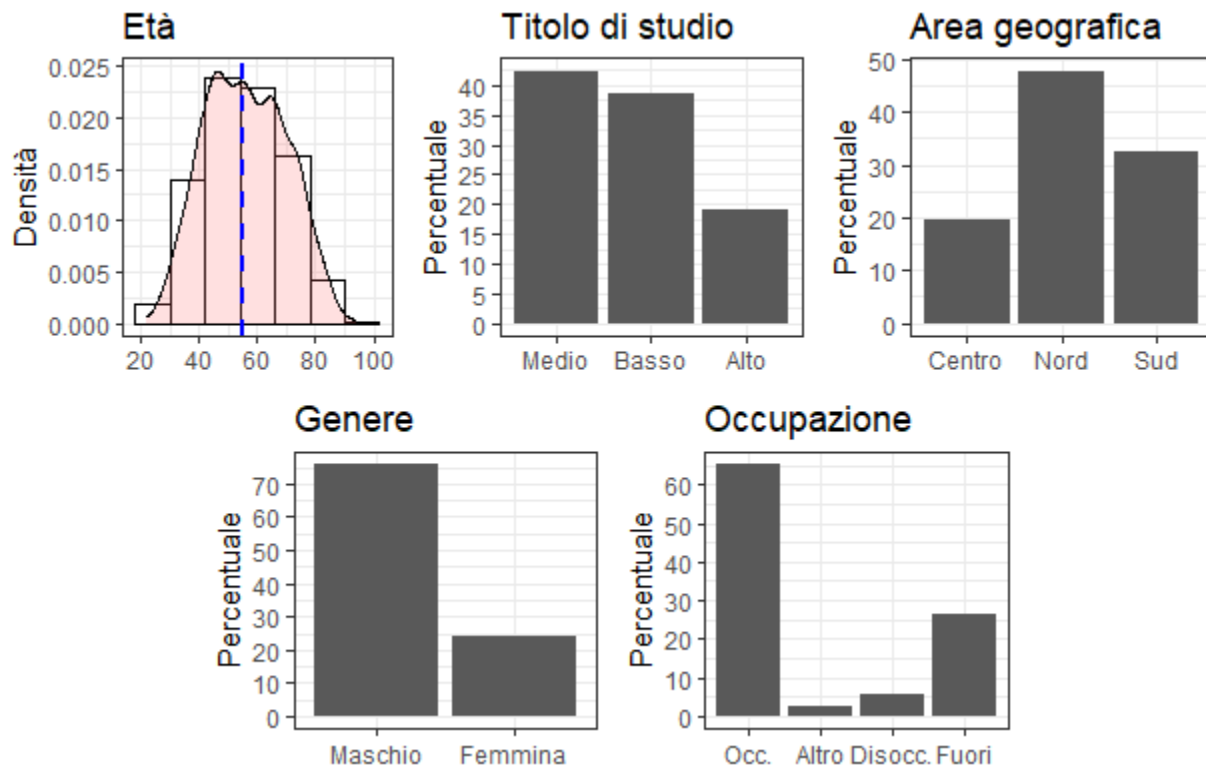


Figura 1.1: Caratteristiche socio-economiche
Fonte dati: Banca d'Italia (2021)

na. Sar  quindi importante capire come e se questo fattore influenzer  le variabili che verranno analizzate.

Titolo di studio

L'istruzione   stata categorizzata in 3 gruppi in base al livello: Basso, Medio e Alto, che coincidono rispettivamente con l'ottenimento del diploma di scuola media, di scuola superiore e in ultimo di un titolo di studio universitario. Il diploma di scuola superiore, ossia il livello "Medio", sar  la categoria di riferimento.

Si pu  notare nel secondo grafico a barre nella figura 1.1 che tra i rispondenti circa il 35% possiede al massimo il diploma di scuola media, mentre solo il 19% circa possiede almeno un titolo di studio universitario.

Area geografica

Un altro fattore da tenere in considerazione tra le caratteristiche socio-economiche delle famiglie italiane   la regione geografica di appartenenza. Lo studio ha considerato l'indi-

viduo come appartenente al nord (Nord) , al centro (Centro) oppure al sud e isole (Sud). La categoria di riferimento per le stime dei modelli sarà il Centro.

Genere

Verrà incluso anche il genere per cogliere eventuali fattori socio-culturali che potrebbero essere significativi. Siccome i rispondenti coincidono con il capofamiglia, la predominanza del genere maschile, essa stessa una delle conseguenze di questi fattori, rende necessariamente le due classi sbilanciate. Come analizzato da Salas-Eljatib et al. (2017), si dimostra empiricamente che dati sbilanciati sono associati a varianze delle stime dei parametri maggiori, rendendole più incerte e imprecise. Sarà quindi un fattore da considerare durante l'interpretazione dei risultati dei modelli.

Occupazione

L'occupazione del rispondente è stata raggruppata in 4 categorie principali: occupato (Occ.), fuori dal mercato del lavoro, ovvero studenti e pensionati (Fuori), disoccupati (Disocc.), Altro. Gli studenti e i pensionati sono stati raggruppati in un'unica categoria poiché la presenza degli studenti nel campione è estremamente bassa, come si evince dalla tabella 1.1 riportante la distribuzione di frequenze per ogni modalità, e quindi si è ritenuto che potesse portare a risultati non significativi.

Tabella 1.1: Distribuzione di frequenze per occupazione
Fonte dati: Banca d'Italia (2021)

Occupazione	n
Occupato	1249
Altro	53
Disoccupato	112
Fuori	500
Studente	3

È importante notare come anche in questo caso la categoria disoccupato sia particolarmente sbilanciata, nonostante sia quella di maggiore interesse per l'analisi. Come per il genere, potrebbe causare problemi di significatività delle stime.

Reddito

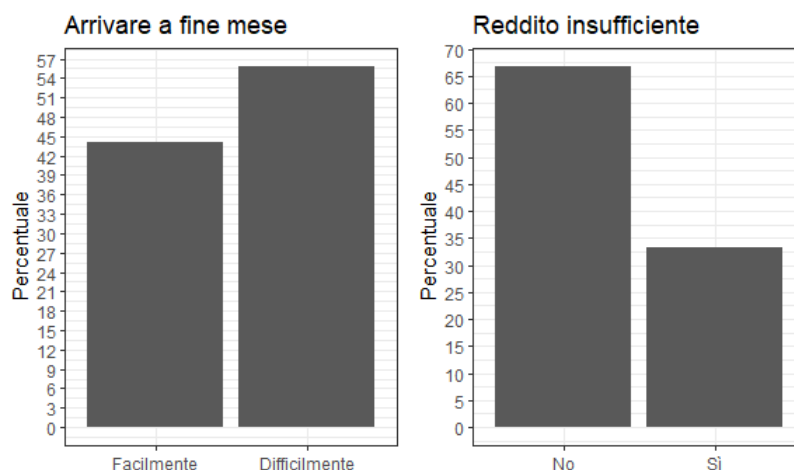


Figura 1.2: Difficoltà nell'arrivare a fine mese e
Insufficienza del reddito a coprire le spese mensili
Fonte dati: Banca d'Italia d'Italia (2021)

Non è stata rilevata alcuna variabile che misuri esplicitamente il reddito dell'individuo, ma sono stati raccolti dati generici sulla capacità di spesa. Si è scelto di concentrarsi in particolare su due domande:

1. Il reddito della sua famiglia le permette di arrivare a fine mese?
2. Negli ultimi 12 mesi le è capitato che il reddito familiare non fosse sufficiente a coprire le spese?

La prima domanda prevedeva una risposta con un grado di intensità da "Con molta difficoltà" a "Molto facilmente", ma la risposta è stata raggruppata in due categorie: "Difficilmente" e "Facilmente", al fine di ridurre il numero di parametri da stimare.

La seconda domanda invece prevedeva una semplice risposta binaria "Sì" o "No".

Si può notare nel primo panel della figura 1.2 che le persone che hanno difficoltà ad arrivare a fine mese superano le persone che ci arrivano facilmente. Il secondo panel mostra come circa il 30% degli intervistati almeno una volta non ha avuto reddito sufficiente per arrivare a fine a mese.

Capitolo 2

Determinanti dei canali informativi

Questo capitolo si concentra sul tema dei canali informativi in materia economica. Ciò che ne motiva lo studio è la plausibile rilevanza che questi potrebbero avere nel processo di formulazione delle aspettative di inflazione. Prima di analizzare questo aspetto è quindi essenziale comprendere cosa spinga un individuo a scegliere un canale informativo piuttosto che un altro.

Per fare ciò, l'analisi si avvarrà delle risposte alla seguente domanda posta nel questionario:

”Quale fra le seguenti fonti di informazione è quella che la Sua famiglia utilizza maggiormente per reperire le informazioni economico-finanziarie rilevanti per voi (ad es. per l'acquisto di beni durevoli, scelte di investimento finanziario, accensione mutuo, etc...)?”

I possibili canali erano qualificati nelle risposte come: Tradizionali, Istituzionali, Social, Contatti con conoscenti, Nessuno. La categoria Tradizionali” è un raggruppamento delle possibili risposte ”Quotidiani” e ”Telegiornali”.

L'analisi confronterà i risultati con quelli ottenuti da Conrad et al. (2022) in merito ad un'analisi simile da loro svolta. La domanda del sondaggio da loro analizzato, condotto da Bundesbank, chiedeva quali fossero i mezzi maggiormente utilizzati per informarsi in merito alla politica monetaria della Banca Centrale Europea. La domanda del sondaggio svolto da Banca d'Italia è più generale e non riguardava nel dettaglio la politica monetaria. Nonostante ciò, si è ritenuto interessante confrontare i risultati ottenuti.

2.1 Si informa?

Inizialmente si cercherà di individuare quali siano i fattori principali nel determinare se l'individuo si informa oppure no. A tale scopo, si sono raggruppate le categorie della risposta originaria in modo da avere una variabile dicotomica con le risposte ”Sì” e ”No”.

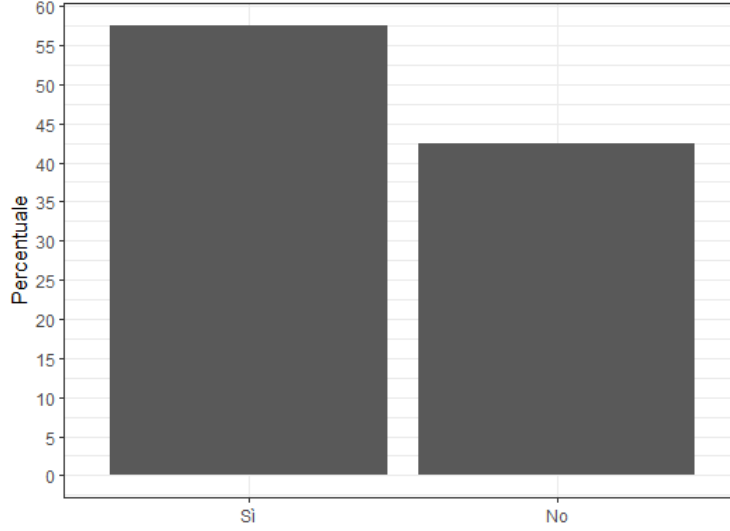


Figura 2.1: Si informa?
Fonte dati: Banca d'Italia (2021)

Nella figura 2.1 si nota come circa il 40% degli individui dichiarati di non informarsi. Le osservazioni sulla variabile risultano quindi abbastanza bilanciate.

2.1.1 Modello di regressione logistica

Per l'analisi di questa variabile si è scelto di utilizzare il modello di regressione logistica, la cui forma analitica è la seguente:

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \alpha + \sum_{i=1}^p \beta_i X_i \quad (2.1)$$

dove $\pi = P(Y = 1 | X_1 = x_1, \dots, X_p = x_p)$ è la probabilità dell'evento di interesse condizionata ai livelli delle altre variabili esplicative. L'evento di interesse in questo caso è costituito dall'evento "Non si informa". π coincide anche con il valore atteso per la generica variabile casuale Bernoulliana Y presa in esame dal modello.

Il modello appartiene alla famiglia dei modelli lineari generalizzati (GLM). Questo tipo di modelli assume che la variabile risposta Y appartenga al Modello di dispersione esponenziale (EDM), definito da un parametro θ , detto canonico, e $\phi > 0$, detto fattore di dispersione. Formalmente, quanto detto si riassume nella seguente espressione: $Y \sim EDM(\theta, \phi)$. La media della variabile casuale Y , μ , è una funzione del parametro canonico. Oltre la Bernoulli, un'altra variabile casuale appartenente al modello di dispersione esponenziale è la variabile casuale Normale. La funzione di probabilità di una variabile che appartiene al modello di dispersione esponenziale è del tipo seguente:

$$\mathcal{P}(y; \theta, \phi) = a(y, \phi) \exp \left\{ \frac{y\theta - k(\theta)}{\phi} \right\} \quad (2.2)$$

Si può derivare per un modello di dispersione esponenziale che:

$$E[Y] = \mu = \frac{dk(\theta)}{d\theta}$$

$$V(Y) = \phi V(\mu) = \phi \frac{d\mu}{d\theta} = \phi \frac{d^2 k(\theta)}{d^2 \theta}$$

Con $V(\mu)$ detta funzione di varianza. Ogni variabile casuale ha la sua $V(\mu)$. Nel caso di una Normale, si deriva che $\phi = \sigma^2$ e $V(\mu) = 1$. Nel caso Bernoulliano, $\phi = 1$ e $V(\mu) = \mu(1 - \mu) = \pi(1 - \pi)$

Un generico GLM ha la seguente forma analitica:

$$g(\mu) = \alpha + \sum_{i=1}^p \beta_i X_i \quad (2.3)$$

La funzione $g(\cdot)$ è detta funzione link, ed è supposta essere una funzione monotona, derivabile e invertibile. Se la funzione link coincide con θ , allora la funzione link si definisce canonica. Nel caso della Normale, $\theta = y$, quindi la funzione link canonica è la funzione identità. Nel caso della Bernoulli, $\theta = \log\left(\frac{\pi}{1-\pi}\right) = \text{logit}(\pi)$. La scelta della funzione link dipende dal caso particolare. Potrebbe risultare migliore la scelta di una funzione link diversa da quella canonica. Un'altra funzione link comunemente usata per un GLM nel caso di variabili Bernoulliane è $g(\cdot) = \Phi^{-1}$, con Φ la funzione di ripartizione della Normale. In quel caso si parla però di regressione probit e non logistica.

