

POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Civile



Tesi di Laurea Magistrale

**Analisi del legame fra protocollo di indagine
(CATI vs. CAWI) e frequenze dichiarate d'uso dei modi di
trasporto da parte dei residenti nell'area metropolitana di Torino**

Relatore:

Prof. Marco Diana

Correlatore:

Ing. Riccardo Ceccato

Candidato:

Salvatore Amato

Dicembre 2018

ABSTRACT

È stato condotto un lavoro di analisi con lo scopo di capire in che misura le risposte alle domande effettuate nell'indagine multiprotocollo DEMONSTRATE, riguardanti i nuovi servizi di mobilità, la diversione modale, la co-modalità e le applicazioni tecnologiche nel sistema di trasporto passeggeri dell'Area Metropolitana di Torino, sono influenzate dalla tipologia di indagine (CATI - CAWI), oltre che da altri fattori socio-economici del campione. La difficoltà è data dal fatto che tali fattori, a loro volta, influenzano la modalità di indagine stessa.

Attraverso un'analisi di inferenza statistica, effettuata tramite lo studio di test parametrici e non, sono state riscontrate delle differenze statisticamente significative tra le frequenze medie di utilizzo di un determinato mezzo di trasporto in funzione della tipologia di sondaggio. Gli intervistati con modalità web hanno dichiarato di utilizzare con maggiore frequenza tutte le modalità di trasporto prese in esame rispetto a coloro che, invece, hanno risposto al sondaggio telefonico. Tale risultato può indicare uno stile di vita più attivo e meno sedentario degli intervistati che hanno risposto al sondaggio con modalità CAWI.

Soffermarsi su un'analisi che valuta la differenza delle risposte sulla frequenza di utilizzo di un determinato mezzo di trasporto in funzione della sola modalità di indagine, risulta essere poco chiara e poco utile nell'ambito dello studio della mobilità del campione considerato. Per questo motivo, è stato condotto uno studio mirato ad individuare il protocollo di indagine come proxy delle variabili socio-economiche per spiegare la scelta di mobilità. L'analisi di correlazione univariata e

la calibrazione di diversi modelli multivariati di regressione lineare e logistica, hanno permesso di valutare il grado di sostituzione delle variabili socio-economiche con la variabile cati/cawi. Successivamente, con la tecnica statistica dell'Anova a due vie, è stata condotta un'analisi più dettagliata, mirata ad individuare quale sia l'effetto predominante sulla frequenza di utilizzo di determinati mezzi di trasporto tra le variabili proxy e la variabile indicante la modalità di indagine e, inoltre, come esso influenzi le scelte di mobilità.

In accordo con i risultati di questa analisi di studio, gli individui che hanno risposto ad un'indagine telefonica o web rappresentano due campioni distinti nell'utilizzo di determinati mezzi di trasporto a causa di una serie di fattori correlati, definiti dalle caratteristiche socio-economiche e dalla tipologia di indagine. In sostanza, queste differenze di risposta tra il sondaggio web e telefonico, riflettono le differenze socio-economiche e trasportistiche degli intervistati.

Analysis of the relationship between survey protocol (CATI vs. CAWI) and stated frequencies in the levels of use of different travel modes by residents in the Turin metropolitan area.

ABSTRACT

An analytic work was conducted with the purpose to understand how the answers to the questions of multi-protocol survey DEMONSTRATE, concerning the new mobility services, modal diversion, co-modality and technology applications in passenger transport systems of the metropolitan area of Turin, are influenced by the type of survey (CATI “Computer Assisted Telephone Interviewing” – CAWI “Computer Assisted Web Interviewing”), as well as other social-economic characteristics of the sample. The problem is that, in their turn, these factors have a concrete effect on the mode of survey itself.

A statistical inference analysis, based on the evaluation of parametric and non-parametric tests, showed statistically significant differences between the mean frequencies of use of a particular mode of transport depending on the type of survey. People who took part of a web interview declared to use more frequently every means on transport studied, than people who participate to a phone survey. This result can indicate a more active lifestyle of respondents with CAWI mode.

To explain the different choices of mobility, a study was led to identify to which extent the survey protocol can be considered as a proxy of some socio-economic variables, which are often not available in innovative and passive mobility data

collection method such as GPS traces. The level of substitution of socio-economic variables with cati/cawi has been evaluated with an analysis of univariate correlation and with the calibration of different multivariate linear and logistic regression models.

Subsequently, through the use of two-way analysis of variance, an in-depth analysis was conducted looking for the predominant effect between proxy and cati/cawi variables on mobility choices.

According to the results of this study, people who answered to a web or phone survey, represent two different samples in the use of specific transport means because of a range of self-related factors, which are defined by socio-economic characteristic and by the kind of survey. So, these different answers between web and phone surveys, highlight socio-economic and mobility differences of interviewed.

INDICE

Introduzione	1
1 Revisione dello stato dell'arte	5
1.1 Sondaggi online vs sondaggi telefonici	5
1.2 Fattori che influenzano le risposte degli intervistati	8
1.3 Studi nel campo dei trasporti	12
2 Introduzione alle analisi statistiche effettuate e loro esecuzione tramite SPSS 17	
2.1 Ipotesi nulla e livello di significatività del test.....	17
2.2 Test T-Student dei campioni indipendenti.....	23
2.3 Test ANOVA a una via	26
2.4 Test ANOVA a due vie	29
2.5 Test U di Mann-Whitney.....	29
2.6 Test H di Kruskal-Wallis.....	31
2.7 La correlazione tra due variabili: coefficiente punto-biserial e Indice Phi 32	
2.8 Regressione lineare multivariata.....	33
2.9 Regressione logistica	35
2.10 Realizzazione dei test precedenti in IBM® SPSS® Statistics.....	36
2.10.1 Test t-Student dei campioni indipendenti	37
2.10.2 ANOVA a una via	38
2.10.3 ANOVA a due vie	39
2.10.4 Test U di Mann-Whitney	40
2.10.5 Test H di Kruskal-Wallis.....	42
2.10.6 Coefficiente punto-biserial e Indice Phi.....	43
2.10.7 Regressione lineare multivariata e regressione logistica	45
3 Descrizione dell'indagine DEMONSTRATE e ipotesi preliminari.....	47
3.1 Indagine di partenza	47
3.2 Ipotesi di lavoro per le successive analisi.....	53
3.3 Caratteristiche socio-demografiche del campione.....	55
4 Studio della significatività nelle differenze fra risposte CATI e CAWI	59
4.1 Piano di analisi	59
4.2 Variabile dipendente e relativa ricodifica.....	59

4.3	Statistiche descrittive: tabelle di contingenza	61
4.4	Test parametrico o test non parametrico: controllo della normalità nella distribuzione dei dati	67
4.5	Significatività nella differenza delle medie	69
4.6	Il ruolo del protocollo di indagine nella significatività delle differenze	73
5	Il protocollo di indagine come “proxy” per spiegare le scelte di mobilità	79
5.1	L'importanza delle variabili socio-economiche.....	79
5.2	Elenco completo delle variabili potenzialmente sostituibili	81
5.3	Correlazione di tutte le variabili considerate con CATI/CAWI.....	89
5.4	Modelli di regressione lineare multivariata per le frequenze di utilizzo di determinati mezzi di trasporto	93
5.4.1	Formulazione e stima dei modelli.....	93
5.4.2	Il tipo di intervista come proxy delle variabili considerate	97
5.5	Modello di regressione logistica con variabile endogena CATI/CAWI.....	106
5.6	Riepilogo del grado di sostituzione delle variabili socio-economiche con CATI/CAWI	111
6	L'EFFETTO della variabile proxy sulle frequenze di utilizzo dei mezzi di trasporto.....	115
6.1	Considerazioni di partenza per l'analisi della varianza a due vie	115
6.2	La valutazione degli effetti di interazione sulle frequenze delle variabili considerate.....	117
6.2.1	Stima degli effetti significativi nelle singole macro celle	121
6.2.2	Stima complessiva degli effetti significativi	125
6.3	Comparazione tra l'Anova a due vie e l'analisi di regressione.....	127
7	Conclusioni e Sviluppi futuri.....	131
	Bibliografia	136

ALLEGATI	140
Allegato 1: Modelli di regressione lineare.....	142
1.1 Variabile dipendente: FREQ_AUCO	143
1.1.1 Mod.1.....	143
1.1.2 Mod.2.....	144
1.1.3 Mod. alternativo: Caso 3.....	144
1.2 Variabile dipendente: FREQ_BICI	150
1.2.1 Mod.1.....	150
1.2.2 Mod.2.....	151
1.3 Variabile dipendente: FREQ_AUPA	151
1.3.1 Mod.1.....	151
1.3.2 Mod.2.....	152
1.3.3 Mod. alternativo: Caso 3.....	153
1.4 Variabile dipendente: FREQ_BUS	154
1.4.1 Mod.1.....	154
1.4.2 Mod.2.....	155
1.5 Variabile dipendente: FREQ_METRO	155
1.5.1 Mod.1.....	155
1.5.2 Mod. 2.....	156
1.6 Variabile dipendente: FREQ_METRO	157
1.6.1 Mod.1.....	157
1.6.2 Mod.2.....	158
1.6.3 Mod. alternativo: Caso 3.....	158
Allegato 2: Analisi delle Varianze a due vie	160
2.2 Variabili indipendente: TIPO_INTERV e TP_ABBON	162
2.2.1 Variabile dipendente: FREQ_AUCO	162
2.2.2 Variabile dipendente: FREQ_BICI	164
2.2.3 Variabile dipendente: FREQ_AUPA	166
2.2.4 Variabile dipendente: FREQ_BUS	168
2.2.5 Variabile dipendente: FREQ_METRO	170
2.2.6 Variabile dipendente: FREQ_TRENO	172
2.3 Variabili indipendenti: TIPO_INTERV e CLASSE_ADULTI	174
2.3.1 Variabile dipendente: FREQ_AUCO	174

2.3.2	Variabile dipendente: FREQ_BICI.....	176
2.3.3	Variabile dipendente: FREQ_AUPA.....	178
2.3.4	Variabile dipendente: FREQ_BUS.....	180
2.3.5	Variabile dipendente: FREQ_METRO.....	182
2.3.6	Variabile dipendente: FREQ_TRENO.....	184
2.4	Variabili indipendenti: TIPO_INTERV e AUCS_ABBON.....	186
2.4.1	Variabile dipendete: FREQ_AUCO.....	186
2.4.2	Variabile dipendete: FREQ_BICI.....	188
2.4.3	Variabile dipendete: FREQ_AUPA.....	190
2.4.4	Variabile dipendente: FREQ_BUS.....	192
2.4.5	Variabile dipendente: FREQ_METRO.....	194
2.4.6	Variabile dipendente: FREQ_TRENO.....	196

INTRODUZIONE

Nel campo della mobilità e della pianificazione dei trasporti, i sondaggi risultano essere uno strumento fondamentale nel processo di ricerca. Essi rappresentano uno dei più importanti metodi utilizzati per ottenere dati sulla domanda di mobilità. Esistono diversi modi per eseguire un sondaggio. Nel linguaggio tecnico, è definita con l'acronimo CATI (Computer Aided Telephone Interview) la modalità di indagine nella quale il questionario viene somministrato telefonicamente e l'intervistatore è provvisto di un videoterminale nel quale provvede a registrare le risposte; invece, l'acronimo CAWI (Computer Assisted Web Interviewing) sta ad indicare un'altra modalità di sondaggio nella quale gli intervistati compilano direttamente online il questionario ed un computer remoto registra in tempo reale le risposte.

Per molti anni, i sondaggi telefonici sono stati i più utilizzati, ma al giorno d'oggi, con l'avvento di Internet e della tecnologia che si sono sostanzialmente impadroniti non solo del modo in cui gli individui vivono ma anche del modo in cui le aziende operano, sempre più istituti di ricerca si affidano a sondaggi online per ottenere i dati di cui hanno bisogno.

Ma il diverso protocollo di indagine influenza la risposta data per quanto concerne la frequenza di utilizzo di un determinato mezzo di trasporto? Si può affermare che, effettivamente, la modalità di indagine definisce due gruppi di persone distinte che non solo hanno differenti caratteristiche socio-economiche ma anche diverse attitudini di mobilità?

Nel presente lavoro di tesi si è cercato di rispondere a queste domande, andando a verificare se coloro i quali hanno risposto alla stessa indagine di mobilità, ma amministrata via web o al telefono, tendono ad essere due gruppi di persone distinte che utilizzano con frequenze diverse determinati mezzi di trasporto. In particolare, si terrà conto delle caratteristiche socio-demografiche degli intervistati, al fine di capire se esse hanno un ruolo importante sulle indagini di mobilità. Ciò contribuirà ad una migliore comprensione delle differenze nelle risposte riguardanti la frequenza d'uso dei mezzi tra chi è stato intervistato con sondaggio telefonico e chi, invece, via web.

L'analisi sperimentale che è stata condotta in questo studio, fa parte di un progetto più ampio, chiamato DEMONSTRATE, riguardante i nuovi servizi di mobilità, la diversione modale, la co-modalità e le applicazioni tecnologiche nel sistema di trasporto passeggeri. In tale progetto, svolto presso il Politecnico di Torino, è stata progettata un'indagine di mobilità multiprotocollo la cui somministrazione è stata affidata all'istituto SWG, al fine di definire la co-modalità ottimale per gli utenti del sistema di trasporto nell'area metropolitana di Torino.

Le risposte a questa indagine riguardanti le caratteristiche socio-economiche, gli spostamenti e le abitudini di mobilità degli intervistati, sono state inserite e codificate in un dataset del software SPSS Statistics, in modo tale da sviluppare alcune analisi di statistica descrittiva e inferenziale utili allo studio. In particolare, è stato possibile svolgere, con i dati a disposizione, delle analisi statistiche mediante test di significatività parametrici e non parametrici e, inoltre, calibrare modelli di regressione multivariata lineare e logistica. Le differenze inferenziali

ottenute considerando la modalità di sondaggio, la frequenza di utilizzo di determinati mezzi di trasporto e le caratteristiche socio-economiche del campione, hanno permesso di effettuare valutazioni descrittive sui due gruppi di persone contraddistinti dal diverso protocollo di indagine.

L'elaborato è suddiviso in una serie di capitoli, nei quali sono esposti i contenuti riguardanti le analisi trattate, il metodo utilizzato, gli strumenti adoperati e i risultati ottenuti in un modo il più possibile esaustivo. Si rimanda al capitolo 7 *Conclusioni e Sviluppi futuri* per le considerazioni finali sul lavoro svolto. Per una migliore comprensione dei valori ottenuti dalle diverse analisi statistiche, si faccia riferimento agli *Allegati*, posti in fondo all'elaborato per rendere più scorrevole la trattazione.

1 REVISIONE DELLO STATO DELL'ARTE

1.1 Sondaggi online vs sondaggi telefonici

Per molti anni, gli studiosi della mobilità e le agenzie di pianificazione dei trasporti hanno fatto affidamento sui sondaggi telefonici come mezzo principale per raccogliere dati sulle abitudini di spostamento delle famiglie, in particolare per effettuare un'analisi di mobilità sulla popolazione in larga scala. Nonostante la rapidità e l'efficienza delle indagini telefoniche che fanno di esse un approccio interessante e preferito, negli ultimi anni si sono riscontrate alcune problematiche. Secondo Dillman (2000) circa un terzo della popolazione degli Stati Uniti non aveva il numero di telefono nell'elenco telefonico all'epoca di questo studio, il 38% aveva la segreteria telefonica e in molti avevano il blocco delle chiamate verso le agenzie che effettuano sondaggi. Inoltre il passaggio dai telefoni fissi verso i telefoni cellulari ha reso l'indagine statistica più difficile ed elaborata per l'ottenimento di un campione veramente rappresentativo della popolazione.

Con l'avvento del computer e con la diffusione della rete internet, i sondaggi online hanno offerto un'alternativa interessante in particolare per il loro costo relativamente basso. Ma anche per questa tipologia di indagine, il campionamento risulta difficile. Couper (2000), nella rivista "*Public Opinion Quarterly*"¹, afferma che se, ad esempio, si considerano famiglie di una determinata area di studio, risulta difficile contattare alcune di esse per e-mail perché non esiste una banca dati di posta elettronica come può esser un elenco telefonico. Per questo motivo, in

¹Rivista accademica fondata nel 1937 dalla "Oxford University Press" per conto dell'"American Association for Public Opinion Research".

alcuni casi, i ricercatori inviano delle lettere di posta ordinaria per invitare le famiglie a compilare il test online oppure per comunicare che è possibile richiedere un sondaggio cartaceo, ma questo pone un ulteriore onere per l'intervistato che può scoraggiare la partecipazione. La non partecipazione, e quindi la non risposta da parte dell'intervistato, è un elemento di criticità, non solo per i sondaggi web ma anche per quelli telefonici. Se molti di coloro che rientrano nel campione non partecipano all'indagine oppure, prima di averla completata, la abbandonano, si otterrà un valore di percentuale di risposta² molto basso. Secondo Bradburn (2016), solitamente, si ottiene una percentuale di risposta molto bassa quando l'indagine risulta all'occhio dell'intervistato troppo lunga, se non si usa un linguaggio familiare oppure se si invia il sondaggio a persone sbagliate.

In passato, diversi studi hanno analizzato le differenze, in termini di vantaggi e svantaggi, tra un sondaggio telefonico e un sondaggio online su un campione rappresentativo di una popolazione. Studiosi come Al-Subaihi (2008), Fricker et al. (2005), Chang and Krosnick (2009) dimostrarono che i sondaggi online hanno una copertura della popolazione più scarsa rispetto ai sondaggi telefonici con selezione casuale (RDD³). Inoltre, i sondaggi online risultano, in alcuni casi, inferiori in numero perché non esiste una "mailing list" rappresentativa di una determinata area come può essere, ad esempio, l'elenco telefonico.

² Percentuale di persone che rispondono al sondaggio e lo completano. Tale percentuale è definita nel seguente modo:

$$\frac{N. di persone che rispondono all'indagine}{N. di persone a cui è stata inviata l'indagine} \times 100$$

³ Tecnica che permette di selezionare casualmente numeri di telefono rappresentativi di una popolazione. Ogni numero ha un'uguale probabilità di esser scelto.

Alcuni studi hanno confrontato le risposte delle indagini telefoniche RDD e online, in funzione del profilo socio-demografico e delle caratteristiche descrittive del campione. Ad esempio, Al-Subaihi (2008) ha scoperto che in Arabia Saudita, nonostante la copertura di internet non fosse un problema in quanto il numero di utenti internet è più elevato rispetto agli abbonati alle rete fissa telefonica, il Web Survey ha un tasso di risposta significativamente inferiore (circa il 70%) rispetto alle interviste telefoniche. Ciò indica che la carenza di utilizzo del sondaggio online è dovuta a fattori culturali piuttosto che alle infrastrutture tecniche e alla copertura di rete internet. Inoltre, sempre per questo motivo, i risultati di questo studio mostrano che gli uomini interagiscono significativamente di più delle donne con il sondaggio web, perlomeno nel paese dove il menzionato studio è stato condotto. Ciò dimostra che esiste un'interazione significativa tra la tipologia di sondaggio e il genere degli intervistati.

Fricker (2005) ha analizzato un sondaggio che valuta l'atteggiamento dell'intervistato nei confronti della scienza e la conoscenza delle nozioni scientifiche di base. Il target⁴ della popolazione esaminata per il campione è costituito da persone con età superiore ai 18 anni residenti negli Stati Uniti d'America. Innanzitutto, partendo dal confronto tra intervistati web e telefonicamente si aspettava di trovare le solite differenze che costituiscono il "digital divide"⁵: cioè, si aspettava di trovare che gli utenti del web sarebbero stati

⁴ Rappresenta il gruppo completo su cui si desidera raccogliere informazioni. Gli elementi di questo gruppo possono essere persone, famiglie, veicoli, aree geografiche o qualsiasi altra unità discreta. La definizione della popolazione target è definita, in molti casi, dagli obiettivi dell'indagine

⁵ Rappresenta il divario esistente tra chi ha accesso effettivo alle tecnologie dell'informazione (in particolare personal computer e Internet) e chi ne è escluso, in modo parziale o totale. I motivi

più giovani e più istruiti di quelli senza accesso al web. I risultati furono in linea con queste aspettative. In secondo luogo, Fricker riscontrò una maggiore partecipazione al sondaggio telefonico rispetto all'indagine sul Web: i tassi di risposta telefonici erano quasi il doppio di quelli degli utenti web. In termini di risposte date al sondaggio, si riuscì a dimostrare che nonostante la differenza d'età e d'istruzione degli intervistati nelle due diverse tipologie di sondaggio, non vi era una differenza significativa nel risultato finale dell'indagine. Questo risultato finale è stato anche ottenuto da Vehovar, Batagelj e Lozar (1999), i quali analizzarono un sondaggio fatto a un campione della popolazione slovena su argomenti riguardanti gli usi e le abitudini degli intervistati.

1.2 Fattori che influenzano le risposte degli intervistati

Alcune tipologie di studio dei sondaggi, hanno anche riscontrato che, in alcuni casi, le risposte a domande attitudinali non differiscono significativamente tra la tipologia di sondaggio web e telefonico. Infatti, in uno studio condotto da Schonlau (2003), si può notare come le risposte alle domande riguardanti le abitudini e le propensioni dell'intervistato non fossero significativamente diverse in funzione della tipologia di sondaggio.

Tuttavia, Greene, Speizer e Wiitala (2008) hanno sperimentato che coloro che venivano intervistati telefonicamente erano più propensi a fornire risposte socialmente desiderabili alle domande sullo stile di vita personale di quanto non lo fossero gli intervistati online. Si può notare che entra in gioco un concetto

di esclusione comprendono diverse variabili: condizioni economiche, livello d'istruzione, qualità delle infrastrutture, differenze di età o di sesso, appartenenza a diversi gruppi etnici, provenienza geografica.

fondamentale per l'analisi statistica che è quello del **Social Desirability Responding** (SRD), ovvero risposta di desiderabilità sociale.

“La desiderabilità sociale (SD) è la valutazione, socialmente condivisa, che in una certa cultura viene data ad un certo atteggiamento o comportamento individuale. Se uno di questi è valutato positivamente o negativamente, una domanda che abbia questo come oggetto può dare risposte distorte, perché l'intervistato può essere riluttante a rivelare opinioni o comportamenti che ritiene indesiderabili e può essere tentato di dare di sé la migliore immagine possibile, anche se non veritiera, in modo volontario o anche involontario” (Corbetta). In sostanza, la desiderabilità sociale rappresenta una tendenza degli intervistati a fornire risposte molto positive quando vengono poste domande su di sé, al fine di risultare positivamente all'occhio dell'intervistatore. Questo tipo di atteggiamento, può dar luogo a *bias*, ovvero a degli *errori* nelle risposte: esistono infatti numerosi studi che affermano che le persone tendono a dare delle risposte che non rappresentano il loro reale pensiero quando esse rispondono a interviste e questionari relativi a comportamenti socialmente non condivisi (uso di sostanze stupefacenti) o condivisi (intenzione di voto). Questa modalità di risposta può non essere considerata soltanto come una tendenza dell'individuo a dare di sé agli altri una immagine positiva. L'altro fattore che sta alla base della Social Desirability Responding è l'*autoinganno*: rappresenta la tendenza inconsapevole a dare risposte positive con l'intento di proteggere la stima di sé, la predisposizione a vedere se stessi sotto una luce positiva.

Oltre alla desiderabilità sociale, si può considerare l'**Acquiescence Response Bias** (errore di acquiescenza nelle risposte, ARB) come elemento che influenza negativamente i dati ottenuti da un sondaggio. L'acquiescenza esprime "la tendenza delle persone di scegliere risposte che esprimono accordo piuttosto che negative" (Corbetta). Questo accade quando viene posta una domanda con risposte sbilanciate del tipo: accordo/disaccordo o sì/no. È una forma comune di errore di misurazione nei sondaggi e può avere effetti sostanziali. Watson (1992) affermò che questo tipo di risposta può esser dovuto alla cortesia, perché le persone tendono ad esser d'accordo (esser educate) a meno che non abbiano sentimenti negativi molto forti su un argomento. Oppure, in presenza di un lungo sondaggio, gli intervistati tendono a rispondere "d'accordo" senza nemmeno leggere le domande per finire il prima possibile l'intervista. Secondo Watson, sebbene le domande con risposte bilanciate del tipo "sì/no" o "accordo/disaccordo" siano facili da amministrare, richiedano meno tempo e suscitano meno risposte "non so" rispetto alla formulazione di domande specifiche, è opportuno non utilizzare queste tipologie di domande in quanto permettono di avere una quantità significativa di Acquiescence Response Bias.

Dillman et al. (1996) hanno delineato tre differenze principali tra un sondaggio telefonico e un sondaggio web. La prima riguarda la presenza o l'assenza di un intervistatore. Al telefono l'intervistatore crea un'interazione sociale che spesso porta gli intervistati a concentrarsi nel fornire risposte positive per dare un'immagine positiva di se, in particolare per le domande su argomenti sensibili. Secondo Dillman, gli utenti contattati telefonicamente danno risposte più

socialmente desiderabili a un intervistatore che a un computer. La seconda, invece, evidenzia che ci sono delle differenze tra la presentazione visiva delle domande nei sondaggi web e la comunicazione fonica delle indagini telefoniche. Gli intervistati con modalità web sono in grado di rivedere le domande e le relative risposte, mentre gli intervistati telefonicamente, in genere, ascoltano le domande solo una volta. L'onere cognitivo richiesto per elaborare la domanda del sondaggio, formulare una risposta e identificare la categoria di risposta più appropriata è sostanzialmente maggiore per gli intervistati telefonici. Una manifestazione di questa differenza è che gli intervistati telefonici scelgono più spesso l'ultima delle varie opzioni offerte (tendenza che si chiama "Effetto Recency"), rispetto agli intervistati di posta, che sono più propensi a selezionare la prima opzione elencata. La terza differenza riguarda il modo in cui viene somministrato un sondaggio. Nei sondaggi Web si possono utilizzare modelli di salto, offrire visualizzazioni grafiche personalizzate e limitare le risposte alle domande. In genere, questo porta l'intervistatore a controllare quali risposte precedentemente date possono esser riviste e modificate.

In quale misura la modalità di indagine web o telefonica influenza le risposte a domande relative a comportamento, atteggiamento e abitudini sociali? Greene, Speizer e Wiitala (2008) hanno cercato di rispondere a questa domanda analizzando un sondaggio effettuato sia con modalità web che con modalità telefonica. Il target della popolazione è costituito da 751 dipendenti di una grande azienda americana operante nel settore manifatturiero. Le domande dell'indagine riguardano lo stato di salute, l'utilizzo dell'assistenza sanitaria, i comportamenti

relativi alla salute e la conoscenza del sistema di assistenza medica. Da questa analisi, si può notare come in alcune domande (riguardanti, ad esempio, la ricerca di informazioni sanitarie da parte dell'intervistato) vi è una tendenza verso risposte socialmente desiderabili tra gli intervistati telefonicamente rispetto a chi ha compilato il sondaggio online. Nonostante ciò, la differenza risulta lieve e generalmente non statisticamente significativa. Dove si possono osservare, invece, grandi differenze modali è sui temi legati allo stile di vita del mangiare sano, dell'attività fisica e dello stato di salute degli intervistati. Quindi le differenze modali dovuti a una risposta di desiderabilità sociale variano considerevolmente in funzione della tipologia di domanda: quanto più la domanda è personale maggiore sarà la desiderabilità sociale tra chi risponde telefonicamente rispetto a chi risponde online.

1.3 Studi nel campo dei trasporti

Tradizionalmente, i questionari online e le interviste telefoniche sono state le modalità principali di raccolta dei dati nella ricerca sui trasporti. Anche in questo ambito, le problematiche riscontrate da Couper (2000) e Dillman (2000) viste nei precedenti paragrafi sono del tutto valide. Alla luce di ciò, è comprensibile che diversi ricercatori abbiano sollecitato ad esplorare il potenziale di mezzi alternativi di raccolta dei dati utili ad un'indagine di mobilità. Ad esempio, i sistemi di posizionamento globale (Draaijer et al. 2000, Murakami et al. 2000, Wolf et al. 2001, Stopher et al. 2003), il monitoraggio dei telefoni cellulari (Guensler e Wolf, 1999) e dei sistemi di realtà virtuale (Tan e Timmermans, 2004) sono stati suggeriti ed esaminati come mezzi per raccogliere dati sul comportamento di

viaggio delle persone. Ci si può aspettare che tali nuove tecnologie siano molto utili per registrare il comportamento reale riguardante lo spostamento degli individui. Tuttavia, a volte, alcune informazioni devono essere completate da dati di giudizio o attitudinali, che non possono essere raccolti attraverso tali tecnologie. Adler et al. (2000, 2002), Lee et al. (2002), Marca (2003), hanno suggerito di utilizzare i metodi tradizionali di indagine, ovvero i sondaggi web e telefonici, come mezzi per raccogliere tali tipi di dati.

Negli ultimi anni sono stati condotti diversi studi per integrare più metodi di indagine al fine di ottenere database di migliore qualità. Verreault e Morency (2018), hanno combinato un diario di viaggio condotto tra la popolazione studentesca di una regione del Quebec e un'indagine telefonica condotta tra diversi nuclei familiari residenti nella medesima area. Oltre ad aumentare la frequenza di campionamento, l'integrazione di entrambi i sondaggi ha migliorato alcuni indicatori: si è riscontrato una riduzione dell'eterogeneità dei tassi di campionamento tra i segmenti di popolazione che si traduce in un incremento dei tassi di viaggio giornalieri e di transito in alcune fasce d'età dei campioni. Inoltre, Verreault e Morency hanno notato che l'integrazione ha portato ad un altro effetto positivo: si ha una riduzione dei fattori di ponderazione correlati all'aumento della dimensione del campione. Quindi, l'integrazione consente al sondaggio di essere più rappresentativo dell'intera popolazione di riferimento raggiungendo gli intervistati esclusi dal tipico schema di campionamento.

Diana (2012) ha utilizzato i dati di un'indagine svolta dall'ISTAT, riguardanti le frequenze di viaggio e l'abilità nell'uso del computer, per valutare le differenze in

termini di mobilità tra coloro che possono rispondere a un sondaggio online e coloro che invece non possono. Erano presenti, in primo luogo, delle correlazioni tra gli intervistati che utilizzano maggiormente strumenti tecnologici e le frequenze di viaggio: quest'ultime, secondo i mezzi di viaggio presi in considerazione, presentano una percentuale più alta tra coloro che utilizzano tecnologie web e digitali. Inoltre, si è osservato che vi sono delle differenze nelle frequenze di guida dell'auto a causa di una serie di fattori riguardanti le attitudini e le capacità tra gli intervistati con sondaggio web e il resto del campione. Diana, ha riscontrato, inoltre, che l'utilizzo del trasporto pubblico può essere legato alla diffusione di internet nella popolazione, in quanto si presume che chi utilizza internet abbia uno stile di vita più attivo anche in termini di mobilità.

Bonsall e Shires (2009), hanno riscontrato alcune differenze principali tra un sondaggio telefonico e un sondaggio web analizzando un questionario sulle aspettative dei datori di lavoro su argomenti riguardanti il livello di pendolarismo dei dipendenti e la disponibilità a viaggi d'affari in una serie di futuri scenari di ampliamento delle tecnologie digitali. Coloro che hanno risposto alle due versioni del questionario erano diversi sotto alcuni punti di vista: rispetto agli intervistati con versione cartacea, i rispondenti alla versione web tendono ad essere più giovani, a venire da organizzazioni aziendali più grandi e ad avere un'esperienza più recente di videoconferenza. Queste differenze sembrano aver portato quest'ultimi a dare maggior peso ai fattori legati alla tecnologia nella stima di futuri livelli di spostamento a casa e di lavoro in diversi scenari; secondo Bonsall e Shires, i rispondenti del web, essendo più orientati all'innovazione tecnologica, avevano

più probabilità di vedere la tecnologia come un pre-requisito importante ed efficace per il lavoro da casa e per la sostituzione di viaggi lavorativi con mezzi di tecnologie avanzate come la videoconferenza. Inoltre, coloro che hanno risposto ad un sondaggio online hanno opinioni meno solide sui futuri viaggi d'affari di un'azienda, dovuto forse alla giovane età di chi ha risposto al sondaggio web e quindi alla meno esperienza sulle questioni che influenzano queste tipologie di viaggio.

Le caratteristiche socio-demografiche degli intervistati hanno un ruolo molto importante sulle indagini di mobilità per capire le differenze nelle risposte tra chi è intervistato con indagine telefonica e chi invece via web. Xing e Handy (2014) hanno analizzato la frequenza d'uso della bici utilizzando due diversi metodi di indagine: una tradizionale indagine telefonica RDD e un sondaggio online. Le caratteristiche socio-demografiche degli intervistati in ciascun campione differivano significativamente: gli intervistati online sono più giovani, sono studenti o lavoratori, possiedono un'auto e non hanno limitazioni fisiche alla guida con bici. Ciò portava a giustificare il perché di chi veniva intervistato online dichiarava di usare con alte frequenze la bici come mezzo di spostamento.

Molteplici sono gli studi che si concentrano sulle differenze inferenziali tra i campioni dei sondaggi online e telefonici nel campo dei trasporti. Chi lavora nell'ambito della pianificazione dei trasporti si affida ai dati dell'indagine raccolti tramite sondaggi telefonici o online o con altri metodi, per stimare modelli esplicativi che permettono di valutare l'effetto di diversi fattori sul comportamento di mobilità. Ciò permette di costruire modelli che prevedono, ad esempio, la futura

domanda di trasporto. Quindi studiare e comprendere le risposte in base ai diversi metodi di indagine risulta un ambito abbastanza importante per ottemperare a quelli che sono gli obiettivi di chi studia la mobilità.

2 INTRODUZIONE ALLE ANALISI STATISTICHE EFFETTUATE E LORO ESECUZIONE TRAMITE SPSS

2.1 Ipotesi nulla e livello di significatività del test

Un importante contributo all'analisi di dati oggetto del presente lavoro è fornito dall'*inferenza statistica*, un procedimento che permette di ottenere informazioni e caratteristiche di una popolazione attraverso lo studio di un campione rappresentativo estratto secondo principi di casualità.

L'inferenza statistica non è da confondere con il calcolo delle probabilità; in quest'ultimo sono note le caratteristiche della popolazione e si è in grado di stabilire la probabilità dei diversi possibili risultati di un esperimento, invece nell'inferenza statistica non si conoscono gli aspetti relativi all'intera popolazione e, attraverso tecniche di indagine, si usa una parte di essa (il campione) per individuarne le caratteristiche.

Per tale motivo, vengono utilizzati i *test statistici di significatività* che permettono di esprimere la probabilità che un singolo campione noto sia estratto da una popolazione con specifiche caratteristiche oppure che due o più campioni derivino da una stessa popolazione. Questo permette inoltre di stabilire se esiste una connessione tra due variabili. In pratica si va a quantificare la significatività di risposta di un test, ovvero la probabilità che un determinato dato meriti una interpretazione e non sia dovuto al caso.

I test statistici considerano inizialmente una *ipotesi zero* H_0 , definita anche *ipotesi nulla*. Quando si esegue l'analisi di due o più campioni, l'ipotesi nulla considera che

non esiste nessuna differenza tra i campioni rispetto al parametro analizzato. Quindi accettare l'ipotesi nulla porta ad affermare che i due o più campioni provengono dalla stessa popolazione e le eventuali differenze presenti nei campioni siano dovute al solo caso. D'altro canto, rifiutare l'ipotesi nulla, può essere sia una decisione *probabilmente* giusta che errata, in quanto i test statistici non sono rappresentativi di una certezza matematica. La probabilità di effettuare tale errore viene espressa dal *livello di significatività* α del test. Solitamente il livello di significatività viene fissato al 5% (0,05) o all'1% (0,01) ma comunque può essere un valore scelto a piacere dallo sperimentatore. La decisione di accettare o rigettare l'ipotesi nulla viene effettuata usando il P-Value: "esprime la probabilità di ottenere un risultato altrettanto estremo o più estremo di quello osservato se la diversità è interamente dovuta alla sola variabilità campionaria, assumendo quindi che l'ipotesi nulla sia vera" (Signorelli). In sostanza se tale valore è minore del livello di significatività α , allora si rifiuta l'ipotesi nulla e quanto più è basso il P-Value, più significativo sarà il risultato del test statistico. In particolare, viene definito significativo se il P-Value è compreso tra 0,01 e 0,05, molto significativo se $0,001 \leq P < 0,01$, estremamente significativa se $P < 0,001$. Infine, è fondamentale porre attenzione su un concetto di particolare rilevanza: «statisticamente significativo». Esso non vuol dire di considerevole importanza o essenziale, significa soltanto che ciò che è stato studiato (o osservato) è «difficilmente dovuto al caso».

In contrapposizione all'ipotesi zero H_0 , si esprime l'*ipotesi alternativa* H_1 o H_a , definita anche *ipotesi sostantiva*. Rifiutare l'ipotesi nulla comporta, quindi, sia

accettare l'ipotesi alternativa e sia affermare che esistono alcune differenze tra i campioni analizzati rispetto ad alcuni parametri considerati. In sostanza i due o più campioni esaminati appartengono alla medesima popolazione e le eventuali differenze non sono dovute al caso.

Si può affermare, quindi, che le due ipotesi sono complementari, ossia considerate insieme esauriscono tutte le possibilità riguardanti il valore che può assumere il parametro in esame.

Con un test statistico di significatività si possono commettere sostanzialmente due tipi di errori [Tab. 1]:

- Errore α o di I tipo: il risultato del test statistico è tale da rifiutare l'ipotesi nulla H_0 quando essa in realtà è vera. In questo caso il risultato del test è definito falso positivo. La probabilità di compiere un errore di I tipo è in funzione del livello di significatività α scelto inizialmente. Quindi alla fine di un qualsiasi test statistico, se le evidenze portano a rifiutare l'ipotesi nulla, non si può sapere con certezza se si è commesso un errore di questo tipo oppure no. Si può solo affermare che la probabilità di aver commesso un errore di I tipo, nel caso in cui fosse vera l'ipotesi H_0 , sarebbe molto bassa e comunque pari al livello di significatività α . Invece, la probabilità complementare $1-\alpha$ definisce il *livello di protezione* del test e rappresenta la probabilità di accettare l'ipotesi zero quando essa è vera. Un test che presenta un alto livello di protezione è definito *molto conservativo*.

- Errore β o di II tipo: il risultato del test statistico è tale da accettare l'ipotesi nulla H_0 quando essa in realtà è falsa. In questo caso il risultato del test è definito falso negativo. La probabilità di compiere un errore β dipende dal numero di soggetti facenti parte del campione analizzato e anche dal livello di significatività α : aumentare il numero di osservazioni e il valore di α prescelto può portare a una riduzione di un errore di II tipo. Invece, la probabilità complementare $1-\beta$ definisce il *livello di potenza* del test e rappresenta la probabilità di rifiutare l'ipotesi zero quando essa è falsa. Quindi un test con un alto livello di potenza porta ad aver maggiori certezze nel rifiutare l'ipotesi H_0 quando essa è effettivamente falsa.

IPOTESI	H_0 VERA	H_0 FALSA
NON RIFIUTO H_0	Buona decisione Vero Positivo ($1-\alpha$)	Errore di II tipo β Falso Negativo
RIFIUTO H_0	Errore di I tipo α Falso Positivo	Buona decisione Vero Negativo ($1-\beta$)

Tab. 1 - Errori di primo e secondo tipo e loro probabilità.

I test statistici possono esser classificati in due gruppi: test *a una coda* (o unilaterale) e test *a due code* (o test bilaterale). Quando la regione di rifiuto è costituita da solo un intervallo (a destra se il segno è positivo, a sinistra se è negativo), il test si dice a una coda. Quando invece la regione di rifiuto è costituita

da due intervalli, ossia da due code della distribuzione, il test si dice a due code. Scegliendo, ad esempio, $\alpha = 0,05$ si hanno due zone di rifiuto collocate ai due estremi, ognuna con un'area del 2,5%, per un test a due code, invece si ha una sola zona di rifiuto in un test a una coda con un'area del 5% [Fig. 1]. Un semplice modo per stabilire di che tipo è un test statistico, senza dover conoscere la/le regioni di rifiuto è il seguente: per un test a due code nell'ipotesi alternativa compare il segno \neq , mentre per un test a una coda compare il segno del maggiore oppure del minore ($>$ o $<$). Il test unilaterale è più conservativo, invece il test bilaterale è più potente e restrittivo: con un test ad una coda è maggiore la probabilità di dimostrare l'esistenza di una differenza significativa (quindi maggiore probabilità di rifiutare l'ipotesi nulla) rispetto a un test a due code.

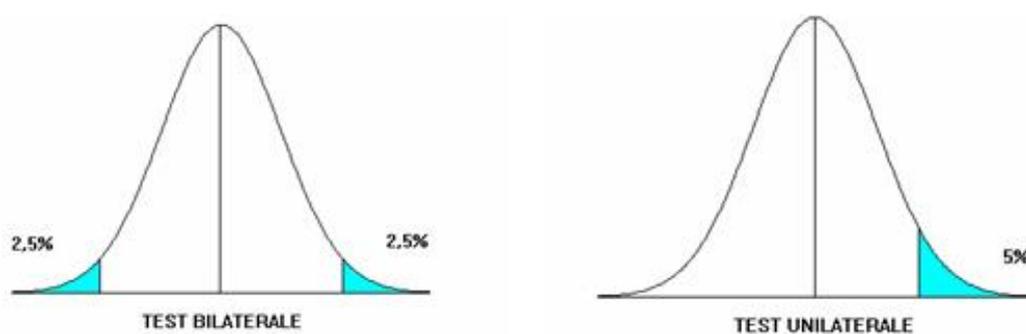


Fig. 1 - Distribuzione normale in cui sono presenti in blu le aree di rifiuto dell'ipotesi nulla rispettivamente per una test a due e a una coda e per un livello di significatività α pari al 5%.

In funzione del tipo di distribuzione della variabile aleatoria, si può scegliere quale test applicare a una serie di dati. In sostanza, si possono dividere i test statistici in due grandi categorie: i test parametrici e i test non parametrici. I primi sono utilizzati nel momento in cui si hanno campioni dei quali si conoscono i parametri statistici relativi alla distribuzione (di norma rappresentata con la curva di Gauss)

come la media, la deviazione standard, la varianza, etc. Quando non è nota la tipologia di distribuzione, è opportuno utilizzare dei test alternativi come quelli non parametrici, in quanto non richiedono la stima di parametri statistici. Nonostante ciò, esistono alcune eccezioni: alcuni test parametrici come il test F ed il test T-Student sono definiti *robusti* rispetto alla condizione di non-normalità dei dati presi in esami, infatti anche quando la distribuzione è influenzata da forte asimmetria, il calcolo della significatività può esser valido in quanto influenzato da errori abbastanza marginali. Alcuni autori affermano che una forte asimmetria può fissare la variazione del P-Value tra 0.04 e 0.07 quando il livello di significatività α è 0.05, mentre esso varia tra 0.005 e 0.02 quando α è pari all' 1%. È un'approssimazione che consente di non esser molto critici sulla validità dei risultati e quindi di accettarli, anche perché i test non parametrici non sempre portano ad ottenere stime più precise.

Si può riassumere nel seguente schema i passi in cui si articola un test statistico di significatività:

1. Si scelgono l'ipotesi nulla H_0 e l'ipotesi alternativa H_a ;
2. Si sceglie il livello di significatività α a cui si vuole eseguire il test;
3. In funzione del tipo di test (una coda o due code) e del valore di α scelto, si determinano i valori critici e la regione di rifiuto;
4. Si sceglie l'ampiezza campionaria, si raccoglie il campione, si calcola dai dati del campione il valore della statistica test e si vede se appartiene o no alla regione di rifiuto;

5. Si prende la decisione: rifiutare o non rifiutare l'ipotesi nulla al livello di significatività α stabilito.

2.2 Test T-Student dei campioni indipendenti

Il test t dei campioni indipendenti viene utilizzato per determinare se esiste una differenza tra le medie di due popolazioni su una variabile dipendente continua. Più specificamente, permette di determinare se la differenza tra questi due gruppi è statisticamente significativa. Questo test è anche noto con un certo numero di nomi diversi, tra cui il t-test indipendente, il t-test delle misure indipendente, il t-test tra soggetti, il t-test non appaiato e il t-test di Student.

Si usa la distribuzione della variabile aleatoria t per confrontare le medie di due campioni indipendenti. Tale distribuzione è definita come:

$$t = \frac{\text{differenza fra le medie campionarie}}{\text{errore standard della differenza fra le medie campionarie}}$$

La forma della distribuzione di t è simmetrica e a campana, ma con una dispersione che aumenta con il numero di gradi di libertà [Fig. 2] . Di conseguenza non vi è una sola curva t ma esiste un intero gruppo di distribuzioni t, una per ogni grado di libertà.

