

LIBERA UNIVERSITÀ MARIA SS. ASSUNTA

**DIPARTIMENTO DI GIURISPRUDENZA, ECONOMIA,
POLITICA E LINGUE MODERNE**

**CORSO DI LAUREA IN
TECNICHE INFORMATICHE PER LA GESTIONE DEI DATI
CLASSE L-31**

**CATTEDRA DI
DATA MINING & ANALITYCS**

Tabagismo in Italia e influenza dei dispositivi senza combustione

Smoking in Italy and influence of devices without combustion

RELATORE

Antonello Maruotti

CANDIDATO

Luca Cefaloni
27505/410

ANNO ACCADEMICO 2022-2023

INDICE

INTRODUZIONE	3
Contesto del Fenomeno del Tabagismo in Italia	3
Rilevanza dello Studio e Obiettivi di Ricerca	3
Struttura della Tesi	4
Spiegazione del tabagismo e possibili rischi	4
Politiche di contrasto al tabagismo	6
ACQUISIZIONE DEI DATI ATTRAVERSO GOOGLE MODULI	8
Vantaggi e Limitazioni dell'Utilizzo di questa Metodologia	8
Procedura di Compilazione del Modulo da Parte dei Partecipanti	9
Strumenti Utilizzati per la Pulizia e la Preparazione dei Dati	9
ANALISI DESCRITTIVA GENERALE	9
Presentazione dei Dati Raccolti	9
Analisi dell'Intero Dataset	10
Fumatori di Sigarette Senza Combustione	11
Fumatori di Sigarette Tradizionali	21
Non Fumatori	32
MCA (ANALISI DELLE CORRISPONDENZE MULTIPLE)	35
Descrizione della metodologia MCA e la sua applicazione nell'ambito dell'analisi dei dati categorici	35
Algoritmo di imputazione multipla mice	38
Approccio Modulare alla Multiple Imputation	38
Algoritmo mice	40
Preparazione dei dataset all'analisi	42
Risultati dell'MCA utilizzando l'intera popolazione	44
Scree Plot (Grafico dello Scarto)	45
Biplot	46
Risultati MCA Sigarette Tradizionali	48
Analisi senza gli outlier	52
Risultati dell'MCA Sigarette Senza Combustione	55
Sottogruppo di ex fumatori di sigarette tradizionali	59
Risultati MCA Non Fumatori	64
Conclusioni	69
Riassunto dei principali risultati ottenuti	69
Discussione delle implicazioni dei risultati per la comprensione del fenomeno del tabagismo in Italia	69
Possibili direzioni future per la ricerca	70
Bibliografia	71
Sitografia	71
Ringraziamenti	72

INTRODUZIONE

CONTESTO DEL FENOMENO DEL TABAGISMO IN ITALIA

Il quadro epidemiologico del tabagismo in Italia è stato indagato attraverso il "Rapporto sul fumo in Italia", presentato durante la Giornata Mondiale contro il Tabacco del 2022 (World No Tobacco Day, 31 maggio). Con una rilevanza significativa, emerge che quasi il 24,2% della popolazione italiana è costituito da fumatori, raggiungendo una percentuale che non si riscontrava dal 2006. Questo rappresenta un incremento di 2 punti percentuali rispetto al 22% dei fumatori nel 2019, ultimo anno di rilevazione pre-pandemica. In contrasto con la tendenza decrescente che si manifestava nel triennio 2017-2019, caratterizzata dalla costante diminuzione delle persone che fumavano, il 2022 ha evidenziato un incremento nella percentuale di fumatori che coinvolge ambedue i sessi.

Parallelamente, si riscontra un aumento anche tra coloro che praticano il consumo di sigarette a tabacco riscaldato, passando dal 1,1% nel 2019 al 3,3% nel 2022. Va sottolineato che oltre un terzo delle persone (il 36,6%) considera tali sigarette meno dannose rispetto a quelle tradizionali.