La stima dei parametri nei modelli GLM avviene tramite la massimizzazione della verosimiglianza. Si dimostra che nel caso $Y \sim N(\mu, \sigma^2)$, la stima di massima verosimiglianza coincide con la stima OLS. Il modello di regressione lineare è quindi un caso particolare dei GLM.

I coefficienti del modello 2.1 sono in logits, ovvero indicano l'incremento nei log-odds all'aumentare di un'unità della variabile indipendente (se continua), oppure all'appartenere ad una specifica classe rispetto a quella di riferimento (se la variabile è categorica), mantenuti costanti gli altri regressori. Con log-odds si intende il logaritmo degli odds, e con odds si intende il seguente rapporto di probabilità:

$$\text{Odds}(Y = 1 | X_1 = x_1, \dots, X_p = x_p) = \frac{P(Y = 1 | X_1 = x_1, \dots, X_p = x_p)}{1 - P(Y = 1 | X_1 = x_1, \dots, X_p = x_p)} \quad (2.4)$$

I coefficienti della regressione logistica coincidono anche con i logaritmi degli Odds Ratios (OR), cioè il rapporto degli Odds al variare di una variabile esplicativa, mantenute costanti le altre. Nell'equazione 2.5 è riportata la formula per un generico OR associato al variare di una generica variabile X_i :

$$OR = \frac{Odds(Y = 1|X_1 = x_1, \dots, X_i = x_i + 1, \dots, X_p = x_p)}{Odds(Y = 1|X_1 = x_1, \dots, X_i = x_i, \dots, X_p = x_p)} \quad (2.5)$$

Ad un coefficiente pari a zero corrisponde un OR pari a 1. Ad un coefficiente positivo un OR maggiore di 1 e minore di 1 nel caso in cui il coefficiente sia negativo. Coefficienti positivi o OR maggiori di 1 indicano un'associazione positiva tra le variabili, mentre vale l'opposto se sono rispettivamente minori di 0 e di 1. Coefficienti nulli e OR pari a 1 indicano assenza di associazione.

Dal risultato di una regressione logistica è sempre possibile estrarre le probabilità dell'evento di interesse invertendo la funzione link canonica, ottenendo il risultato 2.6:

$$\pi = P(Y = 1|X_1 = x_1, \dots, X_p = x_p) = \frac{\exp(\alpha + \sum_{i=1}^p \beta_i x_i)}{1 + \exp(\alpha + \sum_{i=1}^p \beta_i x_i)} \quad (2.6)$$

La funzione di probabilità ricavata non ha un andamento lineare, bensì sigmoidale. Ciò implica che al variare della generica variabile X_i la funzione non varia in maniera costante su tutto il dominio. In particolare, nel caso di un solo regressore continuo X , con coefficiente $\beta > 0$, il suo grafico è del tipo riportato in figura 2.2.

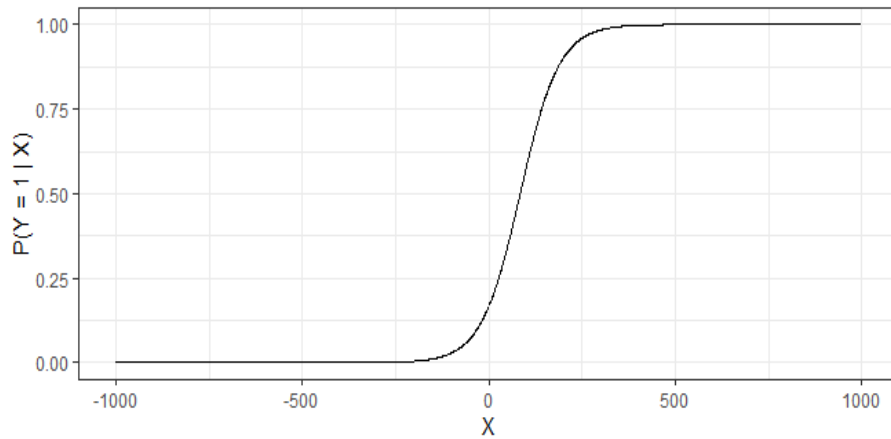


Figura 2.2: Funzione di probabilità sigmoidale

2.1.2 Diagnostica

I modelli lineari generalizzati (GLM) si fondano su delle ipotesi essenziali ai fini della stima dei parametri. Come ricordano Dunn and Smyth (2018), sono le seguenti:

- Assenza di "outliers", valori anomali
- Funzione link corretta
- Linearità: il predittore lineare è corretto e include tutte le variabili rilevanti
- Le realizzazioni della variabile risposta Y sono indipendenti tra di loro
- La variabile risposta Y segue la distribuzione ipotizzata. Nel caso della regressione logistica su variabili binarie, una variabile casuale Bernoulliana.
- La funzione di varianza $V(\mu)$ ipotizzata è corretta.
- Il parametro ϕ è costante

Uno strumento per verificare se le ipotesi sono valide, o quanto ci si discosta dalla loro validità, sono i residui. Tra tutti i tipi, i più semplici sono quelli definiti come $y_i - \hat{\mu}_i$, con $\hat{\mu}_i$ la previsione operata dal modello. Questi residui sono utilizzati sovente per la diagnostica di regressioni lineari con variabile risposta distribuita normalmente, dove sono assunti avere media zero e varianza costante. Per i modelli GLM però questi residui si dimostrano non essere adatti, in quanto la varianza di una variabile appartenente ad EDM è generalmente non costante. Un'eccezione è il caso della Normale, in quanto la sua varianza è costante e uguale a σ^2 . Ciò giustifica infatti l'utilizzo di questi residui per il modello di regressione lineare su variabili Normali.

Un tipo di residui sovente utilizzati nella diagnostica dei GLM, e che ovviano a questo problema, sono i residui definiti, in inglese, *deviance residuals*.

I *deviance residuals* sono, intuitivamente, interpretabili come il peso che ha ogni osservazione nel determinare la devianza residua del modello, definita come:

$$D_r = 2[\ell_{saturato}(\hat{\beta}) - \ell_{proposto}(\hat{\beta})]$$

Dove $\ell_{saturato}(\hat{\beta})$ è la log-verosimiglianza del modello saturo. La devianza residua rappresenta una misura di bontà d'adattamento del modello. Minore è la devianza, maggiore è la bontà d'adattamento.

Si può verificare che $D_r = \sum_{i=1}^n r_{Di}^2$, dove r_{Di} sono i *deviance residuals* associati alle i -esime unità.

Se le ipotesi sono verificate, i *deviance residuals* dovrebbero seguire una distribuzione normale, con media zero e varianza costante. Valori anomali dei *deviance residuals* indicano la probabile presenza di outliers.

Un'alternativa ai *deviance residuals* è costituita dai residui quantile, proposti sempre da Dunn and Smyth (1996) e definiti come:

$$r_Q = \Phi^{-1}[\mathcal{F}(y; \hat{\mu}, \hat{\phi})]$$

Dove \mathcal{F} è la funzione di ripartizione della variabile casuale di interesse e Φ la funzione di ripartizione di una variabile casuale normale standardizzata. Nel caso di variabili casuali discrete, viene scelto un valore della funzione di ripartizione in modo che sia casuale, con distribuzione uniforme in un intervallo compreso tra i due valori che la funzione di ripartizione \mathcal{F} assume per y . Questo permette di avere residui quantili continui anche per variabili casuali discrete (Dunn and Smyth (1996)).

Questi residui possiedono il vantaggio di essere distribuiti normalmente per costruzione, e quindi permettono di ovviare ad un problema che si verifica sovente nel caso di variabili risposta discrete, come il caso Bernoulliano, Binomiale o di Poisson. Il problema in questione riguarda la situazione in cui tutti i tipi di residui, tra cui i *deviance residuals*, si sviluppano su rette parallele tante quante sono le modalità della variabile, rendendone l'interpretazione e la diagnostica difficili da eseguire. Quando tale situazione accade, conviene passare ai residui quantili, rendendoli la scelta consigliata per variabili casuali discrete. Dunn and Smyth (2018) suggeriscono di utilizzarli soprattutto quando si vuole verificare la presenza di trend o andamenti non lineari, indici di possibile mal specificazione del modello. Suggeriscono inoltre di riprodurre i residui quantile almeno 4 volte nel caso di variabili casuali discrete, per verificare quali elementi sono dettati dall'aleatorietà e quali invece non sono aleatori.

2.1.3 Analisi

Nella tabella 2.1 è riportato il risultato del modello di regressione logistica applicato a tutte le variabili analizzate nel capitolo precedente. Dai risultati del modello si evince che le variabili più significative sono: Titolo di studio, Età, Arrivare a fine mese difficilmente.

Come ci si potrebbe aspettare, si nota che un livello di istruzione basso è associato ad una maggiore probabilità di non informarsi. Anche chi ha difficoltà ad arrivare a fine mese, sintomo di reddito basso, è maggiormente probabile che non si informi.

Come ci si aspettava dall'analisi preliminare, inoltre, i coefficienti associati alla modalità Femmina della variabile Genere e alla modalità Disoccupato della variabile Occu-

Tabella 2.1: Modello logistico generale
Fonte dati: Banca d'Italia (2021)

Variabile	Coefficiente	Std. Error	p-value	
Intercetta	-1.389	0.262	0.000	***
Titolo: Basso	0.321	0.108	0.003	**
Titolo: Alto	-0.474	0.140	0.001	***
Età	0.014	0.004	0.002	**
Femmina	0.157	0.114	0.169	
Area: Nord	-0.096	0.129	0.454	
Area: Sud	0.114	0.138	0.407	
Occ: Altro	0.111	0.291	0.703	
Occ: Disoccupato	-0.175	0.208	0.401	
Occ: Fuori	0.253	0.138	0.067	.
Reddito insufficiente almeno una volta	0.083	0.114	0.464	
Arriva a fine mese difficilmente	0.264	0.111	0.017	*
AIC: 2532.6				

pazione sono risultati non significativi, molto plausibilmente per la distribuzione particolarmente sbilanciata.

Anche l'area geografica sembra non essere significativa in questo modello.

La variabile Reddito insufficiente almeno una volta è particolarmente associata alla variabile Arrivare a fine mese difficilmente, quindi la presenza di un regressore analogo potrebbe essere una delle motivazioni per la quale risulta non significativa.

Successivamente si è sfruttato un algoritmo che potesse aiutare a scegliere un sottoinsieme delle variabili senza che il modello perdesse potenziale esplicativo. In particolare, è stato usato un algoritmo di backward stepwise selection al fine di minimizzare l'Akaike Information Criteria. La motivazione che spinge a farlo è generalmente la riduzione del rischio di overfitting del modello.

Il risultato coincide sostanzialmente con quanto si sarebbe potuto raggiungere includendo le variabili maggiormente significative. Infatti, le variabili scelte dall'algoritmo sono state il Titolo di studio, l'Età e Arrivare a fine mese difficilmente. Nella tabella 2.2 è riportato il modello con le stime dei coefficienti per le variabili sopra menzionate.

Lo stesso algoritmo è stato utilizzato per testare l'esistenza di effetti d'interazione significativi tra le variabili. Per l'applicazione, in particolare, si è partiti dal sottoinsieme estratto con il modello ridotto, inserendo anche come possibile effetto d'interazione l'area geografica, nonostante non fosse risultata significativa inizialmente.

Il modello con le interazioni ha ridotto l'AIC, ma di 0.1 unità. Ciò implica che il miglioramento nella bontà d'adattamento è trascurabile, considerando l'aggiunta di com-

Tabella 2.2: Modello ridotto
Fonte dati: Banca d'Italia (2021)

Variabile	Coefficiente	Std. Error	p-value	
Intercetta	-1.606	0.220	0.000	***
Titolo: Basso	0.328	0.104	0.002	**
Titolo: Alto	-0.464	0.139	0.001	***
Età	0.019	0.003	0.000	***
Arriva a fine mese difficilmente	0.318	0.097	0.001	**
AIC: 2529				

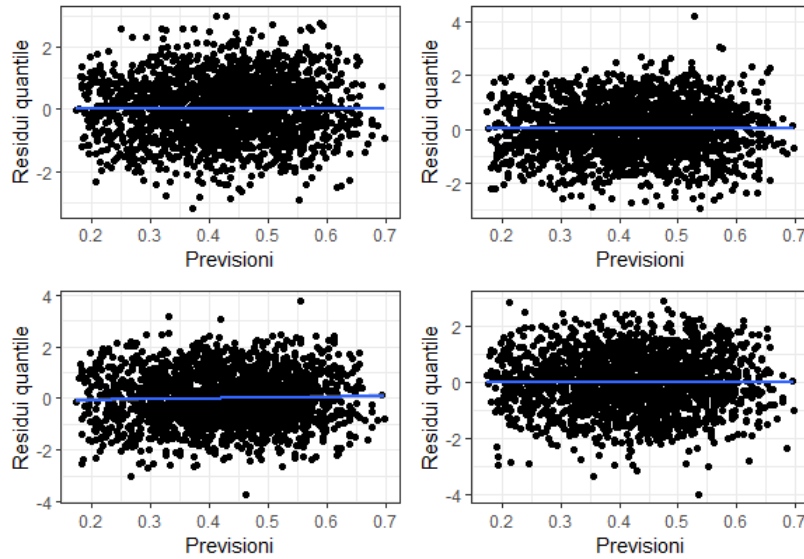


Figura 2.3: Residui quantile del modello logistico ridotto

pietà data dalla stima di nuovi coefficienti. Si è ritenuto perciò opportuno considerare solo il modello senza interazioni.