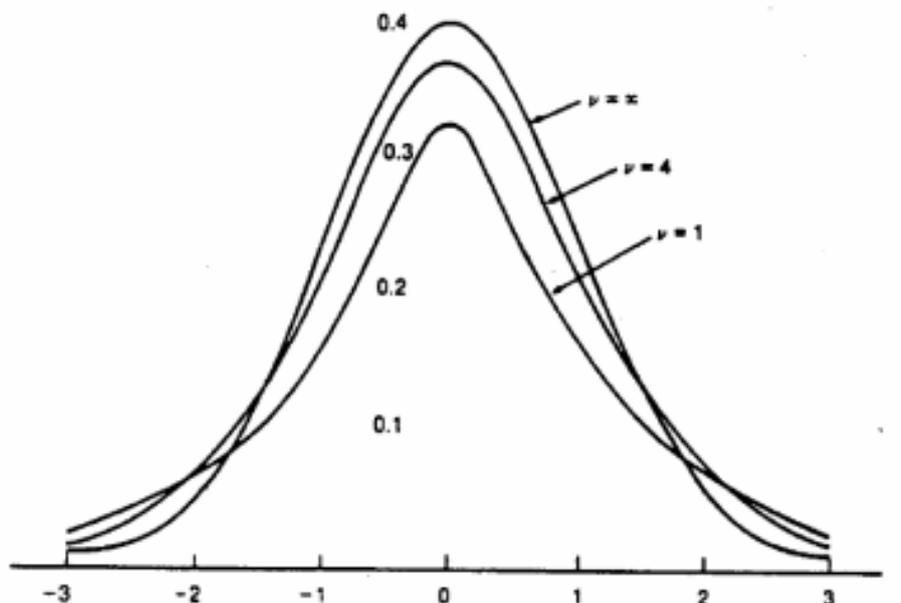


Fig. 2 - Distribuzione di Student. Sull'asse x è espressa la variabile aleatoria t , mentre sull'asse y la densità di probabilità. Le tre curve si riferiscono agli esperimenti con un numero di gradi di libertà pari a 1, 4 ed ∞ .

L'ipotesi nulla e alternativa per un t-test di campioni indipendenti sono le seguenti:

H_0 : le medie dei due campioni sono uguali (cioè $\mu_1 = \mu_2$)

H_a : le medie dei due campioni non sono uguali (cioè $\mu_1 \neq \mu_2$)

Con il t-test si calcolerà un p-value, che rappresenta la probabilità di ottenere un test statistico almeno estremo come quello che è stato effettivamente osservato, supponendo che l'ipotesi nulla sia vera. Se la probabilità è sufficientemente piccola (in genere $p < 0.05$) è possibile concludere che le medie della popolazione dei due gruppi non sono uguali e che si può accettare l'ipotesi alternativa e rifiutare l'ipotesi nulla. In tal caso il risultato viene definito come statisticamente significativo. In alternativa, si accetta l'ipotesi nulla se la probabilità è maggiore del livello di significatività α (di solito $p > 0.05$).

Per eseguire un t-test di campioni indipendenti, ci sono sei ipotesi che devono essere considerate. Le prime tre ipotesi riguardano il design delle misurazioni da analizzare, mentre le altre tre si riferiscono alle caratteristiche dei dati in possesso.

Queste ipotesi sono:

1. La variabile dipendente deve essere di tipo continuo;
2. La variabile è di tipo dicotomico (cioè si compone di due categorie indipendenti);
3. Si dovrebbe avere l'indipendenza delle osservazioni. Se così non fosse, è più appropriato utilizzare un test t per campioni appaiati;
4. Non ci dovrebbero essere valori anomali significativi nei due gruppi della variabile indipendente in termini di variabili dipendenti;
5. La variabile dipendente dovrebbe essere approssimativamente distribuita normalmente per ciascun gruppo della variabile indipendente. Tuttavia il test t dei campioni indipendenti è considerato "robusto" per la violazione della normalità. Ciò significa che alcune violazioni di questa ipotesi possono essere tollerate e il test fornirà comunque risultati validi. Pertanto, si afferma che questo test richiede solo approssimativamente dati con una distribuzione normale. Un test che permette di testare questa ipotesi è il test di Shapiro-Wilk per la normalità;
6. Condizione di omoschedasticità, ossia la varianza dei gruppi è la stessa di quella della popolazione. Un test che permette di verificare questa ipotesi è quello di Levene sull'uguaglianza delle varianze: se il valore di significatività "Sign.", fornito da questo test, è minore del livello di

significatività α , allora la condizione di omoschedasticità non è rispettata. Per superare questo problema, si può ricorrere al Test di Welch (analogo non parametrico del Test t-Student). Tale test non prende in analisi la media come parametro, bensì la mediana, perché essa, in caso di eteroschedasticità, risulta esser un indicatore più robusto; infatti la mediana è un parametro più stabile nei confronti di valori o troppo grandi o troppo piccoli che possono esser presenti in una distribuzione.

2.3 Test ANOVA a una via

Se si desidera determinare se vi siano differenze statisticamente significative tra le medie di due o più popolazioni, è possibile utilizzare come tecnica statistica l'analisi della varianza (Anova, Analysis Of VAriance). Tale tecnica statistica è di tipo parametrico e presuppone che la numerosità campionaria sia rilevante in numero e che si abbia l'indipendenza delle osservazioni.

In base al numero di variabili dipendenti e indipendenti, l'ANOVA può esser classificata con diversi modelli:

- Disegni a una via: è prevista una sola variabile indipendente;
- Disegni fattoriali: sono previste due o più variabili indipendenti;
- Analisi univariata: si considera una sola variabile dipendente;
- Manova (Multivariate Analysis of Variance): si considerano due o più variabili dipendenti.

Le variabili dipendenti sono di tipo quantitativo, invece le variabili indipendenti possono essere: quantitative, note anche come covariate, oppure categoriali.

Quest'ultime, a loro volta, sono distinte in variabili tra soggetti (o fattori between) e entro soggetti (o fattori within).

Per eseguire il test ANOVA a una via, oltre alle condizioni sulle variabili viste precedentemente, le ulteriori ipotesi che bisogna considerare riguardano la condizione di omoschedasticità nel confronto tra più campioni e la distribuzione della variabile dipendente approssimativamente normale. Tuttavia se i dati violano l'ipotesi di omogeneità delle varianze, è necessario utilizzare l'Anova modificata di Welch.

L'ipotesi nulla e alternativa per un ANOVA Test sono le seguenti:

H_0 : le medie di due o più campioni sono uguali (cioè $\mu_1 = \mu_2 = \dots = \mu_k$)

H_a : almeno due medie dei campioni considerati non sono uguali (cioè $\mu_1 \neq \mu_2 \neq \dots \neq \mu_k$)

Dalle ipotesi si può notare che si fa inferenza sulle medie, nonostante ciò si lavora con la scomposizione della varianza.

L'ANOVA prende in analisi la varianza delle popolazioni considerate, in particolare si analizza la distribuzione F (dal suo creatore Fisher) data dal rapporto tra la varianza delle medie delle popolazioni a confronto e la media delle varianze di ogni popolazione:

$$F = \frac{n \cdot s_x^2}{s_p^2} = \frac{\text{varianza tra gruppi}}{\text{varianza entro i gruppi}}$$

dove n è il numero di campioni analizzati. Tale distribuzione è continua, compresa tra 0 e infinito e dipende dai gradi di libertà del numeratore e quelli del denominatore [Fig. 3].

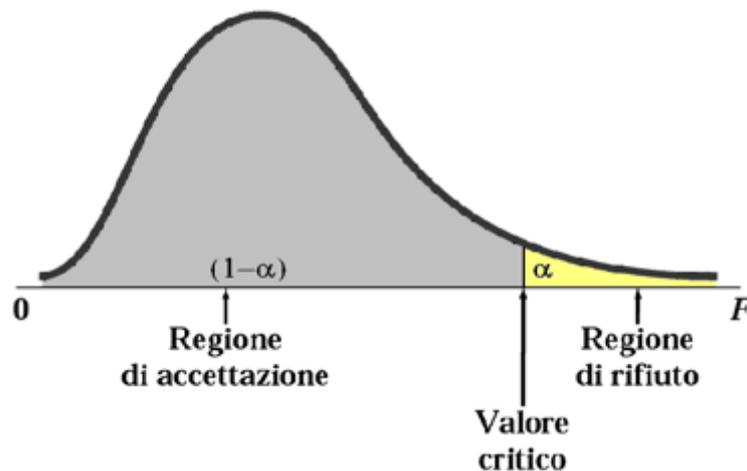


Fig. 3 - Distribuzione F di Fischer. Sull'asse x è espressa la variabile aleatoria F , mentre sull'asse y la densità di probabilità.

Per un livello di significatività α prefissato, la distribuzione F indica quando l'ipotesi nulla può essere accettata e quando, invece, è rifiutata. Se l'ipotesi nulla è vera, allora il valore di F è prossimo all'unità perché si confrontano due stime della varianza della stessa popolazione. In caso contrario, la distribuzione F avrà valori alti, in quanto il numeratore esprime la varianza di medie differenti. Quest'ultimo caso non afferma che le medie sono tutte significativamente diverse l'una dall'altra ma, piuttosto, che c'è almeno una coppia di medie la cui differenza è statisticamente significativa. È quindi necessario individuare quali sono le medie diverse tra loro, procedendo quindi a verificare l'uguaglianza tra tutte le possibili coppie di medie attraverso opportuni test. In tali test, detti test post-hoc, per ogni coppia di medie l'ipotesi nulla è che la differenza tra queste sia pari a zero, mentre

l'alternativa è che le due medie differiscano significativamente tra loro. Il più comune test post hoc è quello di Tukey (nel caso di condizione di omoschedasticità non rispettata, si considera lo stesso modello corretto da Welch).

2.4 Test ANOVA a due vie

L'ANOVA a due vie, definita anche come ANOVA bidirezionale o fattoriale, è un test parametrico che viene utilizzato per determinare se esiste un effetto di interazione tra due variabili indipendenti su una variabile dipendente continua. In sostanza, l'ANOVA a due vie è un test che permette di valutare come l'effetto su una variabile dipendente cambi a seconda dei livelli delle variabili indipendenti considerate. In tal modo, si mettono in evidenza le eventuali interazioni tra le variabili. In molti casi, l'ANOVA a due vie può essere considerata come un'estensione dell'ANOVA a una via. Per questo motivo, le ipotesi e le procedure d'analisi viste in quest'ultimo test sono valide anche per l'ANOVA fattoriale, facendo attenzione nel considerare le stratificazioni che si vengono a formare in questo tipo di analisi. Si hanno due variabili indipendenti in cui ognuna di esse è costituita da due o più categorie o gruppi. Una variabile indipendente con solo due gruppi è definita come variabile dicotomica mentre una variabile indipendente con tre o più gruppi è definita politomica.

2.5 Test U di Mann-Whitney

Il test U di Mann-Whitney, definito anche test dell'ordine robusto dei ranghi, viene spesso presentato come alternativa non parametrica al test t dei campioni indipendenti, in quanto la variabile in studio non rispetta le condizioni di

applicabilità dei test parametrici. Il test U viene quindi utilizzato quando si hanno dati non distribuiti normalmente oppure quando la variabile dipendente è ordinale. In sostanza il test U permette di verificare la significatività della differenza tra le mediane.

Le ipotesi considerate dal seguente test sono:

H₀ : le mediane delle due popolazioni da cui sono estratti i campioni sono uguali

$$(Me_1 = Me_2)$$

H_a : le mediane delle popolazioni da cui sono estratti i campioni non sono uguali

$$(Me_1 \neq Me_2)$$

La procedura nel test U di Mann-Whitney può essere schematizzata nel seguente modo:

1. Si ordinano i dati in maniera crescente, tenendo separati i due campioni analizzati, ad esempio campione X e campione Y, che hanno rispettivamente un numero di osservazioni pari a n ed m.
2. Si assegnano i ranghi che tengono conto dei ties, ovvero dei valori uguali a i quali viene assegnato il valore medio del rango;
3. Si calcola il valore di U in funzione di n che rappresenta la dimensione del campione usato come riferimento e di S che esprime la somma dei ranghi del medesimo campione:

$$U = S - \frac{n(n+1)}{2}$$

4. Si confronta il valore di U trovato nel punto 3 con il valore U critico tabellato funzione del livello di significatività e del numero di osservazioni dei singoli campioni. Si rifiuta l'ipotesi nulla H_0 se il valore calcolato di U risulta rispettivamente maggiore del valore positivo di U critico o minore del valore negativo di U critico.

Il significato statistico del test U di Mann-Whitney è espresso, anche in questo test, dal P-value. Considerando questo valore di probabilità, è possibile prendere una decisione riguardo al mantenimento dell'ipotesi nulla o all'accettazione dell'ipotesi alternativa. Quando il valore di p è maggiore di .05 non si può rifiutare l'ipotesi nulla e il risultato non è statisticamente significativo. Al contrario, quando il valore di p è minore di .05, si accetta l'ipotesi alternativa e il risultato risulta esser statisticamente significativa.

2.6 Test H di Kruskal-Wallis

Il test H di Kruskal-Wallis è un metodo non parametrico per verificare l'uguaglianza delle mediane di diversi campioni, ovvero permette di controllare se tali campioni provengono da una stessa popolazione (o da popolazioni con uguale mediana) oppure no. Il test H è generalmente considerato l'alternativa non parametrica all'ANOVA a una via, che può essere utilizzato quando i dati non superano le ipotesi dell'ANOVA a una via. Questo potrebbe accadere se si hanno dati non distribuiti normalmente, oppure se si analizza una variabile dipendente ordinale (invece l'ANOVA a una via richiede una variabile dipendente continua).

Anche per il test H di Kruskal-Wallis, come per i precedenti, bisogna garantire l'indipendenza delle osservazioni e la variabile indipendente deve avere al minimo due variabili categoriche.

L'ipotesi nulla e l'ipotesi alternativa per il seguente test sono:

H_0 : le mediane di due o più popolazioni da cui sono estratti i campioni sono uguali

(cioè $Me_1 = Me_2 = \dots = Me_k$)

H_a : almeno due mediane delle due o più popolazione da cui sono estratti i

campioni non sono uguali (cioè $Me_1 \neq Me_2 \neq \dots \neq Me_k$)

La metodologia di analisi statistica anche in questo caso prevede che se il P-Value ha un valore inferiore al livello di significatività α scelto, allora il test H risultata statisticamente significativo e quindi la mediana di almeno un campione è diversa dalla mediana di un altro campione. Per scoprire quali mediane differiscono tra i diversi campioni è necessario eseguire un test post hoc. Nel caso del test Kruskal-Wallis H, i confronti a coppie vengono effettuati usando il criterio di Dunn.

2.7 La correlazione tra due variabili: coefficiente punto-biseriale e Indice Phi

Nella correlazione si ipotizza che le variazioni di una variabile Y possano essere associate, in una certa misura, alle variazioni di una variabile X e, allo stesso tempo, che le variazioni di X possano essere associate alle variazioni di Y. Dunque, non è possibile parlare di variabile indipendente e di variabile dipendente, né di relazione "causa-effetto". Il fatto che due variabili varino insieme non vuol dire che

esse siano sicuramente in una relazione di casualità tra loro: potrebbero esserlo come non esserlo.

Per calcolare la correlazione tra una variabile metrica e un'altra variabile dicotomica, si ricorre all'uso del coefficiente punto-biserial r_{pb} .

Invece, l'Indice Phi r_{phi} si usa quando si è interessati a misurare la relazione tra due variabili dicotomiche.

Entrambi i coefficienti possono assumere valori compresi tra -1 e 1. Nello specifico, se il coefficiente di correlazione è maggiore di zero si ha una correlazione positiva, se è minore di zero si ha una correlazione negativa, se è uguale a zero, le due variabili considerate non sono correlate. Inoltre, se il valore del coefficiente di correlazione è compreso tra 0 e 0,3 si ha una correlazione debole, se è compreso tra 0,3 e 0,7 si ha una correlazione moderata, infine, se è compreso tra 0,7 e 1, la correlazione può essere definita forte.

2.8 Regressione lineare multivariata

La teoria della regressione lineare multivariata risponde all'obiettivo di studiare la dipendenza di una variabile indipendente Y da un insieme di m variabili esplicative quantitative X_1, X_2, \dots, X_m , dette *regressori*, mediante un modello lineare del tipo:

$$Y = \beta + \beta_1 X_1 + \beta_2 X_2 \dots + \beta_m X_m + \varepsilon$$

dove il termine " $\beta + \beta_1 X_1 + \beta_2 X_2 \dots + \beta_m X_m$ " rappresenta la componente sistematica del modello, la variabile casuale " ε ", invece, è la componente d'errore del modello. Nello specifico, il parametro " β " rappresenta l'intercetta, ovvero il valore atteso di Y quando le altre variabili esplicative sono nulle. Invece, i termini

“ $\beta_1, \beta_2, \dots, \beta_m$ ” rappresentano i coefficienti di regressione o di pendenza, ovvero l'effetto medio su Y di una variazione unitaria in X_m , tenendo costante il valore delle altre variabili esplicative presenti.

La teoria della regressione lineare multivariata ammette variabili esplicative metriche che abbiano un andamento lineare in correlazione alla variabile indipendente considerata. Tuttavia, in molti casi si rende necessaria l'introduzione anche di variabili categoriche. In tal caso, possono essere inserite in un modello di regressione mediante l'uso delle variabili dummy. Esse assumono in genere solo i valori (0,1) a seconda che il fattore di interesse abbia assunto una delle sue modalità. Se il fattore ha solo due modalità, basta una variabile dummy, altrimenti per k modalità servono k - 1 dummy.

Nel modello di regressione multivariata, dal momento che si è in presenza di almeno due variabili esplicative, la proporzione di variabilità della Y spiegata dalle variabili esplicative è rappresentata dal coefficiente di determinazione R^2 . In sostanza, esso è una misura dell'adattabilità del modello ai dati osservati. Tale coefficiente è definito dal rapporto tra la somma dei quadrati della regressione (SQR) e la somma totale dei quadrati (SQT): rappresenta una misura dell'adattabilità del modello ai dati osservati.

$$R^2 = \frac{SQR}{SQT}$$

Il coefficiente di determinazione R^2 ha valori compresi tra 0 e 1. Se è pari a 1 allora esiste una perfetta relazione lineare fra il fenomeno analizzato e la sua retta di regressione. Se è pari a 0 non esiste alcuna relazione fra le due variabili, mentre i

valori compresi fra 0 e 1 forniscono una indicazione sulla efficacia della retta di regressione nel sintetizzare l'oggetto dell'analisi.

Per sviluppare tale modello, si può ricorrere semplicemente alle funzionalità del software Excel: cliccando su Dati > Analisi Dati > Regressione, si apre una finestra in cui è possibile scegliere la variabile indipendente e le variabili esplicative del modello.

2.9 Regressione logistica

Il modello di regressione logistica nasce nell'ottica di spiegare la probabilità di possesso di un attributo, o di accadimento di un evento, in relazione ad una serie di possibili determinanti, di variabili esogene. In sostanza, il modello di regressione logistica viene utilizzato quando si è interessati a studiare o analizzare la relazione causale tra una variabile dipendente dicotomica e una o più variabili indipendenti quantitative. Convenzionalmente, per le variabili dicotomiche, ad una modalità viene associato il valore 0, all'altra il valore 1, per cui la sua distribuzione risulta esser binomiale. In questi casi, quindi, sebbene sia ugualmente possibile applicare il modello della regressione semplice, da un punto di vista matematico, un modello non lineare sarebbe più appropriato. Infatti, nella sua formulazione ($Y = a + bX$), il modello lineare implica che i valori della variabile Y possano andare da $-\infty$ a $+\infty$. A tal scopo possiamo applicare la trasformazione logistica che consente di controllare i valori e restringerli nel range della probabilità (0; 1). Il modello viene descritto dalla seguente equazione:

$$\text{logit}(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots + \beta_k X_k$$

dove $\text{logit}(Y) = \ln\left(\frac{Y}{1-Y}\right)$ rappresenta la probabilità che l'evento Y si verifichi.

2.10 Realizzazione dei test precedenti in IBM® SPSS® Statistics

Praticare un'analisi statistica senza l'aiuto di strumenti software adeguati renderebbe molto difficile l'ottenimento dei risultati in breve tempo e l'elaborazione dei dati stessa. L'analisi statistica ha beneficiato solo in tempi relativamente recenti di software e hardware utili a esaminare una gran mole di dati e di conseguenza l'uso in grande scala di queste tecniche avviene in tempi successivi a quelli di formulazione e divulgazione. Uno dei software di statistica più popolare e inoltre maggiormente intuitivo e rapido nella rielaborazione dei dati è *SPSS, Statistical Package for Social Science*. È possibile elaborare dati provenienti da quasi tutti i formati di estensione dei file e generare rapporti con report di dati, tabelle, grafici di distribuzione e trend, analisi statistiche complesse e statistiche descrittive. Le finestre di dialogo, i menu e l'editor dei comandi permettono di effettuare analisi elaborate senza dover digitare una singola riga di sintassi. La visualizzazione e l'immissione dei dati di lavoro risulta esser una procedura facile ed efficiente grazie alla somiglianza dell'interfaccia grafica a un semplice foglio di calcolo Excel. Nella finestra principale dei dati, nella parte inferiore sinistra, sono presenti due importanti azioni di comando: "vista dati" e "vista variabili". La prima è dove vengono digitati i dati ovvero le variabili e possono esser rappresentate sia con valori numerici che con stringhe. La finestra vista variabili invece ha 10 colonne che indicano al software diversi aspetti dei valori di misurazione, ad esempio se una variabile è qualitativa o quantitativa.

2.10.1 Test t-Student dei campioni indipendenti

Per eseguire un test-T Student dei campioni indipendenti in SPSS, dal menu in alto della finestra di editor, cliccare su *Analizza > Confronta medie > Test T per campioni indipendenti*. Si apre la finestra *Test T per campioni indipendenti* [Fig. 4] in cui verranno specificate le variabili da utilizzare nell'analisi. È possibile spostare una o più variabile su due aree: variabili di raggruppamento (variabili indipendenti) e variabili del test (variabili dipendenti).

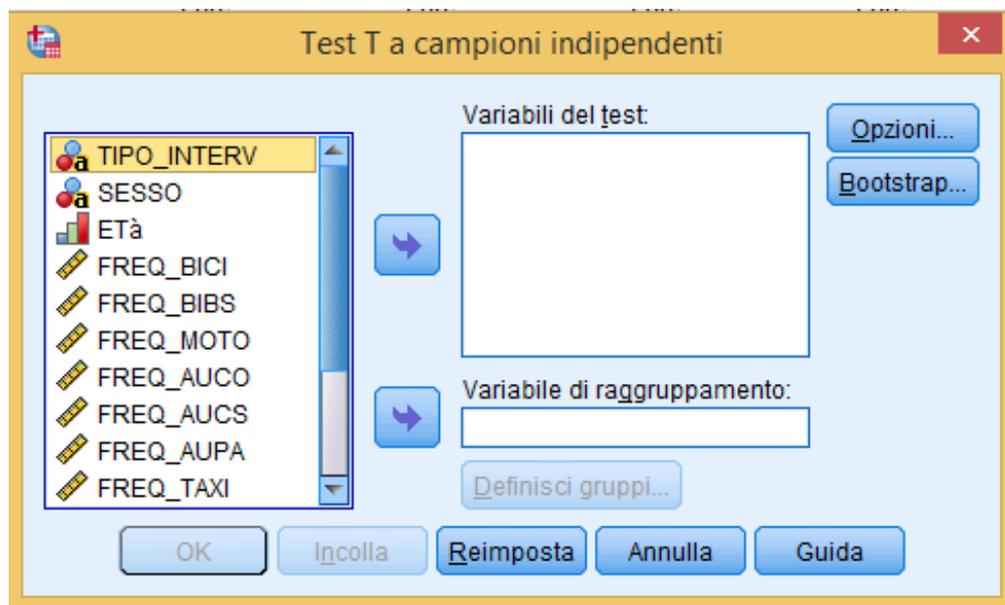


Fig. 4 - Finestra *Test T a campioni indipendenti* in SPSS.

Cliccando sul pulsante *Opzioni*, si aprirà una finestra in cui è possibile inserire la percentuale di intervallo di confidenza. Questa impostazione non influisce sulla statistica del test o sul P-Value o sull'errore standard ma solo sul calcolo dei limiti superiori e inferiori dell'intervallo di confidenza. Si può inserire qualsiasi valore compreso tra 1 e 99, ma come abbiamo visto precedentemente il livello di

significatività è di solito pari a 0.05 o 0.01 e quindi avrebbe senso inserire rispettivamente come valore 95 o 99%.

Nella finestra di Output, verrà visualizzata una tabella con due righe: la prima fa riferimento al test T dei campioni indipendenti standard, ovvero al caso in cui l'ipotesi di omoschedasticità è rispettata; la seconda, invece, al Test di Welch, in cui le varianze presunte non sono uguali.

2.10.2 ANOVA a una via

Per eseguire ANOVA a una via in SPSS, fare clic su *Analizza > Confronta medie > ANOVA a una via*. Si apre una finestra [Fig. 5] in cui verranno specificate le variabili dipendenti e indipendenti da utilizzare nell'analisi statistica. È possibile spostare una o più variabili in una delle due aree: *Elenco dipendenti* o *Fattore* che rappresentano rispettivamente le variabili dipendenti e indipendenti.

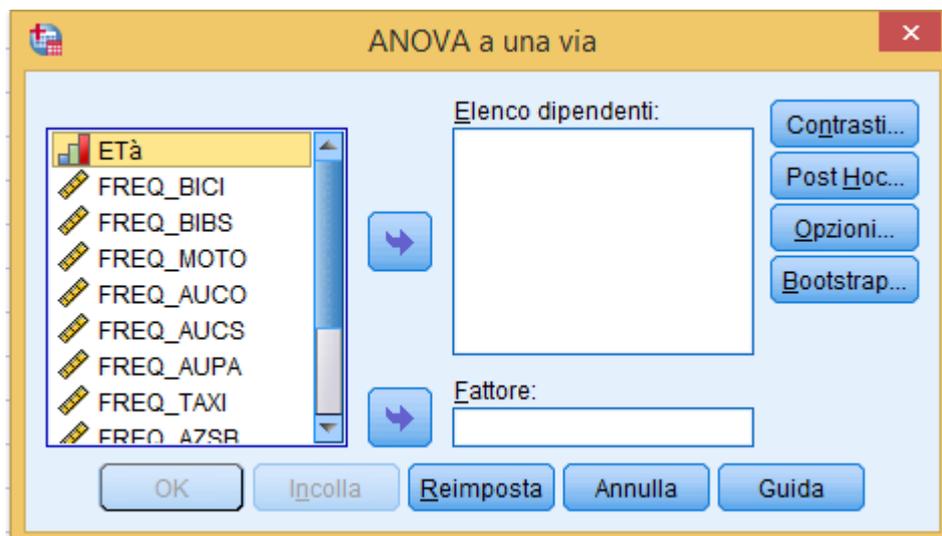


Fig. 5 - Finestra ANOVA a una via in SPSS.

Cliccando sul pulsante Opzioni verrà visualizzata una finestra in cui è possibile specificare quali statistiche includere nell'output, se includere un grafico delle medie e, infine, scegliere come l'analisi affronterà i valori mancanti.

Cliccando invece sul pulsante Post Hoc, si aprirà una finestra in cui è possibile scegliere quali test di confronti multipli è possibile utilizzare.

2.10.3 ANOVA a due vie

In SPSS, per eseguire l'ANOVA a due vie fare clic su *Analizza > Modello Lineare Generale > Univariata*. Si apre una finestra [Fig. 6] in cui è possibile scegliere le variabili dipendenti e indipendenti da utilizzare, la tipologia di grafici da ottenere e le interazioni multiple attraverso i test post-hoc da visualizzare. In quest'ultimo caso la procedura richiede uno script di sintassi. Facendo clic su *Incolla* dalla finestra *Univariata*, si apre una nuova finestra di *Editor di Sintassi* [Fig. 7] in cui sarà possibile aggiungere, oltre alla dicitura già presente che rappresenta le opzioni precedentemente selezionate, un nuovo insieme di funzioni (o riga di comando) che permette di effettuare dei test post hoc, e quindi dei confronti multipli tra i vari gruppi. Una volta finita la scrittura dei comandi, basta cliccare dall'*Editor di Sintassi* su *Esegui > Tutto* per far comparire una nuova finestra di *Output* che permette di visualizzare tutte le analisi e opzioni scelte con questo test.



Fig. 6 – Finestra Univariata data dalla procedura dell'ANOVA a due vie in SPSS.

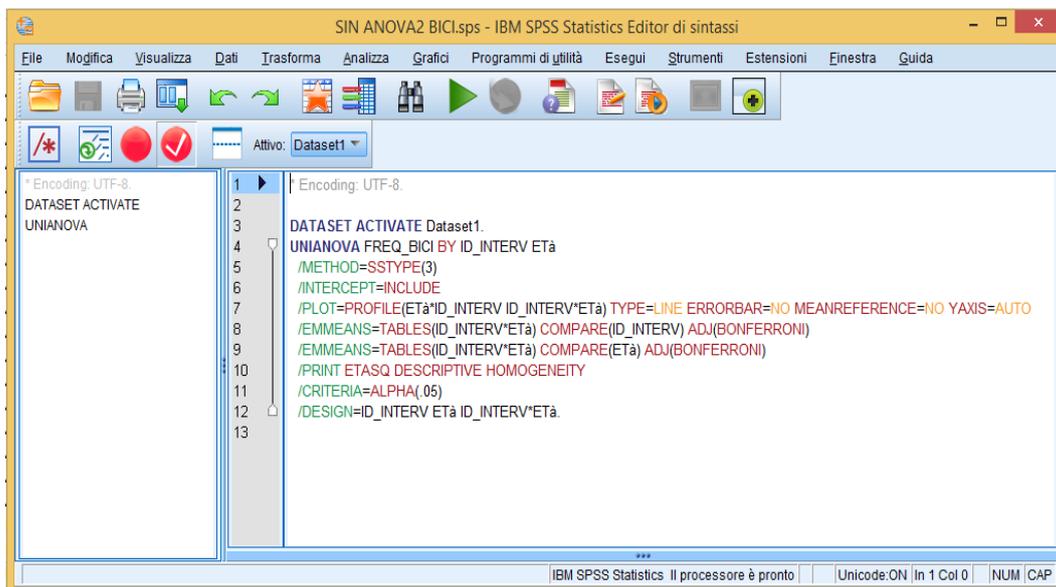


Fig. 7 - Finestra di Editor di Sintassi per ANOVA a due vie in SPSS.

2.10.4 Test U di Mann-Whitney

Esistono due diverse procedure in SPSS per eseguire un test U di Mann-Whitney: una procedura nuova e una procedura legacy. La prima viene utilizzata nel caso in cui la variabile indipendente ha solo due gruppi, invece la seconda quando la variabile indipendente ha tre o più gruppi. Nel caso di una nuova procedura, fare clic su *Analizza > Test non parametrici > Campioni indipendenti*. Si aprirà una

finestra [Fig. 8] in cui è possibile scegliere l'obiettivo del test, le variabili indipendenti e dipendenti da analizzare e l'intervallo di confidenza.

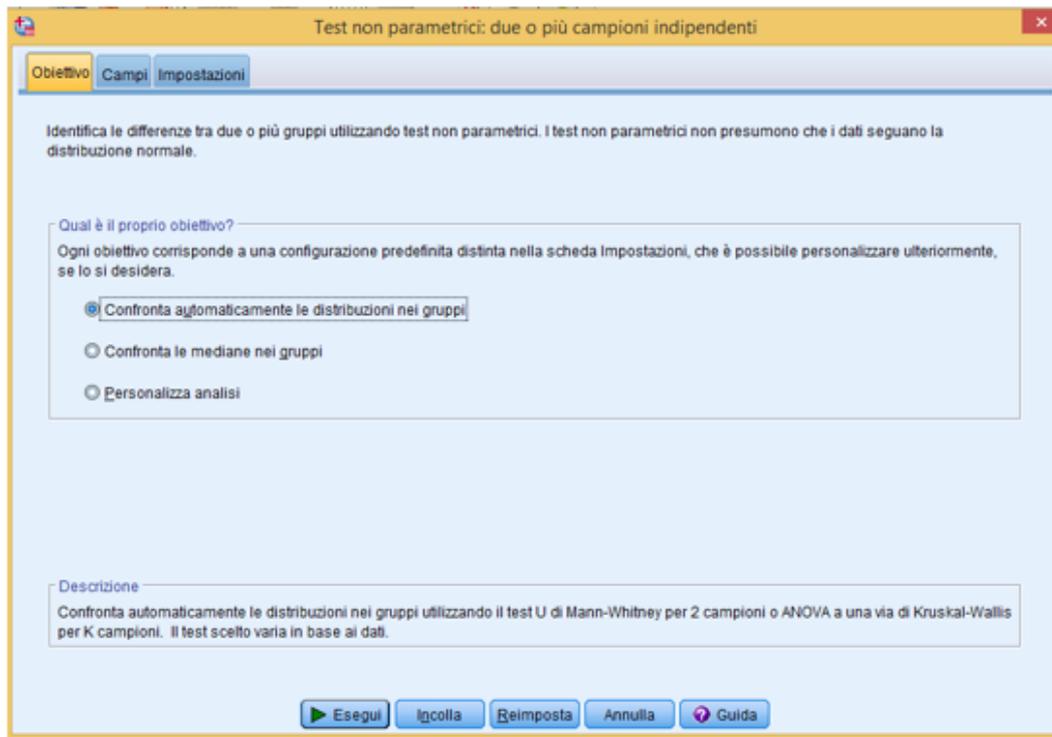


Fig. 8 - Finestra Test non parametrici: due o più campioni indipendenti in SPSS- Test U di Mann-Whitney, Procedura normale.

Nel caso di procedura legacy, invece, cliccare su Analizza > Test non parametrici > Finestre di dialogo legacy > 2 campioni indipendenti. Verrà presentata la finestra di dialogo Test a due campioni indipendenti [Fig. 9] in cui è possibile scegliere le variabili indipendenti e dipendenti e la tipologia di test da analizzare.

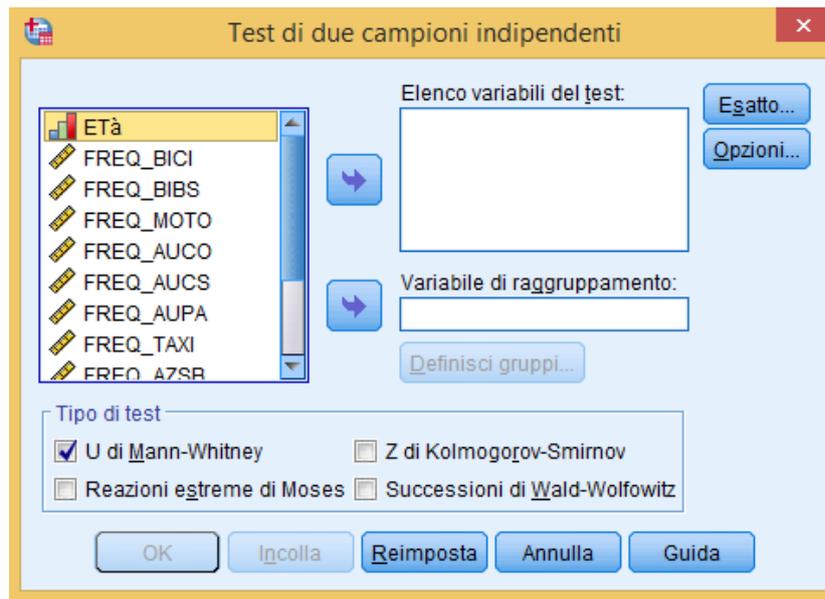


Fig. 9 -Finestra Test a due campioni indipendenti in SPSS- test U di Mann-Whitney, Procedura Legacy.

Cliccando sul pulsante Opzioni verrà visualizzata una finestra in cui è possibile specificare quali statistiche includere nell'output (descrittive e/o quartili) e scegliere come l'analisi affronterà i valori mancanti.

2.10.5 Test H di Kruskal-Wallis

In SPSS, il test Kruskal-Wallis H, viene analizzato allo stesso modo nella nuova procedura del test U di Mann-Whitney, con la sola eccezione che nella sezione Impostazioni [Fig. 10] bisogna cliccare su Personalizza i test e spuntare il comando ANOVA a una via di Kruskal-Wallis.

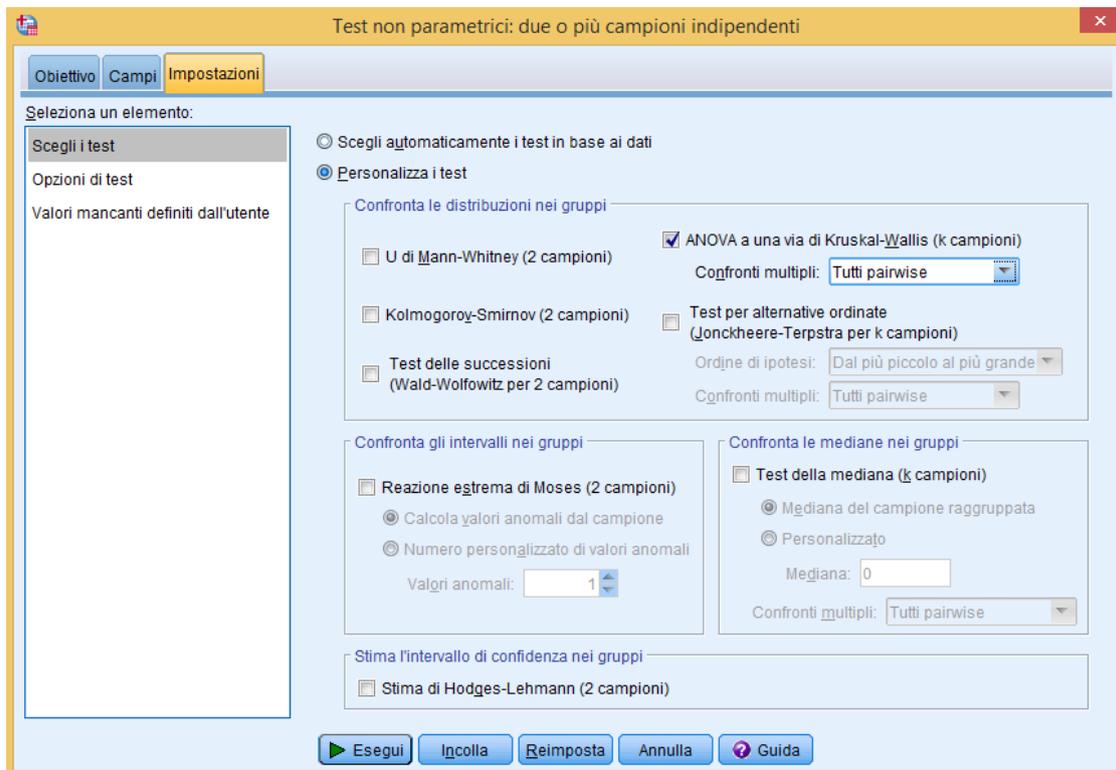


Fig. 10 - Finestra Test non parametrici: due o più campioni indipendenti in SPSS- Test Kruskal-Wallis H.

2.10.6 Coefficiente punto-biseriale e Indice Phi

Per eseguire in SPSS un'analisi di correlazione definita dal coefficiente punto-biseriale, fare clic su *Analizza > Correlazione > Bivariata*. Si apre una finestra [Fig. 11] in cui verranno specificate la variabile dicotomica e le variabili metriche da utilizzare. Per tale tipo di correlazione, bisogna spuntare nella sezione "coefficienti di correlazione" la dicitura "Pearson". Nonostante sia selezionata un'analisi di Pearson, il software SPSS riconosce in automatico che sono state scelte due variabili di cui una metrica e una dicotomica e procede, quindi, in automatico, alla

valutazione della correlazione in funzione del coefficiente di punto-biserial e non di Pearson.

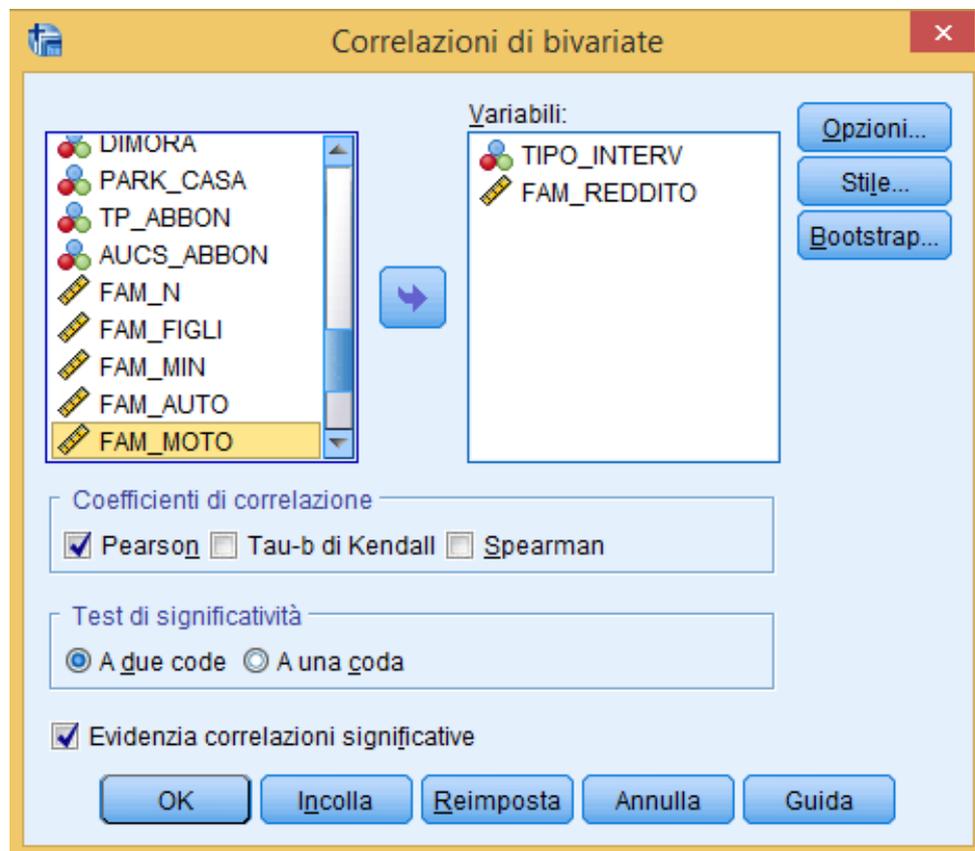


Fig. 11 - Finestra Correlazioni di bivariate data dalla procedura della correlazione definita dal coefficiente punto-biserial in SPSS.

Per eseguire, invece, in SPSS un'analisi di correlazione definita dall'Indice Phi, fare clic su Analizza > Statistiche Descrittive > Tabelle di contingenza. Si apre una finestra [Fig. 12] in cui si può selezionare, per righe e per colonne, le variabili dicotomiche da utilizzare. Cliccando sul pulsante Statistiche verrà visualizzata una finestra in cui è possibile spuntare, nella sezione Nominale, la dicitura "Phi e V di Creamer" per effettuare un'analisi di correlazione definita dall'Indice Phi.



Fig. 12 - Finestra Tabelle di contingenza data dalla procedura dell'Indice Phi di correlazione in SPSS.

2.10.7 Regressione lineare multivariata e regressione logistica

Per eseguire una regressione lineare multivariata e una regressione logistica in SPSS, dal menu in alto della finestra di editor, cliccare su *Analizza > Regressione* e, nel primo caso di analisi, cliccare su *Lineare*, invece, nel secondo caso, cliccare su *Logistica binaria*. Si apre, rispettivamente, la finestra *Regressione Lineare* e *Regressione Logistica* [Fig. 13 Fig. 14] in cui è possibile specificare le variabili endogene ed esogene da utilizzare nel modello. Inoltre, sempre nelle finestre di visualizzazione sopracitate, è possibile selezionare il metodo di analisi denominato *INSERISCI* che permette di creare un unico modello di regressione costituito da tutte le variabili esogene prese in esame. Nel caso di regressione logistica con

variabili esogene di tipo categorico, bisogna cliccare su Catoriale e selezionare tali tipi di variabili.