RILEVANZA DELLO STUDIO E OBIETTIVI DI RICERCA

A livello dell'Unione Europea (UE), il tabagismo è correlato a un tasso di mortalità annuo che sfiora i 700.000 decessi, un numero comparabile all'intera popolazione delle città di Bologna e Firenze messe insieme. Inoltre, milioni di individui nell'UE soffrono di patologie strettamente legate al consumo di tabacco, quali tumori, malattie cardiovascolari e affezioni respiratorie. Di conseguenza, l'UE sta compiendo sforzi significativi per ridurre il consumo di tabacco, soprattutto tra i giovani. Infatti, il 94% dei fumatori inizia a fumare prima dei 25 anni e questi individui hanno una probabilità del 50% in meno di raggiungere la stessa età dei non fumatori e in media vivono 14 anni in meno. Nel corso della loro vita, i fumatori affrontano inoltre una condizione di salute compromessa per un periodo maggiore. L'indagine Eurobarometro del 2017 rileva che il 26% degli europei fuma, con una prevalenza più elevata negli uomini (30%) rispetto alle donne (22%). L'Italia, con un tasso del 24%, si posiziona al decimo posto, inferiore alla

media europea. Questi dati forniscono la base iniziale per gli obiettivi primari della presente ricerca. L'analisi intrapresa fornirà comprensioni approfondite delle abitudini dei fumatori, delle loro spese e, soprattutto, dell'impatto dei dispositivi senza combustione.

STRUTTURA DELLA TESI

La struttura di questa tesi è organizzata seguendo una serie di tecniche apprese durante il percorso accademico. Verrà inizialmente eseguita una fase di pulizia e, se necessario, manipolazione dei dati utilizzando il software R. Successivamente, saranno implementate un'Analisi a Corrispondenza Multipla e una Imputazione Multipla per fornire risposte esaustive alle domande di ricerca.

SPIEGAZIONE DEL TABAGISMO E POSSIBILI RISCHI

Il tabagismo, una consuetudine di consumo di fumo di tabacco che si manifesta come una necessità ineluttabile e quotidiana, costituisce una condizione patologica rilevante all'interno del panorama delle abitudini umane. Il fumo di tabacco è intrinsecamente composto da una vasta gamma di sostanze chimiche, tra cui la nicotina, un alcaloide con attività neurotrofa che incide sia sul sistema cardiovascolare che sul metabolismo. Di particolare interesse è il suo potenziale nell'indurre una dipendenza sia di natura psicologica che fisica, quest'ultima mostrando una variabilità individuale, possibilmente più intensa tra i fumatori abituali e forti, associata all'assunzione regolare di nicotina. Pertanto, il tabagismo non solo comporta una dimensione comportamentale, ma rivela anche una componente biochimica cruciale nel sostenere questa pratica.

Oltre alla relazione strettamente interconnessa tra il fumo di tabacco e la sfera dell'assuefazione, va altresì considerata la presenza di più di 4.000 componenti dannosi all'interno dei prodotti derivanti dalla combustione del tabacco. La potenziale nocività del fumo deriva dalla produzione di sostanze chimiche risultanti dalla combustione stessa. Nel contesto delle implicazioni per la salute umana, le conseguenze cancerogene del fumo di tabacco possono essere principalmente attribuite al catrame, un residuo secco, scuro e viscoso che si forma dopo che la nicotina e i composti gassosi sono stati rimossi dal processo di

combustione. Questo catrame è intrinsecamente ricco di composti chimici notoriamente cancerogeni, come gli idrocarburi aromatici policiclici, tra cui il benzo[a]pirene e il dimetilbenzantracene, così come le amine aromatiche, tra cui le nitrosamine.

L'approccio analitico a tali componenti del tabagismo non solo evidenzia la complessità di questa consuetudine, ma solleva anche preoccupazioni significative in termini di impatto sulla salute pubblica. La comprensione delle interazioni tra gli agenti chimici presenti nel fumo di tabacco e l'organismo umano costituisce un ambito di ricerca critico, in quanto fornisce la base per sviluppare strategie di prevenzione e intervento mirate. In sintesi, il fenomeno del tabagismo rivela un intreccio di aspetti comportamentali, chimici e biologici che richiedono una valutazione approfondita e un'azione continua da parte sia degli individui che della società nel suo complesso.