Per il motivo già evidenziato, i deviance residuals di tale modello sono risultati poco interpretabili, in quanto si distribuivano lungo 2 rette parallele.

Si riportano perciò i risultati di 4 iterazioni dei residui quantile nella figura 2.3. I residui sembrano indicare una corretta specificazione del modello, in quanto non si osservano andamenti non lineari o caratteristiche anomale nella loro distribuzione.

Per misurare come al variare dei livelli dei regressori, mantenuti costanti gli altri ad un livello di riferimento, vengano influenzate le probabilità di non informarsi, sono state predette le probabilità per diverse classi di individui, utilizzando la formula 2.6. I livelli di riferimento scelti per i regressori sono i seguenti:

- Età = 30

- Titolo = Medio
- Difficoltà ad arrivare a fine mese = Facilmente

Tabella 2.3: Probabilità di non informarsi per titolo di studio
Arriva a fine mese facilmente, Età = 30

Titolo	Non si informa	Si informa
Alto	0.183	0.817
Medio	0.262	0.738
Basso	0.330	0.670

Nella tabella 2.3 si nota come la differenza di probabilità di non informarsi per individui dal titolo di studio basso, rispetto al titolo di studio alto, è di circa il 15%. Questo è in accordo con Conrad et al. (2022), in quanto anche l'aumentare della loro variabile Anni di istruzione è associato ad una maggiore probabilità di informarsi.

Tabella 2.4: Probabilità di non informarsi per condizione economica
Titolo = Medio, Età = 30

Arriva a fine mese	Non si informa	Si informa
Facilmente	0.262	0.738
Difficilmente	0.328	0.672

Nella tabella 2.4 si nota che persone che fanno fatica ad arrivare a fine mese è più probabile che non si informino. Anche questo risultato è in accordo con Conrad et al. (2022), siccome la variabile rappresenta una proxy per il livello di reddito della famiglia, che hanno riscontrato essere positivamente associato alla probabilità di informarsi.

Tabella 2.5: Probabilità di non informarsi per età
Arriva a fine mese facilmente, Titolo = Medio

Età	Non si informa	Si informa
20	0.227	0.773
30	0.262	0.738
40	0.301	0.699
50	0.343	0.657
60	0.387	0.613
70	0.433	0.567
80	0.480	0.520

Nella tabella 2.5 si riscontra che all'aumentare dell'età, la probabilità di non informarsi aumenta. La differenza tra un individuo di 20 anni e un individuo di 70 è circa del 20%.

Questo risultato è in disaccordo con quello ottenuto da Conrad et al. (2022), i quali hanno riscontrato associazione negativa tra età e probabilità di non informarsi

2.2 Scelta del canale informativo

La sezione che segue ha lo scopo di analizzare le stesse variabili, al fine di comprendere se siano in grado di spiegare la scelta del canale informativo utilizzato dagli individui.

La distribuzione delle risposte è riassunta nella figura 2.4. Da queste è stata rimossa la modalità Nessuno, in modo da soffermarsi solo sui mezzi specifici e non ripetere i risultati ottenuti con la sezione precedente. La modalità Tradizionali risulta essere la più frequente, mentre le altre risultano essere abbastanza bilanciate tra di loro

Si noti che la variabile Mezzo informativo si distribuisce come una variabile casuale multinomiale con $J = 4$ modalità. Più in particolare, le modalità sono state assunte come nominali, ovvero non ordinabili. Si è ritenuto, infatti, che i tipi di canali informativi non presentassero un ordine naturale secondo il quale fossero facilmente ordinabili.

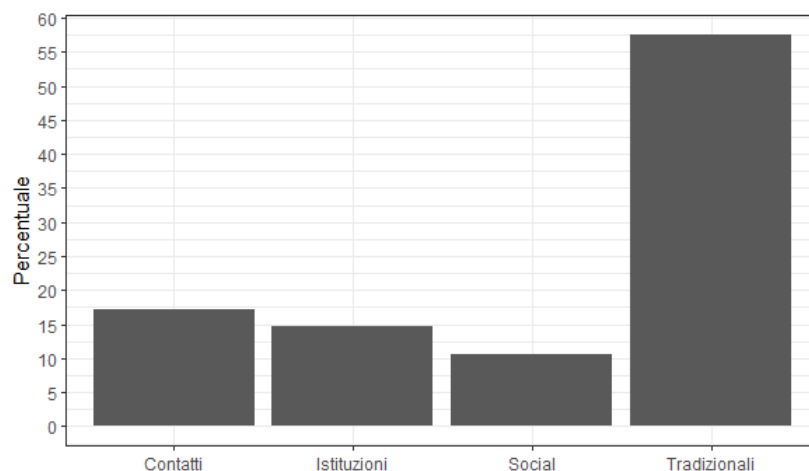


Figura 2.4: Mezzi informativi
Fonte dati: Banca d'Italia (2021)

2.2.1 Modello di regressione multinomiale logistica

Uno dei modelli più comuni per l'analisi della suddetta classe di variabili è il modello di regressione multinomiale logistica, anch'esso appartenente alla famiglia dei modelli lineari generalizzati. Si tratta di una generalizzazione del modello di regressione logistica, applicata a casi in cui le variabili risposta qualitative presentino $J > 2$ possibili modalità.

In particolare, per ogni j -esima modalità il modello stima una regressione logistica in cui i logits sono calcolati rispetto alla J -esima modalità, ovvero quella di riferimento.

La scelta della modalità di riferimento è arbitraria, ma per convenzione viene scelta l'ultima modalità tra quelle presentate oppure la più frequente (Agresti (2002)). Alcuni software seguono questa convenzione; altri, come il software R utilizzato per l'analisi, prendono come riferimento la prima categoria nell'ordine con cui si presentano. La categoria può essere sempre cambiata in funzione dell'interesse del ricercatore, e quindi non rappresenta un vincolo reale. Nella notazione che segue, per definire il modello si è scelto di considerare la categoria di riferimento come l'ultima (J), allineandosi alla notazione scelta da Agresti (2002).

Il modello stimato per ogni j è quindi quello riportato nella formula 2.7, dove π_j è la probabilità di appartenere alla j -esima categoria e π_J di appartenere alla J -esima categoria di riferimento, entrambe condizionate ai livelli delle variabili esplicative.

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \sum_{i=1}^p \beta_{ji} X_i, \quad j \in [1, J-1] \quad (2.7)$$

L'interpretazione dei coefficienti è quindi simile al modello 2.1: l' i -esimo coefficiente β_{ji} è l'incremento del logaritmo del rapporto di probabilità π_j/π_J all'aumentare di un'unità della variabile X_i , mantenendo le altre variabili esplicative costanti.

Come per ogni regressione logistica, si possono estrarre le probabilità π_j con l'inversione della funzione link, ottenendo il risultato 2.8:

$$\pi_j = P(Y = j | X_1 = x_1, \dots, X_p = x_p) = \frac{\exp(\alpha_j + \sum_{i=1}^p \beta_{ji} X_i)}{1 + \sum_{h=1}^{J-1} \exp(\alpha_h + \sum_{i=1}^p \beta_{hi} X_i)} \quad (2.8)$$

Si ricorda che per costruzione $\sum_{j=1}^J \pi_j = 1$. Si può derivare quindi che:

$$\pi_J = P(Y = J | X_1 = x_1, \dots, X_p = x_p) = 1 - \sum_{j=1}^{J-1} \pi_j \quad (2.9)$$

2.2.2 Test del rapporto di verosimiglianza

Un'applicazione utile del test del rapporto di verosimiglianza è la verifica della significatività di una variabile (o più) in un modello proposto dal ricercatore.

Si definisce modello *nested* il modello senza la variabile (o le variabili) di cui si vuole testare la significatività. Le variabili incluse al suo interno devono quindi essere un sottoinsieme delle variabili incluse nel modello proposto.

Si vuole testare l'ipotesi nulla H_0 che la bontà d'adattamento del modello proposto e quella del modello *nested* siano uguali. L'ipotesi alternativa H_1 è che il modello proposto abbia una bontà d'adattamento maggiore.

Se si rifiuta l'ipotesi nulla, allora la variabile rimossa nel modello *nested* è significativa, ovvero il modello proposto è migliore. Se non si rifiuta l'ipotesi nulla, allora la variabile rimossa non è significativa, quindi è preferibile usare il modello *nested*. Questo perché a parità di bontà d'adattamento, il modello *nested* stima meno coefficienti.

La 2.10 definisce il rapporto di verosimiglianza, costituito dal rapporto delle due funzioni di verosimiglianza per il modello proposto e il modello *nested*.

$$\Lambda = \frac{L(\text{Modello proposto})}{L(\text{Modello nested})} = \frac{L_{\text{proposto}}}{L_{\text{nested}}} \quad (2.10)$$

La statistica LR utilizzata dal test è riportata nella 2.11. LR si distribuisce asintoticamente come una variabile casuale χ^2 , con gradi di libertà q pari alla differenza dei gradi di libertà dei due modelli.

Per un p-value minore di α , con α la soglia di significatività, si rifiuta l'ipotesi nulla.

$$LR = 2 \log(\Lambda) = 2 (\log(L_{\text{proposto}}) - \log(L_{\text{nested}})) \sim \chi_q^2 \quad (2.11)$$

2.2.3 Analisi

L'analisi della scelta del canale informativo si svilupperà in due parti: la prima si concentra sui fattori che spiegano la scelta tra i media Tradizionali e gli altri, accorpati in un'unica modalità Altro; la seconda invece indagherà sui fattori rilevanti nella scelta tra gli altri mezzi di informazione, ignorando a quel punto i media Tradizionali.

Per questa analisi, al fine di facilitare la visualizzazione di alcuni grafici, la variabile età è stata scalata dividendola per 5, al fine di ottenere coefficienti che si riferiscono all'effetto dovuto al variare di 5 anni d'età. Questo moltiplicherà i coefficienti per 5, rendendo le visualizzazioni più interpretabili.

Tradizionali e Altro

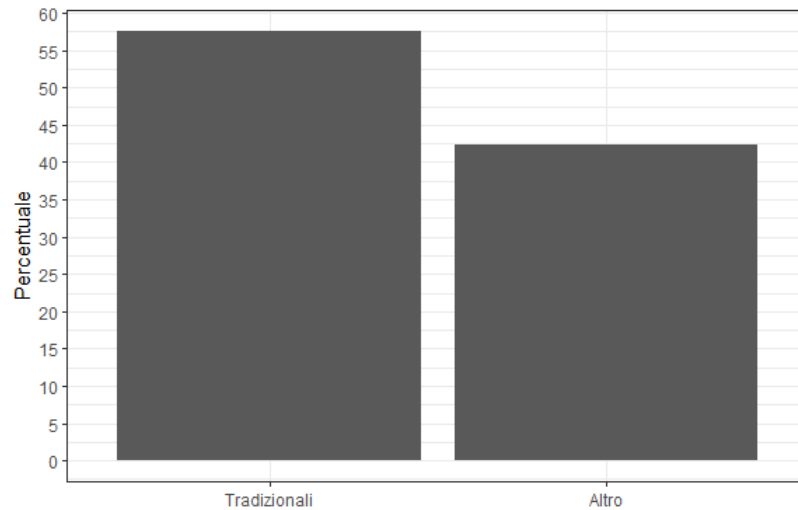


Figura 2.5: Distribuzione di frequenza per i mezzi Tradizionali e non
Fonte dati: Banca d'Italia (2021)

La distribuzione della variabile è riassunta nella figura 2.5. La categoria scelta come riferimento è Tradizionali.

Tabella 2.6: Modello di regressione logistica per Tradizionali e Altro
Fonte dati: Banca d'Italia (2021)

Variabile	Coefficiente	Std. Error	p-value	
Titolo basso	-0.082	0.148	0.579	
Titolo alto	-0.168	0.160	0.295	
Età (5 anni)	-0.123	0.028	0.000	***
Arriva a fine mese difficilmente	0.092	0.145	0.528	
Femmina	0.333	0.148	0.024	*
Occ: Altro	-0.076	0.406	0.851	
Occ: Disoccupato	-0.036	0.276	0.897	
Occ: Fuori	0.324	0.192	0.091	.
Area: Nord	-0.192	0.163	0.239	
Area: Sud	-0.310	0.182	0.088	.
Reddito insufficiente almeno una volta	-0.153	0.156	0.325	
AIC: 1496.2				

Il modello utilizzato per questa variabile è il medesimo sfruttato nella sezione precedente, ovvero il modello di regressione logistica 2.1.