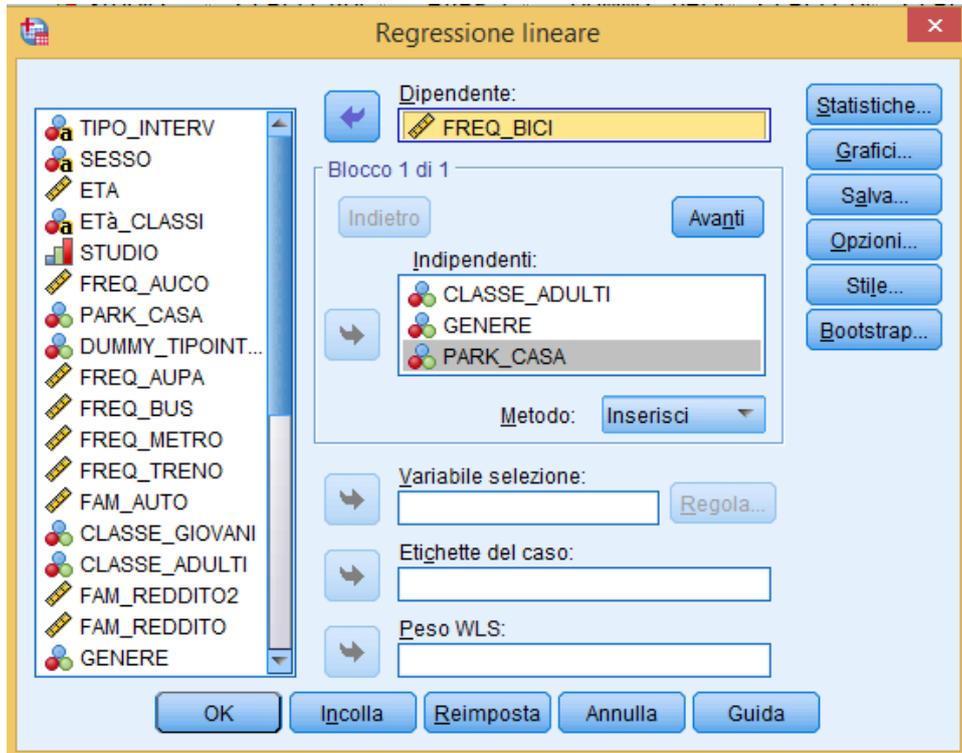


Fig. 13 Finestra regressione lineare in SPSS



Fig. 14 - Finestra Regressione logistica in SPSS

3 DESCRIZIONE DELL'INDAGINE DEMONSTRATE E IPOTESI PRELIMINARI

3.1 Indagine di partenza

L'analisi sperimentale condotta in questa tesi fa parte di un progetto più ampio riguardante i nuovi servizi di mobilità nelle aree urbane, la diversione modale, la co-modalità e le applicazioni tecnologiche nel sistema di trasporto passeggeri. In questo progetto, chiamato DEMONSTRATE, svolto presso il Politecnico di Torino, è stata progettata un'indagine di mobilità multiprotocollo la cui somministrazione è stata affidata all'azienda SWG⁶. L'obiettivo principale del progetto è stato quello di chiarire la relazione tra il mix di interventi congiunti derivanti da diverse discipline (ingegneria dei trasporti, psicologia sociale, economia applicata) e la massima diversione modale che può essere ragionevolmente raggiunta in un contesto urbano, al fine di definire la co-modalità ottimale per gli utenti del sistema di trasporto nell'area di indagine.

L'area di studio è l'Area Metropolitana di Torino (AMT) [Fig. 15]. Con 1.735.631⁷ abitanti, essa rappresenta una conurbazione dovuta alla continuità urbana (prima cintura) ed all'interazione sociale ed economica di alcuni comuni della seconda cintura torinese e la stessa città di Torino. Non va confusa con la Città Metropolitana di Torino, ente territoriale molto vasto che ha sostituito dal primo gennaio 2015 la Provincia di Torino. L'Area Metropolitana torinese comprende

⁶ Azienda Italiana che progetta e realizza ricerche di mercato, di opinione, istituzionali, studi di settore e osservatori, analizzando e integrando i trend e le dinamiche del mercato, della politica e della società.

⁷ Dato aggiornato al 31 Dicembre 2016 e fornito dall' Osservatorio Demografico Territoriale del Piemonte (DeMos).

Torino più 31 comuni della cintura aderenti al consorzio “Agenzia per la Mobilità Piemontese (AMP)”, il cui compito è quello di migliorare e ottimizzare il trasporto pubblico nella suddetta area. Il sistema dei trasporti nell’area metropolitana di Torino ingloba diverse opportunità di mobilità per il cittadino: metropolitana, rete tranviaria, tranvia, rete di filobus e autobus, ferrovie, aeroporti e diversi servizi di car sharing e bike sharing.



Fig. 15 - Area Metropolitana di Torino

A Settembre del 2016 è stata condotta una prima indagine su 1526 individui abitanti nell’Area Metropolitana di Torino. La modalità di indagine ha considerato

sia un sondaggio web (CAWI⁸) che un sondaggio telefonico (CATI⁹). Con distanza periodica di 4 mesi, a febbraio 2017 e giugno 2017, sono state intervistate rispettivamente altre 1460 e 1480 persone residenti nella medesima area di studio della prima indagine. In particolare, considerando tutte e tre le indagini, il 60% è residente nella città capoluogo, invece il restante 40% negli altri comuni dell'Area Metropolitana di Torino [Tab. 2].

	Torino		1 ^a e 2 ^a cintura		AMT (Torino+ 1 ^a e 2 ^a cintura)	
	n.	%	n.	%	n.	%
1 ^a Indagine	922	60	604	40	1526	100
2 ^a Indagine	876	60	584	40	1460	100
3 ^a Indagine	887	60	593	40	1480	100

Tab. 2 - Numero e percentuale di individui contattati nelle tre indagini del progetto *Demonstrate* in funzione dell'area di residenza.

Le domande del sondaggio non sono variate né per tutte e tre le raccolte di dati, né per la tipologia di sondaggio somministrato. In totale, quindi, sono stati contattati 4466 individui, di cui 3390 CATI e 1076 CAWI. Rispettivamente corrispondono al 76% e al 24% del campione totale [Tab. 3].

⁸ Acronimo di "Computer Assisted Web Interviewing". Rappresenta particolari indagini, sondaggi e ricerche via web in cui:

1. il software si occupa dell'invio delle email e di classificare come "Ha risposto" l'unità dopo che ha completato il questionario;
2. il rispondente invitato via email (normalmente) clicca su un link e compila il questionario.

⁹ Acronimo di "Computer Assisted Telephone Interviewing". Rappresenta le indagini, i sondaggi o le ricerche telefoniche in cui:

1. il software fa le telefonate, le passa all'intervistatore quando qualcuno risponde al telefono, "campiona", verifica la qualità delle interviste etc;
2. l'intervistatore legge la domande dallo schermo ed inputa sulla tastiera la risposta dell'intervistato.

	CATI		CAWI		CATI+CAWI	
	n.	%	n.	%	n.	%
1 ^a Indagine	1203	79	323	21	1526	100
2 ^a Indagine	1098	75	362	25	1460	100
3 ^a Indagine	1089	74	391	26	1480	100
Tot. Indagini (1 ^a +2 ^a +3 ^a)	3390	76	1076	24	4466	100

Tab. 3 - Numero e percentuale di individui contattati nelle tre indagini del progetto Demonstrate in funzione della tipologia di intervista.

Considerando la percentuale di risposta, si può notare che i valori precedentemente elencati in Tab. 3 diminuiscono lievemente, perché bisogna tener conto dei casi in cui l'intervista, una volta iniziata, è stata interrotta o per problemi tecnici o perché l'intervistato si è rifiutato di completarla. Il valore del tasso di risposta, in questa analisi, è molto alto ed è pari a 99%: in pochi non hanno completato l'intervista e gli errori tecnici son stati del tutto assenti. Quindi, ai fini dell'indagine, la Tab. 3 può esser riscritta nel seguente modo [Tab. 4]:

	CATI		CAWI		CATI+CAWI	
	n.	%	n.	%	n.	%
1 ^a Indagine	1195	79	309	21	1504	100
2 ^a Indagine	1098	75	362	25	1460	100
3 ^a Indagine	1089	74	391	26	1480	100
Tot. Indagini (1 ^a +2 ^a +3 ^a)	3382	76	1062	24	4444	100

Tab. 4 - Numero e percentuale di interviste complete nelle tre indagini del progetto Demonstrate in funzione della tipologia di intervista.

Il sondaggio è stato strutturato essenzialmente in 5 parti. Ognuna di esse riguarda specifici argomenti:

- A. SEZIONE INTRODUTTIVA: l'intervistato risponde a domande di natura personale, come il genere, l'età e l'occupazione lavorativa;
- B. REPERTORIO DEGLI SPOSTAMENTI: si sono chiesti la frequenza d'utilizzo nell'ultimo mese di alcuni mezzi di trasporto e i diversi luoghi visitati nelle ultime 24 ore; in particolare origine, destinazione, orario di partenza e arrivo, motivo, modo/i, eventuali interscambi;
- C. APPROFONDIMENTO SU UN MACRO-SPOSTAMENTO: il software estrae a sorte uno degli spostamenti elencati nella sezione precedente per poi fare all'intervistato domande più specifiche come, ad esempio, la frequenza settimanale dello spostamento considerato, il numero di persone con cui ha viaggiato l'intervistato, la tipologia di attività compiuta durante lo spostamento, etc;
- D. INDAGINE ATTITUDINALE: parte dell'indagine in cui si valuta il grado di soddisfazione/gradimento e la propensione dell'individuo ad usare mezzi di trasporto alternativi per lo spostamento considerato nella sezione C;
- E. ESPERIMENTI DI SCELTE DICHIARATE: in questa sezione si fanno delle ipotesi di spostamento per vedere l'attitudine dell'individuo basandosi sul totale dei costi, delle attese e del tempo totale specificati nel punto B;
- F. CARATTERIZZAZIONE SOCIO-DEMOGRAFICA: si traccia un profilo personale dettagliato dell'individuo che è stato intervistato. Ad esempio si chiede il titolo di studio, se è in possesso di una patente di guida, se ha un abbonamento ai servizi di trasporto pubblico locale e/o di car sharing, il reddito, etc.

Le domande volte agli intervistati sono state essenzialmente di due tipologie:

1. *Domande Strutturate o A Risposte Chiuse (DS)*: si chiede all'intervistato di scegliere una risposta da un set di alternative. Lo svantaggio di questo tipo di domande è quello di selezionare la serie di risposte da presentare, nonostante non si conoscano a volte tutte le varie forme del fenomeno, o qualora siano note, le modalità con cui procedere al raggruppamento e alla codifica.
2. *Domande con Risposte Bilanciate*: l'intervistato può scegliere solo due alternative di risposta come, ad esempio, accordo/disaccordo oppure sì/no. Questo tipo di domande di solito viene usato per verificare la presenza-assenza di un fenomeno; tali domande hanno spesso la funzione di domande filtro: sono, infatti, utili per individuare eventuali sottogruppi ai quali, successivamente, saranno sottoposte ulteriori domande più specifiche.

Le variabili ottenute dall'indagine sono sia quantitative che qualitative. Le prime sono caratterizzate da valori numerici e riguardano le risposte a domande che si riferiscono, ad esempio, all'età dell'intervistato o al numero di spostamenti nelle ultime 24 ore. Invece, le variabili qualitative assumono valori non numerici e si possono trovare come risposta nelle domande che riguardano, ad esempio, il genere, l'occupazione e il titolo di studio dell'intervistato.

3.2 Ipotesi di lavoro per le successive analisi

Si considerano le 4466 interviste senza tener conto della periodicità, e quindi della distanza di tempo di 4 mesi tra le tre interviste. Questo perché non è utile ai fini dell'analisi trattata in questa tesi.

Non si tiene conto delle interviste con variabile CAMPIONE (indica se l'intervistato fa parte del campione stratificato di progetto) pari a 2,3,4. Questi valori indicano rispettivamente che:

2) l'intervistato è domiciliato fuori dall'area di studio, pertanto non fa parte del campione;

3) l'intervistato è domiciliato fuori dall'area di studio e la variabile TERMINE è pari a 1, 2, 3 o 4, ma non fa parte del campione in quanto fuori quota;

4) L'intervistato è domiciliato nell'area di studio, ma non fa parte del campione in quanto la variabile TERMINE è pari a 5 o 6.

La variabile TERMINE sopracitata rappresenta un codice di terminazione dell'intervista che va da 1 a 6:

- 1) Tutte e 6 le sezioni sono state completate;
- 2) Solo la sezione A, B ed F sono state completate in quanto il soggetto non ha effettuato spostamenti nelle precedenti 24 ore;
- 3) Solo le sezioni A, B ed F sono state completate in quanto tutti gli spostamenti effettuati dal soggetto nelle precedenti 24 ore sono fra due comuni distanti più di 50 km;

- 4) Solo le sezioni A, B ed F sono state completate in quanto tutti gli spostamenti effettuati dal soggetto nelle precedenti 24 ore sono fra due punti esterni all'area di studio;
- 5) Intervista interrotta una volta iniziata in quanto l'intervistatore si è rifiutato di completarla o per mancanza di tempo;
- 6) Intervista interrotta una volta iniziata per problemi tecnici.

Tenendo conto di questa ipotesi, il numero di interviste eliminate dall'analisi sono in totale 22.

Alcune interviste sono ripetute, ovvero l'intervistato ha risposto più volte al sondaggio. Quindi queste sono da considerate una sola volta in quanto non è un'indagine PANEL. Dall'analisi dei dati si riscontra che tutte le interviste ripetute sono presenti nell'indagine CAWI e non CATI.

Considerando le tre ipotesi, il numero di intervistati utile per l'analisi si riduce: 1504 utenti nella prima, 1294 nella seconda e infine 1244 nella terza campagna di indagine, per un totale di 4042 intervistati, di cui l'84% con modalità telefonica e il restante 16% online [Tab. 5].

	CATI		CAWI		CATI+CAWI	
	n.	%	n.	%	n.	%
1 ^a Indagine	1195	79	309	21	1504	100
2 ^a Indagine	1098	85	196	15	1294	100
3 ^a Indagine	1089	88	155	12	1244	100
Tot. Indagini (1 ^a +2 ^a +3 ^a)	3382	84	660	16	4042	100

Tab. 5 - Numero e percentuale di individui utili all'analisi.

3.3 Caratteristiche socio-demografiche del campione

Da un'analisi delle caratteristiche socio-demografiche del campione [Tab. 6], si evince che al sondaggio hanno partecipato maggiormente le donne rispetto agli uomini, con una percentuale del 52 e del 48% rispettivamente. Dal confronto tra intervistati online e telefonicamente ci si aspettava di trovare le solite differenze che costituiscono il "*digital divide*" cioè, che gli utenti del web siano più giovani e più istruiti di quelli che hanno risposto a un'indagine telefonica. I risultati sono in linea con queste aspettative. Infatti, tenendo conto della stratificazione fornita dall'ISTAT per quanto concerne le classi d'età, solo il 5% di coloro i quali hanno più di 65 anni d'età ha risposto ad un'indagine CAWI contro il 95%, invece, che ha risposto ad un'indagine CATI. Inoltre, considerando il livello di istruzione, si può notare che ben oltre la metà degli intervistati ha una licenza/diploma di scuola media superiore o una laurea, con una percentuale rispettivamente del 51 e 22%. I laureati, rispetto a chi è in possesso di altri titoli di studio, sono coloro che hanno maggiormente risposto all'indagine online: essi sono il 29% contro, ad esempio, il 16 e 9% di chi ha rispettivamente una licenza/diploma di scuola media superiore e una licenza media superiore. In sostanza, al crescere del livello di istruzione, si nota un aumento della percentuale di coloro che hanno partecipato a un sondaggio web. Si può osservare anche che tra gli intervistati, l'80% dichiara di esser in possesso della patente di guida per auto e il 20% di avere un abbonamento al TPL. In quest'ultimo caso, nello specifico, circa i quattro quinti di chi ha dichiarato di avere la patente B di guida ha risposto al sondaggio CATI, invece, solo il 5% di chi ha dichiarato di non averla ha risposto al sondaggio CAWI. Per quanto concerne le professioni lavorative, in ordine decrescente, più della metà del totale degli

intervistati è costituito da pensionati (30%), impiegati/addetti al commercio (25%) e, infine, casalinghe (10%). Quest'ultima mansione, nello specifico, ha un valore marginale nell'indagine CAWI (è rappresentata con solo il 9% rispetto, ad esempio, a un 39% di imprenditori e liberi professionisti che hanno risposto ad un sondaggio online). Inoltre la maggior parte di pensionati, impiegati/addetti al commercio e casalinghe ha risposto al sondaggio con modalità CATI piuttosto che CAWI.

	CATI		CAWI		CATI+CAWI	
	N.	%	N.	%	N.	%
Genere						
M	1607	83	328	17	1935	48
F	1775	84	332	16	2107	52
Età						
18-34	733	81	168	19	901	22
35-64	1546	78	436	22	1982	49
65+	1103	95	56	5	1159	29
Dimora						
Torino	1897	80	486	20	2383	59
Fuori Torino	1485	90	174	10	1659	41
Professione						
Imprenditore/Libero Professionista	123	61	79	39	202	5
Funzionario/dirigente	24	36	43	64	67	2
Impiegato/addetto commercio	751	74	262	26	1013	25
Operaio	256	85	44	15	300	7
Insegnante	110	81	26	19	136	3
Rappresentante	41	98	1	2	42	1
Artigiano/commerciante/esercente	49	82	11	18	60	2
Studente	274	86	44	14	318	8
Casalinga	383	91	36	9	419	10
Pensionato	1152	95	60	5	1212	30
In attesa di primo impiego	72	92	6	8	78	2
Disoccupato	143	84	27	16	170	4
Altro	4	16	21	84	25	1
Titolo di studio						
Nessun titolo	7	78	2	0	9	0
Licenza elementare	314	99	3	1	317	8
Licenza media inferiore	709	92	66	9	775	19
Licenza/Diploma media superiore	1726	84	329	16	2055	51
Laurea o dottorato	626	71	260	29	886	22
Patente di guida per automobili						
Sì	2616	81	621	19	3237	80
No	766	95	39	5	805	20
Abbonamento al TPL						
Sì	608	75	209	25	817	20
No	2774	86	451	14	3225	80

Tab. 6 - Caratteristiche socio-demografiche del campione in funzione di alcune variabili e della tipologia di intervista.

4 STUDIO DELLA SIGNIFICATIVITÀ NELLE DIFFERENZE FRA RISPOSTE CATI E CAWI

4.1 Piano di analisi

Nel presente lavoro di tesi si è interessati a capire in che misura le risposte alle domande effettuate nell'indagine DEMONSTRATE sono influenzate dalla modalità di indagine (CATI - CAWI), oltre che da altri fattori socio-demografici del campione. La difficoltà è data dal fatto che a sua volta tali fattori socioeconomici hanno un effetto sulla modalità di indagine, come discusso nel capitolo di revisione dello stato dell'arte. Per capire se i risultati sono significativamente differenti, si usano il test t-Student per variabili indipendenti con due sole categorie e l'ANOVA test nel caso in cui le variabili indipendenti siano più di due. In questo capitolo, passo dopo passo, si analizzano i singoli passaggi che hanno permesso il raggiungimento delle valutazioni finali.

4.2 Variabile dipendente e relativa ricodifica

Al fine di capire se esistano differenze significative nelle risposte in funzione di una indagine CATI o CAWI, e quindi, al fine di valutare l'influenza della tipologia di sondaggio nelle risposte date dagli intervistati, si può considerare, in un primo momento, una qualsiasi variabile dipendente presente nell'indagine di partenza, per poi, successivamente, procedere con l'inferenza statistica. In questo lavoro di tesi, si è deciso di focalizzare l'attenzione sulla variabile dipendente "FREQ", che indica la frequenza di utilizzo in una settimana di determinati mezzi di trasporto. Tale variabile si trova nella parte iniziale del questionario utilizzato nell'indagine

DEMONSTRATE e precisamente nel modulo “B – Repertorio Degli Spostamenti” e la domanda posta all’intervistato è la seguente:

“Ci potrebbe indicare con quale frequenza settimanale ha utilizzato i seguenti mezzi di trasporto per i suoi spostamenti nell’ultimo mese? Risposte possibili: 1 = più di 3 volte alla settimana, 2 = da 1 a 3 volte alla settimana, 3 = meno di una volta alla settimana, 4 = mai.

- BICI - bicicletta privata;
- BIBS - bicicletta del sistema Bike sharing (es. TOBike);
- MOTO - moto o ciclomotore;
- AUCO - autoveicolo privato come conducente (compresi i veicoli aziendali, di cortesia, in comodato d’uso, a noleggio...);
- AUCS - autoveicolo del sistema Car sharing (enjoy, car2go, ioguido, ...);
- AUPA - autoveicolo come passeggero, escluso taxi;
- TAXI - taxi;
- AZSB- Autobus aziendale, scuolabus;
- BUS - bus urbano e suburbano o tram;
- METRO - metropolitana;
- EXTRA - autobus extraurbano;
- TRENO - treno.”

La variabile di risposta fornita dal sondaggio (1,2,3,4) è una variabile quantitativa ordinale, perché i dati si riferiscono a valori non numerici che seguono un certo ordine di valutazione: all’aumentare del numero diminuisce la frequenza di utilizzo, da parte dell’intervistato, del mezzo di trasporto preso in questione. In

alcuni test di significatività parametrici e non, come ad esempio il test t-Student, la variabile ordinale non è sufficiente né a garantire una buona riuscita del test, né a fornire delle informazioni esaustive. Una buona risoluzione di questo tipo di problema può esser affrontata con la ricodifica della variabile quantitativa ordinale in variabile metrica [Tab. 7]. Tale operazione permette la trasformazione della variabile ordinale in modo che le stesse informazioni vengano espresse attraverso una variabile di entità diversa, ovvero metrica, rispetto a quella originale. La ricodifica utilizzata per il seguente caso che considera la variabile *FREQ*, prende in considerazione un valore medio della variabile ordinale: è un'approssimazione grossolana, ma, comunque, i dati metrici forniscono nelle analisi informazioni maggiori rispetto a quelli ordinali.

CODIFICA Variabile Ordinale	RICODIFICA Variabile Metrica
1 = Più di tre volte a settimana	5 = cinque volte a settimana
2 = Da una a tre volte a settimana	2 = due volte a settimana
3 = Meno di una volta a settimana	0,5 = "0,5" volte a settimana
4 = Mai	0 = zero volte a settimana

*Tab. 7 - Ricodifica della variabile *FREQ* da variabile ordinale a variabile metrica*

4.3 Statistiche descrittive: tabelle di contingenza

Per analizzare e rappresentare la relazione tra la variabile indipendente, che esprime la modalità di intervista "TIPO_INTERV", e la variabile dipendente, che fornisce informazioni riguardo la frequenza di utilizzo dei mezzi di trasporto "FREQ", si è analizzata la Tabella di Contingenza [Tab. 8]. In tal modo, si può avere un quadro più chiaro su come gli intervistati hanno risposto, in correlazione alla

tipologia di sondaggio, alla domanda sulla frequenza di utilizzo dei mezzi di trasporto. In sostanza, la tabella di contingenza permette un'analisi descrittiva delle frequenze¹⁰ congiunte delle variabili. Nel software SPSS l'ottenimento della tabella di contingenza è fornito considerando dal menù dell'editor i seguenti passaggi: *Analizza > Statistiche Descrittive > Tabelle di Contingenza*. Verrà visualizzata una finestra in cui è possibile decidere se inserire la variabile dipendente/indipendente nelle righe o nelle colonne della tabella [Fig. 16].



Fig. 16 - Finestra Tabelle di contingenza in SPSS

Da un'attenta lettura e analisi della Tab. 8, si evince che:

- Da un lato, più dell'80% degli intervistati ha dichiarato di non aver utilizzato nemmeno una volta a settimana la bici personale e del servizio di

¹⁰ In statistica, in particolare in statistica descrittiva, dato un carattere oggetto di rilevazione, per frequenza si intende il numero delle unità statistiche su cui una sua modalità (le modalità sono i valori numerici o gli attributi che un carattere può assumere) si presenta.

bike sharing, la moto, l'auto come passeggero e del servizio car sharing, il taxi, l'autobus aziendale, l'autobus extraurbano e il treno; mentre, il 30% non ha mai fatto uso dell'auto privata, il 71,2% dell'autoveicolo come passeggero, il 41,4% del bus e, infine, il 61,3% della metro nell'arco della settimana presa in questione. Dall'altro lato, si può notare che la percentuale di chi ha utilizzato da uno a più di 3 volte (sommatoria dei valori delle variabile metriche 2,00 e 5,00) l'auto privata e il bus è rispettivamente pari al 60,5 e 39,8% del totale degli intervistati. Invece, nelle restanti modalità di trasporto la sommatoria delle variabile metriche 2,00 e 5,00 non supera il 21%. Tali dati riflettono un campione che preferisce l'uso dell'auto privata a scapito del trasporto pubblico e quindi anche dell'ambiente;

- Il numero di osservazioni che si ritrovano nelle diverse categorie delle variabili `FREQ_BUS` e `FREQ_AUCO` non presenta una marcata variabilità, a differenza di quanto avviene considerando altre variabili dipendenti: non vi è una notevole eccedenza di una singola risposta rispetto alle altre, come si può riscontrare invece nei restanti mezzi di trasporto. Il 41,4% del totale degli intervistati ha dichiarato di non utilizzare mai il bus o il tram, il 18,7% meno di una volta a settimana, il 22,7% da una a tre volte a settimana e il 17,1% più di tre volte a settimana. Invece, il 30% del totale degli intervistati ha dichiarato di non utilizzare mai l'auto privata, il 9,5% meno di una volta a settimana, il 18% da una a tre volte a settimana e il 42,5% più di tre volte a settimana;

- Gli intervistati all'indagine con modalità web hanno dichiarato di utilizzare maggiormente il servizio di trasporto pubblico locale (bus, tram, metro, treno, autobus aziendale e extraurbano) rispetto a coloro che ha risposto al sondaggio telefonico. Questo può essere giustificato dal fatto che si presume che chi utilizza internet abbia uno stile di vita più attivo anche in termini di mobilità.

Tavola di contingenza

		FREQ_BICI				
		,00	,50	2,00	5,00	Totale
TIPO_INTERV	Conteggio in CATI	2873	189	203	117	3382
	% in TIPO_INTERV	84,9%	5,6%	6,0%	3,5%	100,0%
	Conteggio in CAWI	383	140	66	71	660
	% in TIPO_INTERV	58,0%	21,2%	10,0%	10,8%	100,0%
Totale	Conteggio	3256	329	269	188	4042
	% in TIPO_INTERV	80,6%	8,1%	6,7%	4,7%	100,0%
		FREQ_BIBS				
		,00	,50	2,00	5,00	Totale
TIPO_INTERV	Conteggio in CATI	3364	11	3	4	3382
	% in TIPO_INTERV	99,5%	0,3%	0,1%	0,1%	100,0%
	Conteggio in CAWI	532	64	42	22	660
	% in TIPO_INTERV	80,6%	9,7%	6,4%	3,3%	100,0%
Totale	Conteggio	3896	75	45	26	4042
	% in TIPO_INTERV	96,4%	1,9%	1,1%	0,6%	100,0%
		FREQ_MOTO				
		,00	,50	2,00	5,00	Totale
TIPO_INTERV	Conteggio in CATI	3216	29	71	66	3382
	% in TIPO_INTERV	95,1%	0,9%	2,1%	2,0%	100,0%
	Conteggio in CAWI	529	52	42	37	660
	% in TIPO_INTERV	80,2%	7,9%	6,4%	5,6%	100,0%
Totale	Conteggio	3745	81	113	103	4042
	% in TIPO_INTERV	92,7%	2,0%	2,8%	2,5%	100,0%
		FREQ_AUCO				
		,00	,50	2,00	5,00	Totale

TIPO_INTERV	Conteggio in CATI	1083	314	609	1376	3382
	% in TIPO_INTERV	32,0%	9,3%	18,0%	40,7%	100,0%
	Conteggio in CAWI	129	70	120	341	660
	% in TIPO_INTERV	19,5%	10,6%	18,2%	51,7%	100,0%
Totale	Conteggio	1212	384	729	1717	4042
	% in TIPO_INTERV	30,0%	9,5%	18,0%	42,5%	100,0%
		FREQ_AUCS				
		,00	,50	2,00	5,00	Totale
TIPO_INTERV	Conteggio in CATI	3343	16	13	10	3382
	% in TIPO_INTERV	98,8%	0,5%	0,4%	0,3%	100,0%
	Conteggio in CAWI	492	103	50	15	660
	% in TIPO_INTERV	74,5%	15,6%	7,6%	2,3%	100,0%
Totale	Conteggio	3835	119	63	25	4042
	% in TIPO_INTERV	94,9%	2,9%	1,6%	0,6%	100,0%
		FREQ_AUPA				
		,00	,50	2,00	5,00	Totale
TIPO_INTERV	Conteggio in CATI	2556	355	418	53	3382
	% in TIPO_INTERV	75,6%	10,5%	12,4%	1,6%	100,0%
	Conteggio in CAWI	321	137	132	70	660
	% in TIPO_INTERV	48,6%	20,8%	20,0%	10,6%	100,0%
Totale	Conteggio	2877	492	550	123	4042
	% in TIPO_INTERV	71,2%	12,2%	13,6%	3,0%	100,0%
		FREQ_TAXI				
		,00	,50	2,00	5,00	Totale
TIPO_INTERV	Conteggio in CATI	3296	62	18	6	3382
	% in TIPO_INTERV	97,5%	1,8%	0,5%	0,2%	100,0%
	Conteggio in CAWI	479	147	22	12	660
	% in TIPO_INTERV	72,6%	22,3%	3,3%	1,8%	100,0%
Totale	Conteggio	3775	209	40	18	4042
	% in TIPO_INTERV	93,4%	5,2%	1,0%	0,4%	100,0%
		FREQ_AZSB				
		,00	,50	2,00	5,00	Totale
TIPO_INTERV	Conteggio in CATI	3332	21	17	12	3382
	% in TIPO_INTERV	98,5%	0,6%	0,5%	0,4%	100,0%
	Conteggio in CAWI	554	49	32	25	660
	% in TIPO_INTERV	83,9%	7,4%	4,8%	3,8%	100,0%
Totale	Conteggio	3886	70	49	37	4042

% in TIPO_INTERV		96,1%	1,7%	1,2%	0,9%	100,0%
		FREQ_BUS				
		,00	,50	2,00	5,00	Totale
TIPO_INTERV	Conteggio in CATI	1504	578	798	502	3382
	% in TIPO_INTERV	44,5%	17,1%	23,6%	14,8%	100,0%
	Conteggio in CAWI	171	179	121	189	660
	% in TIPO_INTERV	25,9%	27,1%	18,3%	28,6%	100,0%
Totale	Conteggio	1675	757	919	691	4042
	% in TIPO_INTERV	41,4%	18,7%	22,7%	17,1%	100,0%
		FREQ_METRO				
		,00	,50	2,00	5,00	Totale
TIPO_INTERV	Conteggio in CATI	2280	514	384	204	3382
	% in TIPO_INTERV	67,4%	15,2%	11,4%	6,0%	100,0%
	Conteggio in CAWI	197	208	132	123	660
	% in TIPO_INTERV	29,8%	31,5%	20,0%	18,6%	100,0%
Totale	Conteggio	2477	722	516	327	4042
	% in TIPO_INTERV	61,3%	17,9%	12,8%	8,1%	100,0%
		FREQ_EXTRA				
		,00	,50	2,00	5,00	Totale
TIPO_INTERV	Conteggio in CATI	3265	57	37	23	3382
	% in TIPO_INTERV	96,5%	1,7%	1,1%	0,7%	100,0%
	Conteggio in CAWI	422	137	65	36	660
	% in TIPO_INTERV	63,9%	20,8%	9,8%	5,5%	100,0%
Totale	Conteggio	3687	194	102	59	4042
	% in TIPO_INTERV	91,2%	4,8%	2,5%	1,5%	100,0%
		FREQ_TRENO				
		,00	,50	2,00	5,00	Totale
TIPO_INTERV	Conteggio in CATI	3261	74	35	12	3382
	% in TIPO_INTERV	96,4%	2,2%	1,0%	0,4%	100,0%
	Conteggio in CAWI	334	229	62	35	660
	% in TIPO_INTERV	50,6%	34,7%	9,4%	5,3%	100,0%
Totale	Conteggio	3595	303	97	47	4042
	% in TIPO_INTERV	88,9%	7,5%	2,4%	1,2%	100,0%

Tab. 8 - Tabella di contingenza delle frequenze per tipologia di sondaggio.

4.4 Test parametrico o test non parametrico: controllo della normalità nella distribuzione dei dati

Prima di scegliere il tipo di test da applicare, è necessario valutare se la variabile dipendente segue o no una distribuzione di probabilità continua, e, quindi, di verificare se i valori tendono a concentrarsi intorno a un singolo valor medio. Se così fosse, non si hanno dubbi o problemi nell'utilizzo di un test di tipo parametrico. Affinché si abbia una distribuzione normale, il grafico di funzione di densità di probabilità associata deve essere simmetrico e deve assumere una forma a campana, conosciuta anche come campana di Gauss. La distribuzione normale è funzione di due parametri associati ai valori della variabile dipendente: la media μ e la varianza σ^2 . Esistono metodi diversi per testare questa ipotesi, sia di natura numerica che di natura grafica. Il metodo utilizzato in questa analisi è rappresentato dal test di Shapiro-Wilk per la normalità. È un metodo numerico e non grafico ed è stato eseguito utilizzando la procedura Analizza > Statistiche Descrittive > Esplora nel software SPSS Statistics. Analizzando tutte e dodici le variabili dipendenti rappresentate dalla frequenza di utilizzo di un determinato mezzo di trasporto [Tab. 9], si è riscontrato che la condizione di normalità, per un livello di significatività α pari al 5%, è stata violata, in quanto il valore di significatività "Sign." è inferiore a 0,05, ovvero, il test risulta significativo ($P\text{-Value} < \alpha$). Quindi i valori non seguono una distribuzione normale di Gauss. Nonostante ciò, i test parametrici (come il test t dei campioni indipendenti) sono considerati "robusti" per la violazione dell'ipotesi di normalità. Ciò significa che il test fornirà comunque risultati validi. Pertanto, si può affermare che i test parametrici richiedono solo approssimativamente dati con una distribuzione normale. Nel

presente lavoro di analisi, si è deciso di utilizzare, nonostante la condizione di normalità non verificata, l'utilizzo prima di un test parametrico e successivamente, come verifica, l'uso di un test non parametrico. Se i risultati tra i due test sono concordi, possiamo ritenere valido il risultato del test parametrico in quanto risulta esser abbastanza robusto per quel campione.

	ID INTERV	Shapiro-Wilk		
		Statistica	gl	Sign.
FREQ_BICI	CATI	,353	3382	,000
	CAWI	,573	660	,000
FREQ_BIBS	CATI	,025	3382	,000
	CAWI	,383	660	,000
FREQ_MOTO	CATI	,189	3382	,000
	CAWI	,410	660	,000
FREQ_AUCO	CATI	,749	3382	,000
	CAWI	,738	660	,000
FREQ_AUCS	CATI	,056	3382	,000
	CAWI	,428	660	,000
FREQ_AUPA	CATI	,480	3382	,000
	CAWI	,668	660	,000
FREQ_TAXI	CATI	,084	3382	,000
	CAWI	,369	660	,000
FREQ_AZSB	CATI	,067	3382	,000
	CAWI	,347	660	,000
FREQ_BUS	CATI	,714	3382	,000
	CAWI	,750	660	,000
FREQ_METRO	CATI	,522	3382	,000
	CAWI	,725	660	,000
FREQ_EXTRA	CATI	,121	3382	,000
	CAWI	,515	660	,000
FREQ_TRENO	CATI	,116	3382	,000
	CAWI	,540	660	,000

Tab. 9 - Test di Shapiro-Wilk per la normalità.

4.5 Significatività nella differenza delle medie

Nella Tab. 10 sono riportati i risultati, ottenuti dall'applicazione sui dati presi in analisi, del test t-Student per il confronto tra le medie di due campioni indipendenti. L'ipotesi nulla H_0 , verificata con un test a due code, prevede che il valore medio della frequenza di utilizzo di un tipo di mezzo di trasporto "k" in CATI sia uguale alla frequenza di utilizzo del medesimo mezzo di trasporto "k" in CAWI:

$$\mu_{k,CATI} = \mu_{k,CAWI}$$

In tal modo, si va a verificare, analizzando la significatività del test, se la differenza delle medie della frequenza di utilizzo di un determinato mezzo di trasporto dipende da fattori casuali oppure dal fatto che gli intervistati abbiano risposto con un sondaggio telefonico piuttosto che web.

		Test di Levene		Test t per l'uguaglianza delle medie						
		F	Sign.	t	gl	Sign. (a due code)	Differenza della media	Differenza errore standard	Intervallo di confidenza della differenza di 95%	
									Inferiore	Superiore
FREQ_BICI	Varianze uguali presunte	234,734	,000	-11,003	4040	,000	-,52298	,04753	-,61616	-,42979
	Varianze uguali non presunte			-8,281	769,772	,000	-,52298	,06315	-,64695	-,39900
FREQ_BIBS	Varianze uguali presunte	1116,418	,000	-17,918	4040	,000	-,333110	,018591	-,369558	-,296662
	Varianze uguali non presunte			-8,545	667,762	,000	-,333110	,038983	-,409654	-,256566
FREQ_MOTO	Varianze uguali presunte	194,231	,000	-8,488	4040	,000	-,30312	,03571	-,37314	-,23310
	Varianze uguali non presunte			-6,183	758,032	,000	-,30312	,04902	-,39936	-,20688
FREQ_AUCO	Varianze uguali presunte	,234	,628	-5,928	4040	,000	-,55914	,09432	-,74406	-,37422
	Varianze uguali non presunte			-6,046	951,929	,000	-,55914	,09248	-,74063	-,37764
FREQ_AUCS	Varianze uguali presunte	738,908	,000	-16,535	4040	,000	-,31834	,01925	-,35609	-,28060
	Varianze uguali non presunte			-9,080	688,388	,000	-,31834	,03506	-,38718	-,24951
FREQ_AUPA	Varianze uguali presunte	455,303	,000	-15,105	4040	,000	-,65606	,04343	-,74122	-,57090
	Varianze uguali non presunte			-10,477	742,167	,000	-,65606	,06262	-,77899	-,53313
FREQ_TAXI	Varianze uguali presunte	498,572	,000	-14,509	4040	,000	-,24026	,01656	-,27272	-,20779
	Varianze uguali non presunte			-8,066	690,506	,000	-,24026	,02979	-,29874	-,18177
FREQ_AZSB	Varianze uguali presunte	603,826	,000	-13,384	4040	,000	-,29259	,02186	-,33545	-,24973
	Varianze uguali non presunte			-7,240	685,860	,000	-,29259	,04041	-,37193	-,21324
FREQ_BUS	Varianze uguali presunte	69,382	,000	-8,326	4040	,000	-,63456	,07622	-,78399	-,48514
	Varianze uguali non presunte			-7,431	851,693	,000	-,63456	,08540	-,80218	-,46695
FREQ_METRO	Varianze uguali presunte	261,393	,000	-15,041	4040	,000	-,88472	,05882	-1,00004	-,76940
	Varianze uguali non presunte			-11,907	790,143	,000	-,88472	,07430	-1,03057	-,73887
FREQ_EXTRA	Varianze uguali presunte	722,748	,000	-18,443	4040	,000	-,50917	,02761	-,56330	-,45505
	Varianze uguali non presunte			-10,594	696,521	,000	-,50917	,04806	-,60354	-,41481
FREQ_TRENO	Varianze uguali presunte	878,975	,000	-23,278	4040	,000	-,57714	,02479	-,62574	-,52853
	Varianze uguali non presunte			-12,426	683,686	,000	-,57714	,04645	-,66833	-,48594

Tab. 10 - Test di Levene per l'uguaglianza delle varianze e Test t per l'uguaglianza delle medie.

Il Test t-Student per il confronto tra due campioni indipendenti e l'analogo Test non parametrico di Welch (quest'ultimo rappresentato in Tab. 10 nella seconda riga "Varianze uguali non presunte"), hanno evidenziato delle situazioni degne di nota. Prima di tutto, vi è una risposta comune per tutte e dodici le frequenze analizzate ad eccezione della `FREQ_AUCO` che presenta, invece, delle differenze di studio in quanto l'uguaglianza delle varianze, solo in questo caso, è rispettata. Tale risultato, presume che gli individui che hanno risposto ad un'indagine web e che utilizzano l'auto privata hanno la stessa variabilità¹¹ di utilizzo dell'auto privata di chi, invece, risponde con indagine telefonica. Nei restanti casi, invece, in cui non è rispettata l'omoschedasticità, non risulta esserci la stessa variabilità di utilizzo del medesimo mezzo di trasporto tra i due gruppi di intervistati con diverse modalità di indagine. Questo può esser dovuto al fatto che l'auto privata, essendo un mezzo di trasporto più comune e più diffuso rispetto agli altri analizzati, può esser indistintamente utilizzato, in termini di frequenza, sia da chi risponde a un sondaggio telefonico sia da chi compila un sondaggio web. Inoltre, come rilevato dall'edizione 2017 dell'Annuario statistico italiano dell'ISTAT¹², l'auto risulta il mezzo di trasporto più utilizzato per andare a lavoro, a studiare, per viaggi.

Per tutte le frequenze di utilizzo dei diversi mezzi di trasporto, ad eccezione dell'auto privata, si è considerato il Test non parametrico di Welch, in quanto la distribuzione dei dati è eteroschedastica, quindi il valore P-Value fornito dal test di Levene è minore del valore di significatività scelto pari a 0,05. Ciò può influire sul

¹¹ La variabilità esprime l'attitudine di un carattere a presentarsi con modalità differenti nelle diverse unità statistiche di un collettivo.

¹² <https://www.istat.it/it/files//2017/12/Asi-2017.pdf>

tasso di errore di I tipo: il risultato del test statistico può essere tale da rifiutare l'ipotesi nulla H_0 , quando essa in realtà è vera. La probabilità di compiere tale tipo di errore è, comunque, molto bassa, perché si è scelto un livello di significatività α pari al 5%.

In secondo luogo, il valore medio delle dodici frequenze ricavate in CATI è significativamente diversa dalle corrispondenti medie di frequenza ottenute con l'indagine CAWI. Infatti, il P-Value ottenuto è sempre pari a ,000¹³. Quindi, si è rifiutata l'ipotesi nulla e si è affermato che le medie sono statisticamente differenti e la loro differenza non dipende da fattori casuali.

Nel caso della $FREQ_AUCO$, essendo le varianze presunte uguali e il test di normalità non rispettato, si è analizzata la differenza delle medie considerando la robustezza del test t-Student standard. Per un'ulteriore verifica sulla validità del risultato si può considerare il corrispondente test non parametrico di U-Mann Whitney [Fig. 17]. In tal caso, l'ipotesi nulla H_0 , non prende in analisi la media, bensì la mediana, perché essa risulta essere un indicatore più robusto (come specificato precedentemente nel paragrafo 2.2). Tale ipotesi prevede che la mediana dei dati della frequenza AUCO in CATI sia uguale alla mediana dei dati della frequenza di utilizzo del medesimo mezzo in CAWI:

$$Me_{auco,CATI} = Me_{auco,CAWI}$$

È stato ottenuto un risultato concorde con quello conseguito con il test t-Student per il medesimo caso e per lo stesso livello di significatività $\alpha = ,05$. Per questo motivo, si è deciso di rifiutare l'ipotesi nulla e di affermare che la differenza tra le

¹³ Non significa che il livello di significatività calcolato sia effettivamente zero, ma che P-Value < 0,0005.

mediane dei due gruppi è statisticamente significativa per le frequenze di utilizzo di tutti i modi di trasporto considerati.

	Ipotesi nulla	Test	Sign.	Decisione
1	La distribuzione di FREQ_METRO è la stessa tra le categorie di TIPO_INTERV.	Test U di Mann-Whitney a campioni indipendenti	,000	Rifiutare l'ipotesi nulla.

Fig. 17 - Test non parametrico U di Mann-Whitney a campioni indipendenti per la variabile dipendente FREQ_AUCO.

4.6 Il ruolo del protocollo di indagine nella significatività delle differenze

Dopo essersi accertati nel precedente paragrafo della significatività nella differenza delle medie, con l'ANOVA test si controllerà ora se tali differenze siano date dal diverso protocollo, ossia dal fatto di aver risposto ad un'indagine CATI, piuttosto che CAWI. Le variabili dipendenti sono sempre la frequenza dei mezzi di trasporto in una settimana codificata in scala metrica, invece, la variabile indipendente è il protocollo di indagine, che è una variabile categorica binaria. L'ipotesi nulla H_0 , verificata con un test a due code, prevede che il valore medio della frequenza di utilizzo di un tipo di mezzo di trasporto "k" in CATI sia uguale alla frequenza di utilizzo del medesimo mezzo di trasporto "k" in CAWI:

$$\mu_{k,CATI} = \mu_{k,CAWI}$$

Come si è visto precedentemente in Tab. 10, il test di Levene è risultato significativo per tutti e dodici i casi di frequenza, ad eccezione del caso in cui si dichiara di utilizzare l'auto privata. Nei casi in cui l'omogeneità della varianza non

è rispettata, si è utilizzato una versione modificata dell'ANOVA test definita da Welch. Invece, nel caso in cui l'omogeneità della varianza non è stata violata, si è utilizzato semplicemente il test ANOVA a una via. Il risultato fornito da SPSS per l'ANOVA di Welch si trova nella tabella "Test robusti dell'uguaglianza delle medie" come mostrato in Tab. 11, invece, il risultato dell'ANOVA a una via standard nella tabella "ANOVA" [Tab. 12]. Inoltre, per entrambi i casi di studio, è stato ritenuto opportuno riportare la Tab. 13 riguardante la frequenza media di un determinato mezzo di trasporto in funzione della tipologia di intervista.