Il consumo di tabacco emerge come uno dei fattori determinanti di decesso, secondo l'ultima revisione della Classificazione Internazionale delle Malattie (International Classification Disease, ICD X). Esso è riconosciuto come una causa sostanziale nella patogenesi di un vasto spettro di neoplasie, tra cui il cancro del polmone, i tumori delle vie respiratorie superiori, epatici, pancreatici, esofagei, gastrici, renali, cervicali e le leucemie. Inoltre, il tabagismo risulta collegato a numerose malattie non neoplastiche, ma altrettanto letali, quali l'insufficienza cardiaca polmonare, la broncopneumopatia cronica ostruttiva (BPCO), l'ictus, le polmoniti, gli aneurismi dell'aorta e le coronaropatie. L'adozione di strategie preventive e di trattamento inerenti al tabagismo rappresenta un elemento chiave nell'alterare il decorso naturale di queste patologie. Studi indicano che l'87% dei decessi per cancro ai polmoni, l'82% dei decessi per BPCO, il 21% dei decessi per patologie coronariche e il 18% dei decessi per ictus sono attribuiti al fumo di tabacco. Anche nei fumatori asintomatici, l'infiammazione cronica delle vie respiratorie inferiori, imputabile al fumo, contribuisce a un declino funzionale accelerato del sistema respiratorio. Da notare che l'esposizione passiva al fumo di tabacco, comunemente chiamata "fumo passivo" o "second hand smoking", è associata a impatti negativi sulla salute analoghi al fumo attivo, sebbene con un livello di rischio inferiore. Tale esposizione passiva si estende anche al periodo

gestazionale, dove le gestanti fumatrici possono trasmettere i danni del fumo al feto. In particolare, i bambini nati da genitori fumatori si trovano ad affrontare un rischio significativo di esposizione al fumo passivo fin dai primi anni di vita.

Il concetto di rischio e dipendenza dal fumo è ulteriormente complesso. Il consumo continuo e prolungato di tabacco ha un impatto non solo sulla durata media della vita, ma anche sulla sua qualità. Ad esempio, si è stimato che fumare 20 sigarette al giorno riduca di circa 4,6 anni la vita media di un individuo che inizia a fumare a 25 anni, equivalente a una perdita di un giorno di vita ogni settimana di fumo. Tra gli uomini adulti fumatori, su 1.000 individui, uno morirà per una causa violenta, sei moriranno in incidenti stradali e 250 perderanno la vita a causa di patologie associate al fumo. Il Center for Disease Control and Prevention (CDC) degli Stati Uniti ha identificato ben 27 malattie direttamente correlate al fumo. L'entità dei danni fisici derivanti dall'esposizione, incluso il fumo passivo, è direttamente proporzionale all'entità complessiva dell'abuso di tabacco. Ulteriormente, fattori quali l'età di inizio, il numero di sigarette consumate al giorno, l'anzianità nel fumo e la profondità dell'inalazione giocano un ruolo fondamentale nel determinare il rischio individuale.

La dipendenza dal fumo è radicata nell'abilità della nicotina di stimolare la secrezione di neurotrasmettitori legati alla regolazione dell'umore e del comportamento, tra cui la dopamina, il neurotrasmettitore coinvolto nella sensazione di piacere. Questo meccanismo sottende la formazione della dipendenza dal tabacco. Sintomi come l'incapacità di smettere, i segni di astinenza in presenza di tentativi di cessazione (ansia, irritabilità, agitazione, difficoltà di concentrazione, cambiamenti dell'umore, frustrazione, rabbia, aumento dell'appetito, insonnia, stipsi o diarrea) e la preferenza per il fumo nonostante problemi di salute emergenti riflettono la complessità della dipendenza dalla nicotina. Si evidenzia che l'unico modo efficace per evitare tale dipendenza è la rinuncia all'abitudine di fumare fin dall'inizio.

POLITICHE DI CONTRASTO AL TABAGISMO

In sintonia con gli sforzi internazionali e nell'ambito del programma "Guadagnare salute" dell'Unione Europea, l'Italia si è impegnata nel contrastare il consumo di

tabacco e la prevenzione del tabagismo, aderendo alla Convenzione Quadro dell'Organizzazione Mondiale della Sanità (OMS) per il controllo del tabacco. Questa Convenzione, approvata dall'Assemblea Mondiale della Sanità nel 2003, è stata tempestivamente sottoscritta dal nostro Paese, riflettendo un impegno concreto a indirizzare le problematiche legate al fumo di tabacco.

La Legge di ratifica n. 54, emanata il 18 marzo 2008, è stata ufficialmente pubblicata sulla Gazzetta Ufficiale il 17 aprile 2008. Tale provvedimento legislativo ha stabilito una serie di principi fondamentali, guidando l'implementazione delle linee guida delineate dalla Convenzione. Questi principi includono:

Informazione sulla Salute: La promozione di una maggiore consapevolezza riguardo alle conseguenze per la salute derivanti dall'esposizione al fumo di tabacco, sia attiva che passiva.