Dalla tabella 2.6 si nota che le uniche variabili significative al 5% sono Età e Genere Femmina. Gli intervalli di confidenza al 10% nella figura 2.6 suggeriscono anche signifi-

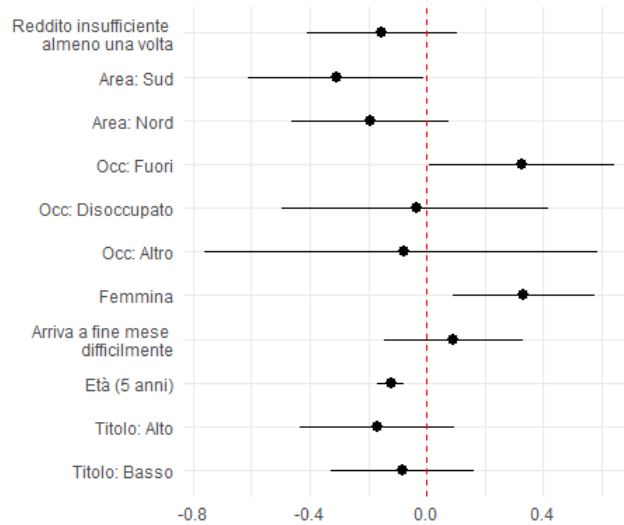


Figura 2.6: Modello di regressione logistica per Tradizionali e Altro.

Stime puntuali e intervalli di confidenza al 10%

Fonte dati: Banca d'Italia (2021)

cattività a questo livello per la modalità Sud della variabile Area e la modalità Fuori della variabile Occupazione.

Il grado di difficoltà ad arrivare a fine mese, risultato significativo nel determinare se l'individuo si informa oppure no, non sembra spiegare la scelta tra l'utilizzo di media tradizionali e non. Anche il titolo di studio, similmente una delle variabili più rilevanti, risulta non essere significativo. Questo sembra essere in contrasto con quanto individuato da Conrad et al. (2022), in quanto l'aumentare del livello di reddito degli individui e degli anni di istruzione sembra essere significativamente associato ad un maggiore utilizzo dei media tradizionali.

Si è scelto di ridurre anche in questo caso il numero di regressori inseriti nel modello, prima di utilizzarlo per predire le probabilità d'utilizzo dei mezzi al variare dei regressori. L'algoritmo di minimizzazione del AIC ha suggerito un modello con le sole variabili Genere ed Età, scartando l'utilizzo delle variabili Occupazione e Area geografica. Anche i LR test, riportati negli script 2.1 e 2.2, suggeriscono che l'utilizzo di tali variabili non migliora significativamente la bontà di adattamento del modello.

Nonostante ciò, si è ritenuto che inserire la variabile Occupazione potesse fornire delle interpretazioni significative in merito alla modalità Fuori, mentre inserire la variabile Area Geografica potesse essere interpretata per la modalità Sud. I coefficienti e gli standard error di tutti i regressori rimangono sostanzialmente invariati se le due variabili sono omesse oppure no, e quindi ciò significa che inserirle non genera distorsioni nelle stime. Entrambe le variabili andranno però analizzate ricordando che non tutte le modalità sono

Script 2.1: LR test su Occupazione

Likelihood ratio test

Model 1: info_mb ~ eta + sesso + occ + area

Model 2: info_mb ~ eta + sesso + area

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	8	-737.17			
2	5	-738.80	-3	3.2698	0.3519

Script 2.2: LR test su Area

Likelihood ratio test

Model 1: info_mb ~ eta + sesso + occ + area

Model 2: info_mb ~ eta + sesso + occ

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	8	-737.17			
2	6	-738.64	-2	2.9397	0.23

risultate significative.

Il risultato del modello, con relativo AIC, è riportato nella tabella 2.7.

Tabella 2.7: Modello di regressione logistica ridotto
Fonte dati: Banca d'Italia (2021)

Variabile	Coefficiente	Std. Error	p-value
Età	-0.024	0.006	0.000
Femmina	0.323	0.145	0.026
Occ: Altro	-0.054	0.405	0.893
Occ: Disoccupato	-0.027	0.272	0.920
Occ: Fuori	0.331	0.191	0.083
Area: Nord	-0.184	0.162	0.256
Area: Sud	-0.308	0.180	0.087
AIC: 1490.3			

Nella figura 2.7 sono riportati 4 risultati dei residui quantile per tale modello, tutti indicanti l'assenza di anomalie e mal specificazioni.

Il modello è stato quindi utilizzato per predire le probabilità associate ai mezzi tradizionali o non tradizionali al variare di uno dei regressori, con gli altri mantenuti ad un livello di riferimento scelto. In questo caso, i riferimenti erano i seguenti: Età = 30, Genere = Maschio, Occupazione = Occupato, Area = Nord. Nel calcolo delle probabilità al variare dell'occupazione, però, si è deciso di considerare l'individuo di riferimento con

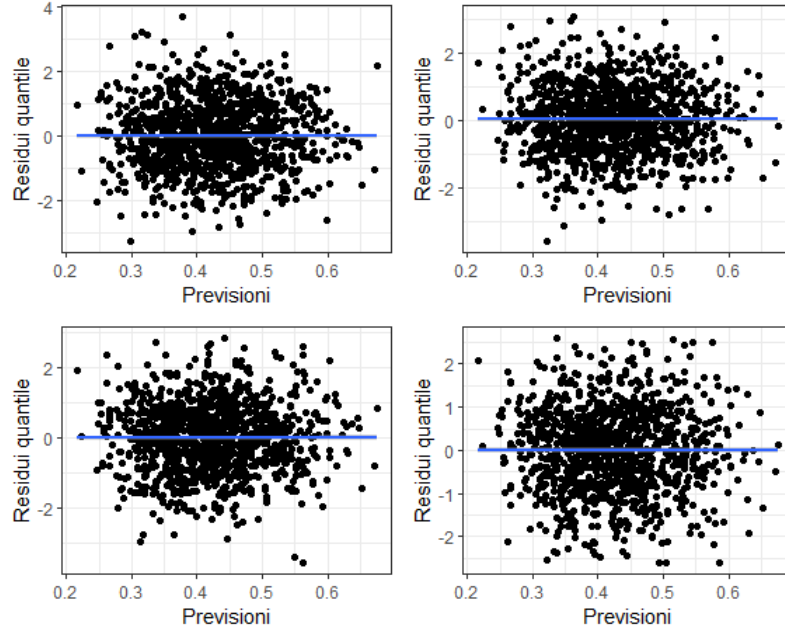


Figura 2.7: Residui quantili per il modello ridotto su Tradizionali e Altro

Età = 60. Ciò è stato motivato dall'impossibilità di interpretare l'effetto di essere Fuori dal mercato dal lavoro considerando un individuo di 30 anni, siccome è una categoria costituita essenzialmente da individui anziani, presumibilmente quindi in pensione.

Nella figura 2.8 è riportata la funzione di probabilità per le due modalità della variabile risposta al variare dell'età. Come si può vedere, la probabilità di utilizzo di mezzi tradizionali aumenta all'aumentare dell'età, seguendo quasi un andamento lineare in questa sezione della funzione sigmoideale. Questo risultato è in accordo con Conrad et al. (2022).

Tabella 2.8: Probabilità di utilizzo dei mezzi d'informazione al variare dell'area geografica

Età = 30, Occupazione = Occupato, Genere = Maschio
Fonte dati: Banca d'Italia (2021)

Area	Altro	Tradizionali
Nord	0.532	0.468
Centro	0.577	0.423
Sud	0.501	0.499

Dalla tabella 2.8 si nota che l'effetto di appartenere al Sud risulta non essere rilevante di per sé nella scelta tra i due mezzi di informazione. Risulta però vero che al Sud ci si informa di più con i media tradizionali rispetto al Nord e al Centro.

Figura 2.8: Funzione di probabilità al variare dell'età
 Genere = Maschio, Occupazione = Occupato, Area = Nord
 Fonte dati: Banca d'Italia (2021)

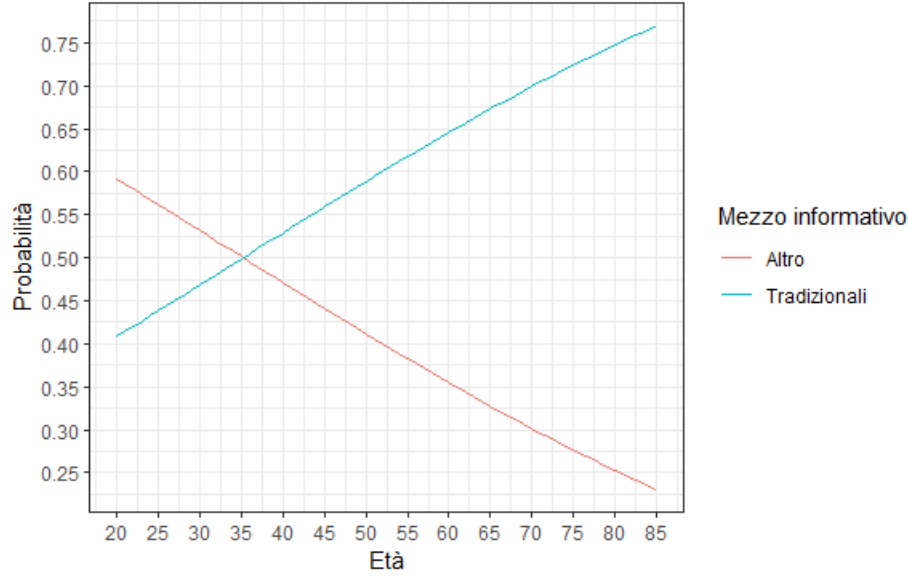


Tabella 2.9: Probabilità di utilizzo dei mezzi d'informazione al variare del genere
 Età = 30, Occupazione = Occupato, Area = Nord
 Fonte dati: Banca d'Italia (2021)

Sesso	Altro	Tradizionali
Maschio	0.532	0.468
Femmina	0.611	0.389

Dalla tabella 2.9 si evince invece che le donne si informano con molta meno probabilità con i mezzi tradizionali, sia in termini assoluti, sia rispetto agli uomini. Anche questo risultato è in accordo con Conrad et al. (2022). La differenza di probabilità per l'utilizzo di questo mezzo è infatti circa l'8% per quanto riguarda il confronto con gli uomini.

Dalla tabella 2.10 si evince che anche gli individui fuori dal mercato del lavoro, che ricordiamo essere essenzialmente pensionati, si informano con meno probabilità tramite media tradizionali rispetto agli altri individui.

Tabella 2.10: Probabilità di utilizzo dei mezzi d'informazione al variare dell'occupazione
 Età = 60, Genere = Maschio, Area = Nord
 Fonte dati: Banca d'Italia (2021)

Occupazione	Altro	Tradizionali
Occ.	0.354	0.646
Fuori	0.433	0.567
Disocc.	0.348	0.652
Altro	0.342	0.658

Social, Istituzioni e Contatti

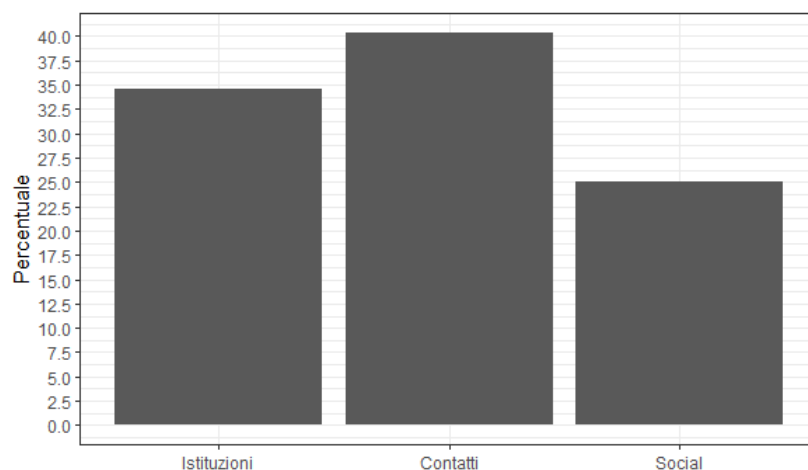


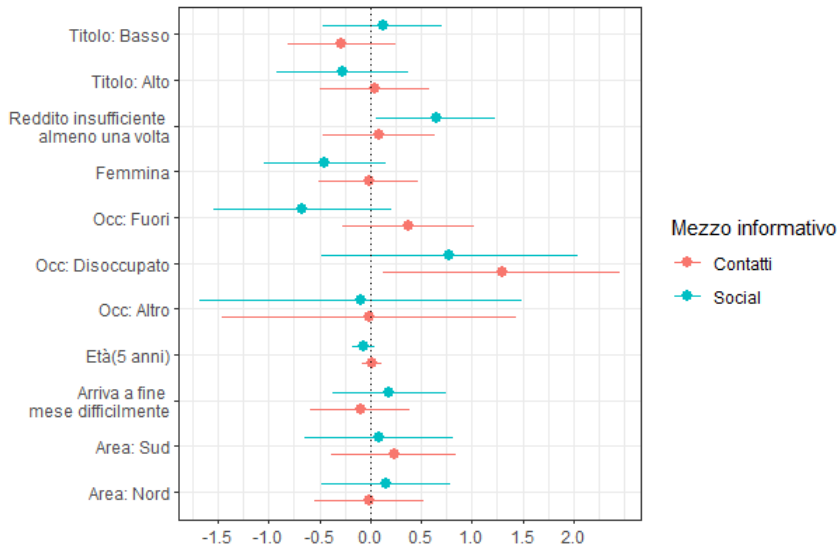
Figura 2.9: Distribuzione di frequenza dei media Istituzioni, dei Social e dei Contatti
 Fonte dati: Banca d'Italia (2021)

Nella figura 2.9 è riportata la distribuzione di frequenza dei restanti mezzi informativi ancora non analizzati. La distribuzione dei mezzi rimuovendo sia Nessuno che Tradizionali risulta ora più bilanciata rispetto a quella visibile nella figura 2.4.