	Statistica	gl1	gl2	Sign.
FREQ_BICI	68,579	1	769,772	,000
FREQ_BIBS	73,017	1	667,762	,000
FREQ_MOTO	38,230	1	758,032	,000
FREQ_AUCS	82,454	1	688,388	,000
FREQ_AUPA	109,767	1	742,167	,000
FREQ_TAXI	65,053	1	695,506	,000
FREQ_AZSB	52,423	1	685,860	,000
FREQ_BUS	55,215	1	851,693	,000
FREQ_EXTRA	141.780	1	790,143	,000
FREQ_TRENO	112.234	1	696,521	,000
FREQ_METRO	141.780	1	790.143	,000

Tab. 11 - Test robusti dell'uguaglianza delle medie - ANOVA di Welch

FREQ_AUCO					
	Somma dei quadrati	gl	Media quadratica	F	Sign.
Tra gruppi	172,646	1	172,646	35,142	,000
Entro i gruppi	19847,673	4040	4,913		
Totale	20020,319	4041			

Tab. 12 - ANOVA a una via

	TIPO_INTERV	
	CATI	CAWI
	Media	Media
FREQ_BICI	,3210	,8439
FREQ_BIBS	,0093	,3424
FREQ_MOTO	,1438	,4470
FREQ_AUCS	,0248	,3432
FREQ_AUPA	,3780	1,0341
FREQ_TAXI	,0287	,2699
FREQ_AZSB	,0309	,3235
FREQ_BUS	1,2995	1,9341
FREQ_EXTRA	,0643	,5735
FREQ_TRENO	,0494	,6265
FREQ_METRO	,6047	1,4894
FREQ_AUCO	2,4409	3,0000

Tab. 13 - Tabella descrittiva relativa ai valori medi di frequenza in funzione della variabile TIPO_INTERV.

Sia negli undici casi dell'ANOVA di Welch, sia nel caso dell'ANOVA a una via standard utilizzata nel caso della FREQ_AUCO, si è riscontrato che il valore "Sign.", ovvero il valore del P-Value, risulta esser pari a ,000 e quindi inferiore a ,0005. In tutti i casi, il test risulta essere fortemente significativo. Per questo motivo, si è rifiutata l'ipotesi nulla H_0 e si è affermato che la differenza delle medie è dovuta alla categoria della variabile indipendente considerata, ovvero il fatto di aver risposto ad un'indagine web oppure ad un'indagine telefonica.

Nel caso della FREQ_AUCO, essendo le varianze presunte uguali e il test di normalità non rispettato, si è analizzata la differenza delle medie, considerando la robustezza del test ANOVA a una via standard. Per un'ulteriore verifica sulla

validità del risultato, è stato successivamente considerato il corrispettivo test non parametrico H di Kruskal-Wallis [Fig. 18]. L'ipotesi nulla H_0 , prevede che la mediana dei dati della frequenza AUCO in CATI sia uguale alla mediana dei dati della frequenza di utilizzo del medesimo mezzo in CAWI:

$$Me_{\text{auco,CATI}} = Me_{\text{auco,CAWI}}$$

Il risultato ottenuto è stato concorde con il risultato ottenuto dal test ANOVA a una via per il medesimo caso e per lo stesso livello di significatività $\alpha = ,05$: si è rifiutata l'ipotesi nulla e si è affermato che la differenza tra le due mediane è statisticamente significativa. Quindi, la differenza delle mediane dei dati riguardanti la frequenza di utilizzo dell'auto privata è dovuta, anch'essa, al diverso protocollo di indagine.

Riepilogo test sull'ipotesi

	Ipotesi nulla	Test	Sign.	Decisione
1	La distribuzione di <code>FREQ_AUCO</code> è la stessa tra le categorie di <code>TIPO_INTERV.</code>	Test di Kruskal-Wallis a campioni indipendenti	,000	Rifiutare l'ipotesi nulla.

Vengono visualizzate le significatività asintotiche. Il livello di significatività è ,05.

Fig. 18 - Test non parametrico H di Kruskal-Wallis a campioni indipendenti per la variabile dipendente `FREQ_AUCO`.

Dai risultati del test ANOVA a una via e degli analoghi test non parametrici, si evince che la variabile `TIPO_INTERV`, contenente le due modalità di intervista (CATI-CAWI), risulta essere una variabile indicatore delle differenze tra le due medie. Ovvero, il diverso protocollo di indagine influenza la risposta data per quanto concerne la frequenza di utilizzo di un determinato mezzo di trasporto. Si potrebbe pensare che, coloro i quali hanno risposto ad un'indagine web o telefonica, tendono ad essere due gruppi di persone distinte che utilizzano con

frequenze diverse un determinato mezzo di trasporto. Infatti, dalla Tab. 13, si evince che le frequenze medie di utilizzo di tutti i mezzi di trasporto considerati in funzione della variabile TIPO_INTERV, risultano esser maggiori nel caso di risposta cawi rispetto al caso di risposta cati. Ciò può indicare che gli intervistati che hanno effettuato il sondaggio con modalità web abbiano uno stile di vita più attivo e meno sedentario, in quanto essi, rispetto a coloro che hanno risposto all'indagine con modalità telefonica, dichiarano di utilizzare maggiormente i diversi mezzi di trasporto per i loro spostamenti.

Inoltre, un altro fattore che potrebbe influenzare la risposta in funzione della diversa modalità di somministrazione dell'indagine, come precedentemente visto nel capitolo 1 par. 2, può esser la desiderabilità sociale. Prendendo in esame la frequenza dell'auto, come si può notare dalla tabella di contingenza precedentemente analizzata [Tab. 8] e dalla tabella descrittive [Tab. 13], gli intervistati web dichiarano di utilizzare maggiormente l'auto rispetto agli intervistati telefonici. Tale differenza può presumere che coloro i quali hanno risposto con modalità CATI per quanto concerne l'utilizzo dell'auto, per dare un'immagine positiva di sé all'intervistatore e per dimostrare di avere uno stile di vita più attivo, abbiano risposto di utilizzare con minore frequenza l'auto rispetto a chi ha risposto con modalità CAWI. In definitiva, risulta di interesse non limitarsi ad accertare il fatto che la differenza fra i due gruppi è associata alla modalità di somministrazione dell'intervista, ma considerare il ruolo di quest'ultima variabile indipendente insieme alle caratteristiche socioeconomiche dell'intervistato. Questa ulteriore analisi è l'oggetto dei prossimi due capitoli.

5 IL PROTOCOLLO DI INDAGINE COME “PROXY” PER SPIEGARE LE SCELTE DI MOBILITÀ

5.1 L'importanza delle variabili socio-economiche

Soffermarsi su un'analisi che valuta la differenza delle risposte sulla frequenza di utilizzo di un determinato mezzo di trasporto in funzione della sola modalità di indagine, risulta essere poco chiara e poco utile nell'ambito dello studio della mobilità del campione considerato. Coloro i quali hanno risposto ad un'indagine CATI piuttosto che CAWI rappresentano infatti due gruppi diversi all'interno dell'universo di riferimento e quindi utilizzano con frequenze diverse determinati mezzi di trasporto. Come si è visto nei precedenti test, la differenza della frequenza di utilizzo di determinati mezzi di trasporto in funzione delle due modalità di indagine, ovvero telefonica e web, è stata significativa in ogni caso analizzato. Quindi la variabile TIPO_INTERV con le due categorie CATI e CAWI risulta essere una variabile indice di questa differenza.

Allo stesso tempo, dal punto di vista concettuale, sarebbe un po' strano fare un modello in cui si predice la frequenza d'uso, ad esempio, dell'autobus, solo in funzione della tipologia di sondaggio somministrato all'intervistatore. Da ora in poi, l'analisi si concentra sullo studio del principio di CAUSA-EFFETTO, ovvero il collegamento tra due o più variabili, per cui al variare di alcune di esse cambia il valore della variabile indipendente. Per studiare questo principio, si è deciso di ricorrere all'analisi di un modello di regressione multivariata in cui bisogna risalire dalla variabile categorica TIPO_INTERV a delle variabili che effettivamente descrivono la causa del fatto per cui un individuo usa più o meno un determinato

mezzo di trasporto. Tali variabili sono definite come variabili socioeconomiche. Non bisogna, infatti, dimenticare che ogni risposta data – sia che essa sia rilevata con intervista telefonica, sia che con indagine web – è, comunque, influenzata dalla vita sociale ed economica del soggetto intervistato e non solo dalla modalità di somministrazione delle domande (come visto e spiegato nei precedenti paragrafi, la tipologia di sondaggio può influire sulle risposte). Le possibili variabili socioeconomiche che si possono estrapolare dal sondaggio DEMONSTRATE, sono, solo per citarne alcune, il genere, l'età, il grado di istruzione e il reddito dell'intervistato. Sono state incluse, inoltre, anche alcune variabili che non sono propriamente socio-economiche ma che comunque possono fornire un buon contributo di valutazione alla nostra analisi; ad esempio, si è deciso di tener conto del fatto che l'intervistato potrebbe avere un box auto oppure possedere un abbonamento ai servizi di trasporto pubblico o car sharing.

Tale ragionamento esprime che, effettivamente, si potrebbe pensare di raggruppare gli intervistati non solo in base alla tipologia di sondaggio che gli è stato somministrato, ma anche in funzione di qualsiasi altra variabile socio-economica. In questo modo di operare, si avrà una maggiore stratificazione del campione che ci permette di condurre ulteriori analisi più approfondite.

Inoltre, in questo capitolo, si terrà conto anche della possibilità di considerare la variabile TIPO_INTERV, costituita dalle due categorie CATI e CAWI, come variabile proxy, ovvero come variabile che può sostituire alcune variabili socio-economiche in dei modelli che permettono di predire la frequenza di utilizzo di determinati mezzi di trasporto. Questa analisi permette di chiarire in quali casi la non

conoscenza di determinate variabili socio-economiche può essere almeno parzialmente compensata dalla considerazione del protocollo di indagine in un ipotetico modello previsionale.

5.2 Elenco completo delle variabili potenzialmente sostituibili

Sono state considerate le 6 variabili di frequenza elencate nel precedente paragrafo e tutte le variabili socio-economiche, fornite dall'indagine DEMONSTRATE, che possono essere potenzialmente sostituite dalla variabile TIPO_INTERV. Nello specifico, si sono considerate le seguenti variabili:

- **FREQ**: variabile indipendente Y che indica con quale frequenza settimanale l'intervistato ha utilizzato specifici mezzi di trasporto per i suoi spostamenti nell'ultimo mese.

- **ETÀ**: variabile esplicativa X_1 di tipo metrica che indica l'età degli individui che hanno partecipato al sondaggio. Essa è una variabile socioeconomica con effetto sulla frequenza di utilizzo di un mezzo di trasporto di tipo non lineare. Valendo la condizione di linearità per il modello di regressione, è stato deciso di formare dei sottogruppi dividendo l'età in classi. Per definire tali classi, sono stati esaminati i grafici di dispersione categorizzati riguardanti l'età in relazione alle corrispettive frequenze di utilizzo dei sei più comuni mezzi di trasporto, ovvero bici, autoveicolo privato come conducente e come passeggero, bus, metro e treno. In sostanza, si è disposto sull'asse delle ascisse l'età metrica (ETÀ_M), invece, sull'asse delle ordinate, la frequenza metrica ricodificata come in Tab. 7, ottenendo, in questo modo, sei grafici di dispersione categorizzati [Fig. 19].

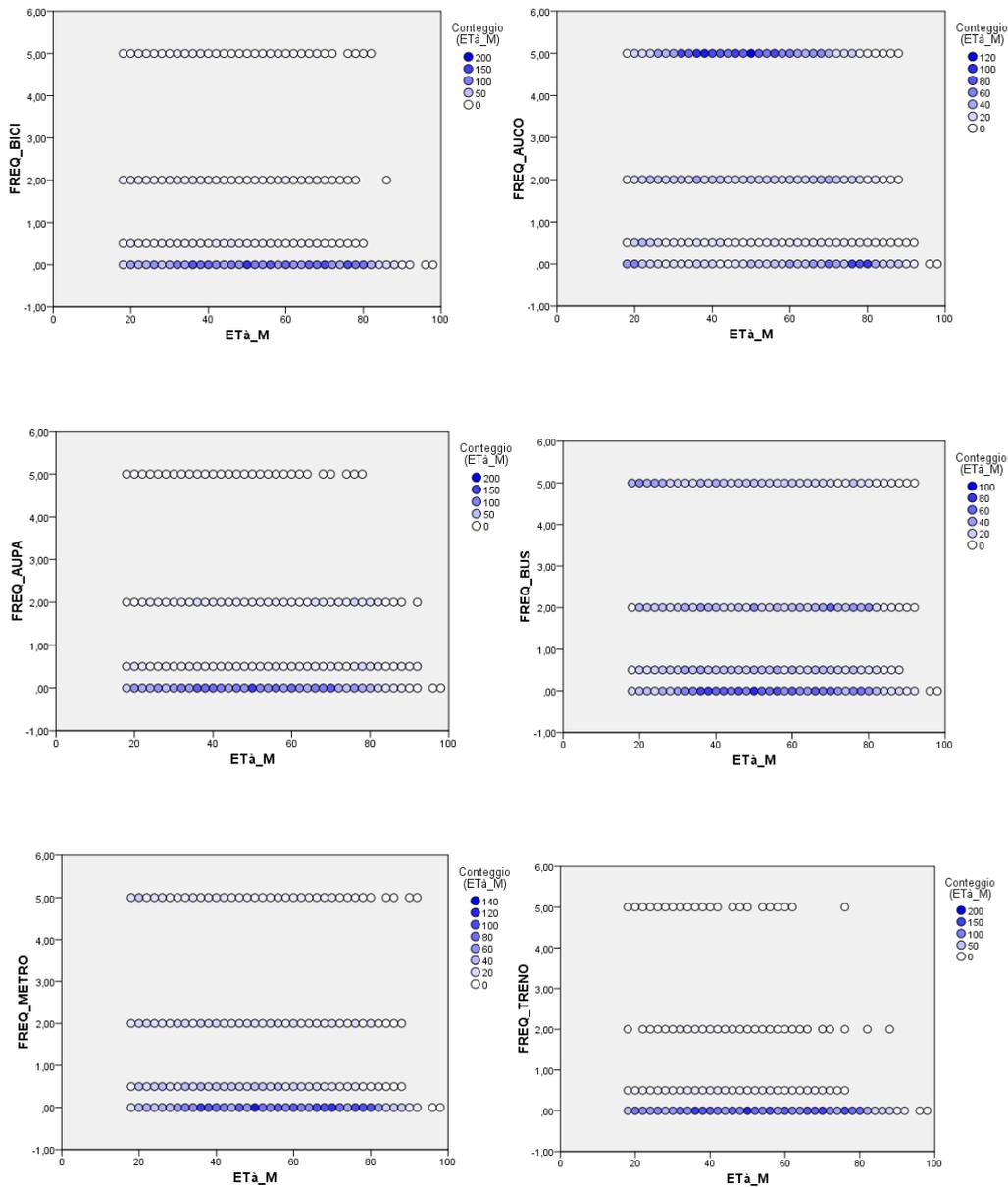


Fig. 19 - Grafici a dispersione categorizzati ETÀ_M – FREQ

Per ogni grafico si è notato che la concentrazione di punti non esprime un andamento lineare della relazione frequenza-età. Per questo motivo, sono state analizzate le zone del grafico che presentano il maggior numero di punti e poi, in funzione di questa densità, sono state determinate le classi d'età. È stato possibile fare ciò, solo nel caso di FREQ_AUCO e FREQ_BUS, come mostrato in Fig. 20.

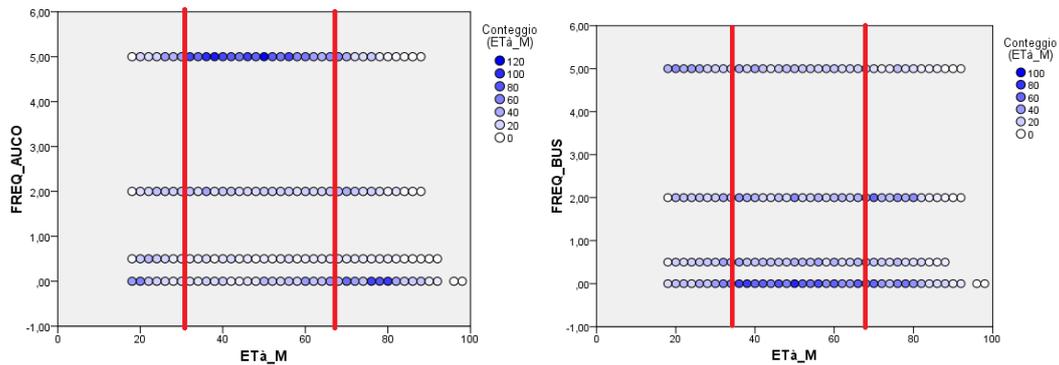


Fig. 20 - Determinazione delle classi d'età

Per questo motivo, considerando congiuntamente i due grafici precedenti, è possibile ottenere tre classi definite dai seguenti range di età [Tab. 14]:

	ETÀ MIN	ETÀ MAX
CLASSE GIOVANI	18	33
CLASSE ADULTI	34	67
CLASSE ANZIANI	68	+

Tab. 14 - Tabella di suddivisione dell'età in classi

Questa classificazione permette di avere per la regressione lineare una variabile categorica derivante da una variabile metrica in cui si evita di suddividere un gruppo di persone aventi età e comportamenti di mobilità approssimativamente paragonabili in diverse categorie. Il campione iniziale sarà quindi suddiviso nelle seguenti tre classi: 851 giovani, 2333 adulti e, infine, 958 anziani.

Nei restanti 4 casi di frequenza, ovvero FREQ_BICI, FREQ_AUPA, FREQ_METRO e, infine, FREQ_TRENO, si è deciso di adottare, per quanto concerne la suddivisione in classi d'età, la stessa classificazione definita precedentemente in Tab. 14 per FREQ_AUCO e FREQ_BUS.

Inoltre, non essendo l'effetto dell'età di tipo lineare si è deciso di utilizzare una codifica dummy, definita anche codifica fittizia. Per fare ciò, si sono considerate le tre classi di età precedentemente definite - giovani, adulti e anziani - e, successivamente, due variabili fittizie D_1 e D_2 definite come in Tab. 15.

	D_1 - CLASSE ADULTI	D_2 - CLASSE GIOVANI
CLASSE GIOVANI	0	1
CLASSE ADULTI	1	0
CLASSE ANZIANI	0	0

Tab. 15 - Codifica Dummy per le classi d'età

Come si può notare, si è selezionato un gruppo di riferimento, la CLASSE ANZIANI, nel quale entrambe le variabili assumono valore nullo mentre negli altri gruppi, in modo alternato, solo una delle due variabili assume il valore unitario. In questo modo si è ottenuto una funzione del modello lineare definita a tratti che cambia secondo la classe d'età e si è potuto studiare le differenze nella frequenza di utilizzo di un determinato mezzo di trasporto fra l'appartenere alla classe adulti piuttosto che a quella anziani oppure a quella giovani in riferimento a quella anziani, valutando la differenza fra l'intercetta nelle diverse forme funzionali.

- TITOLO DI STUDIO: variabile esplicative X_2 che indica il grado di istruzione dell'intervistato. Tale variabile è di tipo qualitativo con cinque fattori: nessun titolo, licenza elementare, licenza media inferiore, licenza o diploma media superiore e, infine, laurea o dottorato. Come se ne evince da Tab. 6 riguardante le caratteristiche socio-demografiche del campione (capitolo 3), gli intervistati per i

primi due livelli sono in numero abbastanza inferiore rispetto agli altri tre di grado superiore. Per tale motivo si è deciso di aggregare le prime tre risposte in un unico fattore denominato UNDER_LICENZAMEDIA. In sostanza tale gruppo comprende gli intervistati che non hanno nessun titolo di studio, chi ha la licenza elementare e chi ha la licenza media inferiore. Successivamente si è deciso di effettuare una codifica dummy. Per fare ciò, si sono considerate le tre classi di titolo di studio – under_licenzamedia, licenzamediasup e laurea – e, successivamente, due variabili fittizie D₁ e D₂ definite come in Tab. 16.

	D ₁ – CLASSE LICENZAMEDIASUP	D ₂ – CLASSE LAUREA
CLASSE UNDER_LICENZA MEDIA	0	0
CLASSE LICENZIAMEDIASUP	1	0
CLASSE LAUREA	0	1

Tab. 16 - Codifica Dummy per livello di istruzione

Come si può notare, si è selezionato un gruppo di riferimento, la CLASSE UNDER_LICENZAMEDIA, nel quale entrambe le variabili assumono valore nullo mentre negli altri gruppi, in modo alternato, solo una delle due variabili assume il valore unitario.

-GENERE: variabile esplicativa X₃ che indica il sesso di una persona. Tale variabile è di tipo qualitativo con due fattori: uomo e donna. Anche in questo caso, si è assunta una codifica dummy ponendo il valore 1 per gli uomini e il valore 0 per le donne.

- DIMORA: variabile esplicativa X₄ che indica il comune dove vive abitualmente l'intervistato. Tale variabile è di tipo qualitativo con due fattori: Torino e 1^a e 2^a

cintura. Anche in questo caso si è assunta una codifica dummy ponendo il valore 1 per coloro che vivono nel comune di Torino e il valore 0 nel restante caso.

- SALUTE: variabile esplicativa X_5 che indica se l'intervistato ha problemi di salute o disabilità. Tale variabile è di tipo qualitativo con due fattori: si e no.

- PATENTE: variabile esplicativa X_6 che indica se l'intervistato è in possesso della patente di tipo B. Tale variabile è di tipo qualitativo con due fattori: si e no.

- PARK_CASA: variabile esplicativa X_7 che indica se l'intervistato ha a disposizione un parcheggio presso la sua dimora. Tale variabile è di tipo qualitativo con due fattori: si e no.

- TP_ABBON: variabile esplicativa X_8 che indica se l'intervistato possiede un abbonamento al trasporto pubblico urbano e suburbano. Tale variabile è di tipo qualitativo con due fattori: si e no.

- AUCS_ABBON: variabile esplicativa X_9 che indica se l'intervistato possiede un abbonamento al servizio di car sharing. Tale variabile è di tipo qualitativo con due fattori: si e no.

- FAM_REDDITO: variabile esplicative X_{10} che indica la somma del reddito medio netto mensile dell'intervistato e di tutti i familiari che dimorano con esso. La variabile di risposta fornita dal sondaggio (1,2,3,...,14) è una variabile quantitativa ordinale, perché i dati si riferiscono a valori non numerici che seguono un certo ordine di valutazione: all'aumentare del numero aumenta il reddito. È stato deciso di effettuare la ricodifica della variabile quantitativa ordinale in variabile metrica [Tab. 17]. Tale operazione permette la trasformazione della variabile ordinale in

modo che le stesse informazioni vengano espresse attraverso una variabile di entità diversa, ovvero metrica, rispetto a quella originale. La ricodifica utilizzata per il seguente caso, prende in considerazione un valore medio della variabile ordinale diviso 1000 per evitare problemi di eteroschedasticità dovuti a fattori di scala diversi tra le variabili prese in esame:

CODIFICA Variabile Ordinale [al mese]	RICODIFICA Variabile Metrica [1000€ al mese]
1 = Fino a 400€	0,20
2 = Da 401€ fino a 600€	0,50
3 = Da 601€ fino a 800€	0,70
4 = Da 801€ fino a 1000€	0,90
5 = Da 1001€ fino a 1200€	1,10
6 = Da 1201€ fino a 1500€	1,35
7 = Da 1501€ fino a 1800€	1,65
8 = Da 1801€ fino a 2000€	1,90
9 = Da 2001€ fino a 2500€	2,25
10 = Da 2501€ fino a 3000€	2,75
11 = Da 3001€ fino a 4000€	3,50
12 = Da 4001€ fino a 6000€	5,00
13 = Da 6001€ a 10000€	8,00
14 = Più di 10001	15,00

Tab. 17 - Ricodifica della variabile FAM_REDDITO da variabile ordinale a variabile metrica

- FAM_AUTO: variabile esplicativa X_{11} di tipo metrica che indica le auto a disposizione dell'intervistato.

- FAM_N: variabile esplicativa X_{12} di tipo metrica che indica il numero di componenti del nucleo familiare dell'intervistato.
- FAM_FIGLI: variabile esplicativa X_{13} di tipo metrica che indica il numero di figli che vivono insieme all'intervistato.
- FAM_MIN: variabile esplicativa X_{14} di tipo metrica che indica il numero di figli minorenni che vivono insieme all'intervistato.
- FAM_MOTO: variabile esplicativa X_{15} di tipo metrica che indica il numero di moto o motorini a disposizione del nucleo familiare dell'intervistato.

In tutti i casi in cui è presente una variabile qualitativa con due fattori del tipo Sì o No si è considerata una codifica dummy ponendo il valore 1 per la risposta affermativa e il valore 0 per la risposta negativa.

Un'altra considerazione, affrontata prima di sviluppare i diversi modelli con l'analisi di regressione lineare e logistica, riguarda il fenomeno di *overfitting* (in italiano: eccessivo adattamento) ovvero la possibilità che il modello statistico si adatti al campione perché ha un numero eccessivo di parametri rispetto al numero di osservazioni. Per evitare tale problema, nella regressione lineare multivariata, viene suggerita da Tabachnick e Fidell (1989) una pianificazione degli esperimenti con rapporto 20:1 fra il numero n delle osservazioni ed il numero k delle variabili esplicative, ossia 20 osservazioni per le 15 variabili esplicative valutate. In tal modo si considera il 9% delle osservazioni totali. La scelta delle osservazioni è stata effettuata casualmente mediante apposito comando "scelta casi" presente nel menu DATI del software SPSS.

5.3 Correlazione di tutte le variabili considerate con CATI/CAWI

Prima di analizzare i diversi modelli di regressione multivariata, è stato ritenuto opportuno studiare quelle che sono le correlazioni tra la variabile TIPO_INTERV e le restanti variabili socio economiche citate nel paragrafo precedente. Lo scopo di questa analisi è quello di capire come al variare di una variabile socio economica varia il fatto di aver risposto ad un'indagine telefonica o web. Di fatto, la correlazione è una condizione necessaria ma non sufficiente per definire un principio di causa-effetto fra due variabili. Al contrario, se le due variabili risultano esser non correlate, risulta difficile affermare che una variabile sia causa-effetto dell'altra.

Dato che la regressione lineare e la regressione logistica, per i motivi detti precedentemente (ovvero per evitare un eventuale fenomeno di overfitting), verranno studiate considerando il 9% delle osservazioni, anche in questa analisi di studio, si è deciso di analizzare le medesime osservazioni.

Essendo la variabile TIPO_INTERV di tipo dicotomica, si è deciso di utilizzare la *Correlazione Punto-Biserial*, nel caso in cui la variabile socio economica è metrica, invece, se la variabile socio economica è dicotomica, si è analizzata la correlazione mediante l'*Indice Phi*. Il risultato fornito da SPSS per la Correlazione Punto-Biserial, si trova nella tabella "Correlazioni" come mostrato in Tab. 18, invece, il risultato dell'Indice Phi nella tabella "Correlazioni con Indice Phi" [Tab. 19] e "Tabella di contingenza" [Tab. 20]. Quest'ultima, rappresenta una tabella a doppia entrata, in cui vengono considerate le percentuali di riga e di colonna delle

frequenze congiunte delle sole variabile che presentano, nell'analisi con indice Phi, una correlazione significativa con la variabile TIPO_INTERV.

Correlazioni		TIPO_INTERV
TIPO_INTERV	Correlazione di Pearson Sign. (a due code) N	1 345
FAM_REDDITO	Correlazione di Pearson Sign. (a due code) N	,151 ,005 345
FAM_N	Correlazione di Pearson Sign. (a due code) N	,177 ,001 345
FAM_FIGLI	Correlazione di Pearson Sign. (a due code) N	,167 ,002 345
FAM_MIN	Correlazione di Pearson Sign. (a due code) N	,143 ,008 345
FAM_AUTO	Correlazione di Pearson Sign. (a due code) N	,103 ,057 345
FAM_MOTO	Correlazione di Pearson Sign. (a due code) N	,039 ,470 345

Tab. 18 - Correlazione punto-biseriale

Correlazioni con indice Phi			Valore	Significatività approssimata
TIPO_INTERV*CLASSE_GIOVANI				
Nominale per nominale	Phi		,028	,075
	V di Cramer		,028	,075
TIPO_INTERV*CLASSE_ADULTI				
Nominale per nominale	Phi		,144	,007
	V di Cramer		,144	,007
TIPO_INTERV*GENERE				
Nominale per nominale	Phi		-,025	,638
	V di Cramer		,025	,638

TIPO_INTERV*CLASSE_LICENZAMEDIASUP			
Nominale per nominale	Phi	,030	,571
	V di Cramer	,030	,571
TIPO_INTERV*CLASSE_LAUREA			
Nominale per nominale	Phi	,150	,005
	V di Cramer	,150	,005
TIPO_INTERV*SALUTE			
Nominale per nominale	Phi	-,141	,009
	V di Cramer	,141	,009
TIPO_INTERV*PATENTE			
Nominale per nominale	Phi	,163	,002
	V di Cramer	,163	,002
TIPO_INTERV*DIMORA			
Nominale per nominale	Phi	,088	,103
	V di Cramer	,088	,103
TIPO_INTERV*PARK_CASA			
Nominale per nominale	Phi	-,061	,260
	V di Cramer	,061	,260
TIPO_INTERV*TP_ABBON			
Nominale per nominale	Phi	,143	,008
	V di Cramer	,143	,008
TIPO_INTERV*AUCS_ABBON			
Nominale per nominale	Phi	,269	,000
	V di Cramer	,269	,000

Tab. 19 - Indice Phi di correlazione

Tabella di contingenza

		TIPO INTERV			
		CATIF		CAWIC	
		% di casi per riga	% di casi per colonna	% di casi per riga	% di casi per colonna
CLASSE_ADULTI	0	91,6%	47,8%	8,4%	27,1%
	1	81,6%	52,2%	18,4%	72,9%
CLASSE_LAUREA	0	88,8%	80,5%	11,2%	62,5%
	1	76,3%	19,5%	23,7%	37,5%
SALUTE	0	84,4%	87,2%	15,6%	100,0%
	1	100,0%	12,8%	0,0%	0,0%
PATENTE	0	97,2%	23,2%	2,8%	4,2%
	1	83,2%	76,8%	16,8%	95,8%
AUCS_ABBON	0	87,1%	100,0%	12,9%	91,7%
	1	0,0%	0,0%	100,0%	8,3%
TP_ABBON	0	88,8%	79,8%	11,2%	62,5%
	1	76,9%	20,2%	23,1%	37,5%

Tab. 20 - Tabella di contingenza relative allo studio dell'Indice Phi

Dalle Tab. 18 e Tab. 19, si evince che in tutti casi, a seconda del segno, esiste una trascurabile tendenza positiva o negativa, in quanto la maggior parte dei valori di correlazione risultano esser inferiori a $\pm 0,300$. Le variabili che presentano un valore di correlazione “maggiore” e risultano esser significative, nel caso di coefficiente punto-biserial, sono: FAM_REDDITO (0,151), FAM_N (0,177), FAM_FIGLI (0,167), FAM_MIN (0,143); invece, nel caso di correlazione definita dall’indice Phi sono: CLASSE_ADULTI (0,144), CLASSE_LAUREA (0,150), SALUTE (-0,141), PATENTE (0,163), TP_ABBON (0,143), AUCS_ABBON (0,269).

Successivamente, si è scelto di analizzare le variabili che presentano un livello di significatività inferiore al 5% in entrambi i casi di studio, in funzione del segno per la correlazione punto-biserial, invece, per la correlazione phi, si è tenuto conto non solo del segno ma anche delle percentuali di riga e di colonna espresse nelle tabella di contingenza. Da quest’ultimo approccio di analisi, si evince che gli adulti rispetto agli anziani, i laureati rispetto a coloro che hanno dichiarato di avere un grado d’istruzione inferiore a quello di licenza media inferiore, chi è in possesso della patente di guida per le automobili e, infine, coloro che sono in possesso di un abbonamento per il trasporto pubblico locale e per il servizio di car sharing, hanno risposto maggiormente ad un’indagine web piuttosto che telefonica. Al contrario, coloro i quali nel sondaggio hanno dichiarato di avere problemi di salute o disabilità, temporanee o permanente, hanno risposto maggiormente ad un sondaggio CATI piuttosto che CAWI. Dall’approccio di analisi dato dalla correlazione punto-biserial, invece, si evince che all’aumentare del reddito, del numero di componenti del nucleo familiare e, infine, del numero di figli minori e

non, l'intervistato tende a rispondere maggiormente ad un'indagine web piuttosto che telefonica.

5.4 Modelli di regressione lineare multivariata per le frequenze di utilizzo di determinati mezzi di trasporto

5.4.1 Formulazione e stima dei modelli

Sulla base dei risultati dell'analisi di correlazione si è deciso di stimare in un primo momento un livello di previsione della frequenza d'uso di un determinato mezzo di trasporto in funzione di sole variabili socio-economiche (modello 1) e poi successivamente aggiungere alla stima la variabile categorica TIPO_INTERV composta da CATI CAWI per valutare se effettivamente il modello migliora oppure no (modello 2). Per fare ciò, si è preso in considerazione un modello di regressione multivariata in cui si fa ricorso a più variabili esplicative X_i per effettuare previsioni su una variabile indipendente Y .

L'equazione che esprime tale modello di regressione è la seguente:

$$\begin{aligned} \text{Mod1: } \text{FREQ}_i = & \beta + \beta_1 \text{CLASSEGIOVANI} + \beta_2 \text{CLASSEADULTI} + \beta_3 \text{LICENZAMEDIASUP} \\ & + \beta_4 \text{CLASSELAUREA} + \beta_5 \text{GENERE} + \beta_6 \text{DIMORA} + \beta_7 \text{SALUTE} + \beta_8 \text{PATENTE} + \\ & \beta_9 \text{PARK_CASA} + \beta_{10} \text{TP_ABBON} + \beta_{11} \text{AUCS_ABBON} + \beta_{12} \text{FAM_REDDITO} + \\ & \beta_{13} \text{FAM_AUTO} + \beta_{14} \text{FAM_N} + \beta_{15} \text{FAM_FIGLI} + \beta_{16} \text{FAM_MIN} + \beta_{17} \text{FAM_MOTO} \end{aligned}$$

Mediante l'uso del software SPSS è stato possibile ricavare, per ogni tipologia di frequenza considerata, un livello di previsione della frequenza d'uso di un determinato mezzo di trasporto in funzione delle variabili socio-economiche espresse nell'equazione qui sopra. Si riporta all'allegato 1 per visualizzare tali tabelle.

In un secondo momento, si è deciso di inserire al precedente modello un'ulteriore variabile esplicativa in modo tale da verificare se l'aggiunta della variabile TIPO_INTERV, che tiene conto delle due modalità di somministrazione dell'indagine, può portare ad avere un modello migliore oppure no. Questa nuova variabile esogena, essendo composta da solo due categorie, assume valore pari a zero o unitario nel caso in cui l'intervistato abbia risposto rispettivamente con indagine telefonica o web.

L'equazione che esprime tale secondo modello di regressione è la seguente:

$$\text{Mod2: } \text{FREQ}_i = \beta + \beta_1 \text{CLASSEGIOVANI} + \beta_2 \text{CLASSEADULTI} + \beta_3 \text{LICENZAMEDIASUP} + \beta_4 \text{CLASSELAUREA} + \beta_5 \text{GENERE} + \beta_6 \text{DIMORA} + \beta_7 \text{SALUTE} + \beta_8 \text{PATENTE} + \beta_9 \text{PARK_CASA} + \beta_{10} \text{TP_ABBON} + \beta_{11} \text{AUCS_ABBON} + \beta_{12} \text{FAM_REDDITO} + \beta_{13} \text{FAM_AUTO} + \beta_{14} \text{FAM_N} + \beta_{15} \text{FAM_FIGLI} + \beta_{16} \text{FAM_MIN} + \beta_{17} \text{FAM_MOTO} + \beta_{18} \text{TIPO_INTERV}$$

Mediante l'utilizzo del software SPSS, tale modello di regressione, per ogni frequenza analizzata, ha fornito gli output presenti nell'allegato 1.

Dai due modelli di regressione, per ogni variabile indipendente considerata, è stato possibile trarre le seguenti considerazioni:

- Tutte le variabili indipendenti di frequenza di utilizzo di un determinato mezzo di trasporto presentano un buon livello di significatività, grazie anche all'elevato numero di gradi di libertà, seppur si siano ridotte le osservazioni per evitare problematiche relative all'overfitting;

- In tutti i modelli, ad eccezione di `FREQ_AUCO` e `FREQ_AUPA`, si è riscontrato che il valore dell'intercetta presente sia nel Mod1 che nel Mod2 non risulta esser significativa. In tali casi, si è optato nel considerare un valore dell'intercetta pari a zero, ottenendo così degli errori standard inferiori;
- Dal confronto del coefficiente di determinazione R^2 ottenuto dal Mod1 e dal Mod2 per ogni singola frequenza presa in considerazione, si può affermare che, ad eccezione di `FREQ_TRENO` e `FREQ_AUPA`, il modello non migliora e l' R^2 può esser paragonabile. Quindi la differenza tra chi ha risposto con modalità CATI e chi invece con modalità CAWI potrebbe esser riconducibile a variabili socio economiche. In sostanza, l'effetto di usare un modo di risposta rispetto ad un altro, è catturato dalle variabili socio-economiche prese in esame e, inoltre, la variabile `TIPO_INTERV` non dà un contributo in più rispetto a quello che è già catturato dalle variabili esogene considerate. Nel caso di `FREQ_TRENO` e di `FREQ_AUPA`, invece, i coefficienti di determinazione di entrambi i modelli non sono paragonabili, anzi, si nota un aumento di tale coefficiente nel Mod2. Ciò può significare che la variabile `TIPO_INTERV` stia dando un contributo alla frequenza che non è rappresentato da tutte le restanti variabili esogene considerate. Si potrebbe optare, nel caso di `FREQ_TRENO`, di considerare una variabile in più che tenga conto della distanza della dimora degli intervistati dalle stazioni del passante ferroviario, supponendo che, chi ci abita vicino, prenda con maggiore frequenza il treno. Non è stato possibile fare ciò, in quanto nel sondaggio non sono forniti dati in merito alla geo localizzazione così precisi. Nel caso di `FREQ_AUPA`, invece, si potrebbe considerare una variabile in più

che tenga conto di ulteriori caratteristiche nell'impedimento dell'utilizzo dell'auto come conducente che probabilmente non sono catturate a sufficienza dalla variabile SALUTE;

- Prendendo in considerazione la variabile esogena DIMORA nel caso di `FREQ_AUCO` e di `FREQ_METRO`, si può notare come, nel primo caso, la variabile indipendente risulti avere il segno del coefficiente negativo, al contrario nel secondo caso. Questo può esprimere il fatto che, coloro che abitano a Torino, hanno la possibilità di prendere la metro rispetto a chi abita fuori Torino; al contrario quest'ultimi sono costretti maggiormente ad usare l'auto avendo un servizio di trasporto pubblico con prestazioni inferiori rispetto a chi abita nel capoluogo piemontese;
- Le variabili `TP_ABBON` e `FAM_AUTO`, invece, risultano avere il segno del coefficiente, rispettivamente, positivo e negativo nel caso di `FREQ_BUS`, al contrario, segno negativo e positivo nel caso di `FREQ_AUCO`. Ciò può significare che, coloro i quali hanno poche auto in famiglia, sono abbonati al servizio di trasporto pubblico locale e usano maggiormente il bus come mezzo di trasporto rispetto all'auto. Invece, coloro che hanno a disposizione un buon numero di auto in famiglia, non sono abbonati al servizio di trasporto pubblico e usano maggiormente l'auto per i loro spostamenti;
- Prendendo in considerazione la variabile `FAM_MIN` nel caso di `FREQ_AUCO`, `FREQ_BUS` e `FREQ_METRO`, si può dedurre che all'aumentare del numero di figli minori, aumenta la frequenza d'uso dell'auto, al contrario, diminuisce la frequenza d'uso della metro e bus. Si può ipotizzare che gli intervistati con figli minori preferiscono spostarsi con l'auto privata invece di usufruire del

trasporto pubblico locale, in quanto l'auto può essere vista non solo come un mezzo più sicuro rispetto ai mezzi del servizio pubblico locale, ma anche un mezzo di trasporto in cui è più facile gestire e stare attenti alle esigenze dei bambini;

- Inoltre, la variabile PATENTE e AUCS_ABBON risultano avere il segno del coefficiente rispettivamente negativo e positivo nel caso di FREQ_AUPA. Si può ipotizzare che gli individui che usano l'auto come passeggero per i loro spostamenti, siano per lo più persone che non sono in possesso della patente di guida per le automobili oppure che non hanno un'auto privata e usufruiscono del servizio di car sharing;
- La variabile PARK_CASA, invece, risulta avere il segno del coefficiente positivo nel caso di FREQ_AUCO, al contrario, negativo nel caso di FREQ_BICI, FREQ_BUS e FREQ_METRO. Questo può significare che coloro che hanno un box auto a casa, preferiscono spostarsi con il proprio veicolo privato, ovvero l'auto, al contrario di chi invece, non avendo a disposizione un parcheggio privato, preferisce utilizzare la bici, il bus o metro, evitando così il problema, ad esempio, di trovare un parcheggio "sotto casa", non essendo quest'ultimo sempre disponibile.

5.4.2 Il tipo di intervista come proxy delle variabili considerate

Sulla base dell'analisi di regressione precedente, si è scelto di valutare se la variabile categorica TIPO_INTERV, costituita dalle due categorie CATI e CAWI, può essere considerata come variabile proxy, ovvero come variabile che può sostituire

alcune variabili socio-economiche, in alcuni modelli che permettono di predire una determinata frequenza di utilizzo di mezzo di trasporto.

Per capire il procedimento che è stato effettuato per determinare se cati/cawi può esser proxy di alcune variabili socio-economiche, si sono effettuate una serie di analisi considerando le variabili in comune tra il primo e il secondo modello e, in funzione della loro significatività, si sono tratte delle conclusioni. Si riporta nella Fig. 21 il diagramma di flusso dei possibili casi che si possono ottenere e i relativi procedimenti. Nella formulazione di tale diagramma si farà riferimento ad una generica "variabile 1" in comune fra i due modelli e per la quale si vuole valutare in che misura potrebbe essere sostituita, in realtà, si ripete questo procedimento per tutte le variabili del modello.

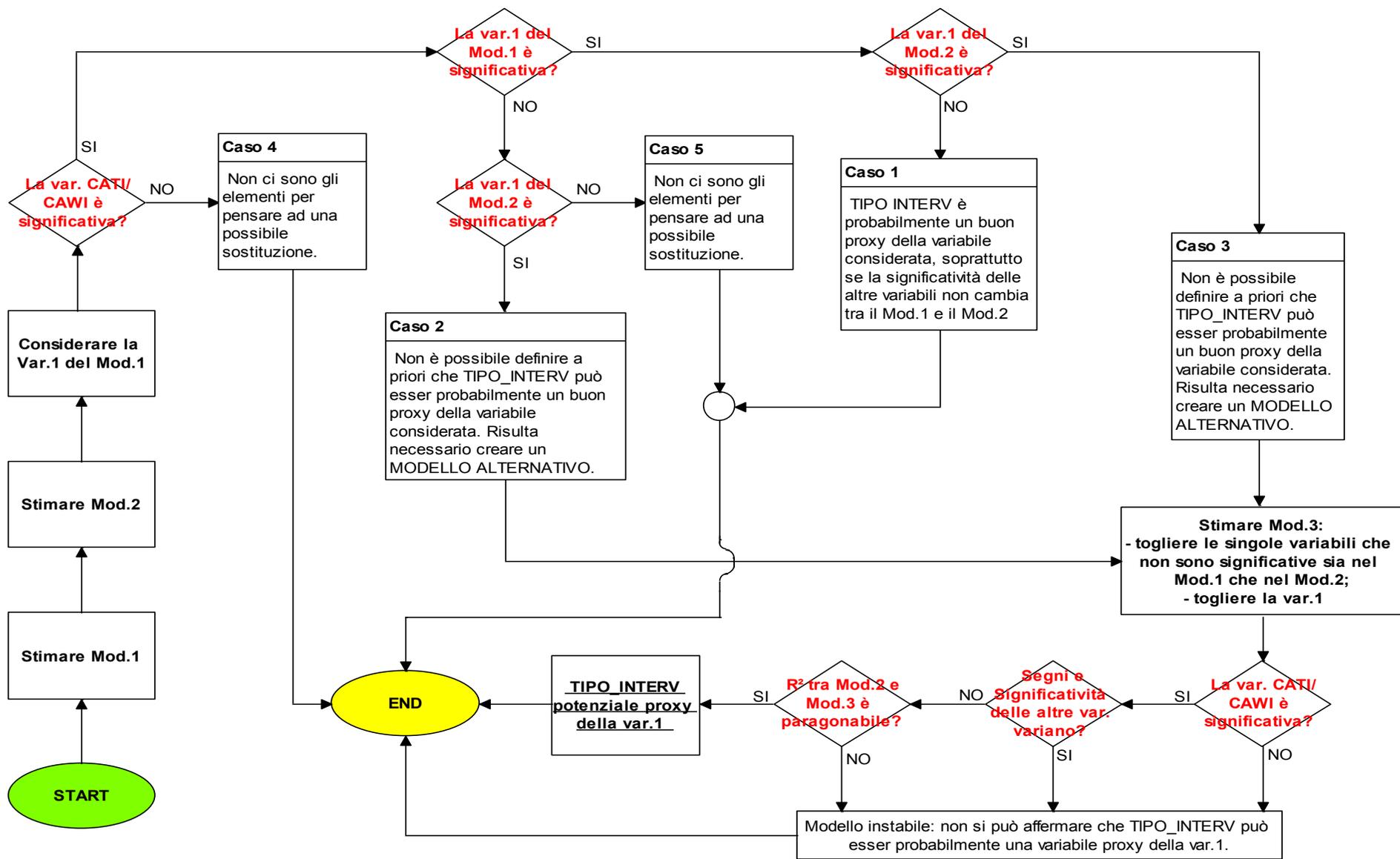


Fig. 21 - Diagramma di flusso utile a determinare se cati/cawi può essere proxy di alcune variabili socio-economiche.