Impegno Plurisettoriale: L'adozione di una strategia politica che incoraggia la formulazione e l'applicazione di misure multisettoriali volte a contrastare l'uso del tabacco e a proteggere dall'esposizione passiva.

Riconversione Economica: Sostenere la transizione economica dei coltivatori di tabacco e dei lavoratori coinvolti nella produzione di prodotti derivati, al fine di mitigare l'effetto delle politiche di controllo del tabacco sulle loro fonti di reddito.

Partecipazione della Società Civile: Favorire un coinvolgimento attivo della società civile in tutte le fasi dell'attuazione delle politiche di controllo del tabacco. Le disposizioni legislative risultano altresì estese alle strategie atte a ridurre la domanda di tabacco, e tra queste sono contemplate:

Misure Finanziarie e Fiscali: Interventi che coinvolgono la dimensione economica del tabacco, al fine di disincentivare l'acquisto e l'uso di prodotti del tabacco attraverso imposte e tariffe.

Ambienti Chiusi Liberi dal Fumo: Interventi volti a proibire il consumo di tabacco nei luoghi pubblici e chiusi, preservando la salute delle persone da esposizioni nocive.

Regolamentazione dei Prodotti del Tabacco: Norme che disciplinano la composizione dei prodotti del tabacco, l'informazione sui prodotti derivati, nonché l'etichettatura e il confezionamento.

La Convenzione Quadro OMS enfatizza in particolare l'importanza dell'educazione e della comunicazione (articoli 12 e 13), sottolineando la rilevanza di una proibizione globale della pubblicità e della sponsorizzazione legate al tabacco. Inoltre, identifica le misure atte a ridurre l'offerta di tabacco, con un'attenzione particolare alla lotta contro il commercio illegale dei prodotti del tabacco. Un capitolo di rilievo è dedicato alla vendita ai minori: l'articolo 16 stabilisce l'obbligo per ogni Paese di adottare misure legislative, esecutive e amministrative per impedire la vendita a coloro che non hanno raggiunto l'età stabilita dalla legge nazionale o dalle disposizioni interne, generalmente fissata a diciotto anni.

ACQUISIZIONE DEI DATI ATTRAVERSO GOOGLE MODULI

L'acquisizione dei dati nell'ambito di questa ricerca è stata effettuata mediante l'utilizzo dei Moduli Google, una soluzione digitale per la raccolta sistematica e strutturata di dati tramite questionari online. Questo strumento ha consentito l'elaborazione delle risposte fornite dai partecipanti, agevolando l'organizzazione, l'archiviazione e l'analisi dei dati raccolti. Il questionario è stato creato con attenzione per garantire la chiarezza delle domande e la coerenza delle risposte, al fine di ottenere dati affidabili e pertinenti al contesto dell'indagine.

VANTAGGI E LIMITAZIONI DELL'UTILIZZO DI QUESTA METODOLOGIA

L'utilizzo dei Moduli Google ha presentato diversi vantaggi. Innanzitutto, l'approccio online ha consentito di raggiungere un campione più ampio e diversificato di partecipanti in modo efficiente. Inoltre, la raccolta digitale dei dati ha ridotto la possibilità di errori di trascrizione, semplificando il processo di inserimento dei dati nel database. La struttura standardizzata del questionario ha favorito la comparabilità dei dati e semplificato la fase di analisi.

Tuttavia, è importante riconoscere alcune limitazioni. La metodologia online potrebbe aver escluso individui che non hanno accesso a Internet o che non sono familiari con il processo di compilazione digitale. Inoltre, esiste il rischio

di risposte incomplete o inaccurate da parte dei partecipanti, che potrebbero non comprendere appieno il significato delle domande o fornire informazioni fuorvianti.

PROCEDURA DI COMPILAZIONE DEL MODULO DA PARTE DEI PARTECIPANTI

I partecipanti sono stati invitati a completare il questionario tramite i Moduli Google. Il questionario è stato accessibile attraverso un link fornito ai potenziali partecipanti via email o altri canali di comunicazione. I partecipanti hanno risposto alle domande selezionando le opzioni appropriate o fornendo risposte testuali dove richiesto. Le risposte sono state raccolte in modo anonimo per preservare la privacy dei partecipanti.