Il modello utilizzato su questa nuova variabile, con $J = 3$ modalità, è il modello di regressione multinomiale logistica 2.7. La modalità scelta come riferimento è Istituzionali.

Dal risultato del modello, rappresentato nella figura 2.10, si nota che poche sono le variabili risultate significative. Tra queste appartengono soltanto le variabili Reddito insufficiente almeno una volta per la modalità Social e Occupazione con la modalità Disoccupato per la modalità Contatti. Anche in questo caso, Titolo di studio risulta non significativa assieme ad Età, variabile che fino a questo momento è sempre risultata significativa.

Figura 2.10: Modello di regressione multinomiale logistica
Stime e intervalli di confidenza al 95%
Fonte dati: Banca d'Italia (2021)



Nello script 2.3 è riportato il risultato del test del rapporto di verosimiglianze per la variabile Età. Anche questo test ne conferma la non significatività.

Script 2.3: LR test su età

Likelihood ratio test

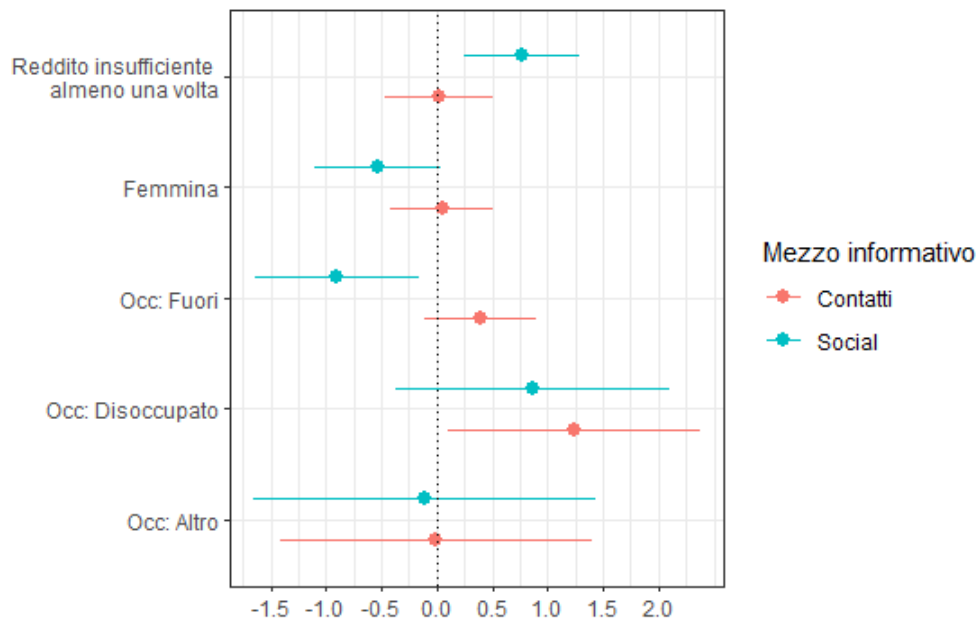
Model 1: `info_m ~ titolo + I(eta/5) + area + occ + condgennow + sesso + Reddito insufficiente almeno una volta`

Model 2: `info_m ~ titolo + area + occ + condgennow + sesso + Reddito insufficiente almeno una volta`

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	24	-484.27			
2	22	-485.31	-2	2.0754	0.3543

Provando a ridurre la dimensionalità del modello, minimizzando l'AIC, otteniamo un modello per cui anche la variabile genere risulta quasi significativa al 5% per la modalità Social. Oltretutto, anche la modalità Fuori della variabile Occupazione acquista maggiore significatività escludendo alcune variabili dal modello. Pare, inoltre, che la significatività di Reddito insufficiente almeno una volta, già oltre la soglia, sia aumentata leggermente. Tutto ciò è visibile nella figura 2.11

Figura 2.11: Modello di regressione multinomiale logistica ridotto
 Stime e intervalli di confidenza al 95%
 Fonte dati: Banca d'Italia (2021)



Negli script 2.4 e 2.5 sono riportati i test del rapporto di verosimiglianze per le variabili Genere e Occupazione. Per la prima, si conferma l'ipotesi di significatività, anche se si rifiuta l'ipotesi nulla di eguale bontà d'adattamento solo al 10%. Si conferma anche per la variabile Occupazione, con un p-value del LR test ben al di sotto della soglia del 5%. Si nota dallo script 2.6 che si conferma anche la non significatività nel modello ridotto dell'età nello spiegare la scelta tra questi mezzi informativi.

Script 2.4: LR test su Genere

Likelihood ratio test

```
Model 1: info_m ~ Reddito insufficiente almeno una volta +
sesso + occ
```

```
Model 2: info_m ~ Reddito insufficiente almeno una volta + occ
#Df  LogLik Df  Chisq Pr(>Chisq)
1   12  -488.44
2   10  -490.87 -2  4.8457    0.08867 .
```


Script 2.5: LR test su Occupazione

Likelihood ratio test

Model 1: info_m ~ Reddito insufficiente almeno una volta +
sesso + occ

Model 2: info_m ~ Reddito insufficiente almeno una volta + sesso
#Df LogLik Df Chisq Pr(>Chisq)
1 12 -488.44
2 6 -498.30 -6 19.717 0.00311 **

Script 2.6: LR test su Età nel modello ridotto

Likelihood ratio test

Model 1: info_m ~ Reddito insufficiente almeno una volta +
sesso + occ + I(eta/5)

Model 2: info_m ~ Reddito insufficiente almeno una volta + sesso + occ
#Df LogLik Df Chisq Pr(>Chisq)
1 14 -487.80
2 12 -488.44 -2 1.2905 0.5245

Anche per questi mezzi informativi si riportano le probabilità associate previste dal modello ridotto, al variare di uno dei regressori e mantenuti costanti gli altri ad un livello di riferimento. I livelli di riferimento per le variabili sono i seguenti: Reddito insufficiente almeno una volta = No, Genere = Maschio, Occupazione = Occupato.

Nella tabella 2.11 si nota che i Disoccupati si informano maggiormente tramite i Contatti, e decisamente meno tramite media Istituzionali. In particolare, sono la categoria di individui che li utilizza meno (circa il 20% in meno rispetto agli altri). Gli individui fuori dal mercato del lavoro, ovvero i pensionati, utilizzano anche loro maggiormente i Contatti. Oltretutto, come ci si potrebbe aspettare, sono la categoria che utilizza di meno i social per informarsi (circa il 10% in meno rispetto agli altri).

Nella tabella 2.12 si nota che gli individui hanno una probabilità maggiore di informarsi tramite i Social se hanno avuto uno o più mesi in cui non sono arrivati a coprire le spese mensili (circa il 15% in più degli altri mezzi). Questi sono anche gli individui che li utilizzano di più, con una differenza di probabilità del 17%.

Nella tabella 2.13 si nota come le donne siano coloro che hanno la minore probabilità di informarsi tramite i Social. Nonostante ciò, anche gli uomini preferiscono informarsi

Tabella 2.11: Probabilità di utilizzo dei mezzi informativi al variare dell'occupazione
 Reddito insufficiente almeno una volta = No, Genere = Maschio

Occupazione	Istituzioni	Contatti	Social
Occupato	0.371	0.368	0.261
Disoccupato	0.164	0.563	0.273
Fuori	0.363	0.533	0.104
Altro	0.383	0.377	0.240

Tabella 2.12: Probabilità di utilizzo dei mezzi informativi al variare della variabile
 Reddito insufficiente (almeno una volta)
 Occupazione = Occupato, Genere = Maschio

Reddito insufficiente	Istituzioni	Contatti	Social
Sì	0.284	0.285	0.431
No	0.371	0.368	0.261

con altro rispetto ai Social. Risulta però che le donne sono associate ad una maggiore probabilità degli uomini di informarsi tramite le Istituzioni e tramite i Contatti.

Tabella 2.13: Probabilità di utilizzo dei mezzi informativi al variare del Genere
 Occupazione = Occupato, Reddito insufficiente almeno una volta = No

Genere	Istituzioni	Contatti	Social
Maschio	0.371	0.368	0.261
Femmina	0.409	0.423	0.168

Capitolo 3

Determinanti delle aspettative di inflazione

In questo capitolo si analizzeranno le determinanti nelle aspettative di inflazione degli individui. Si ricorda che nell'analisi che segue gli individui hanno risposto tra la fine di agosto e l'inizio di settembre 2021.

Per farlo, ci si avvarrà delle variabili utilizzate nel capitolo precedente, assieme alla variabile contenente le scelte degli individui in merito al canale informativo. Oltretutto, si sfrutterà la risposta alla domanda in merito all'impressione sull'andamento dei prezzi nell'anno precedente al sondaggio, la cui distribuzione di frequenza è raffigurata nella figura 3.1.

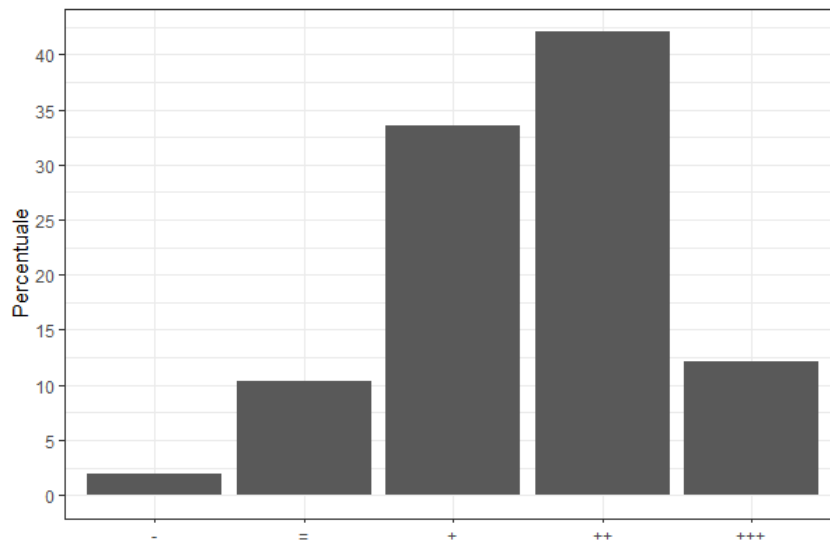


Figura 3.1: Impressione sull'andamento dei prezzi nell'ultimo anno (2020)
Fonte dati: Banca d'Italia (2021)

La maggior parte dei rispondenti sostiene che i prezzi siano aumentati Poco (+) o Abbastanza (++) nell'anno precedente all'indagine. Una parte uguale di persone ritiene che siano aumentati Molto (+++) e che siano rimasti Stabili (=) o Calati (-). Queste ultime due categorie verranno considerate assieme, dato che la percentuale di persone nel campione che ritiene che i prezzi siano calati è estremamente bassa. La categoria verrà chiamata "Non aumentati" e sarà la categoria di riferimento.

Da tempo diversi modelli macroeconomici sostengono l'importanza del livello d'inflazione percepita, in quanto elemento importante nella formulazione delle aspettative di inflazione. Ciò giustifica quindi la scelta di introdurre tale variabile in un possibile modello, almeno preliminarmente.

La variabile oggetto di studio è una risposta in merito alle aspettative di inflazione per l'anno successivo rispetto al precedente, la cui distribuzione di frequenza è raffigurata nella figura 3.2. Le risposte indicano i diversi tassi di aumento rispetto all'aumento percepito l'anno precedente. La quasi totalità degli individui si aspetta che l'inflazione per l'anno successivo sarà positiva, con i vari gradi di aumento (+, ++, +++). Pochi, di nuovo, si aspettano inflazione futura negativa (-).

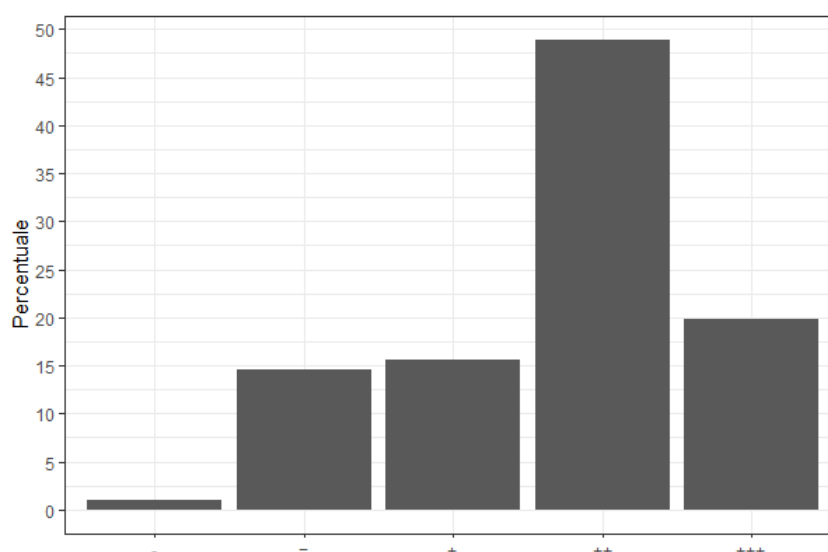


Figura 3.2: Aspettative di aumento dei prezzi per l'anno 2022 rispetto al precedente
Fonte dati: Banca d'Italia (2021)

Elemento essenziale dell'analisi è riconoscere la struttura ordinale della variabile circa le aspettative di aumento dei prezzi. Infatti, ammette molteplici categorie di risposte, come una variabile casuale multinomiale, ma intrinsecamente ordinabili. I motivi che spingono a sfruttare l'ordinabilità sono vari.