Il caso 1 rappresenta il caso più chiaro di sostituzione. Esso è caratterizzato dalla variabile TIPO_INTERV significativa e dalle variabili socio-economiche che sono significative nel modello 1 e che diventano non significative nel modello 2. In tal caso, si può affermare che cati/cawi è probabilmente un buon proxy della variabile socio-economica considerata, soprattutto se la significatività delle altre variabili non cambia.

Il caso 2, invece, è definito dalla variabile TIPO_INTERV significativa e dalle variabili esogene che sono non significative nel modello 1 e che diventano significative nel modello 2. In tal caso, per ciascuna di queste variabili, si è scelto di stimare un modello alternativo definito come nel modello 2 ma senza la variabile considerata e le variabili non significative. Nella stima del nuovo modello, se la variabile TIPO_INTERV rimane significativa, se non cambiano né i segni né la significatività delle variabili esogene rimanenti, e, infine, se i coefficienti di determinazione R^2 del modello 2 e del modello alternativo risultano esser paragonabili, allora cati/cawi può essere considerata come un potenziale proxy della variabile eliminata, anche se andrebbero svolte ulteriori verifiche, ad esempio, sul grado di correlazione fra le variabili. Tale risultato d'altro canto non ha grande rilevanza pratica, in quanto la variabile sotto esame non è significativa nel primo modello, e pertanto non si analizzerà nel dettaglio.

Il caso 3 si differenzia dal precedente in quanto la variabile sotto esame è significativa anche nel modello 1. Anche in questo caso, per ciascuna variabile che risulta esser significativa sia nel modello 1 che nel modello 2, si è scelto di stimare un nuovo modello alternativo senza la variabile considerata e senza le variabili non

significative. Se la variabile TIPO_INTERV rimane significativa e non cambiano né i segni né le significatività delle variabili esogene rimanenti, è possibile valutare se i coefficienti di determinazione R^2 tra il modello 2 e il nuovo modello alternativo possono esser paragonabili. In tal caso, la variabile TIPO_INTERV potrebbe essere proxy della variabile sotto esame, ma occorrerebbe di nuovo effettuare ulteriori indagini.

Nel caso 4, invece, qualunque sia la significatività della variabile in comune considerata sia nel primo che nel secondo modello, essendo la variabile TIPO_INTERV non significativa, allora cati/cawi non è probabilmente proxy di nessuna delle variabili presenti nel modello.

Infine, nel caso 5, la variabile sotto esame risulta esser non significativa sia nel modello 1 che nel modello 2. In tal caso, nonostante la variabile TIPO_INTERV risulti esser significativa, non ci sono degli elementi sufficienti per pensare ad una sostituzione e considerare la variabile TIPO_INTERV come proxy della variabile analizzata.

Per le tabelle riguardanti i diversi modelli stimati per la valutazione di CATI/CAWI come variabile proxy, si rimanda all'allegato 1. Qui di seguito, invece, si riporta una tabella riassuntiva [Tab. 21], in cui una cella viene colorata di rosso quando, nel caso 2 e 3, i nuovi modelli alternativi stimati presentano o la variabile TIPO_INTERV non significativa, o un valore di R^2 non paragonabile, oppure quando vi è una variazione dei segni o della significatività delle restanti variabili rispetto al modello 2, e quindi, per uno di questi motivi, la variabile in comune non può esser

sostituita da cati/cawi. Invece, sono state colorate di verde tutte le variabili che, probabilmente, possono esser sostituite da cati/cawi.

In particolare, si è deciso di evidenziare in grassetto il caso 1 in quanto risulta esser l'unico caso veramente convincente di sostituzione. Per tale motivo, in questo paragrafo, verranno trattate e argomentate solo le variabili socio-economiche che fanno parte di questo caso.

		DIVERSI CASI OTTENUTI PER LA VALUTAZIONE DELLE VARIABILI POTENZIALMENTE SOSTITUIBILI DALLA VARIABILE CANDIDATA PROXY TIPO_INTERV					
		FREQ_AUCO	FREQ_AUPA	FREQ_BICI	FREQ_BUS	FREQ_METRO	FREQ_TRENO
VARIABILI ESOGENE IN COMUNE	CLASSE_GIOVANI	Caso3	Caso5	Caso4	Caso4	Caso4	Caso5
	CLASSE_ADULTI	Caso1	Caso5	Caso4	Caso4	Caso4	Caso5
	FAM_REDDITO	Caso5	Caso5	Caso4	Caso4	Caso4	Caso5
	GENERE	Caso3	Caso5	Caso4	Caso4	Caso4	Caso5
	CLASSE_LICENZAMEDIASUP	Caso5	Caso5	Caso4	Caso4	Caso4	Caso5
	CLASSE_LAUREA	Caso5	Caso5	Caso4	Caso4	Caso4	Caso5
	SALUTE	Caso5	Caso5	Caso4	Caso4	Caso4	Caso5
	PATENTE	Caso3	Caso3	Caso4	Caso4	Caso4	Caso5
	DIMORA	Caso5	Caso3	Caso4	Caso4	Caso4	Caso5
	PARK_CASA	Caso3	Caso5	Caso4	Caso4	Caso4	Caso3
	TP_ABBON	Caso3	Caso5	Caso4	Caso4	Caso4	Caso1
	AUCS_ABBON	Caso5	Caso1	Caso4	Caso4	Caso4	Caso5
	FAM_N	Caso3	Caso5	Caso4	Caso4	Caso4	Caso5
	FAM_FIGLI	Caso3	Caso5	Caso4	Caso4	Caso4	Caso5
	FAM_MIN	Caso3	Caso5	Caso4	Caso4	Caso4	Caso3
	FAM_MOTO	Caso5	Caso5	Caso4	Caso4	Caso4	Caso5
	FAM_AUTO	Caso3	Caso5	Caso4	Caso4	Caso4	Caso5
TIPO_INTERV	Significativa	Significativa	No Significativa	No significativa	No Significativa	Significativa	

Tab. 21 - Tabella riassuntiva del procedimento utilizzato per la valutazione delle variabili sostituibili da TIPO_INTERV.

Dalla precedente tabella è stato possibile effettuare due tipi di valutazione: una riguardante l'analisi dei segni dei coefficienti della variabile TIPO_INTERV e delle relative variabili socio-economiche che possono esser sostituite da cati/cawi, in relazione alla frequenze di utilizzo del mezzo di trasporto, e l'altra, invece, che mira ad individuare, in funzione dei segni dei coefficienti, una relazione tra la variabile potenziale proxy e le relative variabili socio-economiche.

Dalla Tab. 21 si evince che la variabile TIPO_INTERV può esser potenzialmente utilizzata come variabile proxy, per predire la FREQ_AUCO, al posto della seguente variabile socio-economica: CLASSE_ADULTI. Essendo i segni dei coefficienti di entrambe le variabili positivi, si avrà lo stesso effetto sulla frequenza. Ovvero il fatto di esser adulti e di aver risposto ad un sondaggio web porta ad avere un aumento in termini di frequenza dell'utilizzo dell'auto. Tale relazione può esser dovuta al fatto che un maggiore utilizzo dell'auto presume, da parte degli intervistati, uno stile di vita più attivo e meno sedentario, e tale caratteristica comportamentale si può riscontrare sia negli adulti rispetto agli anziani e sia negli individui che, trascorrendo meno tempo a casa, non hanno avuto la possibilità di rispondere al telefono e, quindi, hanno risposto ad un sondaggio web. Inoltre, sempre in funzione dei segni dei coefficienti e della variabile CLASSE_ADULTI che può esser sostituita da CATI/CAWI, si può affermare che gli intervistati adulti, ovvero quelle persone comprese nella fascia d'età che va dai 34 ai 67 anni d'età [Tab. 14], essendo, presumibilmente, più pratici nell'utilizzo di internet e, inoltre, passando maggior tempo d'avanti al computer o a qualsiasi altro strumento informatico rispetto agli anziani, con più probabilità avranno risposto ad un sondaggio web e non telefonico.

Nel caso di `FREQ_AUPA`, dummy `cati/cawi` è proxy di `AUCS_ABBON`. Entrambe le variabili indipendenti hanno il segno del coefficiente positivo, quindi, gli intervistati che hanno dichiarato di esser abbonati al servizio di car sharing, hanno anche, verosimilmente, risposto al sondaggio con modalità web, in quanto, tale servizio di trasporto, essendo principalmente gestito mediante applicazione mobile o sito web, è utilizzato da individui che hanno una certa dimestichezza e familiarità con internet. Inoltre, tenendo sempre conto dei segni dei coefficienti, si può affermare che, gli intervistati che hanno risposto ad un'indagine web e che hanno dichiarato di esser abbonati ad un servizio di car sharing, utilizzano con maggiore frequenza l'auto come passeggero. Tale relazione può esser giustificata dal fatto che gli intervistati abbonati al servizio di car sharing non hanno, presumibilmente, un'auto privata, e quindi per i loro spostamenti, o utilizzano l'auto condivisa o come passeggero, oppure, che chi ha risposto ad un sondaggio web, utilizzi internet anche per i servizi di car pooling, i quali rientrano nella categoria di autoveicoli come passeggero.

Invece, per predire la `FREQ_TRENO`, la variabile `TIPO_INTERV` può sostituire `TP_ABBON`. Entrambe le variabili esogene hanno il segno del coefficiente positivo, quindi, gli intervistati che hanno risposto ad un sondaggio web, sono anche, presumibilmente, degli abbonati al servizio di trasporto pubblico locale. Inoltre, essendo i segni dei coefficienti di entrambe le variabili indipendenti positivi, si ha lo stesso effetto sulla frequenza. Ovvero, il fatto di possedere un abbonamento ai servizi di trasporto pubblico e di aver risposto ad un sondaggio web, porta ad avere un aumento, in termini di frequenza, dell'utilizzo del treno.

Per tutti i casi di studio sopra citati, non è possibile affermare con certezza e decisione se la variabile CATI/CAWI è effettivamente proxy delle variabili socio-economiche trovate. Andrebbero fatte ulteriori valutazioni e considerazioni, come effettuato successivamente nel Paragrafo 5.6

5.5 Modello di regressione logistica con variabile endogena CATI/CAWI

Con il modello di regressione logistica, si cerca di analizzare se la variabile TIPO_INTERV, contenente le due categorie CATI/CAWI, può esser classificata come una variabile che racchiude in sé informazioni riguardanti le caratteristiche socio-economiche degli individui. Pertanto, TIPO_INTERV sarà ora la variabile dipendente del modello. Nel quadro analitico del presente lavoro, la regressione logistica può esser vista come un metodo di analisi confermativo della regressione lineare, per quanto concerne la stima della variabile TIPO_INTERV come variabile proxy. Per questo motivo, si è deciso di fare un'analisi più mirata andando a considerare solo le variabili esogene che, nel capitolo precedente, sono risultate esser potenzialmente sostituibili dalla variabile proxy TIPO_INTERV, ovvero quelle variabili che nella regressione lineare sono state analizzate come caso 1, 2 e 3 (celle verdi della Tab. 21). Così facendo, essendo il caso 2 e il caso 3 quelli che presentano più criticità nel confermare con maggiore certezza la sostituzione di una variabile socio-economica con cati/cawi, si possono ottenere, ulteriori garanzie sulla sostituzione.

Quindi, le variabili indipendenti considerate nella specificazione del modello, sono: CLASSE_GIOVANI, CLASSE_ADULTI, TP_ABBON e AUCS_ABBON.

Anche in questa tipologia di analisi, per evitare un eventuale fenomeno di overfitting, si è scelto di calibrare il modello considerando il medesimo 9% delle osservazioni totali già utilizzate precedentemente nello studio della correlazione e nella regressione lineare.

L'equazione che esprime il modello di regressione logistica è la seguente:

$$\ln\left(\frac{CAWI}{1-CAWI}\right) = \beta_0 + \beta_1 \text{CLASSE_GIOVANI} + \beta_2 \text{CLASSE_ADULTI} + \beta_3 \text{TP_ABBON} + \beta_4 \text{AUCS_ABBON}$$

dove al numeratore, CAWI rappresenta la probabilità di aver risposto ad un'indagine web. Invece, al denominatore, 1-CAWI rappresenta la probabilità di aver risposto ad un'indagine telefonica.

Nell'esecuzione del seguente modello di regressione logistica mediante software SPSS, si è riscontrato che la variabile AUCS_ABBON fornisce dei problemi computazionali nella calibrazione del modello (Errore Standard=20087,76; Sign=,999), dovuti al fatto che sono pochissime le osservazioni di chi dichiara di avere un abbonamento al servizio di car sharing rispetto la totalità del campione analizzato (0 e 4 osservazioni rispettivamente per CATI e CAWI). Per tale motivo, si è deciso di eliminare la variabile AUCS_ABBON per evitare di avere un modello poco stabile. Inoltre si è deciso di tener conto di un valore dell'intercetta pari a zero, in quanto essa risulta non esser significativa.

Pertanto, il modello di regressione logistica analizzato è il seguente:

$$\ln\left(\frac{CAWI}{1-CAWI}\right) = \beta_1 \text{CLASSE_GIOVANI} + \beta_2 \text{CLASSE_ADULTI} + \beta_3 \text{TP_ABBON}$$

L'output fornito dal software SPSS per il seguente modello di regressione è espresso nella Tab. 22:

		B	S.E.	Wald	gl	Sign.
Fase 1	CLASSE_GIOVANI	-1,509	,321	22,081	1	,000
	CLASSE_ADULTI	-1,467	,195	56,302	1	,000
	TP_ABBON	-,112	,307	,134	1	,714

Tab. 22 - Tabella di regressione logistica multivariata.

Dalla seguente tabella, si evince che le variabili considerate, risultano esser tutte significative ad eccezione di TP_ABBON che presenta un valore di significatività maggiore al limite imposto, ovvero del 5%.

Per valutare meglio il ruolo della variabile TP_ABBON che è apparentemente in discrepanza con le analisi precedenti, sulla base di una discussione online¹⁴ presente in un forum di statistica, si è deciso di valutare, sempre in un'analisi di regressione logistica definita, in tal caso, *univariata* per la presenza di una sola variabile esplicativa, la singola variabile TP_ABBON. Il modello analizzato è il seguente:

$$\ln\left(\frac{CAWI}{1-CAWI}\right) = \beta_1 \text{TP_ABBON}$$

¹⁴ <https://stats.stackexchange.com/questions/194855/significant-in-bivariate-regression-but-not-significant-in-multivariate-regress>

L'output fornito dal software SPSS per il seguente modello di regressione è espresso nella Tab. 23:

	B	S.E.	Wald	gl	Sign.
Fase 1 TP_ABBON	-1,204	,269	20,071	1	,000

Tab. 23 - Tabella di regressione logistica univariata.

Si può riscontrare che la variabile TP_ABBON risulta esser non significativa nella regressione multivariata, invece, risulta esser significativa nella regressione univariata. Tale risultato può esser dovuto a eventuali fenomeni di *multicollinearità*. In pratica, la variabile TP_ABBON presa singolarmente può predire la variabile TIPO_INTERV. Ma nel momento in cui vengono introdotte ulteriori variabili (ovvero CLASSE_GIOVANI e CLASSE_ADULTI), esse potrebbero essere delle variabili migliori nel predire la variabile dipendente e, di conseguenza, parte della varianza prevista della variabile TP_ABBON, viene attribuita alle restanti variabili esogene presenti nel modello, ovvero CLASSE_GIOVANI e CLASSE_ADULTI. Per tale motivo, si riduce il significato complessivo della variabile TP_ABBON come variabile sostituibile da TIPO_INTERV.

Quindi si può affermare che la variabile TIPO_INTERV può esser, presumibilmente, proxy delle variabili esogene CLASSE_GIOVANI, CLASSE_ADULTI e TP_ABBON, seppur quest'ultima sia una variabile con un ruolo in subordine rispetto alle altre due variabili. Per affermare con maggiore certezza tale riscontro, è opportuno valutare ulteriori considerazioni, come effettuato nel paragrafo successivo.

Dallo studio dei segni dei coefficienti della regressione logistica, si evince che i giovani e gli adulti, rispetto agli anziani, preferiscono rispondere ad un'indagine web piuttosto che telefonica. Questo risultato può esser giustificato dal fatto che tali categorie di persone intervistate, avendo un'età minore rispetto a quella degli anziani, e quindi, avendo vissuto nei periodi di diffusione in scala mondiale di strumenti informatici e di internet, abbiano, presumibilmente, maggiore dimestichezza nell'utilizzo del computer e del web, e quindi con più probabilità hanno risposto a un'indagine cawi e non cati.

Un'altra ipotesi che può giustificare il seguente risultato, può esser dovuto al fatto che i giovani e gli adulti potrebbero avere uno stile di vita meno sedentario rispetto a quello degli anziani, quindi, trovandosi meno tempo a casa, è meno probabile che abbiano risposto al telefono per effettuare un'indagine di tipo cati.

Inoltre, si può tenere presente che il sondaggio telefonico viene effettuato attraverso chiamata a telefono fisso, il quale è diventato, specialmente negli ultimi decenni con l'avvento del cellulare e degli smartphones, un accessorio sempre meno presente nelle case, specialmente in quelle composte da giovani e adulti. Al contrario, nelle abitazioni composte da anziani, il telefono fisso rimane un mezzo di comunicazione ancora molto utilizzato. Ciò può giustificare il fatto che gli adulti e gli anziani, non avendo un telefono fisso a casa, abbiano, con più probabilità, risposto ad un sondaggio web piuttosto che telefonico.

Spostando l'attenzione, in questo caso, sul segno del coefficiente della variabile TP_ABBON, esso risulta esser positivo. Si evince, quindi, che gli intervistati che hanno dichiarato di esser abbonati al servizio di trasporto pubblico locale, rispondono maggiormente ad un'indagine CAWI e non CATI. Ciò può esser dovuto

al fatto che, tali individui, siano più pratici nell'utilizzo del computer e nella navigazione web in quanto potrebbero non solo acquistare o rinnovare l'abbonamento online, ma anche verificare gli orari dei mezzi in tempo reale tramite app/sito web, oppure, perché, potrebbero condurre una stile di vita più attivo e meno sedentario, e quindi trovandosi meno tempo a casa, non hanno risposto alla chiamata del sondaggio telefonico.

5.6 Riepilogo del grado di sostituzione delle variabili socio-economiche con CATI/CAWI

Dal risultato finale della regressione lineare (Paragrafo 5.5) e della regressione logistica (Paragrafo 5.6), e, inoltre, grazie allo studio della correlazione tra la variabile TIPO_INTERV e le variabili socio-economiche (Paragrafo 5.4), è stato possibile formulare per le variabili considerate, le seguenti considerazioni:

- **CLASSE_GIOVANI:** variabile socio-economica che con l'analisi di regressione lineare risulterebbe poter esser sostituita dalla variabile potenziale proxy TIPO_INTERV per predire la frequenza di utilizzo dell'auto. Tale risultato, rientra nel caso 3, ovvero nel caso meno certo di sostituzione rispetto al caso 1. Inoltre, la regressione logistica ha permesso di confermare quanto trovato prima con la regressione lineare. Invece, si è riscontrato un risultato discordante con l'analisi di correlazione, poiché la relazione tra la variabile TIPO_INTERV e CLASSE_GIOVANI, risulta esser non solo bassa (,055) ma anche non significativa (,305). Ciò implica che le due variabili risultano essere scorrelate. Essendo la correlazione una

condizione necessaria ma non sufficiente per affermare se una qualsiasi variabile è proxy di un'altra, non possiamo affermare che TIPO_INTERV risulta esser proxy di CLASSE_GIOVANI;

CLASSE_ADULTI: variabile socio-economica che con l'analisi di regressione lineare risulta poter esser sostituita dalla variabile potenziale proxy TIPO_INTERV per predire la frequenza di utilizzo dell'auto. La sostituzione, rientra nel caso 1 di tale regressione, e quindi nel caso più certo. Nella regressione logistica, inoltre, la variabile CLASSE_ADULTI, risulta esser significativa e, quindi, conferma quanto trovato con la regressione lineare. Inoltre, dall'analisi di correlazione con l'indice Phi, risulta esserci una discreta correlazione (.144) e una buona significatività (.007) tra la variabile socio-economica e la variabile che indica la tipologia di intervista. Quindi, dallo studio delle tre analisi, si evince che, effettivamente, TIPO_INTERV può esser considerata come variabile proxy di CLASSE_ADULTI. In particolare, se l'intervistato ha risposto ad un'indagine cawi, allora, in un ipotetico modello previsionale, è possibile sostituire la variabile indicante il protocollo di indagine al posto di CLASSE_ADULTI per compensare la mancanza della variabile indicante gli intervistati con un range di anni compreso tra i 34 e 67 anni d'età;

- AUCS_ABBON: variabile che rientra nel caso 1 della regressione lineare, ovvero nel caso più certo di sostituzione, per predire la frequenza di utilizzo dell'autoveicolo come passeggero. Inoltre, presenta un valore di correlazione pari a ,0269, il quale risulta esser fortemente significativo (.000). Non si possono fare ulteriori considerazioni, in quanto, tale

variabile, non è stata analizzata nella regressione logistica per problemi computazionali, come spiegato nel precedente paragrafo. In sostanza, quindi, si può affermare che, sia per l'analisi di regressione logistica che per la correlazione, TIPO_INTERV può essere considerata proxy della variabile AUCS_ABBON. In particolare, se l'intervistato ha risposto ad un'indagine cawi, allora, in un ipotetico modello previsionale, è possibile sostituire la variabile indicante il protocollo di indagine al posto di AUCS_ABBON per compensare la mancanza della variabile indicante gli intervistati che hanno risposto positivamente alla domanda riguardante il possesso dell'abbonamento a un servizio di car sharing;

- TP_ABBON: variabile che con l'analisi di regressione lineare risulterebbe poter essere sostituita dalla variabile potenziale proxy TIPO_INTERV per predire la frequenza di utilizzo del treno. Tale risultato rientra nel caso 1 della regressione lineare, ovvero nel caso più certo di sostituzione. Inoltre, dall'analisi di correlazione, si è ottenuto un risultato pari a ,143, il quale risulta essere significativo (.008). Ciò permette di avvalorare quanto trovato prima con la regressione lineare. La regressione logistica, inoltre, ha confermato quanto detto precedentemente, ovvero che la variabile TIPO_INTERV può essere considerata proxy della variabile TP_ABBON, anche se il risultato è meno netto per i problemi di multicollinearità descritti nel paragrafo precedente. In particolare, se l'intervistato ha risposto ad un'indagine cawi, allora, in un ipotetico modello previsionale, è possibile sostituire la variabile indicante il protocollo di indagine al posto di TP_ABBON per compensare la mancanza della variabile indicante gli

intervistati che sono in possesso di un abbonamento al servizio di trasporto pubblico locale.

Se ne evince che la non conoscenza delle variabili socio-economiche sostituibili da TIPO_INTERV, ovvero CLASSE_ADULTI, AUCS_ABBON e TP_ABBON, può essere almeno parzialmente compensata dalla considerazione del protocollo di indagine in un ipotetico modello previsionale.

Si riporta qui di seguito una tabella riassuntiva [Tab. 24], riguardante la bontà della sostituzione delle singole variabili socio-economiche - ottenute con i tre metodi di analisi considerati - che possono esser sostituite dalla variabile proxy TIPO_INTERV. Si è scelto di inserire le variabili in ordine decrescente di bontà di sostituzione, ovvero dalle variabili che risultano poter esser sostituite con maggiore certezza dalla variabile proxy TIPO_INTERV a quelle meno.

VARIABILE	SIGNIFICATIVITÀ			TIPO_INTERV PROXY
	CORRELAZIONE	REGRESSIONE LINEARE	REGRESSIONE LOGISTICA	
CLASSE_ADULTI	Significativa	Caso 1	Significativa	SI
AUCS_ABBON	Significativa	Caso 1	Non definita	SI
TP_ABBON	Significativa	Caso 1	Significativa*	SI**
CLASSE_GIOVANI	NO Significativa	Caso 3	Significativa	NO
* Significatività della variabile nel caso di regressione logistica univariata.				
** Proxy ma con bontà di significato minore rispetto alle restanti variabili				

Tab. 24 - Tabella di riepilogo del grado di sostituzione delle variabili socio-economiche con CATI/CAWI.

6 L'EFFETTO DELLA VARIABILE PROXY SULLE FREQUENZE DI UTILIZZO DEI MEZZI DI TRASPORTO

6.1 Considerazioni di partenza per l'analisi della varianza a due vie

Se nel capitolo 4, dai risultati del test ANOVA a una via e degli analoghi test non parametrici, si è ricavato che la variabile TIPO_INTERV, contenente le due modalità di intervista (CATI-CAWI), risulta esser una variabile indicatore delle differenze tra le due medie, ovvero il diverso protocollo di indagine influenza la risposta data per quanto concerne la frequenza di utilizzo di un determinato mezzo di trasporto, in questo capitolo, mediante il test ANOVA a due vie, si valuta un eventuale effetto d'interazione di una variabile indipendente su una variabile dipendente a un livello specifico di un'altra variabile indipendente. In sostanza, si determina come l'effetto su una variabile dipendente cambi a seconda dei livelli delle variabili indipendenti considerate.

Considerando le analisi finora presentate, si è deciso di analizzare come alcune variabili indipendenti, rappresentate dalla variabile TIPO_INTERV e le relative variabili proxy trovate nel capitolo 5, variano in relazione alla variabile dipendente, ovvero la frequenza di utilizzo di un determinato mezzo di trasporto. In particolare, si è deciso di valutare quale sia l'effetto predominante tra le coppie di variabili indipendenti, ovvero se la frequenza di utilizzo di un determinato mezzo di trasporto è maggiormente influenzato dalla modalità di somministrazione dell'indagine oppure dalla variabile proxy presa in esame. Quest'analisi completa pertanto quella del capitolo precedente ed è volta a meglio

chiarire in quali casi la non conoscenza di determinate variabili socio-economiche può essere almeno parzialmente compensata dalla considerazione del protocollo di indagine in un ipotetico modello previsionale.

A seguito delle considerazioni fatte precedentemente, sono state considerate come variabili dipendenti le frequenze di utilizzo dei sei più comuni mezzi di trasporto, ovvero bici, autoveicolo privato come conducente e come passeggero, bus, metro e treno. Invece, come variabili indipendenti sono state considerate le variabili che sono risultate, nel capitolo precedente, poter esser sostituite dalla variabile proxy TIPO_INTERV, ovvero CLASSE_ADULTI, TP_ABBON E AUCS_ABBON, in modo tale da verificare come esse influenzano, in relazione a CATI/CAWI, la frequenza media di utilizzo dei mezzi di trasporto considerati. Successivamente, sono stati confrontati i risultati ottenuti con i modelli di regressione lineare con quelli dell'ANOVA a due vie.

Per evitare un'eventuale fenomeno di overfitting, inoltre, si è scelto di effettuare l'analisi dell'ANOVA a due vie considerando il medesimo 9% delle osservazioni totali già trattate precedentemente nello studio della correlazione e nelle analisi di regressione.

Mediante software SPSS è stato possibile effettuare l'analisi della varianza a due vie tra le variabili considerate. Si riporta all'allegato 2 per gli output di tabelle e grafici forniti dal programma.

6.2 La valutazione degli effetti di interazione sulle frequenze delle variabili considerate

Dagli output dell'ANOVA a due vie presenti in allegato 2, si è scelto di valutare e commentare le differenze delle medie di frequenza tra i diversi livelli categorici delle variabili che presentano una significatività inferiore al 5%, poiché, in tal caso, si può affermare che tale differenza non è dovuta al caso ma all'effetto delle variabili prese in esame.

Di seguito, si riportano due tabelle riassuntive [Tab. 25] [Tab. 26], in cui si pone, in ogni cella, il valore della frequenza media dei mezzi di trasporto considerati in funzione dei diversi livelli categorici delle variabili indipendenti. Essendo ogni variabile indipendente di tipo dicotomico, si è scelto di evidenziare di nero le "macro celle", le quali contengono al loro interno quattro celle rappresentative, nell'insieme, dei singoli casi d'analisi presi in esame.

Inoltre, per capire meglio quali sono gli effetti che risultano esser significativi, si è deciso di aggiungere degli asterischi per indicare la significatività delle differenze fra coppie di celle (sia in orizzontale che in verticale) che presentano soddisfatta tale condizione all'interno di ogni singolo caso, e, quindi all'interno di ogni macro cella. Ovvero, in Tab. 25, sono stati aggiunti asterischi per indicare la significatività delle differenze fra coppie di celle verticali che presentano un effetto significativo della variabile o TP_ABBON o CLASSE_ADULTI o AUCS_ABBON sulla frequenza di utilizzo di un determinato mezzo di trasporto a diversi livelli di categoria della variabile TIPO_INTERV. Invece, in Tab. 26, sono stati aggiunti asterischi per indicare la significatività delle differenze fra coppie di celle orizzontali che

presentano un effetto significativo della variabile TIPO_INTERV sulla frequenza di utilizzo di un determinato mezzo di trasporto ai diversi livelli di categoria della variabile o TP_ABBON o CLASSE_ADULTI o AUCS_ABBON.

		TIPO_INTERV											
		CAWI	CATI	CAWI	CATI	CAWI	CATI	CAWI	CATI	CAWI	CATI	CAWI	CATI
		FREQ_AUCO		FREQ_BICI		FREQ_AUPA		FREQ_BUS		FREQ_METRO		FREQ_TRENO	
TP_ABBON	SI	1,89	0,95	0,58	0,23	1,64	0,31	3,33	3,22	2,25	1,65	0,69	0,01
	NO	3,60**	2,70***	0,85	0,32	0,85**	0,34	1,28***	1,13***	0,90***	0,44***	0,53	0,27
CLASSE_ADULTI	SI	3,08	2,90	0,53	0,32	0,77	0,27	1,82	1,52	1,41	0,69	0,63	0,04
	NO	3,04***	1,60	1,34	0,28*	2,15***	0,40	2,50	1,64	1,39	0,68	0,50***	0,01***
AUCS_ABBON	SI	3,12	-	1,75	-	1,25	-	0,88	-	1,88	-	0,13	-
	NO	2,94	2,36	0,66*	0,30	1,14	0,33	2,11	1,58	1,36	0,69	0,64	0,02
		- Frequenza media non calcolata per l'assenza di osservazioni in tale combinazione; * significatività a $p \leq 0.05$ rispetto a "SI" (riga immediatamente superiore); ** significatività a $p \leq 0.01$ rispetto a "SI" (riga immediatamente superiore); *** significatività a $p \leq 0.001$ rispetto a "SI" (riga immediatamente superiore).											

Tab. 25 – Valori delle frequenze medie di utilizzo di determinati mezzi di trasporto nelle combinazioni delle variabili dicotomiche considerate e relativi effetti significativi delle variabil TP_ABBON, CLASSE_ADULTI e AUCS_ABBON.

		TIPO_INTERV											
		CAWI	CATI	CAWI	CATI	CAWI	CATI	CAWI	CATI	CAWI	CATI	CAWI	CATI
		FREQ_AUCO		FREQ_BICI		FREQ_AUPA		FREQ_BUS		FREQ_METRO		FREQ_TRENO	
TP_ABBON	SI	1,89	0,95	0,58	0,23	1,64	0,31***	3,33	3,22	2,25	1,65	0,69	0,01***
	NO	3,60	2,70*	0,85	0,32**	0,85	0,34**	1,28	1,13	0,90	0,44*	0,53	0,27***
CLASSE_ADULTI	SI	3,08	2,90*	0,53	0,32	0,77	0,27**	1,82	1,52	1,41	0,69	0,63	0,04***
	NO	3,04	1,60	1,34	0,28***	2,15	0,40***	2,50	1,64	1,39	0,68	0,50	0,01***
AUCS_ABBON	SI	3,12	-	1,75	-	1,25	-	0,88	-	1,88	-	0,13	-
	NO	2,94	2,36	0,66	0,30*	1,14	0,33***	2,11	1,58	1,36	0,69**	0,64	0,02***
		- Frequenza media non calcolata per l'assenza di osservazioni in tale combinazione; * significatività a $p \leq 0.05$ rispetto a "CAWI" (colonna immediatamente a sinistra); ** significatività a $p \leq 0.01$ rispetto a "CAWI" (colonna immediatamente a sinistra); *** significatività a $p \leq 0.001$ rispetto a "CAWI" (colonna immediatamente a sinistra).											

Tab. 26 – Valori delle frequenze medie di utilizzo di determinati mezzi di trasporto nelle combinazioni delle variabili dicotomiche considerate e relativi effetti significativi della variabile TIPO_INTERV.

È stato possibile effettuare due tipi di valutazione. Per facilità di comprensione si riportano, qui di seguito (par. 6.2.1 e par. 6.2.2), in maniera separata.

6.2.1 Stima degli effetti significativi nelle singole macro celle

La prima valutazione tiene conto degli effetti significativi nelle singole macro celle, e quindi delle coppie di celle che, nelle Tab. 25 Tab. 26, presentano degli asterischi di significatività. Si riportano, di seguito, le relative considerazioni classificate in funzione delle frequenze di utilizzo di determinati mezzi di trasporto:

A) **FREQ_AUCO**

L'effetto della variabile TP_ABBON risulta essere significativo in entrambi i casi, ovvero sia per coloro che hanno risposto ad un sondaggio con modalità CATI e sia per coloro che hanno risposto con modalità CAWI. Invece, l'effetto della variabile TIPO_INTERV risulta esser significativo solo nella categoria di coloro che hanno dichiarato di non avere l'abbonamento al servizio di trasporto pubblico. Inoltre, tale effetto, risulta essere di intensità minore rispetto a quello definito dalla variabile TP_ABBON. In sostanza, il fatto di non avere l'abbonamento al servizio di trasporto pubblico implica un aumento della frequenza dell'auto maggiore rispetto a quello definito dalla modalità di indagine. Quindi, in questa macro cella, risulta che l'effetto dell'abbonamento al trasporto pubblico ha un effetto più pesante rispetto alle caratteristiche socio-economiche che stanno dietro a cati/cawi. Tale valutazione risulta esser logica, in quanto è chiaro che chi non ha l'abbonamento ai mezzi di trasporto pubblico utilizzi maggiormente l'auto per spostarsi.

Gli effetti delle variabili CLASSE_ADULTI e TIPO_INTERV risultano esser significativi rispettivamente solo nel caso di risposta al sondaggio con modalità web e solo nel caso di individui adulti. Inoltre, l'effetto fornito dalla variabile TIPO_INTERV risulta esser di intensità maggiore rispetto a quello definito dalla variabile CLASSE_ADULTI. Si evince, quindi, che gli adulti che hanno risposto al web utilizzano con maggiore frequenza l'auto rispetto agli adulti che hanno risposto ad un sondaggio telefonico. Quindi si può presumere, che i primi hanno uno stile di vita più attivo rispetto ai secondi.

B) FREQ_BICI

L'effetto della variabile TIPO_INTERV risulta esser significativo nelle macro celle contenenti i casi di coloro che non hanno l'abbonamento ai servizi di trasporto pubblico e di car sharing e, infine, nella categoria anziani. Inoltre, si può notare che gli individui che hanno risposto ad un'indagine CAWI utilizzano con maggiore frequenza la bici rispetto a coloro che hanno risposto ad un'indagine CATI.

L'effetto della variabile CLASSE_ADULTI risulta esser significativo tra gli intervistati al telefono. In particolare, si evince che l'effetto predominante tra le due variabili indipendenti è quello della variabile TIPO_INTERV. Quindi si può affermare che gli anziani che hanno risposto ad un sondaggio web utilizzano con maggiore frequenza la bici rispetto agli anziani che hanno risposto ad un sondaggio telefonico.

L'effetto della variabile AUCS_ABBON, il quale risulta esser significativo tra coloro che hanno risposto ad un sondaggio web, è predominante rispetto all'effetto dato dalla variabile TIPO_INTERV, il quale, invece, risulta esser significativo tra coloro

che hanno dichiarato di non possedere un abbonamento di car sharing. Per tale motivo, si evince che, tra gli intervistati che hanno risposto ad un sondaggio cawi, coloro che sono abbonati al servizio di car sharing utilizzano maggiormente la bici rispetto a coloro che non lo sono.

C) FREQ_AUPA

Nelle macro celle contenenti TP_ABBON e CLASSE_ADULTI, risulta che l'effetto di tali variabili è significativo solo tra gli intervistati con modalità CAWI. Invece, l'effetto della variabile TIPO_INTERV risulta esser significativo in tutte le categorie delle variabile socio-economiche considerate, ad eccezione di coloro che dichiarano di non avere un abbonamento al servizio di car sharing. Inoltre, l'effetto della variabile TIPO_INTERV risulta essere quello predominante rispetto all'effetto dato da TP_ABBON e CLASSE_ADULTI. Quindi, l'aumento della frequenza di utilizzo dell'auto come passeggero risulta esser maggiormente influenzata dalle caratteristiche socio-economiche riconducibili alla modalità di indagine rispetto alle caratteristiche delle altre variabili considerate. Essendo, anche in questo caso, la frequenza del web maggiore rispetto la frequenza del telefono, si può pensare che gli intervistati che hanno risposto ad un sondaggio CAWI, utilizzino internet anche per i servizi di car pooling, i quali rientrano nella categoria di autoveicoli come passeggero.

D) FREQ_BUS

Nella macro cella riguardante la variabile TP_ABBON, risulta che tale variabile ha un effetto significativo sui diversi livelli di categoria della variabile TIPO_INTERV.

Inoltre, risulta che coloro che hanno l'abbonamento al servizio di trasporto pubblico utilizzano con maggiore frequenza il bus rispetto coloro che non lo hanno. Tale valutazione risulta esser ovvia anche da un punto di vista economico, in quanto agli individui che utilizzano frequente il bus è più conveniente avere un abbonamento al servizio di trasporto pubblico piuttosto che utilizzare il singolo biglietto di corsa semplice.

E) `FREQ_METRO`

L'effetto della variabile `TP_ABBON` risulta essere significativo in entrambi i casi, ovvero sia per coloro che hanno risposto ad un sondaggio con modalità `CATI` e sia per coloro che hanno risposto con modalità `CAWI`. Invece, l'effetto della variabile `TIPO_INTERV` risulta esser significativo solo nel caso delle categorie di coloro che dichiarano di non avere l'abbonamento al `tpl` e `car sharing`. Inoltre, l'effetto della variabile `TIPO_INTERV` risulta essere di intensità minore rispetto a quello definito dalla variabile `TP_ABBON`. Quindi, si evince che, la maggiore frequenza di utilizzo della metro è dovuta principalmente al possesso oppure no dell'abbonamento al servizio di trasporto pubblico rispetto alle caratteristiche socio-economiche riconducibili a `CATI/CAWI`. In sostanza, si può affermare che coloro che utilizzano maggiormente la metro sono per lo più degli individui che sono in possesso di un abbonamento al servizio di trasporto pubblico locale. Tale valutazione risulta esser logica, in quanto coloro che utilizzano la metro con maggiore frequenza sono motivati ad avere un abbonamento al servizio di trasporto pubblico locale sia per motivi di risparmio economico (utilizzare l'abbonamento è più vantaggioso di

utilizzare più biglietti di corsa semplice) sia per l'impossibilità di usufruire di tale servizio senza biglietto.

F) `FREQ_TRENO`

L'effetto della variabile `CLASSE_ADULTI` risulta esser significativo non solo tra gli intervistati al web ma anche tra gli intervistati al telefono. Invece, l'effetto della variabile `TIPO_INTERV` risulta esser significativo in tutte le categorie delle variabili socio-economiche analizzate, ad eccezione di coloro che dichiarano di non avere un abbonamento al servizio di car sharing. Inoltre, nella macro cella riguardante la `CLASSE_ADULTI`, l'effetto della variabile `TIPO_INTERV` risulta essere quello predominante. Quindi, l'aumento della frequenza di utilizzo del treno risulta esser maggiormente influenzato dalle caratteristiche socio-economiche complessive che stanno dietro a cati/cawi rispetto alla sola caratteristica di esser adulti o anziani.

6.2.2 Stima complessiva degli effetti significativi

Nella seconda valutazione, invece, si analizzano nel complesso le Tab. 25 e Tab. 26, tenendo conto di quelli che sono gli effetti significativi e delle frequenze medie di utilizzo dei singoli mezzi di trasporto, per avere una visione globale dell'interpretazione dei risultati dell'ANOVA a due vie. E' possibile formulare le seguenti considerazioni:

- Coloro i quali hanno condotto il sondaggio con modalità web utilizzano con maggiore frequenza tutti i mezzi di trasporto presi in esame rispetto a coloro che, invece, hanno condotto l'indagine con modalità telefonica. Tale risultato risulta esser concorde con quello trovato con l'Anova a una via. Ciò

può indicare che gli intervistati che hanno effettuato il sondaggio con modalità web abbiano uno stile di vita più attivo e meno sedentario, in quanto essi, rispetto a coloro che hanno risposto all'indagine con modalità telefonica, dichiarano di utilizzare maggiormente i diversi mezzi di trasporto per i loro spostamenti;

- Gli intervistati che hanno dichiarato di non avere un abbonamento al servizio di trasporto pubblico locale, indipendentemente dalla modalità di indagine, rispetto a coloro che dichiarano di avere l'abbonamento, utilizzano con maggiore frequenza i mezzi personali disponibili rispetto a quelli di pubblico utilizzo. Ovvero, tale tipologia di intervistati utilizza con maggiore frequenza l'auto e la bici personale. Al contrario, chi dichiara di avere un abbonamento al trasporto pubblico, utilizza con maggiore frequenza la metro, il treno e l'autoveicolo come passeggero. In sostanza, l'effetto significativo di avere o non avere l'abbonamento al servizio di trasporto pubblico spinge gli individui ad utilizzare per i loro spostamenti, nel primo caso, i mezzi privati, nel secondo, invece, quelli disponibili alla collettività definiti dal trasporto pubblico locale;
- Da una valutazione complessiva degli effetti significativi della variabile CLASSE_ADULTI, si evince che gli adulti rispetto gli anziani, utilizzano con maggiore frequenza l'auto come conducente, la bici e il treno. Invece, gli anziani, l'autoveicolo come passeggero. Questo può esser dovuto al fatto che gli anziani, ovvero quegli individui che hanno più di 68 anni d'età [Tab. 14], potrebbero non aver rinnovato la patente oppure avere degli impedimenti fisici dovuti all'età che portano ad utilizzare meno l'auto, la bici e il treno

per spostarsi rispetto agli adulti, privilegiando così l'utilizzo dell'auto come passeggero;

- Considerando, invece, la variabile AUCS_ABBON, risulta che coloro che hanno dichiarato di avere un abbonamento al servizio di car sharing utilizzano con maggiore frequenza la bici rispetto a coloro che dichiarano di non avere un abbonamento. Non è possibile fare ulteriori considerazioni, in quanto, il software SPSS, non ha valutato la significatività degli effetti che includono la relazione tra AUCS_ABBON-SI e TIPO_INTERV-CATI per assenza di osservazioni in tale combinazione.