STRUMENTI UTILIZZATI PER LA PULIZIA E LA PREPARAZIONE DEI DATI

Una volta raccolti i dati attraverso i Moduli Google, è stata condotta un'attenta fase di pulizia e preparazione dei dati. Sono stati utilizzati strumenti software specializzati per verificare la coerenza delle risposte, individuare valori anomali e trattare eventuali dati mancanti. Questo processo ha garantito la qualità e l'integrità complessiva dei dati prima dell'analisi.

In conclusione, l'utilizzo dei Moduli Google si è dimostrato un metodo efficace per acquisire dati da un campione diversificato di partecipanti.

ANALISI DESCRITTIVA GENERALE

Nel contesto della nostra ricerca sul tabagismo in Italia, questa sezione svolge un ruolo cruciale nell'offrire una visione approfondita dei dati raccolti e nelle statistiche descrittive chiave delle variabili rilevanti. Questa fase analitica mira a gettare luce sulle complesse sfaccettature dei comportamenti di consumo di tabacco e sulle caratteristiche demografiche della nostra popolazione di studio.

PRESENTAZIONE DEI DATI RACCOLTI

Il nostro dataset è una vasta e ricca fonte di informazioni, che spaziano da dati personali come la data di nascita e il genere, a informazioni socio-demografiche come la regione di

residenza e l'occupazione, senza tralasciare dettagli specifici sul consumo di tabacco. Tali informazioni rappresentano la spina dorsale della nostra indagine e forniscono la base fondamentale per comprendere l'ampiezza dei comportamenti e delle caratteristiche dei partecipanti.

ANALISI DELL'INTERO DATASET

Iniziando dall'analisi globale, il nostro dataset comprende un totale di 402 individui che hanno gentilmente partecipato compilando il questionario. La mediana dell'età all'interno del dataset è di 22 anni, e questa varietà demografica si riflette nei 267 partecipanti di sesso femminile e nei 130 di sesso maschile. Le occupazioni sono altamente diverse, con 1 soggetto disoccupato, 4 lavoratori, 2 pensionati e 291 studenti.

Ciò che emerge in modo più distintivo all'interno del dataset è la suddivisione in tre categorie principali di partecipanti:

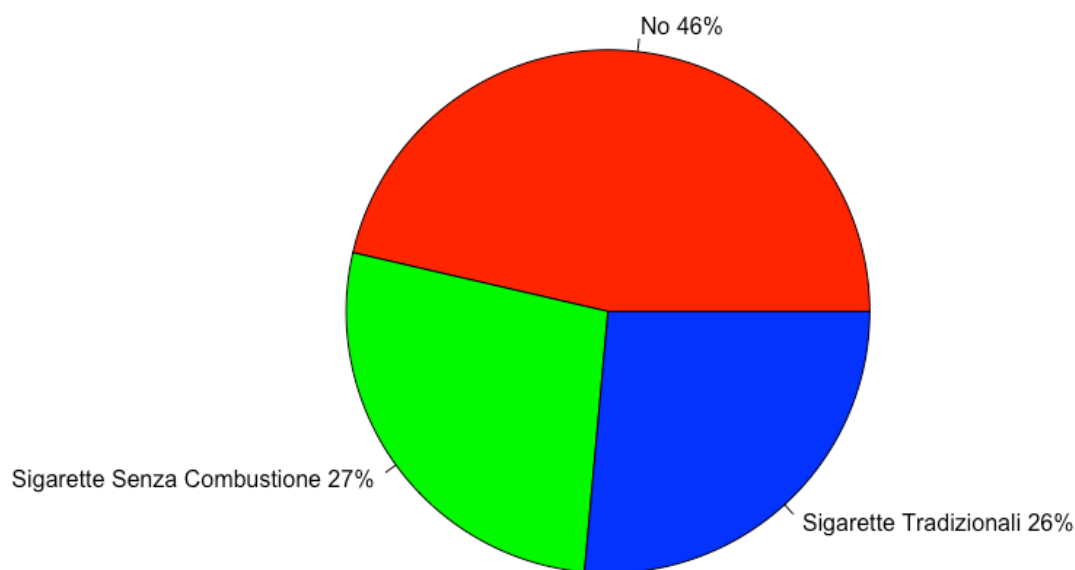


Grafico 1 Composizione in percentuale della popolazione presa in esame Fonte: Creato dall'autore

Il grafico a torta qui presentato rappresenta la distribuzione percentuale dei tre gruppi di consumatori di tabacco all'interno del dataset di riferimento.