Innanzitutto, come si vedrà nell'analisi e come ricordato da Agresti (2002), modelli che sfruttano l'ordinabilità richiedono di stimare un numero di parametri generalmente minore, ad esempio rispetto al modello multinomiale logistico. Questa parsimonia rende tali categorie di modelli particolarmente appetibili.

Oltretutto, come ricorda sempre Agresti (2002), modelli che prendono in considerazione l'ordine, nel caso in cui l'ordine esista, presentano vantaggi in termini di potenza dei test e precisione nelle stime.

3.1 Modello a logit cumulati

Un modello che prende in considerazione l'ordinalità della variabile risposta è il modello a logit cumulati, detto anche Modello ad Odds Proporzionali. Similmente al modello di regressione logistica 2.1, il modello prende in considerazione i logit, cioè i $\log(\text{Odds})$, ma non più per la probabilità che Y sia uguale ad un livello j , bensì per la probabilità cumulata fino a tale livello. I logit cumulati saranno quindi così definiti:

$$\text{logit}[P(Y \leq j)] = \log \left(\frac{P(Y \leq j)}{P(Y > j)} \right) \quad (3.1)$$

Il modello ad Odds Proporzionali può avere parametrizzazioni differenti, ma una delle più comuni e quella di default del software R è la seguente:

$$\text{logit}[P(Y \leq j)] = \alpha_j - \beta' \mathbf{x}, \quad j \in [1, J - 1] \quad (3.2)$$

Il modello stima per ogni logit le intercette α_j e il vettore di coefficienti β . Quest'ultimo non varia al variare del logit preso in considerazione, ovvero si assume che ogni variabile esplicativa abbia lo stesso effetto per ogni categoria della variabile Y . Da qui il termine Odds Proporzionali. (Perry (2018)).

Si noti che questa parametrizzazione prevede un segno negativo dopo l'intercetta. Questo permette, nel caso in cui β sia positivo, di intendere l'incremento della variabile esplicativa (oppure il passaggio ad un'altra modalità se qualitativa) come una riduzione dei logit, ovvero una riduzione delle probabilità cumulate (Agresti (2002)). Questo implica che viene associata minore probabilità alle categorie minori di j e maggiore probabilità alle categorie successive a j . L'interpretazione di β è quindi tale per cui ad un segno positivo si associa una relazione diretta, mentre ad un segno negativo si associa una relazione inversa.

3.2 Modello logit a categorie adiacenti

Un'alternativa al modello a logit cumulati è rappresentato dal modello logit a categorie adiacenti.

Si definiscono logit adiacenti i seguenti logaritmi dei rapporti di probabilità:

$$\text{logit}[P(Y = j|Y = j \vee Y = j + 1)] = \log \left(\frac{\pi_j}{\pi_{j+1}} \right), \quad j \in [1, J - 1] \quad (3.3)$$

Il modello logit a categorie adiacenti è quindi così definito:

$$\log \left(\frac{\pi_j}{\pi_{j+1}} \right) = \alpha_j + \beta' \mathbf{x}, \quad j \in [1, J - 1] \quad (3.4)$$

Analogamente al modello ad Odds proporzionali, il modello stima per ogni logit le intercette α_j , mentre β è uguale per ogni logit. Anche in questo modello quindi si assume proporzionalità tra gli odds.

Come sottolinea Agresti (2002), la scelta tra questo modello e il modello citato precedentemente dipende essenzialmente dall'interesse del ricercatore: se si ha interesse ad osservare gli effetti sulle singole categorie, conviene utilizzare il modello logit a categorie adiacenti. Nel caso invece si fosse interessati ad utilizzare tutta la scala di possibili valori, ipotizzando quindi la presenza di una variabile latente continua sottostante alla variabile, allora conviene utilizzare il modello a logit cumulati.

Siccome gli effetti si riferiscono a tutta la scala di possibili valori con il modello a logit cumulati, i coefficienti saranno generalmente maggiori, ma il rapporto coefficiente e standard error dovrebbe essere simile tra i due modelli (Agresti (2002)). Sempre Agresti (2002) sottolinea che il modello a logit cumulati gode di invarianza delle stime rispetto al numero e alla scelta delle categorie della variabile risposta, proprietà non rispettata dall'altro modello.

3.3 Ipotesi di Odds Proporzionali e modelli logit ordinali generalizzati

L'assunzione che gli odds siano proporzionali si è visto essere importante sia che si utilizzi un modello con logit cumulati, sia con logit a categorie adiacenti. L'ipotesi, infatti, permette di ridurre sensibilmente il numero di parametri da stimare, in quanto si assume che ogni variabile abbia lo stesso effetto per ogni logit.

Risulta perciò ovvia la necessità di testare questa ipotesi. Per il modello a logit cumulati è stato sviluppato in R il test di Brant. Il test si basa sul confronto tra i coefficienti di diverse regressioni logistiche effettuate separatamente su ogni logit cumulato. Se le differenze risultano essere statisticamente significative, allora si rifiuta l'ipotesi di Odds Proporzionali. Il test permette anche di verificare quale delle variabili inserite viola l'ipotesi. Nel caso dei logit a categorie adiacenti, siccome su R non è presente una funzione che svolge automaticamente il test, se le variabili sono poche, è possibile verificare se l'ipotesi di odds proporzionali è rispettata svolgendo il test LR tra il modello rispettante l'ipotesi e un modello meno restrittivo. Se il test LR rifiuta l'ipotesi nulla, allora l'ipotesi di odds proporzionali non si può ritenere valida.

Nel caso in cui l'ipotesi sia violata, è necessario identificare tale modello meno restrittivo, che deve considerare molteplici effetti, tanti quanti sono i logit, per ogni variabile che non rispetta l'ipotesi.

È quindi il caso dei modelli logit ordinali generalizzati. Il modello con i logit cumulati si parametrizza nel modo seguente:

$$\text{logit}[P(Y \leq j)] = \alpha_j - \beta'_j \mathbf{x} \quad (3.5)$$

Mentre il modello logit a categorie adiacenti:

$$\log \left(\frac{\pi_j}{\pi_{j+1}} \right) = \alpha_j + \beta'_j \mathbf{x} \quad (3.6)$$

Come preannunciato, può essere che solo alcune delle variabili non rispettino l'ipotesi, mentre altre che non presentino differenze significative al variare dei logit. Nel caso ciò fosse vero, è utile generalizzare in maniera parziale i modelli, ovvero per alcune variabili avere un coefficiente per ogni logit, mentre per altre avere un coefficiente unico. Ciò permette di ridurre ove possibile la dimensionalità.

3.4 Analisi con regressione logistica

A scopo illustrativo, si riporta in questa sezione il risultato di un'analisi preliminare con la variabile risposta accorpata in una variabile dicotomica, dove le categorie dell'andamento dei prezzi sono "Non aumentano" e "Aumentano". Si sono quindi accorpate le modalità relative ad un'aspettativa d'inflazione negativa e positiva. La modalità "Non aumentano" è quella di riferimento.

Il risultato del modello di regressione logistica con tutte le variabili è riportato nella tabella 3.1.

Variabile	Stima	std.error	statistic	p-value	
Età(5 anni)	0.088	0.034	2.590	0.010	**
Titolo: Basso	-0.065	0.168	-0.386	0.699	
Titolo: Alto	0.299	0.203	1.474	0.140	
Area: Nord	0.229	0.195	1.173	0.241	
Area: Sud	-0.090	0.204	-0.440	0.660	
Mezzo informativo: Contatti	0.034	0.249	0.138	0.890	
Mezzo informativo: Istituzioni	0.718	0.309	2.326	0.020	*
Mezzo informativo: Social	-0.182	0.281	-0.646	0.518	
Mezzo informativo: Tradizionali	0.222	0.172	1.288	0.198	
Impressione sui prezzi: +	2.032	0.177	11.457	0.000	***
Impressione sui prezzi: ++	3.082	0.204	15.085	0.000	***
Impressione sui prezzi: +++	3.565	0.371	9.603	0.000	***
Arriva a fine mese difficilmente	0.382	0.166	2.297	0.022	*
Occ: Altro	-0.972	0.383	-2.536	0.011	*
Occ: Disoccupato	-0.790	0.281	-2.810	0.005	**
Occ: Fuori	-0.218	0.218	-1.000	0.317	
Reddito insufficiente almeno una volta	0.377	0.185	2.040	0.041	*

Tabella 3.1: Modello di regressione logistica sull'aspettativa d'inflazione
Fonte dati: Banca d'Italia (2021)

Dal risultato del modello si nota che la variabile Titolo di studio e Area geografica non sono risultate significative. Risultano invece significative l'Età, la variabile Impressione di aumento dei prezzi (tutte le modalità), e poi in particolare la modalità Istituzioni della variabile "Mezzo informativo", le due variabili sul livello di reddito e alcune delle modalità della variabile Occupazione.

Chi ha avuto un'impressione di aumento dei prezzi è più probabile che si aspetti inflazione positiva, in accordo con le teorie economiche.

Questa relazione si ripete per tutte le variabili significative, eccetto per le due modalità della variabile Occupazione. Emerge, infatti, che chi è disoccupato è più probabile che non si aspetti inflazione.

Questo modello soffre però di forte sbilanciamento della variabile risposta, siccome la quasi totalità degli individui si aspetta inflazione positiva. Ciò si nota ad esempio se si osservano i residui (quantile) riportati nella figura 3.3. Infatti, si osserva una forte presenza di probabilità predette nell'intervallo di valori vicini ad 1. Per questo motivo, tale modello non risulterebbe affidabile nella previsione delle probabilità, a meno di intervenire con ribilanciamenti artificiali della variabile risposta.

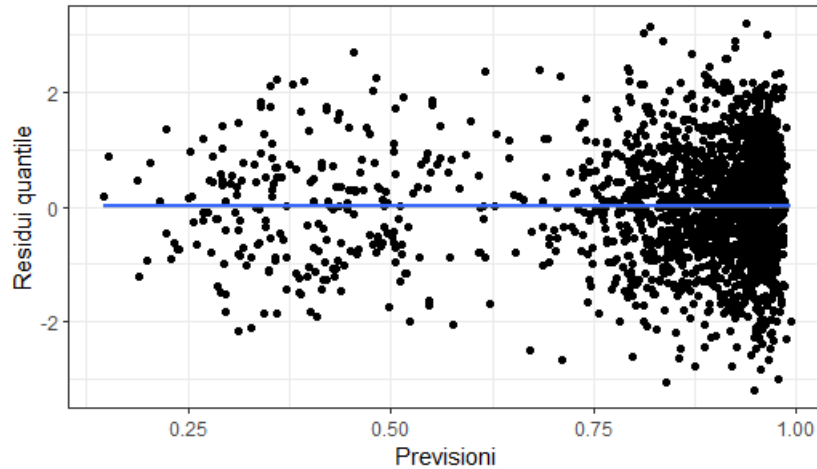


Figura 3.3: Residui quantile del modello logistico per la variabile Aspettative di inflazione

3.5 Analisi con modelli ordinali

In questa sezione si analizzerà la variabile Aspettative di inflazione sfruttando la struttura ordinale della variabile. Il primo modello che si è scelto di utilizzare è il modello a logit cumulati.

Il risultato del modello con tutte le variabili è riportato nella figura 3.4. Dal grafico si nota che le variabili maggiormente significative sono: Reddito insufficiente almeno una volta, Impressioni sui prezzi passati (tutte le modalità) e l'Età (anche se il riscaldare l'età per avere un effetto ogni 5 anni non è stato sufficiente a far comparire gli IC nel grafico).