6.3 Comparazione tra l'Anova a due vie e l'analisi di regressione

Nel Capitolo 5, mediante la valutazione e la calibrazione di modelli di regressione lineare, è stato individuato che TP_ABBON, CLASSE_ADULTI e AUCS_ABBON sono variabili potenzialmente sostituibili dalla variabile proxy TIPO_INTERV per predire la frequenza di utilizzo, rispettivamente, dell'auto come conducente, dell'auto come passeggero e, infine, del treno. Ovvero, tali variabili possono esser sostituite da CATI/CAWI per predire le frequenze di utilizzo di determinati mezzi di trasporto. Paragonando tale valutazione con ciò che è emerso dall'analisi dell'Anova a due vie, è stato possibile formulare le seguenti considerazioni:

- CLASSE_ADULTI: variabile che, nell'analisi della varianza a due vie e, in particolare, nella macro cella di FREQ_AUCO, risulta avere un effetto minore rispetto a quello fornito dalla variabile TIPO_INTERV. Quindi la frequenza di utilizzo dell'auto è influenzata maggiormente dall'effetto dato

da CATI/CAWI piuttosto che dall'effetto fornito da CLASSE_ADULTI. Tale risultato risulta esser concorde con quanto trovato precedentemente con l'analisi di regressione lineare, ovvero che TIPO_INTERV risulta esser proxy di CLASSE_ADULTI per predire la frequenza di utilizzo dell'auto. Dalla valutazione di entrambe le analisi, si evince che il fatto di esser adulti e di aver risposto ad un sondaggio web porta ad avere un aumento in termini di frequenza dell'utilizzo dell'auto. Tale relazione può esser dovuta al fatto che un maggiore utilizzo dell'auto presume, da parte degli intervistati, uno stile di vita più attivo e meno sedentario, e tale caratteristica comportamentale si può riscontrare sia negli adulti rispetto agli anziani e sia negli individui che, trascorrendo meno tempo a casa, non hanno avuto la possibilità di rispondere al telefono e, quindi, hanno risposto ad un sondaggio web;

- TP_ABBON: variabile che, nella macro cella di FREQ_TRENO dell'Anova a due vie, risulta avere un effetto non significativo al contrario della variabile TIPO_INTERV. Quindi la frequenza di utilizzo del treno è influenzata dall'effetto, ai diversi livelli della variabile TP_ABBON, riconducibile alle caratteristiche socio-economiche che a loro volta sono rappresentate da CATI/CAWI. Da un punto di vista operativo, tale risultato risulta esser concorde con quanto trovato precedentemente con l'analisi di regressione lineare, ovvero che TIPO_INTERV risulta esser proxy di TP_ABBON per predire la frequenza di utilizzo del treno. Si può dedurre, quindi, che il fatto di possedere un abbonamento ai servizi di trasporto pubblico e di

aver risposto ad un sondaggio web, porta ad avere un aumento, in termini di frequenza, dell'utilizzo del treno;

- AUCS_ABBON: variabile che, nella macro cella di FREQ_AUPA dell'Anova a due vie, risulta non avere un effetto significativo al contrario di TIPO_INTERV. Quindi la frequenza di utilizzo dell'auto come passeggero è influenzata dall'effetto dato da CATI/CAWI. Tale risultato risulta esser concorde con quanto trovato precedentemente con l'analisi di regressione lineare, ovvero che TIPO_INTERV risulta esser proxy di AUCS_ABBON per predire la frequenza di utilizzo dell'auto come passeggero. Dalla valutazione di entrambe le analisi, si evince che gli intervistati che hanno risposto ad un'indagine web e che hanno dichiarato di esser abbonati ad un servizio di car sharing, utilizzano con maggiore frequenza l'auto come passeggero. Tale relazione può esser giustificata dal fatto che gli intervistati abbonati al servizio di car sharing non hanno, presumibilmente, un'auto privata, e quindi per i loro spostamenti, o utilizzano l'auto condivisa o come passeggero, oppure, che chi ha risposto ad un sondaggio web, utilizzi internet anche per i servizi di car pooling, i quali rientrano nella categoria di autoveicoli come passeggero.

7 CONCLUSIONI E SVILUPPI FUTURI

Nel corrente capitolo conclusivo, si cerca di dare un quadro sintetico dell'analisi trattata e di dare spazio agli obiettivi di questa ricerca, cercando di riassumere e comprendere, a livello più pratico, i risultati raggiunti e i possibili sviluppi futuri.

In questo studio, i risultati delle indagini riguardanti le frequenze di utilizzo di determinati mezzi di trasporto, ottenuti utilizzando due diversi metodi di sondaggio – un tradizionale sondaggio telefonico e un sondaggio online –, sono stati confrontati rispetto alle differenze nelle caratteristiche socio-economiche del campione. Le differenze sono state testate anche valutando se il protocollo di indagine stesso influenzi le risposte date e testimoni il fatto che l'aver risposto ad un sondaggio cati o cawi vada a distinguere due gruppi di persone distinte.

Le caratteristiche socio-economiche degli intervistati sono risultate differenti in ciascun campione. Dal confronto tra intervistati online e telefonicamente si sono trovate le differenze note in letteratura che costituiscono il "digital divide": gli intervistati online sono più giovani, sono più istruiti, hanno un lavoro, possiedono la patente di categoria B, hanno un abbonamento al servizio di trasporto pubblico e non hanno limitazioni fisiche sulla guida rispetto agli intervistati del sondaggio telefonico. Tali diversità socio-demografiche hanno permesso di spiegare alcune delle differenze osservate nelle risposte alle domande attitudinali, in particolare quelle relative alla frequenza di utilizzo di determinati mezzi di trasporto.

È stato ottenuto, prima con il test t-Student e poi con l'Anova a una via, che la differenza di utilizzo di determinati mezzi di trasporto in funzione della modalità

di indagine, è risultata esser significativa. Quindi, la variabile, contenente le due modalità di intervista cati/cawi (TIPO_INTERV), risulta esser una variabile indicatore delle differenze tra le due medie di frequenza. Ovvero, il diverso protocollo di indagine influenza la risposta data per quanto concerne la frequenza di utilizzo di un determinato mezzo di trasporto. In sostanza, coloro i quali hanno risposto ad un'indagine web o telefonica, tendono ad essere due gruppi di persone distinte che utilizzano con frequenze diverse un determinato mezzo di trasporto.

Si è ottenuto che le frequenze medie di utilizzo di tutti i mezzi di trasporto considerati in funzione della variabile indicante il protocollo di indagine (TIPO_INTERV), risultano esser maggiori nel caso di risposta cawi rispetto al caso di risposta cati. Ciò testimonia che gli intervistati che hanno effettuato il sondaggio con modalità web hanno uno stile di vita più attivo e meno sedentario, in quanto essi, per l'appunto, rispetto a coloro che hanno risposto all'indagine con modalità telefonica, hanno dichiarato di utilizzare maggiormente i diversi mezzi di trasporto per i loro spostamenti. È anche possibile che tali differenze attitudinali sulle frequenze derivino dal modo in cui gli individui rispondono alla stessa domanda, ovvero se viene posta da una persona al telefono o da un computer (social desiribility responding).

Successivamente, mediante calibrazione di modelli di regressione multivariata, è stato possibile verificare che effettivamente c'è un principio di causa-effetto tra le variabili socio-economiche e le variabili indicanti le frequenze di viaggio, in quanto, quest'ultime sono influenzate non solo dalla modalità di somministrazione delle domande ma anche dalle condizioni sociali ed economiche del soggetto

intervistato. Ad esempio, gli intervistati che hanno poche auto in famiglia, sono abbonati al servizio di trasporto pubblico locale e usano maggiormente il bus come mezzo di trasporto rispetto all'auto. Invece, coloro che hanno a disposizione un buon numero di auto in famiglia, non sono abbonati al servizio di trasporto pubblico e usano maggiormente l'auto per i loro spostamenti. Oppure, gli intervistati che abitano nella 1° e 2° cintura usano l'auto con maggiore frequenza, avendo un servizio di trasporto pubblico con prestazioni inferiori rispetto a chi abita nel capoluogo piemontese. O anche, gli intervistati con figli minori preferiscono spostarsi con l'auto privata invece di usufruire del trasporto pubblico locale, in quanto l'auto può esser vista non solo come un mezzo più sicuro rispetto ai mezzi del servizio pubblico locale, ma anche un mezzo di trasporto in cui è più facile gestire e stare attenti alle esigenze dei bambini.

Inoltre, è stato rilevato che, per predire la frequenza di utilizzo di determinati mezzi di trasporto, la variabile indicante il protocollo di indagine può esser classificata come variabile potenziale proxy di alcune variabili socio-economiche. Tali variabili che son risultate poter esser sostituite dalla modalità di sondaggio sono: le variabili che indicano se l'intervistato possiede un abbonamento al servizio di trasporto pubblico locale (TP_ABBON) o al servizio di car sharing (AUCS_ABBON) e, infine, la variabile che considera gli intervistati con un range di età compreso tra i 34 e 67 anni (CLASSE_ADULTI),. Tale valutazione risulta esser molto utile, in quanto, in un ipotetico modello previsionale, la non conoscenza di tali variabili socio-economiche, può essere almeno parzialmente compensata dalle considerazioni sul protocollo di indagine. L'effetto combinato tra la variabile potenziale proxy (TIPO_INTERV) e le variabili che possono esser sostituite da essa,

hanno dimostrato che, effettivamente, esistono delle differenze significative nella frequenza di utilizzo di determinati mezzi di trasporto in funzione dei diversi livelli categorici delle variabili. Ad esempio, gli intervistati adulti che hanno risposto ad un sondaggio web usano con maggiore frequenza l'auto rispetto agli altri mezzi di trasporto. Oppure, si è ottenuto che gli intervistati che hanno risposto ad un'indagine web e che hanno dichiarato di esser abbonati ad un servizio di car sharing, utilizzano con maggiore frequenza l'auto come passeggero rispetto le altre forme di mobilità.

Nel complesso, questo studio suggerisce che i metodi di sondaggio online e telefonico hanno il potenziale per produrre risultati significativamente diversi, sia da un punto di vista descrittivo che inferenziale. Infatti, il campione che ha risposto ad un sondaggio telefonico è risultato differente dal campione che ha risposto ad un sondaggio web, in quanto, sono differenti non solo le caratteristiche socio-economiche tra i due gruppi, ma anche le caratteristiche attitudinali sulla frequenza di utilizzo di determinati mezzi di trasporto. Quindi, anche se le interviste web sono una modalità di indagine sempre più utilizzata dagli istituti di ricerca per via del loro costo economico più basso rispetto ad altre modalità di indagine, come, ad esempio, l'intervista telefonica o l'intervista "Face to Face", non si può prescindere da esse per avere un campione sempre più rappresentativo della popolazione. Inoltre, alcune tipologie di indagine, nel rappresentare al meglio la popolazione in analisi, presentano degli svantaggi che possono esser mitigati dai vantaggi di altre modalità di intervista. Per analisi di studio simili a quelle condotte in questo lavoro di tesi, si suggerisce, pertanto, di elaborare delle valutazioni su un questionario basato almeno su due modalità di sondaggio.

Questo elaborato, può esser ampliato andando a valutare con maggiore attenzione quelli che sono gli errori derivanti dal Social Desirability Responding. Ciò permetterà di fare un'ulteriore analisi di ricerca per valutare, appunto, se le risposte al sondaggio possono essere influenzate da tale tipo di errore e, inoltre, se ci sono motivi validi che hanno portato l'intervistatore a non esser veritiero nelle risposte.

Ulteriori approfondimenti possono riguardare la valutazione di altre variabili dipendenti presenti nel questionario di partenza "Demonstrate", ad esempio, il tempo di viaggio a bordo o relativo tempo di attesa dei mezzi di trasporto pubblico locale, oppure il costo complessivo del viaggio. Ciò permetterà di effettuare una valutazione complessiva sulla stessa impronta di quella effettuata in questo lavoro di tesi.

In sostanza, questo lavoro di tesi si presta a fornire ulteriori spunti sul potenziale effetto del protocollo di indagine e sul comportamento di viaggio degli individui abitanti nell'area metropolitana di Torino.

BIBLIOGRAFIA

- ADLER, T., RIMMER, L. AND CARPENTER, D. (2002). Use of internet-based household travel diary survey instrument. *Transportation Research Record*, 1804, 134-143.
- ADLER, T., RIMMER, L., BANDY, G. AND SCHELLINGER, D. (2000). Use of respondent-interactive geocoding in the Baltimore, Maryland mode choice survey. *Transportation Research Record*, 1719, 154-158.
- AL-SUBAIHI, A. A. (2008). Comparison of Web and Telephone Survey Response Rates in Saudi Arabia. *The Electronic Journal of Business Research Methods* 6: 123-132.
- BONSALL, P., J. SHIRES. (2009). Estimating the robustness of questionnaire results: lessons from a mixed-mode survey of expectations for tele-working and road-based business travel, *Transportation*, pp. 47-64.
- BRADBURN N. M. (2016). A Pioneer in Social Indicators and Quality of Life Research, *Research in Quality of Life*, Volume 11, Issue 1, pp.325-327.
- CHANG, L. C., AND J. KROSNICK. (2009). National Surveys via RDD Telephone Interviewing vs. the Internet: Comparing Sample Representativeness and Response Quality. *Public Opinion Quarterly* 73 (4): 641-678.
- CORBETTA, P. (2014). *Metodologia e tecniche della ricerca sociale, i paradigmi della ricerca sociale*. Vol 1. Seconda edizione, Il Mulino.
- COUPER, M. P. (2000). Web Surveys: A Review of Issues and Approaches, *Public Opinion Quarterly* 64: 464-494.
- DIANA, M. (2012). Internet Mobility Survey Sampling Biases in Measuring Frequency of Use of Transport Modes, *Transportation Research Record*, 2285, 66-73.
- DILLMAN D, SANGSTER R, TARNAI J, ROCKWOOD T. (1996). Understanding Differences in People's Answers to Telephone and Mail Surveys. In: Braverman M, Slater J, editors. *Advances in Survey Research*. San Francisco: Jossey-Bass; pp. 45-62.
- DILLMAN, D. A. (2000). *Mail and Internet Survey. The Tailored Design Method*. 2nd ed. New York: John Wiley and Sons.
- DRAAIJER, G., KALFS, N. AND PERDOK, J. (2000). Global positioning system as data collection method for travel research. *Transportation Research Record*, 1719, 147-153.

- FRICKER, S., M. GALESIC, R. TOURANGEAU, AND T. YAN (2005). "An Experimental Comparison of Web and Telephone Surveys." *Public Opinion Quarterly* 69: 370–392.
- GREENE, J., H. SPEIZER, AND W. WIITALA (2008). Telephone and Web: Mixed-mode Challenge. *Health Services Research* 43: 230–248.
- GUENSLER, R. AND WOLF, J. (1999). Development of a handheld electronic travel diary for monitoring individual trip making behaviour. Paper presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C.
- LEE, M.S., DOHERTY, S.T., SABETIASHRAF, R. AND MCNALLY, M. (2002). An internet computerized household activity scheduling elicitor survey. Paper presented at the 79th Annual Meeting of the Transportation Research Board, Washington, D.C.
- MARCA, J.E. (2003). The design and implementation of an on-line travel and activity survey. Paper presented at the 82nd Annual Meeting of the Transportation Research Board, Washington, D.C.
- MURAKAMI, E., WAGNER, D.P. AND NEUMEISTER, D.M. (2000). Using global positioning systems and personal digital assistants for personal travel surveys in the United States. *Transport Surveys: Raising the Standard*, Transportation Research Circular E-008, TRB, National Research Council, Washington, D.C., pp. III-B/1-21.
- SCHONLAU, M., K. ZAPERT, L. P. SIMON, K. SANSTAD, S. MARCUS, J. ADAMS, M. SPRANCA, H. KAN, R. TURNER, AND S. BERRY (2003). A Comparison between Responses from a Propensity-weighted Web Survey and an Identical RDD Survey. *Social Science Computer Review* 21: 1–11.
- STOPHER, P., BULLOCK, P. AND HORST, F. (2003). Conducting a GPS survey with a time-use diary. Paper presented at the 82nd Annual Meeting of the Transportation Research Board, Washington, D.C.
- TABACHNICK, B. G. AND FIDELL, L. S. (1989). *Using multivariate statistics* (2nd edition). New York: HarperCollins.
- TAN, A. AND TIMMERMANS, H.J.P. (2004). Paper-and-pencil retrospective activity-travel diaries versus virtual reality re-enactment sessions to collect activity-travel pattern data: A validation study. Paper presented at the 83rd Annual Meeting of the Transportation Research Board, Washington D.C.
- VEHOVAR, V., Z. BATAGELJ, AND K. LOZAR (1999). Web Surveys: Can the Weighting Solve the Problem? *Proceedings of the Survey Research Methods Section*, American Statistical Association, Alexandria, VA, pp. 962–967.
- VERREAULT, H. AND MORENCY, C. (2018) Integration of a phone-based household travel survey and a web-based student travel survey. *Transportation*. 45: 89-103.

WATSON, DOROTHY (1992). Correcting for Acquiescent Response Bias in the Absence of a Balanced Scale: An Application to Class Consciousness". *Sociological Methods & Research*. 21 (1): 52-88.

WOLF, J., MYERS, J. AND ARCE, C. (2001). Trip rate analysis in GPS-enhanced personal travel surveys. Paper presented at the 5th International Conference on Transport Survey Quality and Innovation, Kruger Park, South Africa.

XING Y., HANDY S. (2014). Online versus phone surveys: comparison of results for a bicycling survey, *Transportation Planning and Technology*, 37:6, pp. 554-567.

ALLEGATI

**Analisi del legame fra protocollo di indagine
(CATI vs. CAWI) e frequenze dichiarate d'uso dei modi di
trasporto da parte dei residenti nell'area metropolitana di Torino**

ALLEGATO 1: MODELLI DI REGRESSIONE LINEARE

Nel seguente allegato sono riportate le analisi riguardanti la calibrazione di modelli di regressione lineare multivariata. Essi hanno permesso non solo di valutare un livello di previsione della frequenza d'uso di un determinato mezzo di trasporto in funzione delle variabili indipendenti considerate, ma anche di stabilire l'eventuale condizione della variabile TIPO_INTERV come proxy di alcune variabili socio-economiche.

Tale allegato è composto da 6 paragrafi, tanti quante sono le frequenze d'utilizzo dei mezzi di trasporto considerati. Ognuno di esso, a sua volta, contiene due sotto paragrafi: nel primo e nel secondo è trattata, rispettivamente, la calibrazione del Mod.1 e del Mod.2, i quali differiscono l'uno dall'altro per la presenza/assenza della variabile TIPO_INTERV. Inoltre, alcuni paragrafi presentano un ulteriore sotto paragrafo di livello 3 in cui vengono calibrati ulteriori modelli per la valutazione della variabile socio-economica rimossa – secondo il caso preso in esame - che può esser sostituita dalla variabile candidata proxy TIPO_INTERV.

Il software SPSS Statistics, utilizzato per la calibrazione dei modelli di regressione lineare multivariata, fornisce un output costituito da tre tabelle:

- “Riepilogo del modello”: tabella in cui si è tenuto conto del coefficiente di determinazione R^2 ;
- “Coefficienti”: tabella nella quale sono stati considerati il segno, il valore e la significatività delle singole variabili indipendenti;

- “Anova”: tabella in cui è riportato il livello di significatività del modello di regressione preso in esame, la somma dei quadrati e il test F di Fischer. Nell'allegato 1 non si riporta per brevità tale tabella.

1.1 Variabile dipendente: FREQ_AUCO

1.1.1 Mod.1

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,733	,537	,513	1,56050

Coefficienti

Modello		Coefficienti non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore standard	Beta		
1	(Costante)	1,311	,370		3,542	,000
	CLASSE_GIOVANI	,904	,378	,166	2,393	,017
	CLASSE_ADULTI	,615	,306	,137	2,011	,045
	FAM_REDDITO	-,050	,102	-,024	-,484	,629
	GENERE	,957	,182	,214	5,269	,000
	CLASSE_LICENZAMEDIASUP	-,076	,280	-,017	-,270	,787
	CLASSE_LAUREA	,294	,333	,055	,884	,377
	SALUTE	-,497	,323	-,070	-1,541	,124
	PATENTE	,791	,285	,143	2,777	,006
	DIMORA	-,332	,182	-,073	-1,826	,069
	PARK_CASA	,668	,244	,132	2,737	,007
	TP_ABBON	-1,167	,213	-,219	-5,466	,000
	AUCS_ABBON	,331	,823	,016	,402	,688
	FAM_N	-,880	,160	-,398	-5,509	,000
	FAM_FIGLI	,575	,240	,182	2,396	,017
	FAM_MIN	,631	,236	,160	2,672	,008
	FAM_AUTO	1,103	,195	,369	5,660	,000
	FAM_MOTO	-,419	,220	-,079	-1,907	,057

1.1.2 Mod.2

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,735	,541	,516	1,55634

Coefficienti

Modello		Coefficients non standardizzati		Coefficients standardizzati	t	Sign.
		B	Errore standard	Beta		
1	(Costante)	1,338	,370		3,622	,000
	CLASSE_GIOVANI	,865	,378	,159	2,292	,023
	CLASSE_ADULTI	,566	,306	,126	1,845	,066
	FAM_REDDITO	-,060	,102	-,029	-,589	,556
	GENERE	,962	,181	,215	5,307	,000
	CLASSE_LICENZAMEDIASUP	-,084	,279	-,019	-,300	,764
	CLASSE_LAUREA	,279	,332	,052	,841	,401
	SALUTE	-,476	,322	-,067	-1,480	,140
	PATENTE	,733	,286	,133	2,563	,011
	DIMORA	-,341	,182	-,075	-1,881	,061
	PARK_CASA	,735	,247	,146	2,979	,003
	TP_ABBON	-1,222	,215	-,229	-5,672	,000
	AUCS_ABBON	-,016	,847	-,001	-,019	,985
	FAM_N	-,892	,160	-,404	-5,593	,000
	FAM_FIGLI	,573	,239	,182	2,395	,017
	FAM_MIN	,608	,236	,154	2,579	,010
	FAM_AUTO	1,124	,195	,376	5,768	,000
	FAM_MOTO	-,414	,219	-,078	-1,886	,060
	TIPO_INTERV	,747	,327	,139	1,959	,048

1.1.3 Mod. alternativo: Caso 3

Variabile rimossa: CLASSE_GIOVANI

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,717	,514	,501	1,57975

Coefficienti

Modello		Coefficients non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore standard	Beta		
1	(Costante)	,961	,285		3,377	,001
	GENERE	,843	,180	,189	4,680	,000
	PATENTE	1,033	,274	,187	3,771	,000
	PARK_CASA	,856	,245	,170	3,493	,001
	TP_ABBON	-1,269	,213	-,238	-5,946	,000
	FAM_N	-,875	,142	-,396	-6,154	,000
	FAM_FIGLI	,451	,211	,143	2,138	,033
	FAM_MIN	,762	,233	,193	3,263	,001
	FAM_AUTO	1,216	,182	,407	6,670	,000
	TIPO_INTERV	,569	,261	,088	2,185	,030

Variabile rimossa: GENERE

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,696	,485	,471	1,62646

Coefficienti

Modello		Coefficients non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore standard	Beta		
1	(Costante)	1,332	,305		4,368	,000
	PATENTE	1,297	,276	,235	4,703	,000
	PARK_CASA	,971	,251	,192	3,864	,000
	TP_ABBON	-1,288	,220	-,241	-5,843	,000
	FAM_N	-,995	,162	-,450	-6,156	,000
	FAM_FIGLI	,484	,245	,153	1,978	,049
	FAM_MIN	,836	,240	,212	3,487	,001
	FAM_AUTO	1,179	,188	,394	6,264	,000
	TIPO_INTERV	,519	,268	,080	1,934	,054
	CLASSE_GIOVANI	,354	,271	,065	1,303	,193

Variabile rimossa: PATENTE

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,705	,496	,483	1,60817

Coefficienti

Modello		Coefficients non standardizzati		Coefficients standardizzati	t	Sign.
		B	Errore standard	Beta		
		1	(Costante)	1,434		
	PARK_CASA	1,116	,239	,221	4,663	,000
	TP_ABBON	-1,319	,218	-,247	-6,058	,000
	FAM_N	-1,024	,159	-,463	-6,432	,000
	FAM_FIGLI	,675	,242	,214	2,785	,006
	FAM_MIN	,765	,238	,194	3,216	,001
	FAM_AUTO	1,417	,176	,474	8,048	,000
	TIPO_INTERV	,731	,261	,113	2,798	,005
	CLASSE_GIOVANI	,378	,268	,069	1,409	,160
	GENERE	,986	,179	,221	5,503	,000

Variabile rimossa: PARK_CASA

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,706	,499	,485	1,60423

Coefficienti

Modello		Coefficients non standardizzati		Coefficients standardizzati	t	Sign.
		B	Errore standard	Beta		
		1	(Costante)	1,167		
	TP_ABBON	-1,197	,216	-,224	-5,547	,000
	FAM_N	-,931	,159	-,421	-5,840	,000
	FAM_FIGLI	,552	,242	,175	2,277	,023
	FAM_MIN	,747	,237	,189	3,149	,002
	FAM_AUTO	1,380	,178	,461	7,756	,000

TIPO_INTERV	,382	,260	,059	1,468	,143
CLASSE_GIOVANI	,347	,268	,064	1,298	,195
GENERE	,905	,182	,203	4,970	,000
PATENTE	1,294	,267	,234	4,847	,000

Variabile rimossa: TP_ABBON

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,681	,464	,449	1,65934

Coefficienti

Modello		Coefficienti non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore standard	Beta		
1	(Costante)	,941	,315		2,991	,003
	FAM_N	-1,077	,164	-,487	-6,570	,000
	FAM_FIGLI	,604	,251	,191	2,405	,017
	FAM_MIN	,928	,244	,235	3,809	,000
	FAM_AUTO	1,379	,189	,461	7,283	,000
	TIPO_INTERV	,289	,270	,045	1,069	,286
	CLASSE_GIOVANI	,226	,276	,042	,820	,413
	GENERE	,839	,189	,188	4,433	,000
	PATENTE	1,082	,288	,196	3,758	,000
	PARK_CASA	,670	,256	,133	2,622	,009

Variabile rimossa: FAM_N

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,680	,462	,448	1,66204

Coefficienti

Modello		Coefficients non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore standard	Beta		
1	(Costante)	-,188	,228		-,823	,411
	FAM_FIGLI	-,168	,214	-,053	-,785	,433
	FAM_MIN	,677	,246	,171	2,756	,006
	FAM_AUTO	,688	,171	,230	4,013	,000
	TIPO_INTERV	,483	,274	,075	1,761	,079
	CLASSE_GIOVANI	-,342	,251	-,063	-1,363	,174
	GENERE	,890	,189	,199	4,701	,000
	PATENTE	1,196	,287	,217	4,167	,000
	PARK_CASA	,767	,258	,152	2,978	,003
	TP_ABBON	-1,448	,224	-,271	-6,476	,000

Variabile rimossa: FAM_FIGLI

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,712	,507	,494	1,59044

Coefficienti

Modello		Coefficients non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore standard	Beta		
1	(Costante)	,819	,283		2,898	,004
	FAM_MIN	1,088	,191	,275	5,683	,000
	FAM_AUTO	1,190	,184	,398	6,467	,000
	TIPO_INTERV	,581	,262	,090	2,213	,028
	CLASSE_GIOVANI	,037	,234	,007	,156	,876
	GENERE	,796	,180	,178	4,409	,000
	PATENTE	1,081	,275	,196	3,930	,000
	PARK_CASA	,822	,246	,163	3,333	,001
	TP_ABBON	-1,293	,216	-,242	-5,996	,000
	FAM_N	-,757	,135	-,343	-5,614	,000

Variabile rimossa: FAM_MIN

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,708	,502	,488	1,59965

Coefficienti

Modello		Coefficients non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore standard	Beta		
1	(Costante)	1,054	,304		3,466	,001
	FAM_AUTO	1,153	,185	,385	6,244	,000
	TIPO_INTERV	,585	,264	,091	2,217	,027
	CLASSE_GIOVANI	,386	,266	,071	1,450	,148
	GENERE	,889	,182	,199	4,888	,000
	PATENTE	1,040	,278	,188	3,746	,000
	PARK_CASA	,856	,248	,170	3,449	,001
	TP_ABBON	-1,378	,215	-,258	-6,409	,000
	FAM_N	-,939	,159	-,425	-5,903	,000
	FAM_FIGLI	1,042	,197	,330	5,298	,000

Variabile rimossa: FAM_AUTO

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,674	,454	,440	1,67387

Coefficienti

Modello		Coefficients non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore standard	Beta		
1	(Costante)	,788	,315		2,504	,013
	TIPO_INTERV	,568	,276	,088	2,055	,041
	CLASSE_GIOVANI	,483	,278	,089	1,736	,083
	GENERE	,818	,191	,183	4,282	,000
	PATENTE	1,603	,275	,290	5,831	,000
	PARK_CASA	1,316	,249	,261	5,288	,000
	TP_ABBON	-1,523	,224	-,285	-6,807	,000

FAM_N	-,496	,148	-,224	-3,341	,001
FAM_FIGLI	,576	,253	,183	2,277	,023
FAM_MIN	,632	,247	,160	2,560	,011

1.2 Variabile dipendente: FREQ_BICI

1.2.1 Mod.1

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,409	,167	,124	1,05112

Coefficienti

Modello		Coefficients non standardizzati		Coefficients standardizzati	t	Sign.
		B	Errore standard	Beta		
1	CLASSE_GIOVANI	,031	,254	,013	,122	,903
	CLASSE_ADULTI	-,019	,202	-,013	-,095	,924
	FAM_REDDITO	,002	,068	,005	,030	,976
	GENERE	,004	,122	,003	,037	,971
	CLASSE_LICENZAMEDIASUP	,150	,187	,097	,799	,425
	CLASSE_LAUREA	,043	,224	,018	,191	,849
	SALUTE	-,139	,192	-,041	-,720	,472
	PATENTE	,087	,184	,069	,471	,638
	DIMORA	-,018	,118	-,012	-,154	,878
	PARK_CASA	-,154	,164	-,117	-,939	,349
	TP_ABBON	-,102	,144	-,043	-,712	,477
	AUCS_ABBON	1,244	,555	,119	2,243	,026
	FAM_N	,102	,091	,243	1,117	,265
	FAM_FIGLI	-,132	,150	-,098	-,878	,380
	FAM_MIN	-,046	,159	-,025	-,289	,773
	FAM_AUTO	,103	,130	,140	,794	,428
	FAM_MOTO	,183	,147	,075	1,243	,215

1.2.2 Mod.2

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,419	,176	,130	1,04718

Coefficienti

Modello		Coefficienti non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore standard	Beta		
1	CLASSE_GIOVANI	,001	,254	,000	,002	,998
	CLASSE_ADULTI	-,053	,202	-,035	-,263	,793
	FAM_REDDITO	-,005	,068	-,012	-,076	,939
	GENERE	,009	,122	,005	,072	,943
	CLASSE_LICENZAMEDIASUP	,145	,187	,094	,779	,437
	CLASSE_LAUREA	,032	,223	,013	,144	,886
	SALUTE	-,114	,192	-,034	-,596	,552
	PATENTE	,048	,184	,038	,260	,795
	DIMORA	-,022	,117	-,015	-,191	,849
	PARK_CASA	-,102	,166	-,078	-,617	,538
	TP_ABBON	-,145	,145	-,061	-,998	,319
	AUCS_ABBON	,983	,570	,094	1,724	,086
	FAM_N	,097	,091	,233	1,074	,283
	FAM_FIGLI	-,138	,150	-,102	-,924	,356
	FAM_MIN	-,063	,159	-,035	-,395	,693
	FAM_AUTO	,117	,130	,159	,902	,368
	FAM_MOTO	,186	,147	,076	1,268	,206
TIPO_INTERV	,338	,181	,112	1,863	,063	

1.3 Variabile dipendente: FREQ_AUPA

1.3.1 Mod.1

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,354	,125	,080	,98059

		Coefficients				
		Coefficients non standardizzati		Coefficients standardizzati		
Modello		B	Errore standard	Beta	t	Sign.
1	(Costante)	,474	,233		2,039	,042
	CLASSE_GIOVANI	,222	,237	,089	,935	,350
	CLASSE_ADULTI	,205	,192	,100	1,066	,287
	FAM_REDDITO	,104	,064	,111	1,621	,106
	GENERE	-,159	,114	-,078	-1,393	,164
	CLASSE_LICENZAMEDIASUP	-,039	,176	-,019	-,224	,823
	CLASSE_LAUREA	-,065	,209	-,026	-,309	,758
	SALUTE	,164	,203	,050	,811	,418
	PATENTE	-,778	,179	-,308	-4,349	,000
	DIMORA	,023	,114	,011	,197	,844
	PARK_CASA	,362	,153	,157	2,361	,019
	TP_ABBON	,121	,134	,049	,900	,369
	AUCS_ABBON	1,041	,517	,109	2,012	,045
	FAM_N	,066	,100	,066	,662	,508
	FAM_FIGLI	-,291	,151	-,202	-1,927	,055
	FAM_MIN	,220	,148	,122	1,484	,139
	FAM_AUTO	-,099	,123	-,073	-,811	,418
	FAM_MOTO	,024	,138	,010	,173	,863

1.3.2 Mod.2

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,475	,225	,182	,92428

		Coefficients non standardizzati		Coefficients standardizzati		
Modello		B	Errore standard	Beta	t	Sign.
1	(Costante)	,538	,219		2,451	,015
	CLASSE_GIOVANI	,132	,224	,053	,590	,555
	CLASSE_ADULTI	,090	,182	,044	,493	,622
	FAM_REDDITO	,079	,061	,084	1,307	,192
	GENERE	-,149	,108	-,073	-1,381	,168
	CLASSE_LICENZAMEDIASUP	-,058	,166	-,029	-,353	,725

CLASSE_LAUREA	-,100	,197	-,041	-,507	,613
SALUTE	,213	,191	,065	1,113	,266
PATENTE	-,911	,170	-,361	-5,364	,000
DIMORA	,001	,108	,000	,009	,993
PARK_CASA	,518	,147	,224	3,535	,000
TP_ABBON	-,008	,128	-,003	-,066	,947
AUCS_ABBON	,236	,503	,025	,468	,640
FAM_N	,039	,095	,038	,410	,682
FAM_FIGLI	-,295	,142	-,205	-2,075	,039
FAM_MIN	,168	,140	,093	1,197	,232
FAM_AUTO	-,052	,116	-,038	-,452	,652
FAM_MOTO	,037	,130	,015	,281	,779
TIPO_INTERV	1,039	,160	,352	6,486	,000

1.3.3 Mod. alternativo: Caso 3

Variabile rimossa: PATENTE

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,277	,077	,071	,98509

Coefficienti

Modello		Coefficients non standardizzati		Coefficients standardizzati	t	Sign.
		B	Errore standard	Beta		
1	(Costante)	,292	,106		2,756	,006
	PARK_CASA	,053	,120	,023	,438	,662
	TIPO_INTERV	,818	,154	,277	5,330	,000

Variabile rimossa: PARK_CASA

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,401	,161	,156	,93923

Coefficienti

Modello		Coefficienti non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore standard	Beta		
		1	(Costante)	,903		
	TIPO_INTERV	,956	,148	,324	6,455	,000
	PATENTE	-,744	,127	-,295	-5,867	,000

1.4 Variabile dipendente: FREQ_BUS

1.4.1 Mod.1

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,769	,592	,570	1,67398

Coefficienti

Modello		Coefficienti non standardizzati		Coefficienti standardizzati	t	Sign.
		B	Errore standard	Beta		
		1	CLASSE_GIOVANI	-,112		
	CLASSE_ADULTI	,410	,321	,119	1,277	,203
	FAM_REDDITO	,015	,109	,015	,139	,889
	GENERE	-,344	,194	-,094	-1,773	,077
	CLASSE_LICENZAMEDIASUP	,059	,298	,017	,199	,842
	CLASSE_LAUREA	,118	,356	,022	,332	,740
	SALUTE	-,230	,306	-,030	-,750	,454
	PATENTE	,778	,293	,272	2,661	,008
	DIMORA	,651	,188	,194	3,466	,001
	PARK_CASA	-,386	,261	-,129	-1,478	,140
	TP_ABBON	1,878	,229	,350	8,206	,000
	AUCS_ABBON	-1,314	,883	-,055	-1,488	,138
	FAM_N	,658	,145	,693	4,543	,000
	FAM_FIGLI	-,776	,239	-,252	-3,241	,001
	FAM_MIN	-,206	,253	-,050	-,815	,416
	FAM_AUTO	-,591	,207	-,353	-2,853	,005
	FAM_MOTO	-,442	,235	-,080	-1,884	,060

1.4.2 Mod.2

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,769	,592	,569	1,67651

Coefficienti

Modello		Coefficients non standardizzati		Coefficients standardizzati	t	Sign.
		B	Errore standard	Beta		
1	CLASSE_GIOVANI	-,109	,406	-,020	-,268	,789
	CLASSE_ADULTI	,413	,323	,120	1,279	,202
	FAM_REDDITO	,016	,109	,016	,145	,885
	GENERE	-,345	,195	-,094	-1,772	,077
	CLASSE_LICENZAMEDIASUP	,060	,299	,017	,200	,841
	CLASSE_LAUREA	,119	,357	,022	,334	,739
	SALUTE	-,232	,307	-,030	-,754	,451
	PATENTE	,782	,295	,273	2,652	,008
	DIMORA	,651	,188	,194	3,462	,001
	PARK_CASA	-,391	,265	-,131	-1,472	,142
	TP_ABBON	1,882	,232	,350	8,109	,000
	AUCS_ABBON	-1,291	,913	-,054	-1,414	,158
	FAM_N	,658	,145	,693	4,537	,000
	FAM_FIGLI	-,775	,240	-,252	-3,232	,001
	FAM_MIN	-,205	,254	-,050	-,806	,421
	FAM_AUTO	-,592	,208	-,354	-2,850	,005
	FAM_MOTO	-,443	,235	-,080	-1,882	,061
TIPO_INTERV	-,031	,290	-,004	-,105	,916	

1.5 Variabile dipendente: FREQ_METRO

1.5.1 Mod.1

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,617	,381	,349	1,35934

		Coefficients				
		Coefficients non standardizzati		Coefficients standardizzati		
Modello		B	Errore standard	Beta	t	Sign.
1	CLASSE_GIOVANI	-,163	,329	-,045	-,495	,621
	CLASSE_ADULTI	-,101	,261	-,045	-,388	,698
	FAM_REDDITO	,027	,088	,040	,303	,762
	GENERE	-,283	,158	-,118	-1,796	,073
	CLASSE_LICENZAMEDIASUP	,546	,242	,235	2,254	,025
	CLASSE_LAUREA	,564	,289	,157	1,950	,052
	SALUTE	-,490	,249	-,097	-1,972	,049
	PATENTE	-,039	,238	-,021	-,164	,870
	DIMORA	,199	,152	,090	1,304	,193
	PARK_CASA	-,358	,212	-,182	-1,689	,092
	TP_ABBON	1,237	,186	,349	6,653	,000
	AUCS_ABBON	,515	,717	,033	,718	,473
	FAM_N	,173	,118	,277	1,473	,142
	FAM_FIGLI	-,133	,194	-,066	-,685	,494
	FAM_MIN	-,027	,206	-,010	-,130	,896
	FAM_AUTO	,099	,168	,090	,588	,557
	FAM_MOTO	-,066	,191	-,018	-,347	,729

1.5.2 Mod. 2

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,618	,382	,348	1,35937

		Coefficients				
		Coefficients non standardizzati		Coefficients standardizzati		
Modello		B	Errore standard	Beta	t	Sign.
1	CLASSE_GIOVANI	-,184	,329	-,051	-,558	,577
	CLASSE_ADULTI	-,125	,262	-,055	-,476	,634
	FAM_REDDITO	,022	,088	,032	,246	,806
	GENERE	-,280	,158	-,117	-1,778	,076
	CLASSE_LICENZAMEDIASUP	,543	,242	,233	2,241	,026

CLASSE_LAUREA	,557	,290	,155	1,923	,055
SALUTE	-,474	,249	-,093	-1,901	,058
PATENTE	-,066	,239	-,035	-,275	,784
DIMORA	,196	,152	,089	1,284	,200
PARK_CASA	-,322	,215	-,164	-1,499	,135
TP_ABBON	1,207	,188	,341	6,415	,000
AUCS_ABBON	,334	,740	,021	,452	,652
FAM_N	,170	,118	,272	1,448	,149
FAM_FIGLI	-,138	,194	-,068	-,707	,480
FAM_MIN	-,038	,206	-,014	-,186	,852
FAM_AUTO	,108	,168	,098	,644	,520
FAM_MOTO	-,064	,191	-,018	-,337	,737
TIPO_INTERV	,233	,235	,052	,992	,322

1.6 Variabile dipendente: FREQ_METRO

1.6.1 Mod.1

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,387	,149	,105	,47360

Coefficienti

Modello		Coefficients non standardizzati		Coefficients standardizzati	t	Sign.
		B	Errore standard	Beta		
1	CLASSE_GIOVANI	,073	,114	,068	,638	,524
	CLASSE_ADULTI	,066	,091	,098	,728	,467
	FAM_REDDITO	,020	,031	,101	,654	,513
	GENERE	,033	,055	,047	,607	,544
	CLASSE_LICENZAMEDIASUP	,019	,084	,028	,230	,818
	CLASSE_LAUREA	,080	,101	,075	,789	,431
	SALUTE	-,017	,087	-,011	-,195	,846
	PATENTE	,082	,083	,147	,995	,320
	DIMORA	-,048	,053	-,073	-,902	,368
	PARK_CASA	-,229	,074	-,391	-3,094	,002
	TP_ABBON	,144	,065	,137	2,225	,027
	AUCS_ABBON	-,027	,250	-,006	-,107	,915
	FAM_N	-,001	,041	-,005	-,024	,981
	FAM_FIGLI	-,028	,068	-,046	-,410	,682
	FAM_MIN	,234	,072	,291	3,274	,001
	FAM_AUTO	,005	,059	,015	,083	,934
	FAM_MOTO	,008	,066	,007	,116	,907

1.6.2 Mod.2

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,513	,263	,222	,44154

Coefficienti

Modello		Coefficients non standardizzati		Coefficients standardizzati	t	Sign.
		B	Errore standard	Beta		
1	CLASSE_GIOVANI	,024	,107	,022	,226	,821
	CLASSE_ADULTI	,012	,085	,017	,137	,891
	FAM_REDDITO	,009	,029	,043	,296	,767
	GENERE	,040	,051	,056	,784	,434
	CLASSE_LICENZAMEDIASUP	,012	,079	,018	,157	,875
	CLASSE_LAUREA	,062	,094	,059	,664	,507
	SALUTE	,022	,081	,015	,270	,787
	PATENTE	,020	,078	,036	,260	,795
	DIMORA	-,055	,050	-,083	-1,105	,270
	PARK_CASA	-,146	,070	-,249	-2,084	,038
	TP_ABBON	,076	,061	,072	1,245	,214
	AUCS_ABBON	-,447	,240	-,096	-1,859	,064
	FAM_N	-,008	,038	-,041	-,202	,840
	FAM_FIGLI	-,038	,063	-,063	-,602	,548
	FAM_MIN	,208	,067	,258	3,106	,002
	FAM_AUTO	,027	,055	,082	,493	,622
	FAM_MOTO	,012	,062	,011	,201	,841
	TIPO_INTERV	,543	,076	,404	7,097	,000

1.6.3 Mod. alternativo: Caso 3

Variabile rimossa: PARK_CASA

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,485	,235	,230	,43925

Coefficienti

Modello		Coefficients non standardizzati		Coefficients standardizzati	t	Sign.
		B	Errore standard	Beta		
		1	FAM_MIN	,166		
	TIPO_INTERV	,518	,066	,386	7,847	,000

Variabile rimossa: FAM_MIN

Riepilogo del modello

Modello	R	R-quadrato	R-quadrato adattato	Errore std. della stima
1	,442	,196	,191	,45039

Coefficienti

Modello		Coefficients non standardizzati		Coefficients standardizzati	t	Sign.
		B	Errore standard	Beta		
		1	TIPO_INTERV	,595		
	PARK_CASA	-,002	,030	-,004	-,073	,942

ALLEGATO 2: ANALISI DELLE VARIANZE A DUE VIE

Nell'allegato 2 sono presenti gli output ottenuti con l'Anova a due vie. Tale analisi valuta un eventuale effetto d'interazione di una variabile indipendente su una variabile dipendente a un livello specifico di un'altra variabile indipendente. Per tale motivo, sono presenti dei paragrafi, indicanti le variabili indipendenti considerate, e dei sotto paragrafi, per indicare, invece, la variabile dipendente valutata.

Ogni analisi è composta da tre tabelle e due grafici:

- “Stime”: prima tabella in cui sono stati considerati i valori delle medie di frequenza a seconda dei diversi livelli categorici delle variabili indipendenti;
- “Confronti pairwise”: seconda tabella che tiene conto della significatività della variabile TIPO_INTERV sui diversi livelli categorici dell'altra variabile indipendente considerata;
- “Confronti pairwise”: terza tabella che tiene conto della significatività dell'effetto della variabile indipendente considerata sui diversi livelli categorici della variabile TIPO_INTERV;
- “Medie marginali stimate di FREQ_”: primo grafico in cui sono rappresentati graficamente i valori di frequenza media ai diversi livelli delle variabili indipendenti. Tale grafico tiene conto dell'effetto dato dalla variabile TIPO_INTERV;

- “Medie marginali stimate di `FREQ_`”: secondo grafico in cui sono rappresentati graficamente i valori di frequenza media ai diversi livelli delle variabili indipendenti. Tale grafico tiene conto dell’effetto dato dalle variabili indipendente potenzialmente sostituibili da `TIPO_INTERV`.

Per le variabili indipendenti `TP_ABBON`, `CLASSE_ADULTI` e `AUCS_ABBON` si è considerata una codifica dummy ponendo il valore 1 per la risposta affermativa e il valore 0 per la risposta negativa.