Il dataset è stato sottoposto a un'analisi generale al fine di categorizzare i partecipanti in base alle loro abitudini di consumo di tabacco.

I tre gruppi distinti sono stati identificati come segue:

Gruppo "No": Questo gruppo rappresenta i partecipanti che hanno dichiarato di non avere alcuna abitudine di consumo di tabacco.

Gruppo "Sigarette Senza Combustione": Questo gruppo comprende i partecipanti che hanno riferito di utilizzare prodotti a base di tabacco senza combustione, come sigarette elettroniche o dispositivi simili.

Gruppo "Sigarette Tradizionali": Questo gruppo è costituito da coloro che hanno confermato l'uso di sigarette tradizionali, le quali coinvolgono la combustione del tabacco.

Il grafico evidenzia chiaramente che i fumatori, inclusi nel gruppo delle "Sigarette Senza Combustione" e nel gruppo delle "Sigarette Tradizionali", sono più numerosi rispetto ai non fumatori nel campione. Il gruppo delle "Sigarette Senza Combustione" costituisce il 27% del totale dei partecipanti, mentre il gruppo delle "Sigarette Tradizionali" rappresenta il 26%. In contrasto, il gruppo dei "Non Fumatori" rappresenta il 46% del campione, il che indica che coloro che scelgono di non consumare tabacco costituiscono la maggioranza all'interno della popolazione campionata. Questi risultati forniscono una panoramica iniziale delle preferenze di consumo di tabacco all'interno del campione, sottolineando l'abbondanza di fumatori rispetto ai non fumatori.

Per ottenere una comprensione più dettagliata della distribuzione dei consumatori di tabacco all'interno del campione, è opportuno esplorare ulteriormente i sottogruppi all'interno dei tre principali gruppi:

1. FUMATORI DI SIGARETTE SENZA COMBUSTIONE

Questo gruppo comprende 109 individui, di cui 79 sono donne e 30 uomini. Con una mediana di età di 22 anni, questi partecipanti rappresentano coloro che hanno abbracciato la tecnologia delle sigarette senza combustione.

Si parla di dispositivi come: l'Iqos, le Sigarette Elettroniche, il Glo e le Sigarette Usa e Getta.

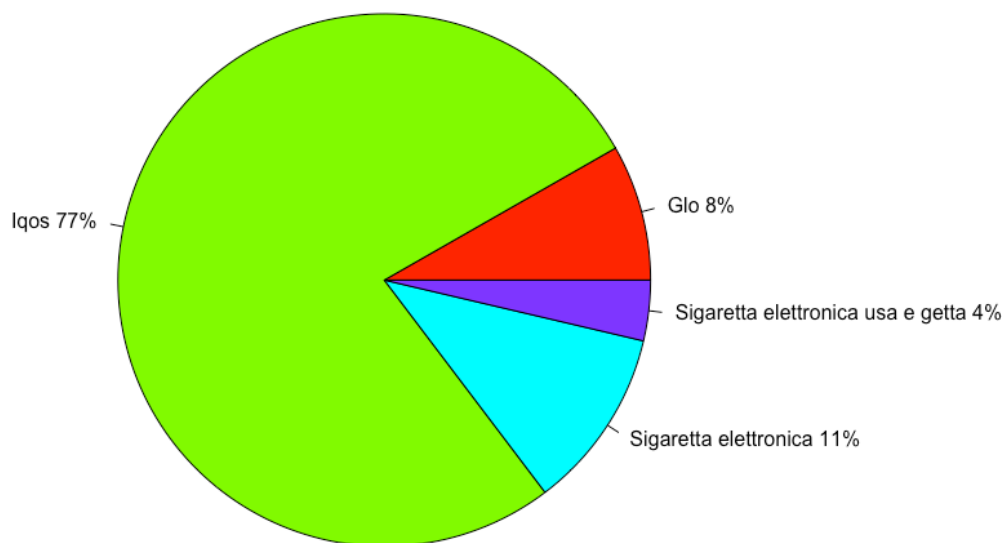


Grafico 2 Percentuali dei dispositivi utilizzati dal gruppo che fuma sigarette senza combustione