Script 3.1: LR test su Mezzi informativi per il modello a logit cumulati

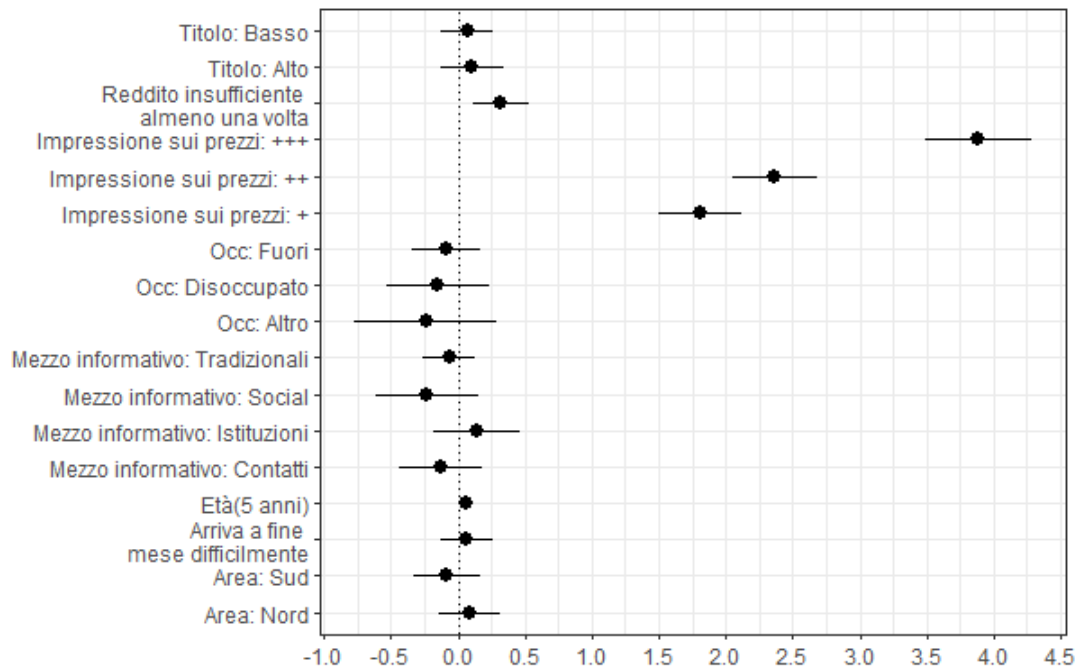
Likelihood ratio test

Model 1: `prezzif ~ eta + titolo + area + info_m + prezzip + condgennow + occ + Reddito insufficiente almeno una volta`

Model 2: `prezzif ~ eta + titolo + area + prezzip + condgennow + occ + Reddito insufficiente almeno una volta`

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	21	-2236.3			
2	17	-2238.0	-4	3.4357	0.4877

Figura 3.4: Modello a logit cumulati. Stime e IC al 95%
Fonte dati: Banca d'Italia (2021)



Emerge invece che i mezzi informativi non sembrano essere significativi nello spiegare le aspettative di inflazione, almeno in questo modello. Si è deciso di condurre un test LR, riportato nello script 3.1, per verificare ulteriormente la significatività della variabile, ma anche tale test conferma la sua non significatività

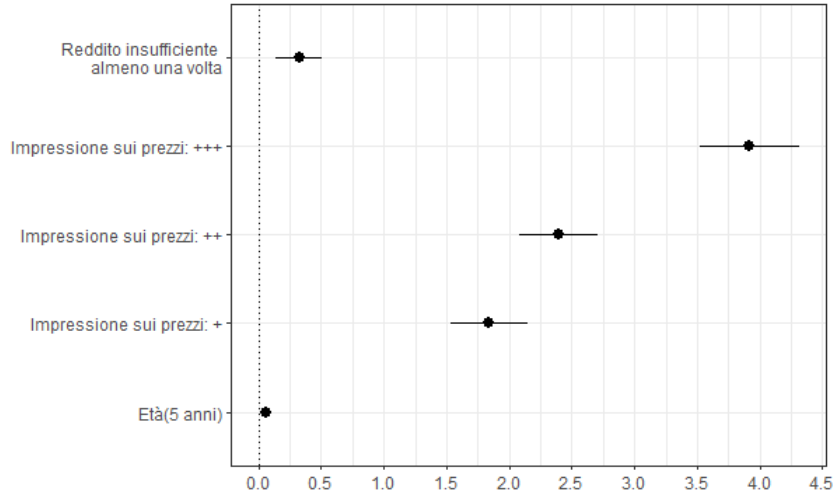
Anche le variabili Titolo di studio, Area geografica e Occupazione sono risultate non significative.

Si è scelto di utilizzare l'algoritmo di minimizzazione dell'AIC per scegliere il migliore sottoinsieme di variabili per questo modello. Il modello ridotto è riportato nella figura 3.5. L'algoritmo di minimizzazione dell'AIC non ha suggerito l'inserimento di alcuna variabile d'interazione.

A questo punto, si è testata l'ipotesi di Odds Proporzionali per il modello ridotto, al fine di verificare se fosse necessario utilizzare una sua versione generalizzata.

Il risultato del test di Brant è risultato inizialmente non attendibile, in quanto il software R ha fatto notare l'assenza di alcune coppie di modalità tra la variabile risposta e i regressori. Si è quindi indagata la presenza di tabelle sparse, ovvero con frequenze congiunte attese minori di 5, tra le coppie di variabili qualitative.

Figura 3.5: Modello a logit cumulati ridotto. Stime e IC al 95%
Fonte dati: Banca d'Italia (2021)



Nella tabella 3.2 si nota che effettivamente tra la variabile Aspettative di inflazione futura e Impressione sull'andamento dei prezzi nell'anno precedente vi sono frequenze attese minori di 5 per la prima riga.

Tabella 3.2: Tabella di frequenze attese per le variabili
Aspettative di aumento dei prezzi nel futuro (righe) e
Impressione sull'andamento dei prezzi nell'anno precedente (colonne)

	Non aumentati	Poco	Abbastanza	Molto
Calano	2.45	6.71	8.41	2.43
Stabili	34.32	93.92	117.73	34.03
Di meno	36.65	100.29	125.71	36.34
Uguale	114.99	314.62	394.38	114.01
Di più	46.58	127.46	159.77	46.19

Per la variabile Reddito insufficiente almeno una volta, si nota dalla tabella 3.3 che non sono presenti frequenze attese tali da rendere la tabella sparsa.

Accorpando le modalità "Calano" e "Stabili" dei prezzi futuri in una sola modalità, "Non aumentano", si ottiene un test di Brant senza avvertimenti circa la sua attendibilità. Il risultato è riportato nella tabella 3.4. Siccome la variabile Impressione sull'andamento dei prezzi nell'anno precedente è risultata l'unica non rispettante l'ipotesi di odds proporzionali, sarà l'unica per la quale si generalizzerà il modello e si stimerà un coefficiente per ogni logit cumulato.

Tabella 3.3: Tabella di frequenze attese per le variabili
Aspettative di inflazione futura (righe) e
Reddito insufficiente almeno una volta (colonne)

	No	Sì
Calano	13.36	6.64
Stabili	187.10	92.90
Di meno	199.80	99.20
Uguale	626.80	311.20
Di più	253.93	126.07

Stimando il modello a logit cumulati generalizzato come appena descritto, si è riscontrato un problema con le stime, in quanto sono state predette probabilità negative.

Non si è trovata una motivazione convincente che potesse spiegare la ragione di questo fenomeno. Si riporta solo che è stato rilevato effetto Hauck-Donner per la variabile Impressione sull'andamento dei prezzi. Il fenomeno è un caso limite in cui la statistica test di Wald non è più funzione monotona della differenza tra la stima e il valore del parametro sotto l'ipotesi nulla, quando la stima tende ai confini dello spazio parametrico. Generalmente, porta ad una distorsione del p-value, impedendo di fare inferenza correttamente, e ad una riduzione della potenza dei test. (Yee (2022))

Con le variabili maggiormente significative individuate con il modello precedente, assieme alla variabile Mezzo informativo, si è scelto di cambiare modello e utilizzare il modello logit a categorie adiacenti, assumendo di nuovo che fosse soddisfatta l'ipotesi di Odds Proporzionali.

Il risultato del modello è riportato nella figura 3.6.

Anche in questo modello sembrano essere significative le variabili inserite, tranne la variabile Mezzo informativo. Lo script 3.2 sottolinea oltretutto che anche in questo modello la variabile Impressione sull'andamento dei prezzi nell'anno precedente non soddisfa l'ipotesi di Odds Proporzionali, poiché è significativo l'incremento della bontà d'adattamento del modello generalizzato per quella variabile. Sono omessi gli script per

	X2	df	p-value
Omnibus	90.28	10.00	0.00
Età	3.98	2.00	0.14
Impressione sui prezzi: +	39.88	2.00	0.00
Impressione sui prezzi: ++	70.08	2.00	0.00
Impressione sui prezzi: +++	5.75	2.00	0.06
Reddito insufficiente almeno una volta	1.39	2.00	0.50

Tabella 3.4: Test di Brant per il modello a logit cumulati

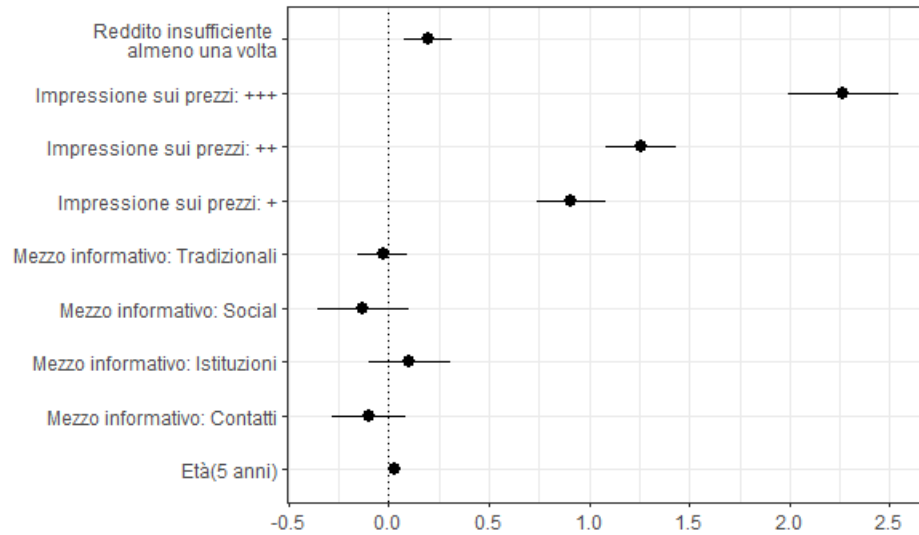


Figura 3.6: Modello logit a categorie adiacenti. Stime e IC al 95%
Fonte dati: Banca d'Italia (2021)

la variabile Età e Reddito insufficiente almeno una volta, ma di nuovo i test indicano che l'ipotesi di odds proporzionali non è rifiutata.

Script 3.2: LR test su Impressioni di inflazione per il modello logit a categorie adiacenti generalizzato contro quello proporzionale

Likelihood ratio test

Model 1: `prezzif2 ~ Reddito insufficiente almeno una volta +
prezzip + info_m + eta`

Model 2: `prezzif2 ~ Reddito insufficiente almeno una volta +
prezzip + info_m + eta`

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	5733	-2136.6			
2	5739	-2177.3	6	81.324	1.903e-15 ***

Nella tabella 3.5 è riportato il risultato del modello generalizzato per la variabile Impressione sull'andamento dei prezzi nell'anno precedente e Mezzo informativo. Il test LR ha riportato per quest'ultima che la bontà d'adattamento non migliora dopo la generalizzazione. Nonostante ciò, si è ritenuto ugualmente interessante riportare i risultati ottenuti poiché alcuni coefficienti sono risultati significativi.

In particolare, chi si informa tramite mezzi istituzionali e tradizionali è più probabile, rispetto a chi non si informa, che si aspetti un'inflazione positiva, anche se modesta (logit 1). Conrad et al. (2022) hanno individuato effetti significativi solo per i media tradizionali, e il loro risultato afferma che chi si informa tramite questi mezzi ha un'inflazione attesa minore rispetto a chi non si informa. Con minore si intende in valore assoluto: possedevano infatti risposte quantitative in merito alla previsione d'inflazione.

Variabile	logit	Stima	std.error	statistic	p-value	
Età (5 anni)	Tutti	0.030	0.010	3.080	0.002	**
Reddito insufficiente almeno una volta	Tutti	0.197	0.059	3.324	0.001	***
Impressione sui prezzi: +	1	1.938	0.253	7.663	0.000	***
	2	0.200	0.263	0.759	0.448	
	3	-0.343	0.291	-1.177	0.239	
Impressione sui prezzi: ++	1	2.631	0.274	9.596	0.000	***
	2	0.630	0.262	2.401	0.016	*
	3	-0.295	0.281	-1.047	0.295	
Impressione sui prezzi: +++	1	2.153	0.479	4.492	0.000	***
	2	1.042	0.378	2.758	0.006	**
	3	1.426	0.301	4.741	0.000	***
Mezzo informativo: Contatti	1	0.242	0.297	0.814	0.415	
	2	-0.256	0.235	-1.086	0.277	
	3	-0.206	0.236	-0.871	0.384	
Mezzo informativo: Istituzioni	1	0.976	0.347	2.810	0.005	**
	2	-0.271	0.239	-1.134	0.257	
	3	-0.125	0.243	-0.514	0.607	
Mezzo informativo: Social	1	0.258	0.341	0.757	0.449	
	2	-0.540	0.292	-1.851	0.064	.
	3	0.164	0.290	0.566	0.571	
Mezzo informativo: Tradizionali	1	0.485	0.206	2.349	0.019	*
	2	-0.270	0.157	-1.717	0.086	.
	3	-0.164	0.148	-1.104	0.270	

Tabella 3.5: Modello logit a categorie adiacenti generalizzato per Impressione di inflazione e Mezzo informativo. 1, 2 e 3 indicano a quale logit adiacente ci si riferisce per il coefficiente in questione. $1 = \log(P[Y = \text{Di meno}] / P[Y = \text{Non aumentano}])$, $2 = \log(P[Y = \text{Uguale}] / P[Y = \text{Di meno}])$, $3 = \log(P[Y = \text{Di più}] / P[Y = \text{Uguale}])$. "Tutti" indica che la variabile ha un coefficiente unico.

Fonte dati: Banca d'Italia (2021)

Dal risultato del modello appare che anche l'età e la variabile Reddito insufficiente almeno una volta hanno un effetto positivo sui logit adiacenti: all'aumento dell'età e delle difficoltà economiche è associato una maggiore probabilità di essere nel logit successivo

invece che nel logit precedente. Questo implica una relazione positiva tra le variabili e la previsione d'inflazione.

Per la variabile Impressione sui prezzi, si hanno degli effetti, tra quelli significativi, tutti positivi, indicando che chi ha avuto un'impressione di aumento dei prezzi è più probabile che si aspetti inflazione positiva nel futuro.