2.2 Variabili indipendenti: TIPO_INTERV e TP_ABBON

2.2.1 Variabile dipendente: FREQ_AUCO

Stime

Variabile dipendente: FREQ_AUCO

TIPO_INTERV	TP_ABBON	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	2,707	,138	2,436	2,978
	1,00	,992	,274	,454	1,530
CAWIC	0	3,600	,387	2,839	4,361
	1,00	1,889	,500	,906	2,871

Confronti pairwise

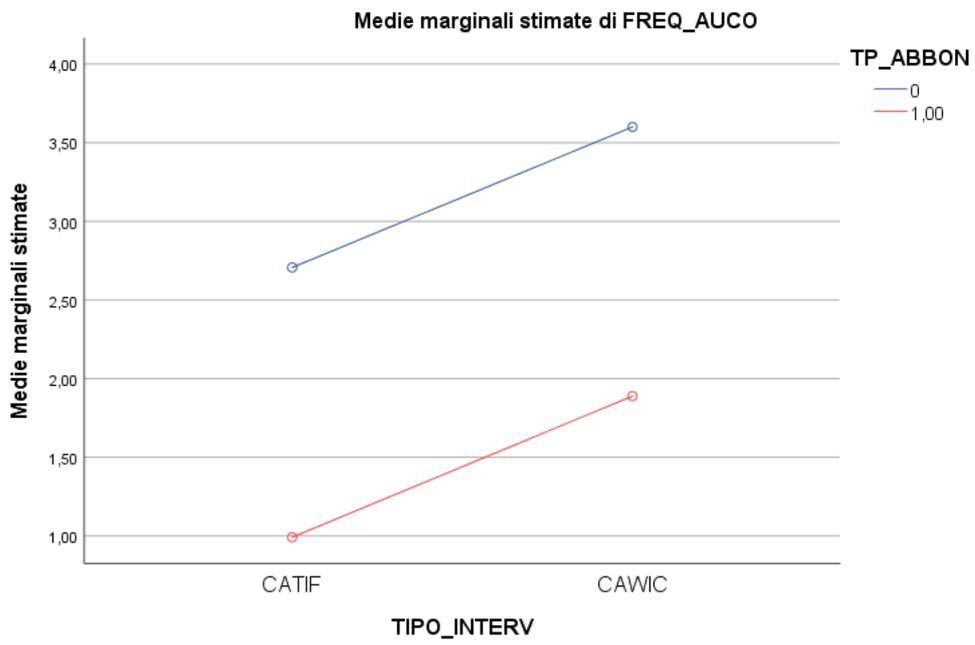
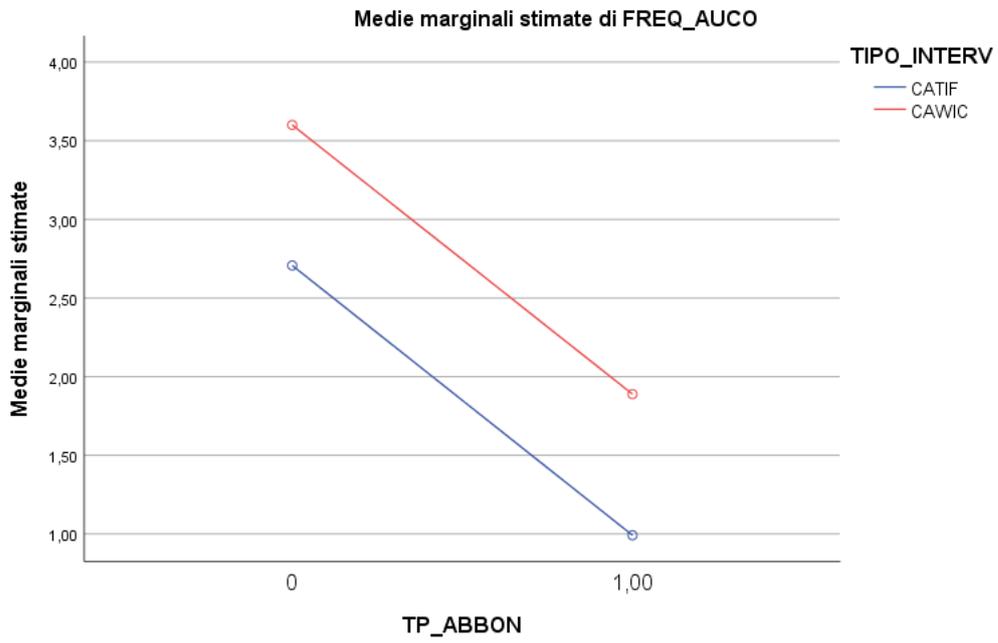
Variabile dipendente: FREQ_AUCO

TP_ABBON	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,893	,411	,030	-1,701	-,085
	CAWIC	CATIF	,893	,411	,030	,085	1,701
1,00	CATIF	CAWIC	-,897	,570	,116	-2,017	,223
	CAWIC	CATIF	,897	,570	,116	-,223	2,017

Confronti pairwise

Variabile dipendente: FREQ_AUCO

TIPO_INTERV	TP_ABBON	TP_ABBON	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1,00	1,715	,306	,000	1,113	2,318
	1,00	0	-1,715	,306	,000	-2,318	-1,113
CAWIC	0	1,00	1,711	,632	,007	,468	2,954
	1,00	0	-1,711	,632	,007	-2,954	-,468



2.2.2 Variabile dipendente: **FREQ_BICI**

Stime

Variabile dipendente: FREQ_BICI

TIPO_INTERV	TP_ABBON	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	,323	,069	,188	,458
	1,00	,225	,136	-,043	,493
CAWIC	0	,850	,193	,471	1,229
	1,00	,583	,249	,094	1,072

Confronti pairwise

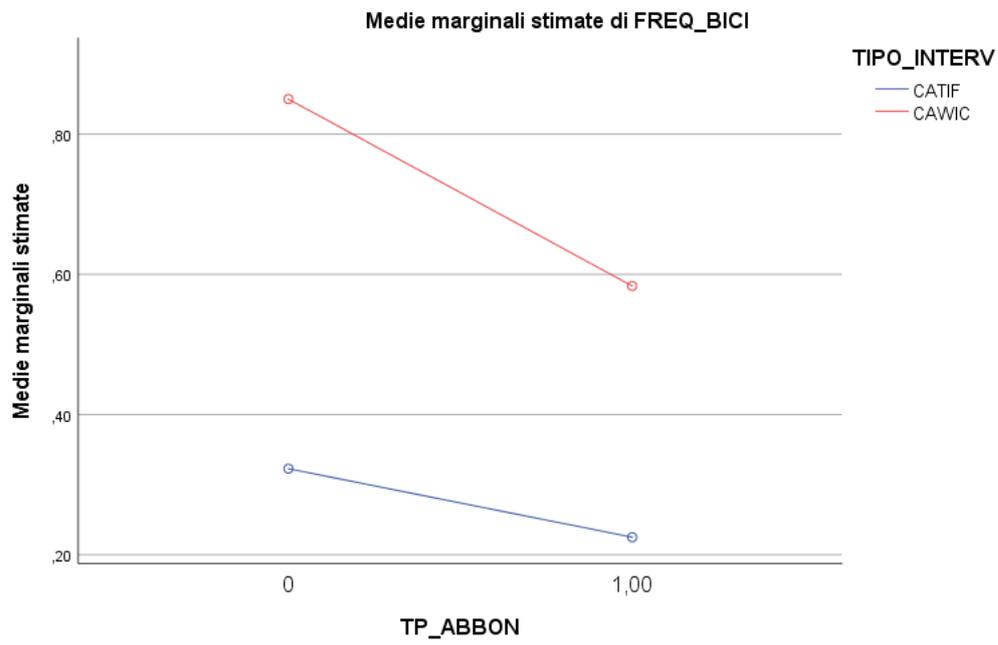
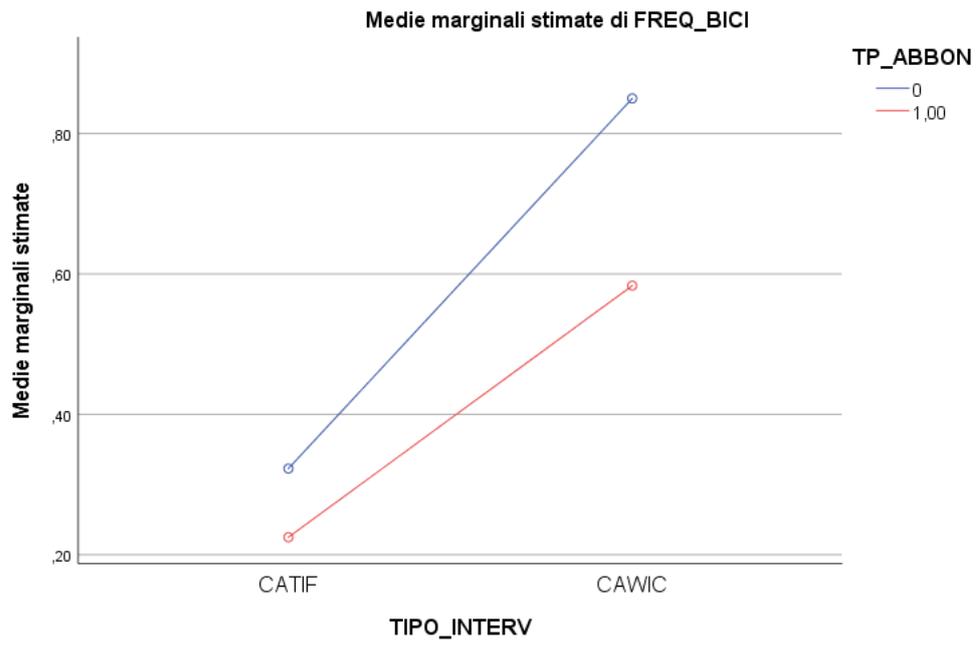
Variabile dipendente: FREQ_BICI

TP_ABBON	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,527	,204	,010	-,929	-,125
	CAWIC	CATIF	,527	,204	,010	,125	,929
1,00	CATIF	CAWIC	-,358	,283	,207	-,916	,199
	CAWIC	CATIF	,358	,283	,207	-,199	,916

Confronti pairwise

Variabile dipendente: FREQ_BICI

TIPO_INTERV	TP_ABBON	TP_ABBON	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1,00	,098	,152	,522	-,202	,398
	1,00	0	-,098	,152	,522	-,398	,202
CAWIC	0	1,00	,267	,315	,397	-,352	,885
	1,00	0	-,267	,315	,397	-,885	,352



2.2.3 Variabile dipendente: **FREQ_AUPA**

Stime

Variabile dipendente: FREQ_AUPA

TIPO_INTERV	TP_ABBON	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	,338	,063	,213	,462
	1,00	,308	,126	,060	,556
CAWIC	0	,850	,178	,499	1,201
	1,00	1,639	,230	1,186	2,092

Confronti pairwise

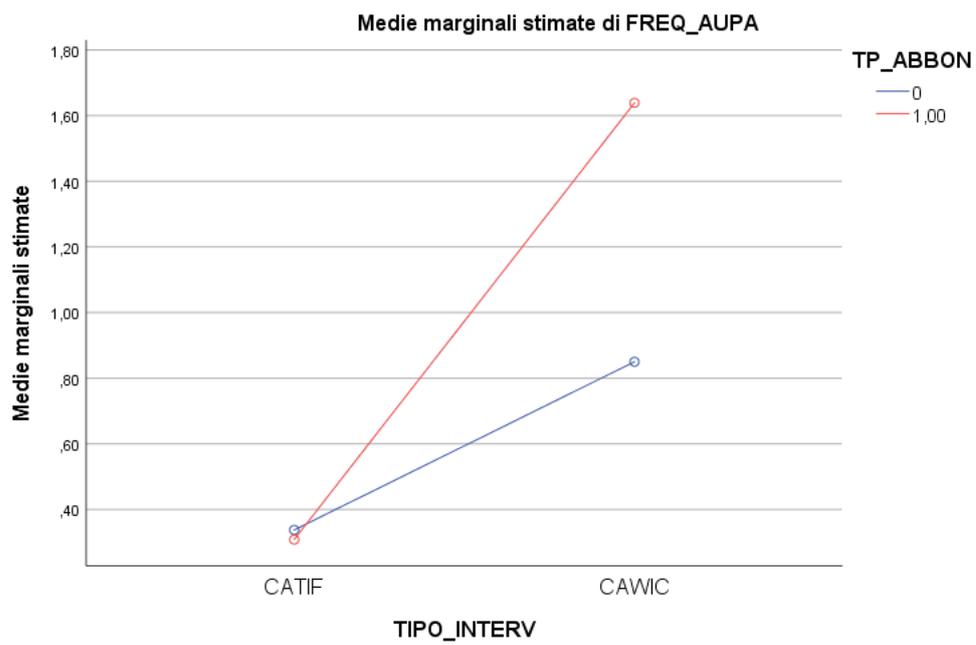
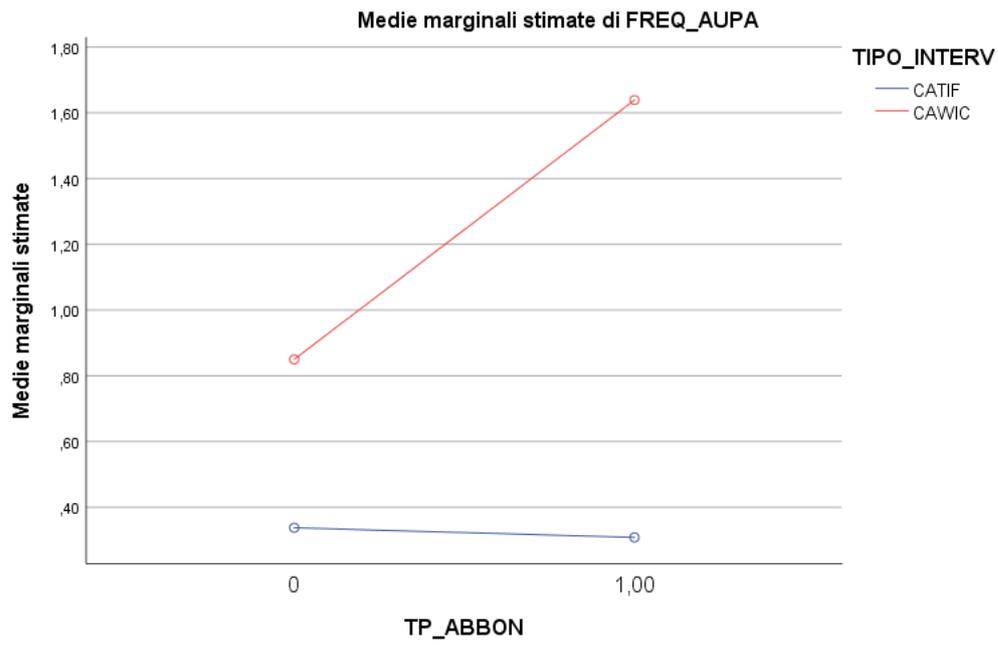
Variabile dipendente: FREQ_AUPA

TP_ABBON	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,512	,189	,007	-,885	-,140
	CAWIC	CATIF	,512	,189	,007	,140	,885
1,00	CATIF	CAWIC	-1,331	,262	,000	-1,847	-,814
	CAWIC	CATIF	1,331	,262	,000	,814	1,847

Confronti pairwise

Variabile dipendente: FREQ_AUPA

TIPO_INTERV	TP_ABBON	TP_ABBON	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1,00	,029	,141	,836	-,248	,307
	1,00	0	-,029	,141	,836	-,307	,248
CAWIC	0	1,00	-,789*	,291	,007	-1,361	-,216
	1,00	0	,789*	,291	,007	,216	1,361



2.2.4 Variabile dipendente: **FREQ_BUS**

Stime

Variabile dipendente: **FREQ_BUS**

TIPO_INTERV	TP_ABBON	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	1,131	,114	,907	1,354
	1,00	3,333	,226	2,889	3,778
CAWIC	0	1,283	,319	,655	1,912
	1,00	3,222	,412	2,411	4,034

Confronti pairwise

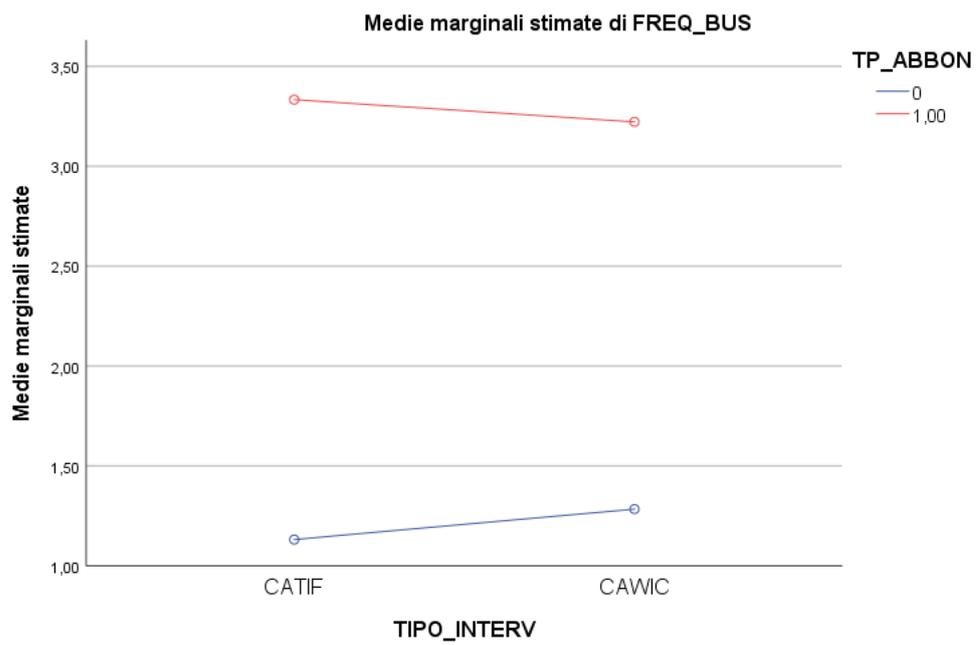
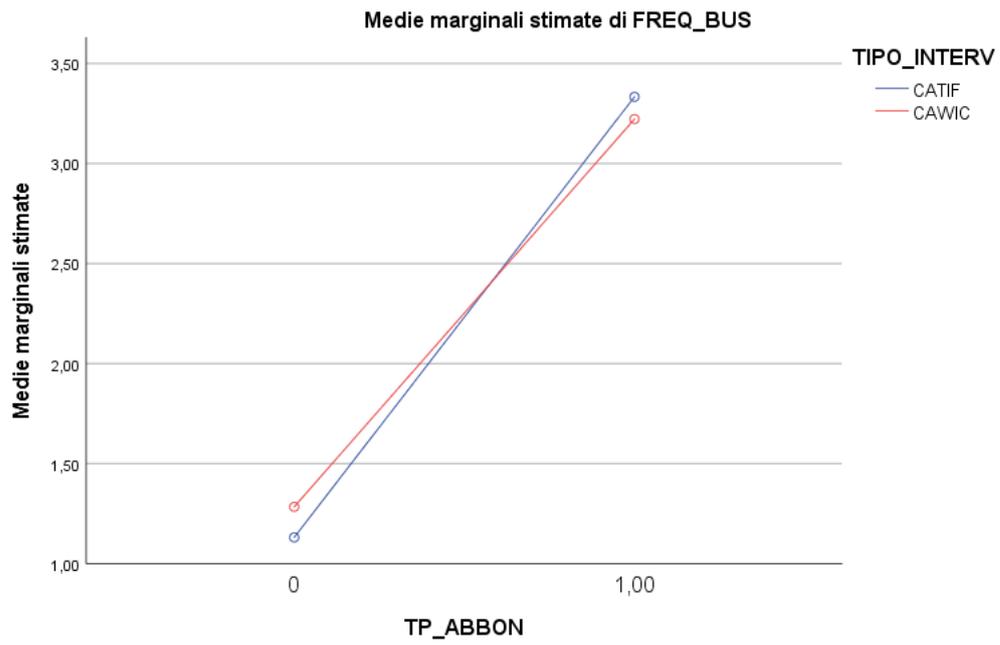
Variabile dipendente: **FREQ_BUS**

TP_ABBON	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,153	,339	,653	-,820	,514
	CAWIC	CATIF	,153	,339	,653	-,514	,820
1,00	CATIF	CAWIC	,111	,470	,813	-,814	1,036
	CAWIC	CATIF	-,111	,470	,813	-1,036	,814

Confronti pairwise

Variabile dipendente: **FREQ_BUS**

TIPO_INTERV	TP_ABBON	TP_ABBON	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1,00	-2,203	,253	,000	-2,700	-1,705
	1,00	0	2,203	,253	,000	1,705	2,700
CAWIC	0	1,00	-1,939	,522	,000	-2,965	-,913
	1,00	0	1,939	,522	,000	,913	2,965



2.2.5 Variabile dipendente: FREQ_METRO

Stime

Variabile dipendente: FREQ_METRO

TIPO_INTERV	TP_ABBON	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	,441	,090	,264	,618
	1,00	1,650	,179	1,298	2,002
CAWIC	0	,900	,253	,403	1,397
	1,00	2,250	,326	1,608	2,892

Confronti pairwise

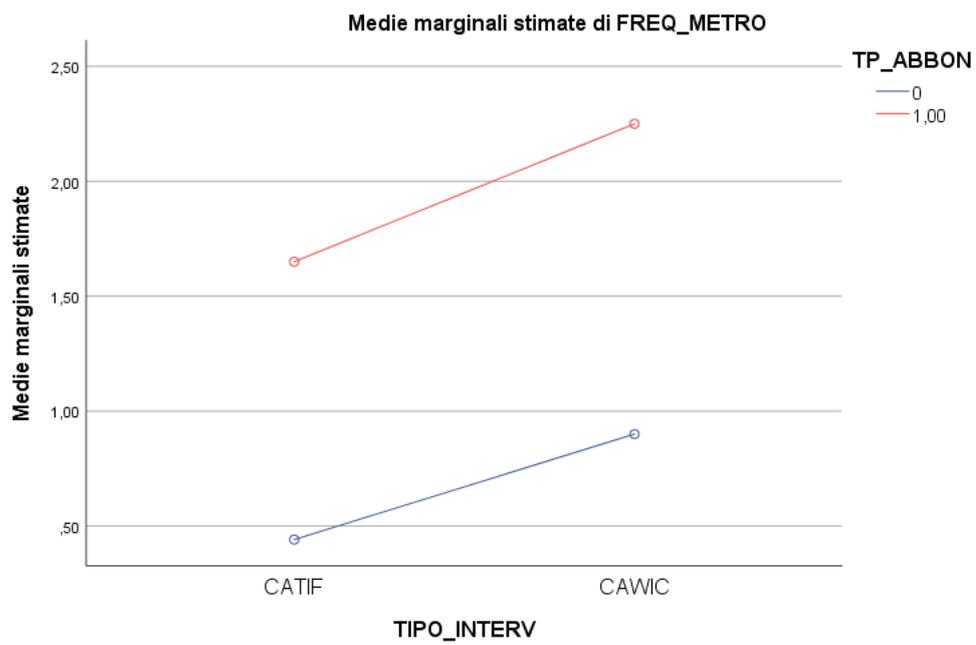
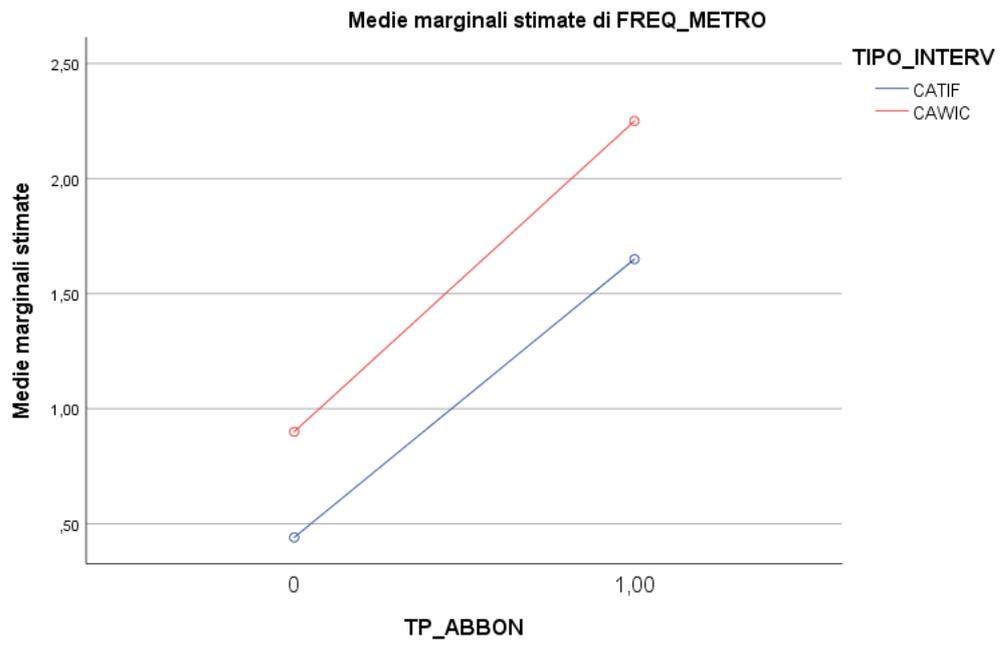
Variabile dipendente: FREQ_METRO

TP_ABBON	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,459	,268	,048	-,987	,069
	CAWIC	CATIF	,459	,268	,048	-,069	,987
1,00	CATIF	CAWIC	-,600	,372	,108	-1,332	,132
	CAWIC	CATIF	,600	,372	,108	-,132	1,332

Confronti pairwise

Variabile dipendente: FREQ_METRO

TIPO_INTERV	TP_ABBON	TP_ABBON	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1,00	-1,209	,200	,000	-1,603	-,816
	1,00	0	1,209	,200	,000	,816	1,603
CAWIC	0	1,00	-1,350	,413	,001	-2,162	-,538
	1,00	0	1,350	,413	,001	,538	2,162



2.2.6 Variabile dipendente: FREQ_TRENO

Stime

Variabile dipendente: FREQ_TRENO

TIPO_INTERV	TP_ABBON	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	,027	,029	-,030	,085
	1,00	,000	,058	-,114	,114
CAWIC	0	,533	,082	,372	,695
	1,00	,694	,106	,486	,903

Confronti pairwise

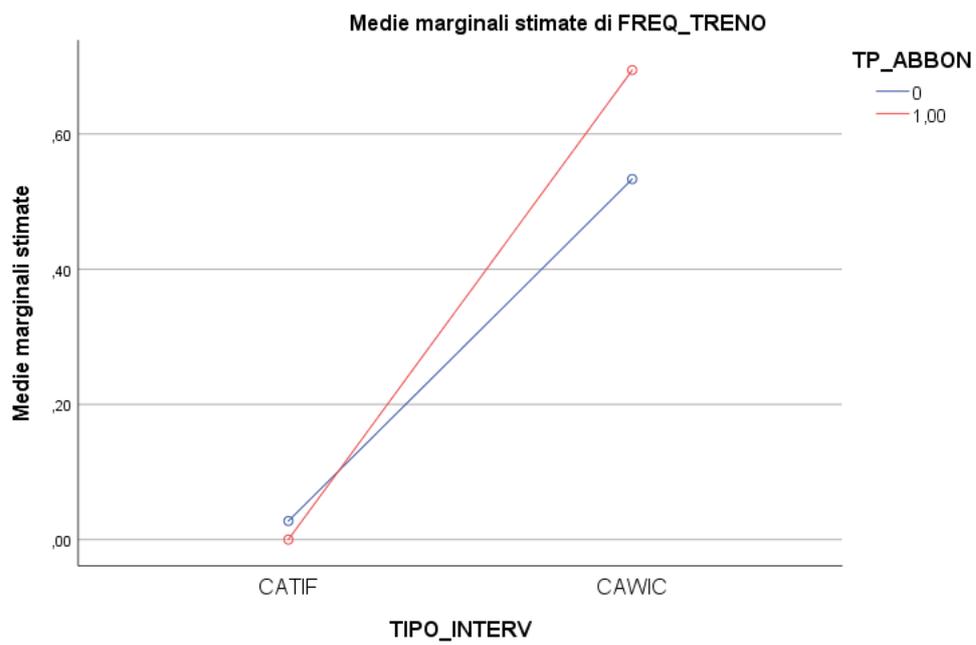
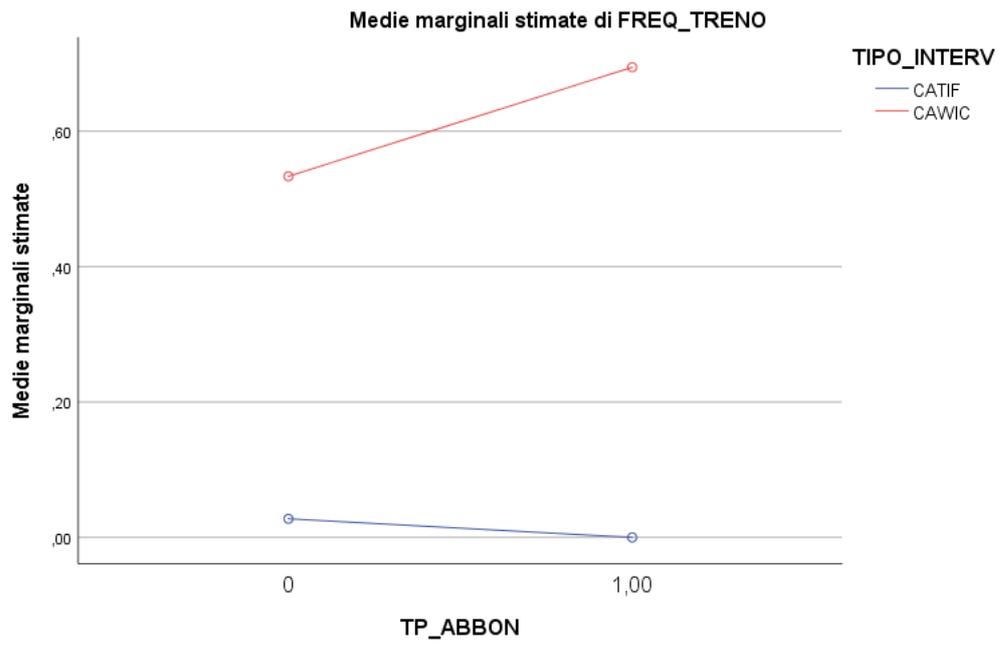
Variabile dipendente: FREQ_TRENO

TP_ABBON	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,506	,087	,000	-,678	-,334
	CAWIC	CATIF	,506	,087	,000	,334	,678
1,00	CATIF	CAWIC	-,694	,121	,000	-,932	-,456
	CAWIC	CATIF	,694	,121	,000	,456	,932

Confronti pairwise

Variabile dipendente: FREQ_TRENO

TIPO_INTERV	TP_ABBON	TP_ABBON	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1,00	,027	,065	,674	-,101	,155
	1,00	0	-,027	,065	,674	-,155	,101
CAWIC	0	1,00	-,161	,134	,231	-,425	,103
	1,00	0	,161	,134	,231	-,103	,425



2.3 Variabili indipendenti: TIPO_INTERV e CLASSE_ADULTI

2.3.1 Variabile dipendente: FREQ_AUCO

Stime

Variabile dipendente: FREQ_AUCO

TIPO_INTERV	CLASSE_ADULTI	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	1,609	,179	1,257	1,961
	1	3,048	,171	2,711	3,385
CAWIC	0	3,077	,592	1,913	4,241
	1	2,914	,361	2,205	3,623

Confronti pairwise

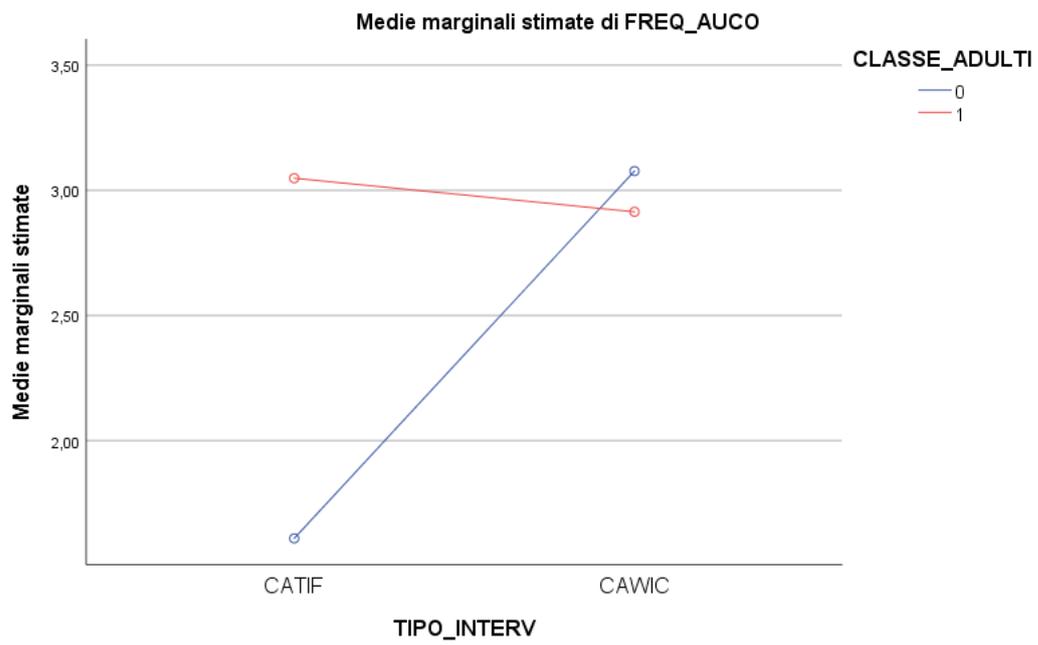
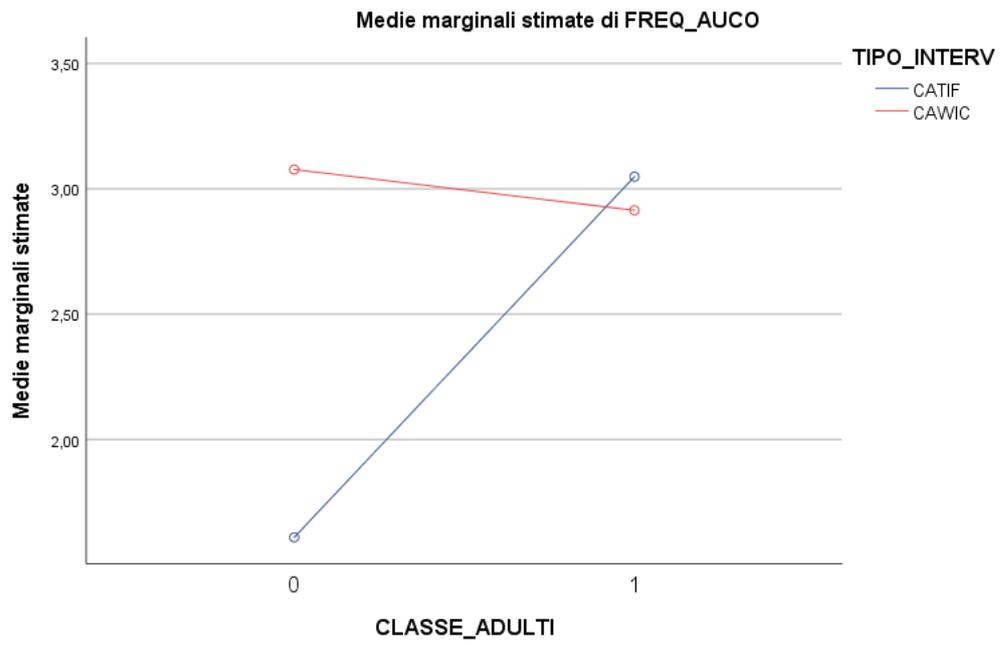
Variabile dipendente: FREQ_AUCO

CLASSE_ADULTI	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-1,468	,618	,018	-2,684	-,252
	CAWIC	CATIF	1,468	,618	,018	,252	2,684
1	CATIF	CAWIC	,134	,399	,737	-,651	,919
	CAWIC	CATIF	-,134	,399	,737	-,919	,651

Confronti pairwise

Variabile dipendente: FREQ_AUCO

TIPO_INTERV	CLASSE_ADULTI	CLASSE_ADULTI	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1	-1,439	,248	,000	-1,927	-,952
	1	0	1,439	,248	,000	,952	1,927
CAWIC	0	1	,163	,693	,815	-1,200	1,525
	1	0	-,163	,693	,815	-1,525	1,200



2.3.2 Variabile dipendente: FREQ_BICI

Variabile dipendente: FREQ_BICI

TIPO INTERV	CLASSE ADULTI	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	,282	,088	,109	,455
	1	,323	,084	,157	,488
CAWIC	0	1,346	,291	,775	1,918
	1	,529	,177	,180	,877

Confronti pairwise

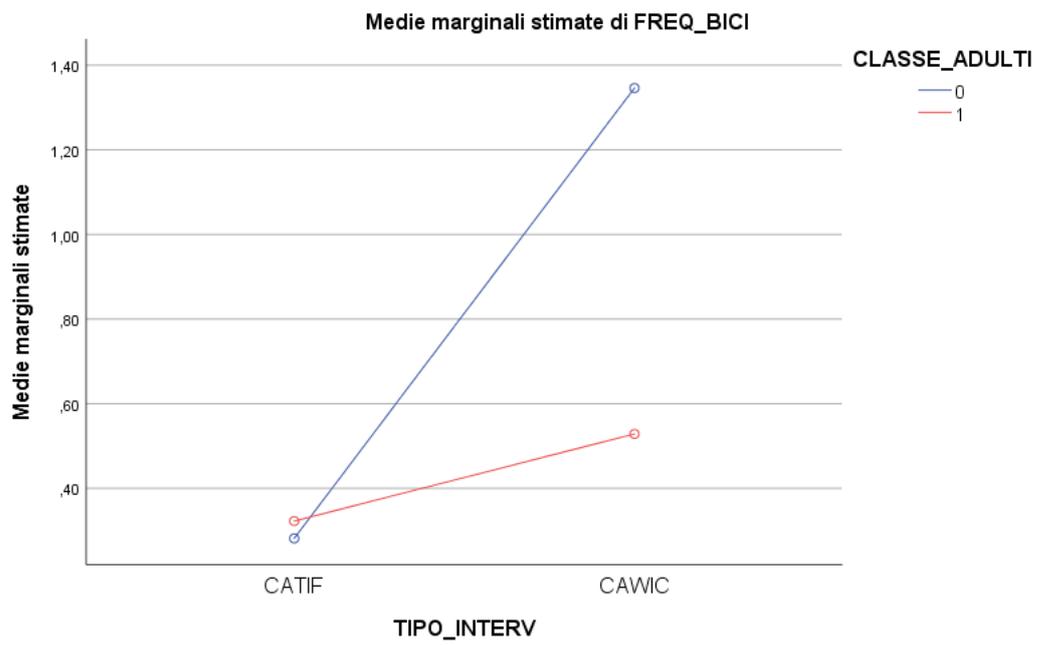
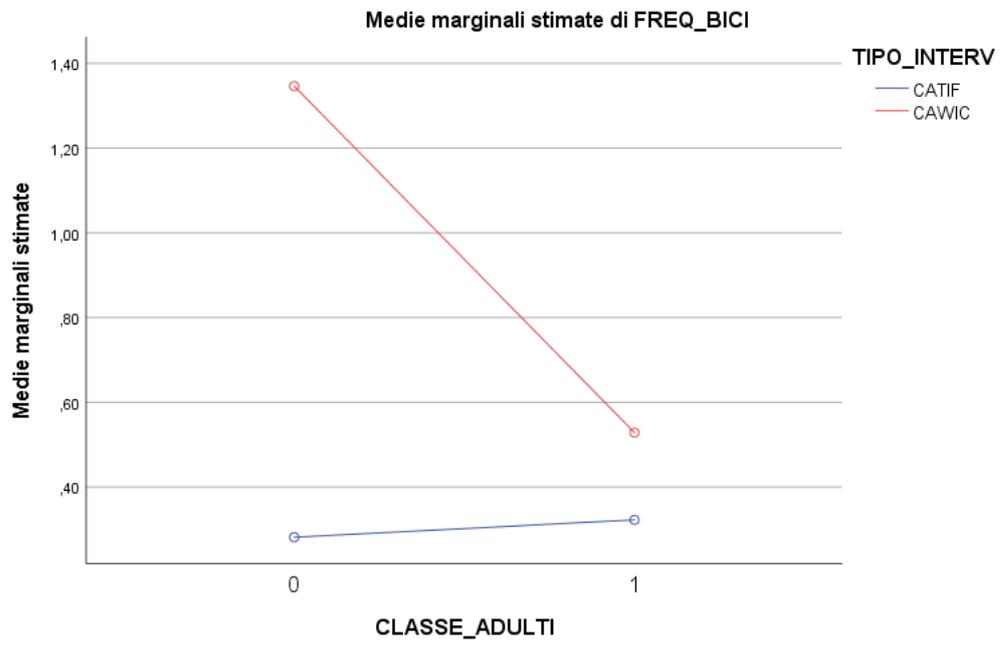
Variabile dipendente: FREQ_BICI

CLASSE ADULTI	TIPO INTERV	TIPO INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-1,064	,304	,001	-1,662	-,467
	CAWIC	CATIF	1,064	,304	,001	,467	1,662
1	CATIF	CAWIC	-,206	,196	,294	-,592	,180
	CAWIC	CATIF	,206	,196	,294	-,180	,592

Confronti pairwise

Variabile dipendente: FREQ_BICI

TIPO INTERV	CLASSE ADULTI	CLASSE ADULTI	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1	-,041	,340	,017	-,280	,198
	1	0	,041	,340	,017	-,198	,280
CAWIC	0	1	,818	,122	,737	,148	1,487
	1	0	-,818	,122	,737	-1,487	-,148



2.3.3 Variabile dipendente: FREQ_AUPA

Stime

Variabile dipendente: FREQ_AUPA

TIPO_INTERV	CLASSE_ADULTI	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	,398	,080	,240	,556
	1	,271	,077	,120	,422
CAWIC	0	2,154	,266	1,631	2,676
	1	,771	,162	,453	1,090

Confronti pairwise

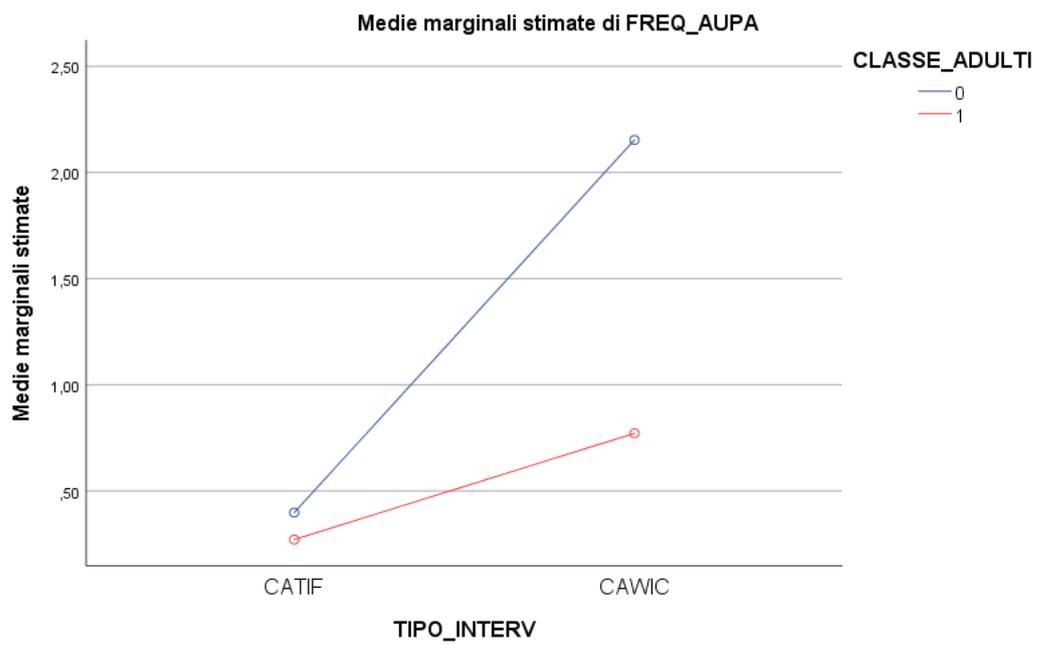
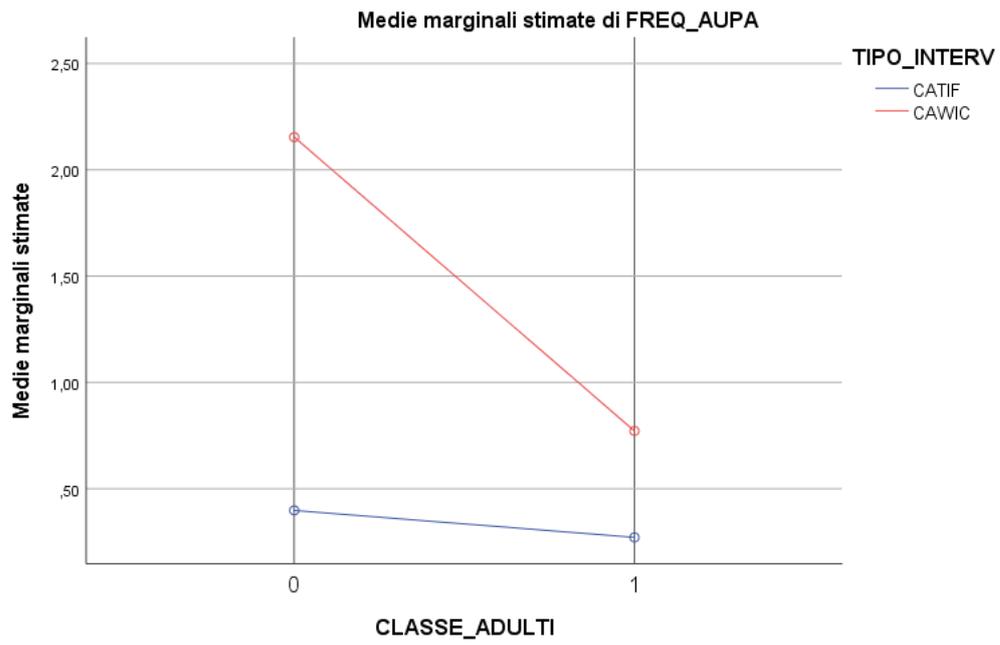
Variabile dipendente: FREQ_AUPA