Inoltre, risulta che maggiore è l'aumento percepito dei prezzi, maggiore è l'aumento atteso degli stessi nel futuro. Ciò si nota dagli effetti per i logit più alti che diventano più significativi al mutare della variabile esplicativa.

3.6 Il ruolo delle esperienze

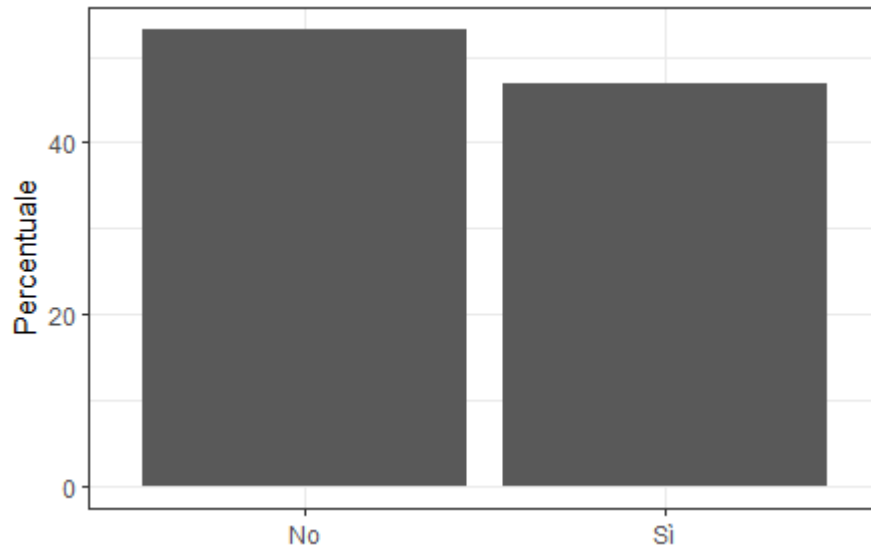


Figura 3.7: Variabile Esperienza, rappresentante le persone che hanno avuto esperienza di inflazione alta durante anni impressionabili (18-25 anni)

Fonte dati: Banca d'Italia (2021)

In questa sezione si è voluto approfondire il tema delle esperienze, specialmente durante gli anni impressionabili nella vita di un individuo.

I risultati ottenuti da Krosnick and Alwin (1989) supportano l'ipotesi per la quale gli individui sono particolarmente soggetti a mutamenti nella personalità, e quindi nelle convinzioni, durante uno specifico intervallo temporale, tra i 18 e 25 anni. L'idea è quindi che eventi particolarmente traumatici potrebbero condizionare l'individuo in maniera permanente.

Ciò giustifica quindi la ricerca in merito alla seguente domanda: Aver avuto esperienze di inflazione alta durante anni impressionabili influenza le aspettative di inflazione?

Nel caso italiano, si è deciso di prendere in considerazione come periodo caratterizzato da inflazione alta il periodo tra il 1972 e il 1986. Questo è infatti il periodo in cui l'inflazione è rimasta, di anno in anno, pari o superiore al 5% circa, considerabile un livello particolarmente alto.

Per cogliere questo effetto tra i rispondenti, si è costruita una variabile dicotomica Esperienza per gli individui che erano tra i 18 e i 25 anni in questo intervallo temporale. Chi, nel 2021, aveva un'età tra i 53 e i 74 anni, per la variabile Esperienza verrà classificato come Sì, altrimenti, fuori da quell'intervallo, verrà classificato come No. Nella figura 3.7 si nota che quasi la metà degli individui rientra nella categoria Sì.

Variabile	logit	Stima	std.error	statistic	p-value	
Età (5 anni)	Tutti	0.033	0.011	2.957	0.003	**
Reddito insufficiente almeno una volta	Tutti	0.194	0.059	3.272	0.001	**
Impressione sui prezzi: +	1	1.933	0.251	7.698	0.000	***
	2	0.214	0.262	0.816	0.414	
	3	-0.355	0.290	-1.220	0.222	
Impressione sui prezzi: ++	1	2.589	0.272	9.535	0.000	***
	2	0.673	0.261	2.579	0.010	**
	3	-0.295	0.281	-1.051	0.293	
Impressione sui prezzi: +++	1	2.114	0.478	4.423	0.000	***
	2	1.072	0.377	2.842	0.004	**
	3	1.424	0.300	4.744	0.000	***
Esperienza: Sì	Tutti	-0.010	0.062	-0.162	0.871	

Tabella 3.6: Modello logit a categorie adiacenti con la variabile Esperienza, assunta rispettare l'ipotesi di Odds Proporzionali
Fonte dati: Banca d'Italia (2021)

In accordo con quanto individuato precedente come modello ordinale per la variabile Aspettativa di aumento dei prezzi, si è deciso di utilizzare un modello logit a categorie adiacenti con le variabili utilizzate nel precedente modello. Si è escluso però Mezzo informativo, in quanto la sua bassa capacità esplicativa non avrebbe condotto a risultati ulteriori oltre a quelli già analizzati. Al suo posto si è inserita la variabile Esperienza, assumendo che rispettasse l'ipotesi di Odds Proporzionali. Anche reinserendo la variabile Mezzo informativo, la stima del coefficiente per Esperienza non muta in maniera significativa, e quindi ciò conferma la scelta di escludere Mezzo informativo dal modello.

Dal risultato riportato nella tabella 3.6 si nota che, in questa configurazione, il modello non ha riscontrato un effetto significativo per la variabile Esperienza. Il test LR riportato nello script 3.3 suggerisce però l'idea di generalizzare anche per questa variabile.

Script 3.3: LR test su Esperienza per il modello logit a categorie adiacenti generalizzato contro quello proporzionale

Likelihood ratio test

Model 1: prezzif ~ Reddito insufficiente almeno una volta +
eta + prezzip + expe

Model 2: prezzif ~ Reddito insufficiente almeno una volta +
eta + prezzip + expe

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	5734	-2135.5			
2	5736	-2138.5	2	5.9577	0.05085

Il risultato del modello generalizzato è riportato quindi nella tabella 3.7. Da questo risulta che aver avuto esperienza di inflazione alta durante gli anni impressionabili ha un effetto positivo sull'aspettativa di inflazione, e in particolare nel avere un'aspettativa di aumento dei prezzi alta (logit 2), anche se non eccessivamente elevata (per il logit 3 ha un effetto debolmente significativo e negativo, indicando che è poco probabile abbia un'aspettativa di inflazione elevata).

Variabile	logit	Stima	std.error	statistic	p-value	
Età (5 anni)	Tutti	0.032	0.011	2.940	0.003	***
Reddito insufficiente almeno una volta	Tutti	0.193	0.059	3.263	0.001	**
Impressione sui prezzi: +	1	1.931	0.251	7.691	0.000	***
	2	0.216	0.263	0.821	0.412	
	3	-0.357	0.291	-1.229	0.219	
Impressione sui prezzi: ++	1	2.581	0.272	9.503	0.000	***
	2	0.683	0.262	2.613	0.009	**
	3	-0.302	0.281	-1.076	0.282	
Impressione sui prezzi: +++	1	2.090	0.478	4.372	0.000	***
	2	1.099	0.378	2.910	0.004	**
	3	1.401	0.300	4.665	0.000	***
Esperienza: Sì	1	-0.237	0.179	-1.323	0.186	
	2	0.278	0.139	2.006	0.045	*
	3	-0.216	0.131	-1.656	0.098	.

Tabella 3.7: Modello logit a categorie adiacenti generalizzato per la variabile Esperienza e la variabile Impressione di inflazione
Fonte dati: Banca d'Italia (2021)

Conrad et al. (2022) hanno svolto un'analisi sull'effetto di essere cittadini tedeschi provenienti dalla Germania dell'est, prima della riunificazione, sulle aspettative di inflazione. Dopo la riunificazione, infatti, l'inflazione nel paese è aumentata considerevolmente. Gli autori non hanno posto vincoli all'età degli individui, tralasciando quindi il fattore legato agli anni impressionabili. Il risultato da loro ottenuto a tal proposito è che aver avuto quell'esperienza porta ad avere aspettative di inflazione positive, rispetto a non aspettarsi inflazione.

Si è provato a replicare la stessa analisi ampliando l'insieme di persone da considerare e inserendole all'interno della nuova variabile Esperienza Generale (ora verranno considerate tutte le persone oltre i 53 anni nel 2021). Utilizzando lo stesso modello, e assumendo la non proporzionalità della variabile Esperienza Generale, si sono ottenuti i risultati riportati nella tabella 3.8.

Variabile	logit	Stima	std.error	statistic	p-value	
Età (5 anni)	Tutti	0.036	0.017	2.063	0.039	***
Reddito insufficiente almeno una volta	Tutti	0.193	0.059	3.265	0.001	**
Impressione sui prezzi: +	1	1.936	0.251	7.712	0.000	***
	2	0.209	0.263	0.796	0.426	
	3	-0.351	0.290	-1.208	0.227	
Impressione sui prezzi: ++	1	2.583	0.272	9.514	0.000	***
	2	0.683	0.262	2.608	0.009	**
	3	-0.301	0.281	-1.074	0.283	
Impressione sui prezzi: +++	1	2.088	0.478	4.367	0.000	***
	2	1.106	0.378	2.925	0.003	**
	3	1.390	0.300	4.629	0.000	***
Esperienza Generale: Sì	1	-0.250	0.193	-1.297	0.195	
	2	0.301	0.157	1.915	0.055	.
	3	-0.305	0.152	-1.999	0.046	*

Tabella 3.8: Modello logit a categorie adiacenti generalizzato per la variabile Esperienza Generale e la variabile Impressione di inflazione
Fonte dati: Banca d'Italia (2021)

Risulta in particolare che questi individui non presentano molte differenze rispetto agli individui presi in considerazione precedentemente, in quanto anch'essi si aspettano un'inflazione positiva e alta ma non eccessiva. Questo porta a concludere che forse il ruolo delle esperienze, in questo caso, è più legato all'esperienza in sé che al periodo della vita durante il quale si vive l'evento d'interesse, in questo caso inflazione elevata.

Conclusione

Di seguito sono riportati e discussi i risultati dell'analisi precedentemente svolta. Innanzitutto, la propensione ad informarsi in ambito economico è maggiore se l'individuo è più giovane, più istruito e con minori difficoltà economiche.

Si è riscontrato che i soggetti più propensi all'utilizzo di media tradizionali sono gli individui più anziani e gli individui provenienti dal Sud. I Social, invece, sono più utilizzati da coloro che hanno difficoltà economiche. Fanno eccezione le donne, che usano con meno probabilità degli uomini sia i media tradizionali, sia i Social. I disoccupati, invece, si informano maggiormente tramite i contatti e meno tramite fonti istituzionali. Coloro che sono fuori dal mercato del lavoro, che nel caso analizzato rappresentano maggiormente gli individui in pensione, prediligono l'uso dei contatti e sono particolarmente avversi all'utilizzo dei Social. Emerge anche una propensione leggermente minore all'utilizzo di media tradizionali.

Circa le aspettative sull'inflazione, persone più anziane e persone con maggiori difficoltà economiche sono più propense ad aspettarsi un aumento dei prezzi. Lo stesso si può dire per coloro che si informano con media tradizionali e istituzionali, rispetto a chi non si informa. Infine, maggiore è la percezione che i prezzi siano aumentati nel passato, maggiore è l'aumento atteso dei prezzi nel futuro.

Anche il ruolo delle esperienze è risultato rilevante nell'influenzare le aspettative di inflazione. In particolare, chi ha avuto esperienza di inflazione elevata nel corso della propria vita è maggiormente incline ad aspettarsi un aumento dei prezzi nel futuro. Lo stesso effetto si riscontra in chi ha avuto la stessa esperienza ma durante gli anni impressionabili, tra i 18 e i 25 anni. Poiché l'effetto sulle aspettative di inflazione non sembra essere particolarmente differente si può dedurre che il fattore dell'impressionabilità è marginale, e che invece il fattore rilevante consiste nell'esperienza in sé.

Bibliografia

Alan Agresti. *Categorical Data Analysis*. John Wiley & Sons, 2002. ISBN 0-471-36093-7.

Christian Conrad, Zeno Enders, and Alexander Glas. The role of information and experience for households' inflation expectations. *European Economic Review*, 143:104015, 2022. ISSN 0014-2921. doi: <https://doi.org/10.1016/j.euroecorev.2021.104015>. URL <https://www.sciencedirect.com/science/article/pii/S001429212100283X>.

Banca d'Italia. Indagine straordinaria sulle famiglie italiane, 2021. URL <https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/indag-straord-famiglie-italiane/index.html>.

Peter K. Dunn and Gordon K. Smyth. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 1996.

Peter K. Dunn and Gordon K. Smyth. *Generalized Linear Models With Examples in R*. Springer Texts in Statistics, 2018. ISBN 978-1-4419-0118-7.

Jon A. Krosnick and Duane E. Alwin. Aging and susceptibility to attitude change. *Journal of Personality and Social Psychology*, 1989.

Stephen Perry. Ordinal logistic regression models and statistical software: What you need to know. *Cornell Statistical Consulting Unit*, 2018.

Christian Salas-Eljatib, Andres Fuentes-Ramirez, Timothy Gregoire, Adison Altamirano, and Valeska Yaitul. A study on the effects of unbalanced data when fitting logistic regression models in ecology. *Ecological Indicators*, 85, 2017. doi: 10.1016/j.ecolind.2017.10.030.

Thomas W. Yee. On the hauck–donner effect in wald tests: Detection, tipping points, and parameter space characterization. *Journal of the American Statistical Association*, 2022.