CLASSE_ADULTI	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-1,756	,278	,000	-2,302	-1,210
	CAWIC	CATIF	1,756	,278	,000	1,210	2,302
1	CATIF	CAWIC	-,500	,179	,006	-,853	-,148
	CAWIC	CATIF	,500	,179	,006	,148	,853

Confronti pairwise

Variabile dipendente: FREQ_AUPA

TIPO_INTERV	CLASSE_ADULTI	CLASSE_ADULTI	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1	,127	,111	,255	-,092	,346
	1	0	-,127	,111	,255	-,346	,092
CAWIC	0	1	1,382	,311	,000	,771	1,994
	1	0	-1,382	,311	,000	-1,994	-,771



2.3.4 Variabile dipendente: **FREQ_BUS**

Stime

Variabile dipendente: **FREQ_BUS**

TIPO_INTERV	CLASSE_ADULTI	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	1,634	,165	1,310	1,958
	1	1,523	,158	1,212	1,833
CAWIC	0	2,500	,544	1,429	3,571
	1	1,829	,332	1,176	2,481

Confronti pairwise

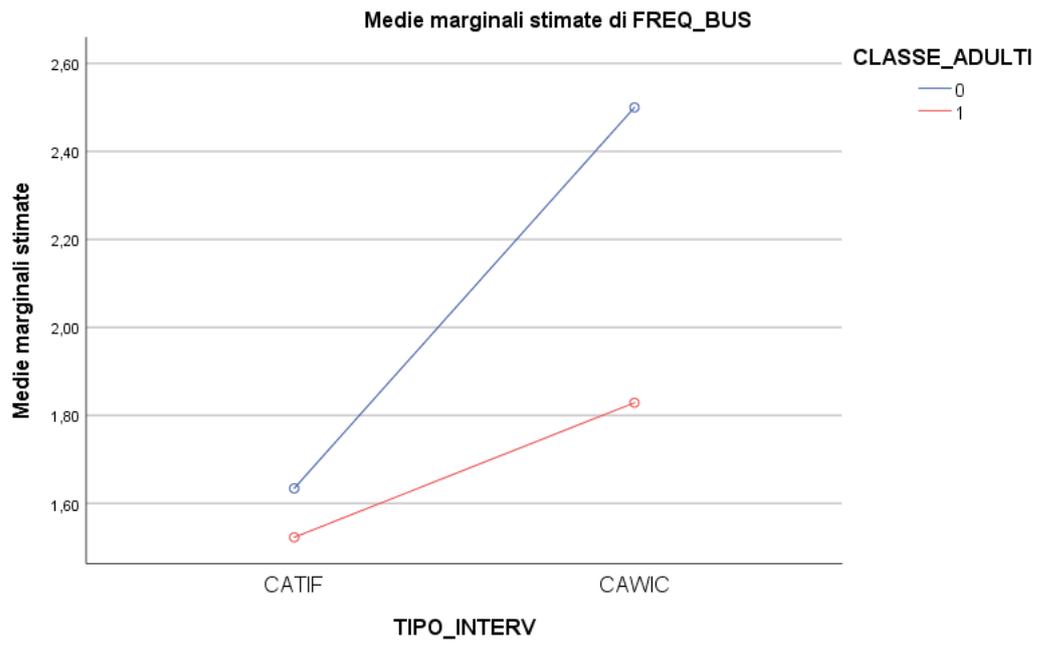
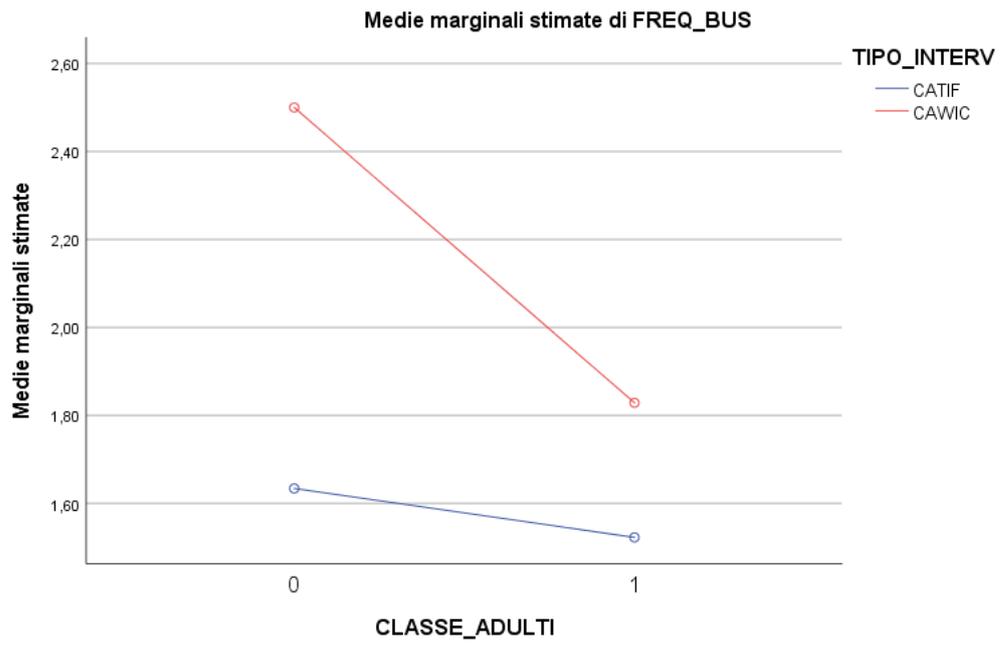
Variabile dipendente: **FREQ_BUS**

CLASSE_ADULTI	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,866	,569	,129	-1,985	,252
	CAWIC	CATIF	,866	,569	,129	-,252	1,985
1	CATIF	CAWIC	-,306	,367	,405	-1,028	,416
	CAWIC	CATIF	,306	,367	,405	-,416	1,028

Confronti pairwise

Variabile dipendente: **FREQ_BUS**

TIPO_INTERV	CLASSE_ADULTI	CLASSE_ADULTI	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1	,111	,228	,626	-,337	,560
	1	0	-,111	,228	,626	-,560	,337
CAWIC	0	1	,671	,637	,293	-,582	1,925
	1	0	-,671	,637	,293	-1,925	,582



2.3.5 Variabile dipendente: FREQ_METRO

Stime

Variabile dipendente: FREQ_METRO

TIPO_INTERV	CLASSE_ADULTI	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	,680	,124	,436	,923
	1	,690	,119	,457	,924
CAWIC	0	1,385	,410	,579	2,190
	1	1,414	,250	,923	1,905

Confronti pairwise

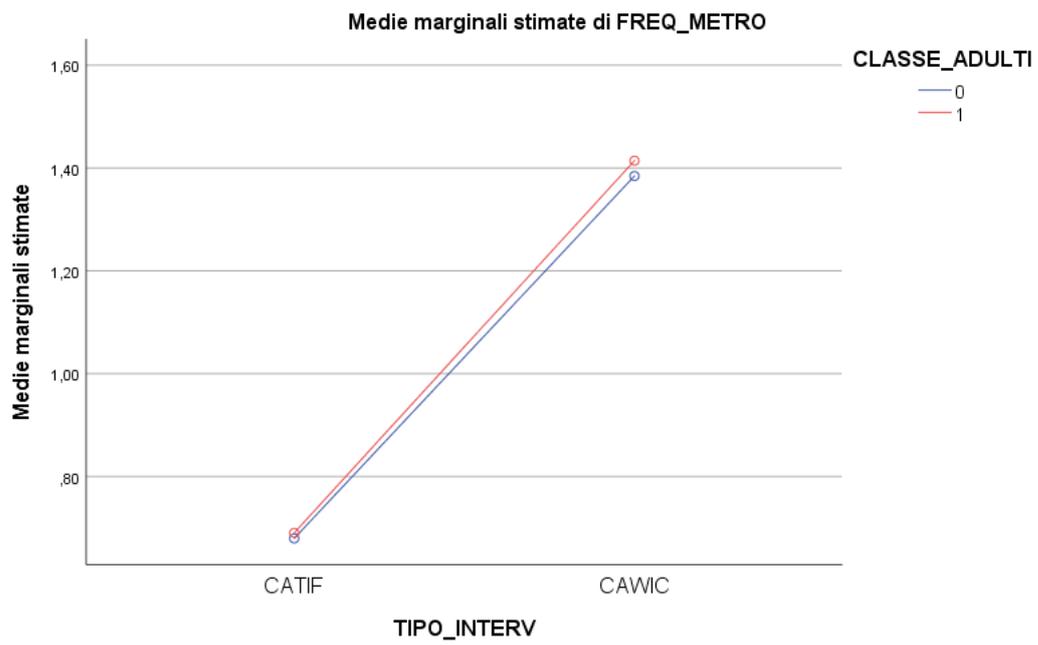
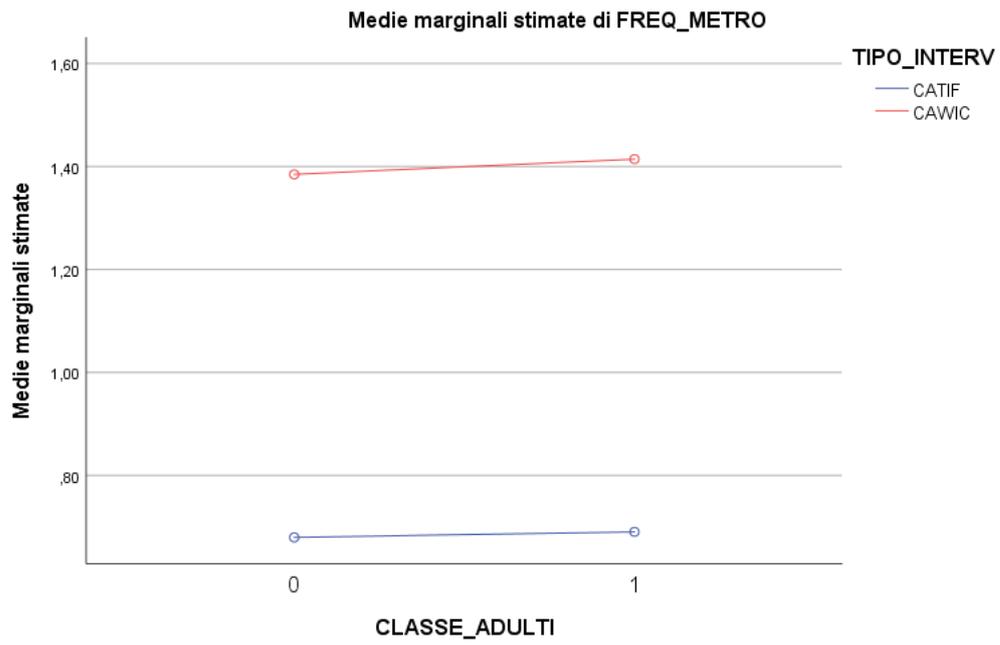
Variabile dipendente: FREQ_METRO

CLASSE_ADULTI	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,705	,428	,100	-1,547	,137
	CAWIC	CATIF	,705	,428	,100	-,137	1,547
1	CATIF	CAWIC	-,724	,276	,069	-1,268	-,180
	CAWIC	CATIF	,724	,276	,069	,180	1,268

Confronti pairwise

Variabile dipendente: FREQ_METRO

TIPO_INTERV	CLASSE_ADULTI	CLASSE_ADULTI	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1	-,011	,172	,950	-,348	,327
	1	0	,011	,172	,950	-,327	,348
CAWIC	0	1	-,030	,480	,951	-,973	,914
	1	0	,030	,480	,951	-,914	,973



2.3.6 Variabile dipendente: FREQ_TRENO

Stime

Variabile dipendente: FREQ_TRENO

TIPO_INTERV	CLASSE_ADULTI	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	,004	,038	-,071	,078
	1	,039	,036	-,032	,110
CAWIC	0	,500	,125	,254	,746
	1	,629	,076	,479	,778

Confronti pairwise

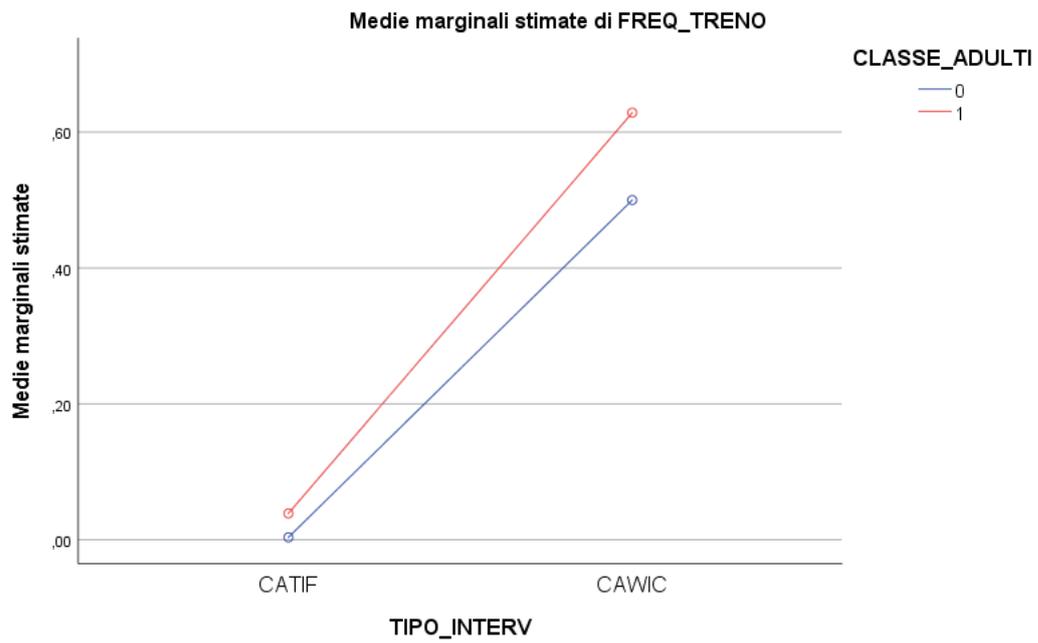
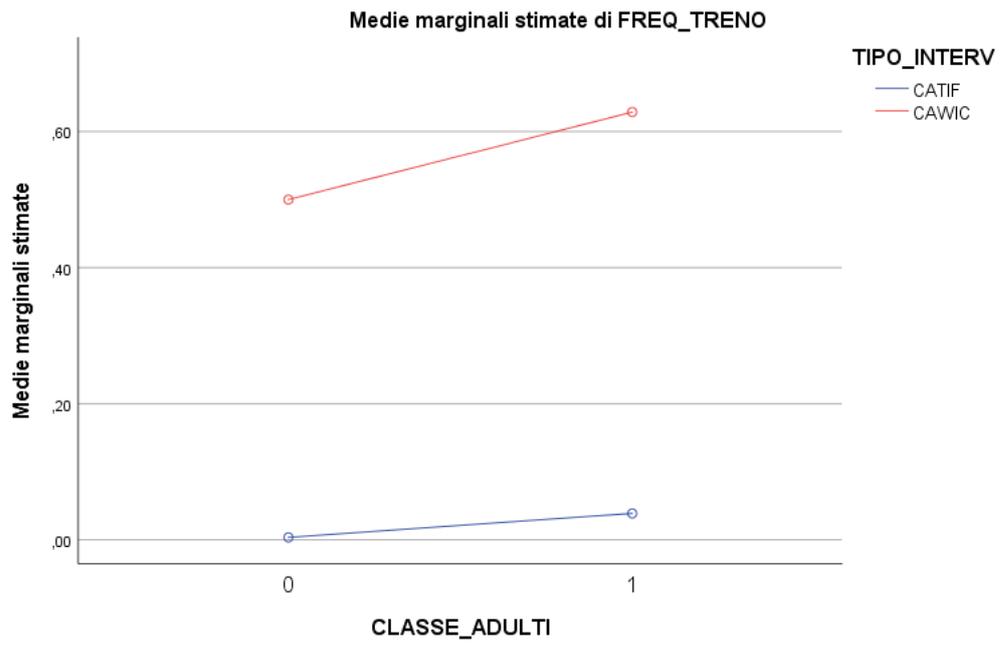
Variabile dipendente: FREQ_TRENO

CLASSE_ADULTI	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,496	,131	,000	-,753	-,240
	CAWIC	CATIF	,496	,131	,000	,240	,753
1	CATIF	CAWIC	-,590	,084	,000	-,756	-,424
	CAWIC	CATIF	,590	,084	,000	,424	,756

Confronti pairwise

Variabile dipendente: FREQ_TRENO

TIPO_INTERV	CLASSE_ADULTI	CLASSE_ADULTI	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1	-,035	,052	,041	-,138	,068
	1	0	,035	,052	,041	-,068	,138
CAWIC	0	1	-,129	,146	,038	-,416	,159
	1	0	,129	,146	,038	-,159	,416



2.4 Variabili indipendenti: TIPO_INTERV e AUCS_ABBON

2.4.1 Variabile dipendente: FREQ_AUCO

Stime

Variabile dipendente: FREQ_AUCO

TIPO_INTERV	AUCS_ABBON	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	2,360	,130	2,105	2,615
	1	.a	.	.	.
CAWIC	0	2,943	,337	2,281	3,605
	1	3,125	1,116	,929	5,321

a. Questa combinazione di livello dei fattori non viene osservata, pertanto la corrispondente media marginale della popolazione non è stimabile.

Confronti pairwise

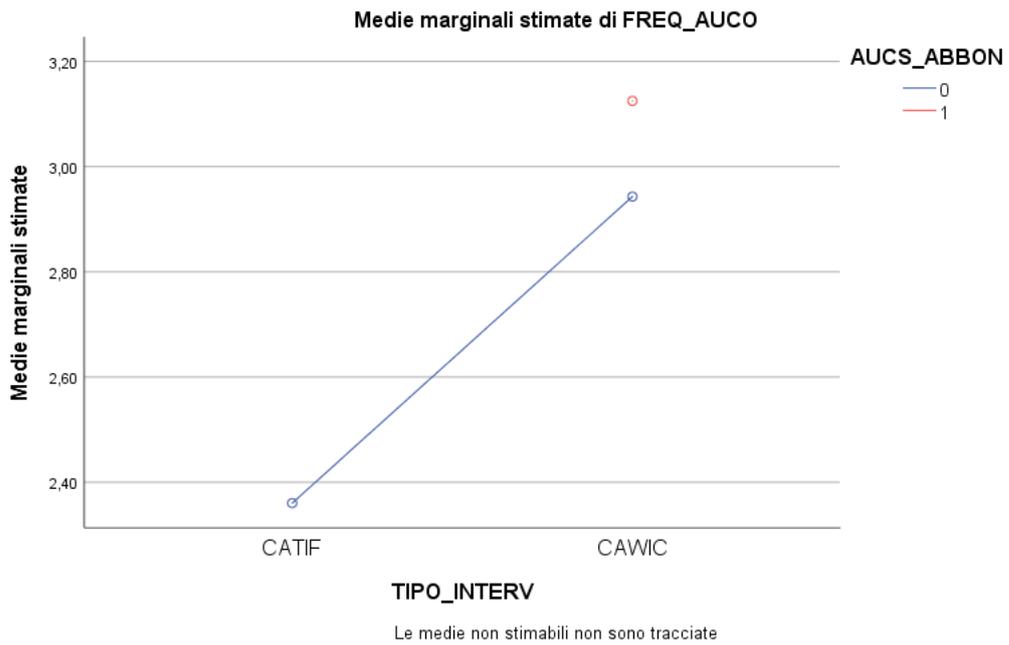
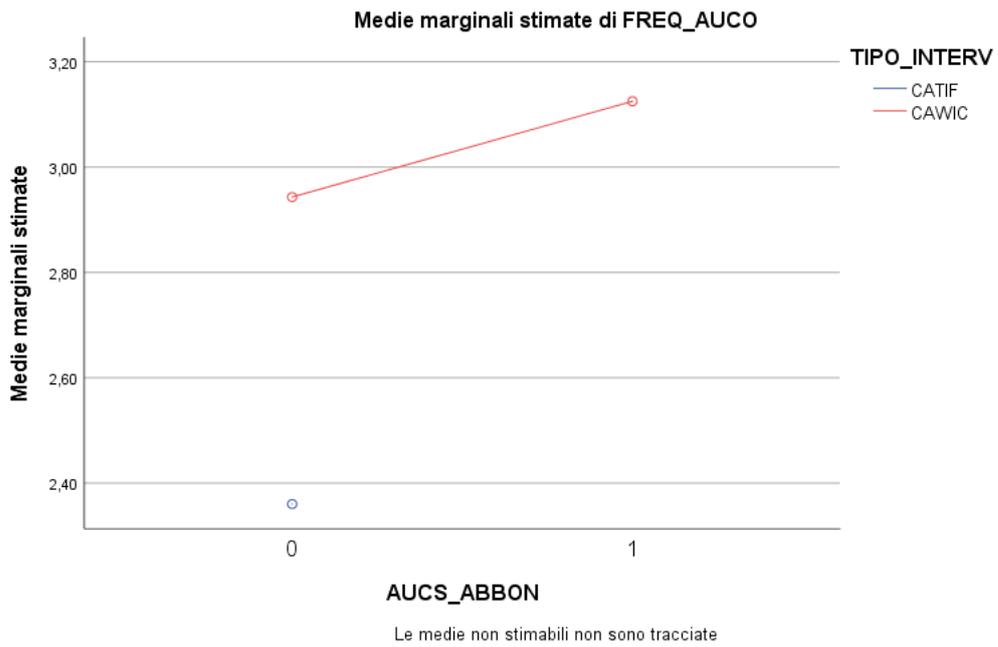
Variabile dipendente: FREQ_AUCO

AUCS_ABBON	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,583	,361	,107	-1,292	,127
	CAWIC	CATIF	,583	,361	,107	-,127	1,292
1	CATIF	CAWIC
	CAWIC	CATIF

Confronti pairwise

Variabile dipendente: FREQ_AUCO

TIPO_INTERV	AUCS_ABBON	AUCS_ABBON	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1
	1	0
CAWIC	0	1	-,182	1,166	,876	-2,475	2,112
	1	0	,182	1,166	,876	-2,112	2,475



2.4.2 Variabile dipendente: FREQ_BICI

Stime

Variabile dipendente: FREQ_BICI

TIPO_INTERV	AUCS_ABBON	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	,303	,061	,183	,423
	1	.a	.	.	.
CAWIC	0	,659	,158	,348	,970
	1	1,750	,525	,718	2,782

a. Questa combinazione di livello dei fattori non viene osservata, pertanto la corrispondente media marginale della popolazione non è stimabile.

Confronti pairwise

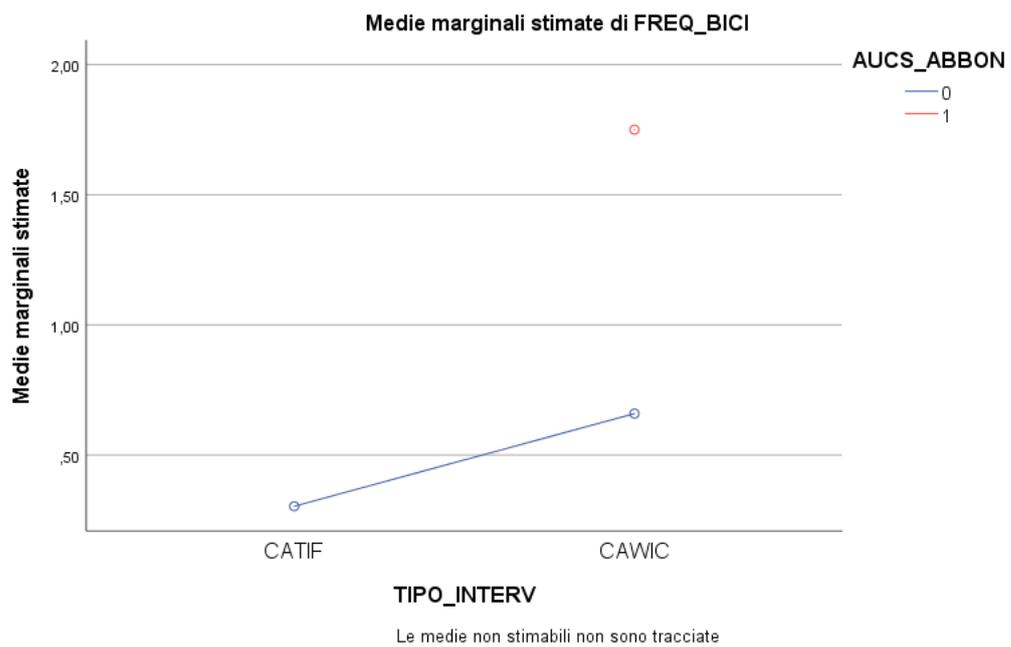
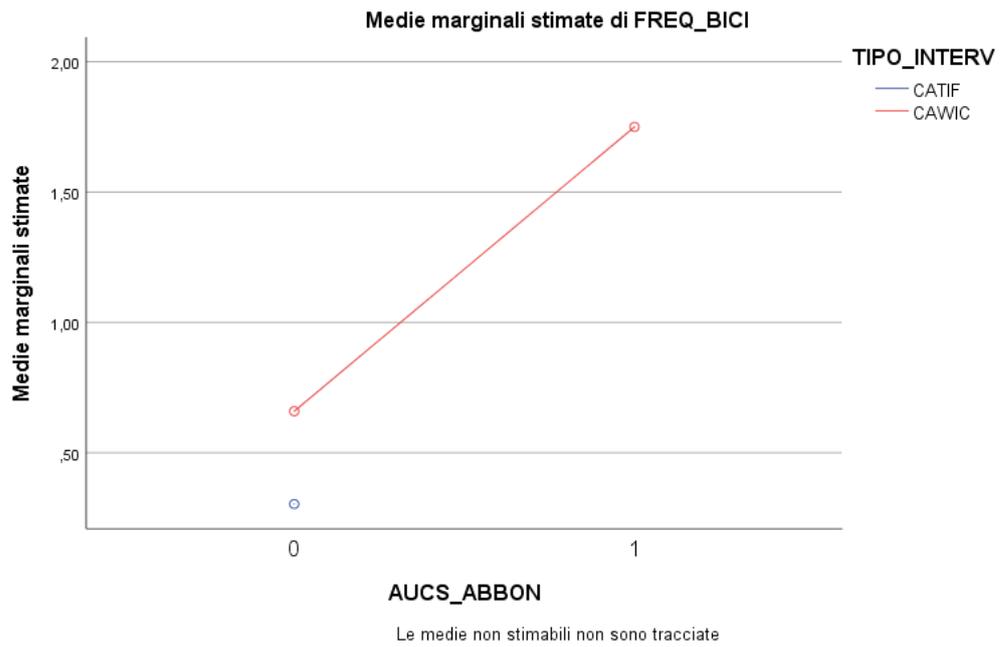
Variabile dipendente: FREQ_BICI

AUCS_ABBON	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,356	,169	,036	-,689	-,023
	CAWIC	CATIF	,356	,169	,036	,023	,689
1	CATIF	CAWIC
	CAWIC	CATIF

Confronti pairwise

Variabile dipendente: FREQ_BICI

TIPO_INTERV	AUCS_ABBON	AUCS_ABBON	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1
	1	0
CAWIC	0	1	-1,091*	,548	,047	-2,168	-,013
	1	0	1,091*	,548	,047	,013	2,168



2.4.3 Variabile dipendente: FREQ_AUPA

Stime

Variabile dipendente: FREQ_AUPA

TIPO_INTERV	AUCS_ABBON	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	,332	,057	,219	,444
	1	.a	.	.	.
CAWIC	0	1,136	,149	,844	1,429
	1	1,250	,493	,281	2,219

a. Questa combinazione di livello dei fattori non viene osservata, pertanto la corrispondente media marginale della popolazione non è stimabile.

Confronti pairwise

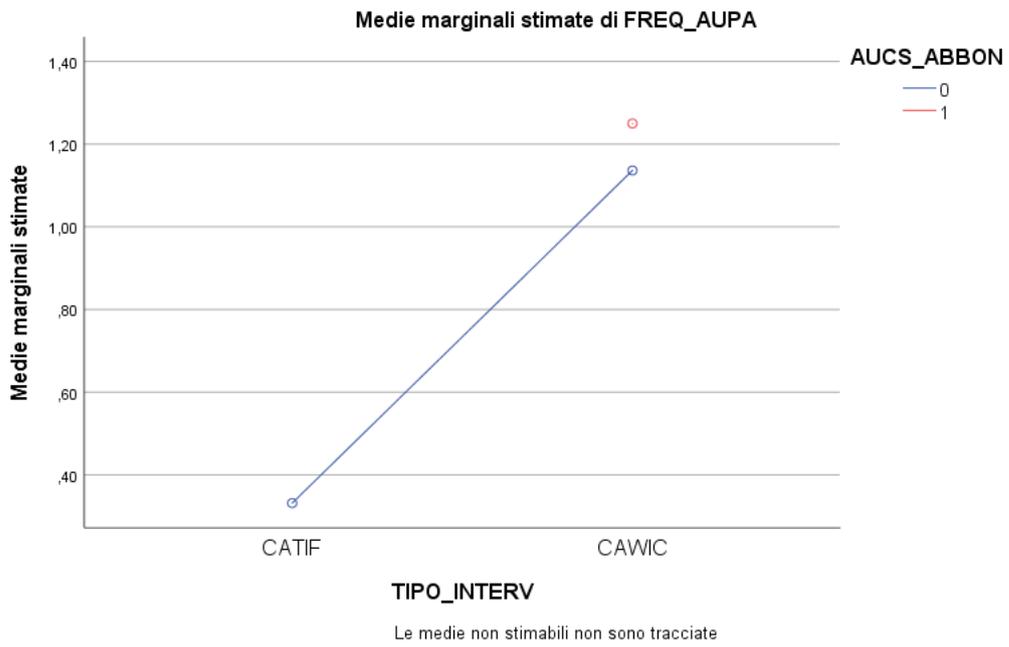
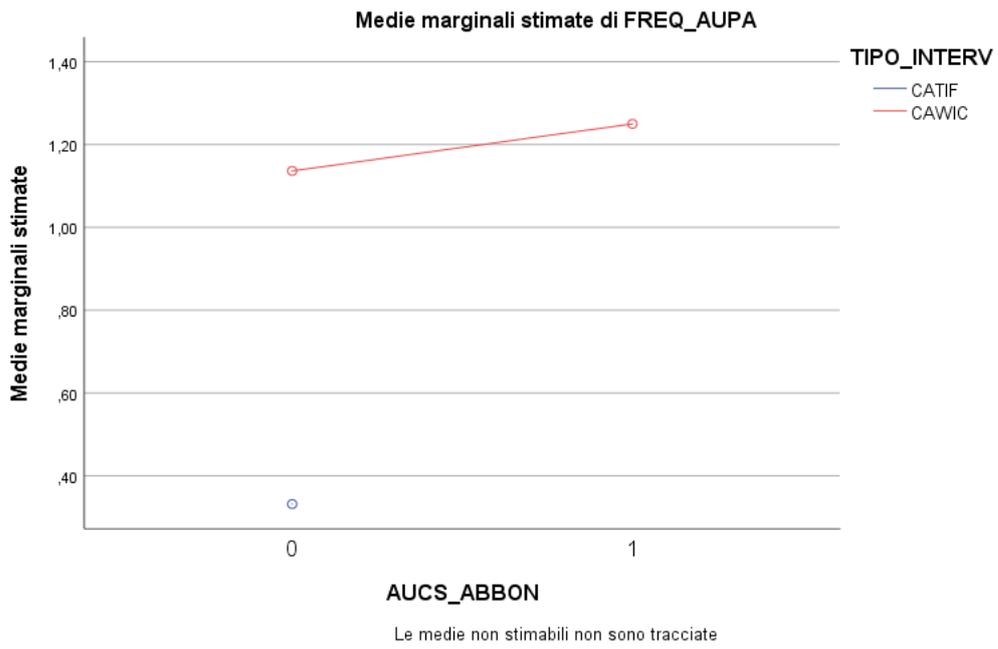
Variabile dipendente: FREQ_AUPA

AUCS_ABBON	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,805	,159	,000	-1,118	-,492
	CAWIC	CATIF	,805	,159	,000	,492	1,118
1	CATIF	CAWIC
	CAWIC	CATIF

Confronti pairwise

Variabile dipendente: FREQ_AUPA

TIPO_INTERV	AUCS_ABBON	AUCS_ABBON	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1
	1	0
CAWIC	0	1	-,114	,515	,825	-1,126	,898
	1	0	,114	,515	,825	-,898	1,126



2.4.4 Variabile dipendente: FREQ_BUS

Stime

Variabile dipendente: FREQ_BUS

TIPO_INTERV	AUCS_ABBON	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	1,576	,114	1,352	1,799
	1	.a	.	.	.
CAWIC	0	2,114	,295	1,533	2,695
	1	,875	,980	-1,052	2,802

a. Questa combinazione di livello dei fattori non viene osservata, pertanto la corrispondente media marginale della popolazione non è stimabile.

Confronti pairwise

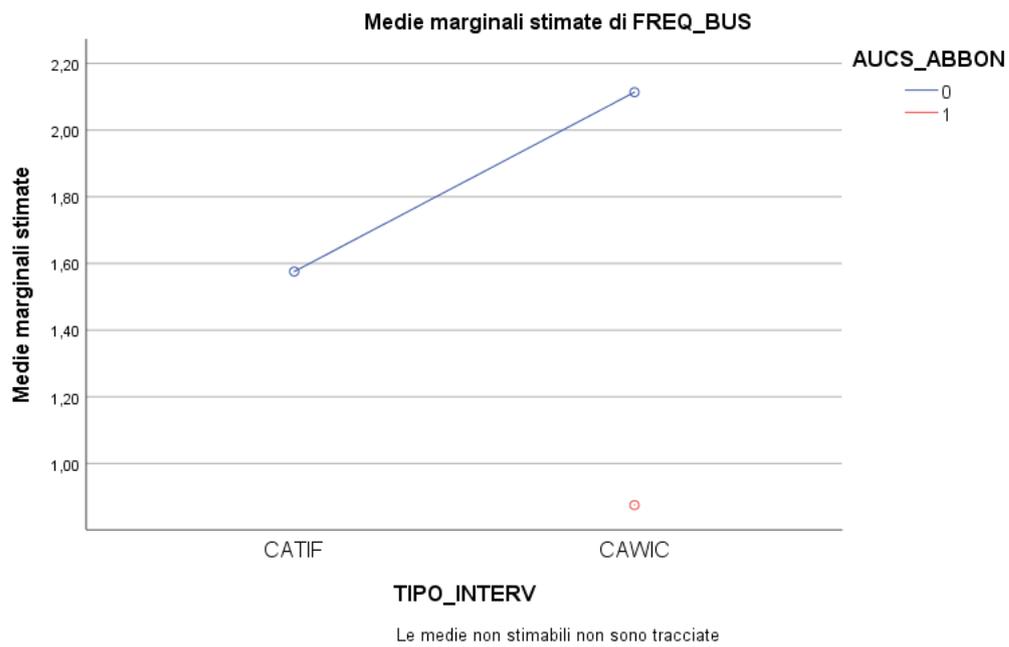
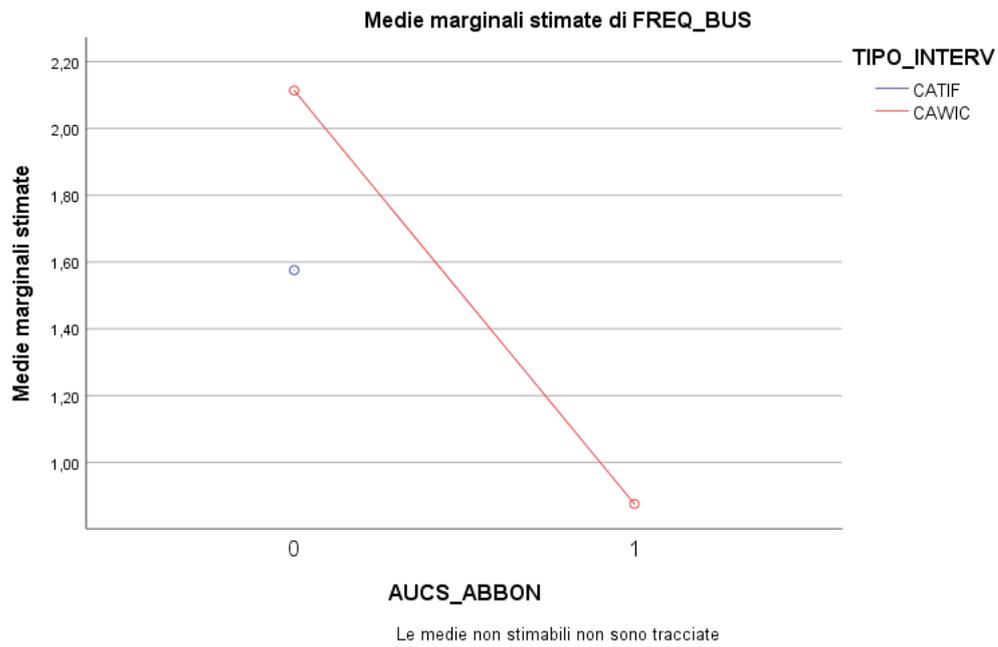
Variabile dipendente: FREQ_BUS

AUCS_ABBON	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,538	,317	,090	-1,160	,085
	CAWIC	CATIF	,538	,317	,090	-,085	1,160
1	CATIF	CAWIC
	CAWIC	CATIF

Confronti pairwise

Variabile dipendente: FREQ_BUS

TIPO_INTERV	AUCS_ABBON	AUCS_ABBON	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1
	1	0
CAWIC	0	1	1,239	1,023	,227	-,774	3,251
	1	0	-1,239	1,023	,227	-3,251	,774



2.4.5 Variabile dipendente: FREQ_METRO

Stime

Variabile dipendente: FREQ_METRO

TIPO_INTERV	AUCS_ABBON	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	,685	,086	,517	,853
	1	. ^a	.	.	.
CAWIC	0	1,364	,222	,927	1,801
	1	1,875	,737	,425	3,325

a. Questa combinazione di livello dei fattori non viene osservata, pertanto la corrispondente media marginale della popolazione non è stimabile.

Confronti pairwise

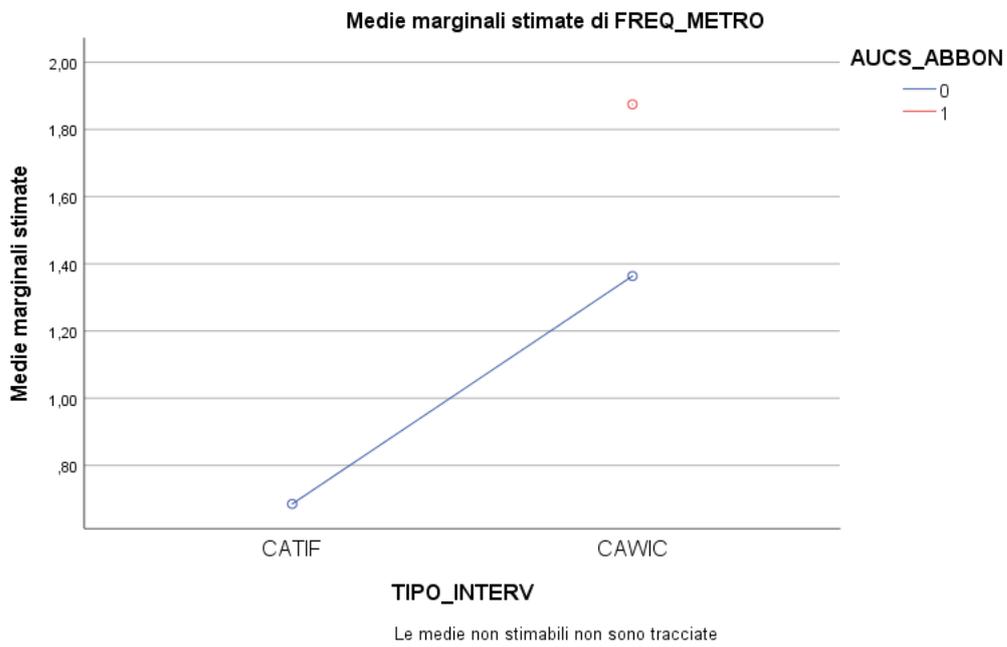
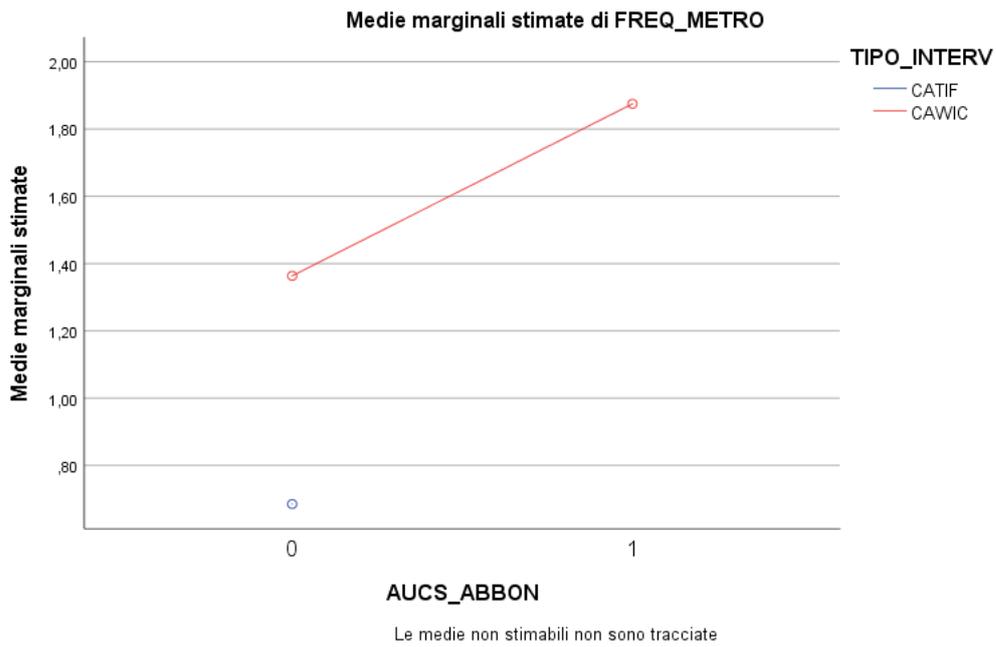
Variabile dipendente: FREQ_METRO

AUCS_ABBON	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,678	,238	,005	-1,147	-,210
	CAWIC	CATIF	,678	,238	,005	,210	1,147
1	CATIF	CAWIC
	CAWIC	CATIF

Confronti pairwise

Variabile dipendente: FREQ_METRO

TIPO_INTERV	AUCS_ABBON	AUCS_ABBON	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1
	1	0
CAWIC	0	1	-,511	,770	,507	-2,026	1,003
	1	0	,511	,770	,507	-1,003	2,026



2.4.6 Variabile dipendente: FREQ_TRENO

Stime

Variabile dipendente: FREQ_TRENO

TIPO_INTERV	AUCS_ABBON	Media	Errore std.	Intervallo di confidenza 95%	
				Limite inferiore	Limite superiore
CATIF	0	,022	,026	-,029	,073
	1	. ^a	.	.	.
CAWIC	0	,636	,067	,504	,769
	1	,125	,224	-,315	,565

a. Questa combinazione di livello dei fattori non viene osservata, pertanto la corrispondente media marginale della popolazione non è stimabile.

Confronti pairwise

Variabile dipendente: FREQ_TRENO

AUCS_ABBON	TIPO_INTERV	TIPO_INTERV	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
0	CATIF	CAWIC	-,614	,072	,000	-,757	-,472
	CAWIC	CATIF	,614	,072	,000	,472	,757
1	CATIF	CAWIC
	CAWIC	CATIF

Confronti pairwise

Variabile dipendente: FREQ_TRENO

TIPO_INTERV	AUCS_ABBON	AUCS_ABBON	Differenza della media	Errore std.	Sign.	95% intervallo di confidenza per differenza	
						Limite inferiore	Limite superiore
CATIF	0	1
	1	0
CAWIC	0	1	,511*	,234	,029	,052	,971
	1	0	-,511*	,234	,029	-,971	-,052