Fonte: Creato dall'autore

Il grafico a torta sottolinea le preferenze dei partecipanti per i dispositivi di fumo, rappresentando la tipologia di dispositivi utilizzati dagli individui intervistati, rivela chiaramente che la categoria "Iqos" è la più predominante tra le tipologie di dispositivi per fumare, con la maggior parte degli individui intervistati (circa il 77% del totale) che preferiscono questo dispositivo. La categoria "Glo" rappresenta la seconda scelta più popolare, ma con una percentuale significativamente inferiore (l'8 %). Le "Sigarette Elettroniche" e le "Sigarette Elettroniche Usa e Getta" costituiscono le opzioni meno comuni, rappresentando rispettivamente il 11% e il 4% del totale. Questi risultati mettono in evidenza la forte preferenza dei partecipanti per l'utilizzo dell'Iqos rispetto ad altri dispositivi per fumare all'interno del campione. Queste informazioni sono rilevanti per comprendere le dinamiche di consumo di tabacco tra gli individui intervistati.

In particolare, il dato che il consumo mediano nei dispositivi senza combustione, come 'Iqos', è di 10 sigarette al giorno rivela un aspetto significativo delle

preferenze dei partecipanti.

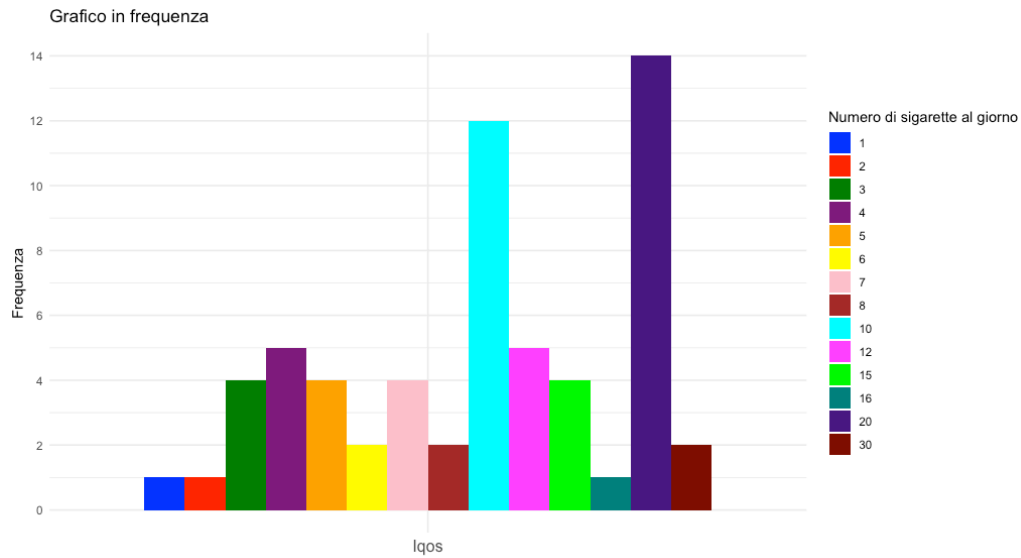


Grafico 3 Distribuzione di frequenza delle sigarette fumate al giorno dalla sotto categoria Iqos

Fonte: Creato dall'autore

Analizzando il grafico a barre, possiamo notare che il consumo di 10 sigarette al giorno è la frequenza più comune tra gli utenti di 'Iqos'.

Questo risultato è in linea con la mediana di consumo e sottolinea ulteriormente la rilevanza di 'Iqos' all'interno del campione.

La concentrazione sull'analisi dettagliata degli utenti di 'Iqos', che rappresentano la maggior parte di questo gruppo, ci ha permesso di identificare in modo chiaro la mediana del consumo quotidiano di 10 sigarette tra questi utenti.

Questa informazione è fondamentale per comprendere le preferenze e le abitudini di consumo specifiche di chi utilizza dispositivi senza combustione, come 'Iqos'.

Alla luce di questi dati, è evidente che le politiche e gli interventi volti a promuovere alternative meno dannose al consumo di tabacco dovrebbero prendere in considerazione l'importanza di dispositivi come 'Iqos' e la frequenza di consumo comune tra gli utenti, contribuendo così a sviluppare strategie più mirate e informate.

Questo gruppo di partecipanti presenta una spesa settimanale mediana di 20 euro per l'acquisto di prodotti legati al tabacco, come indicato dal grafico.

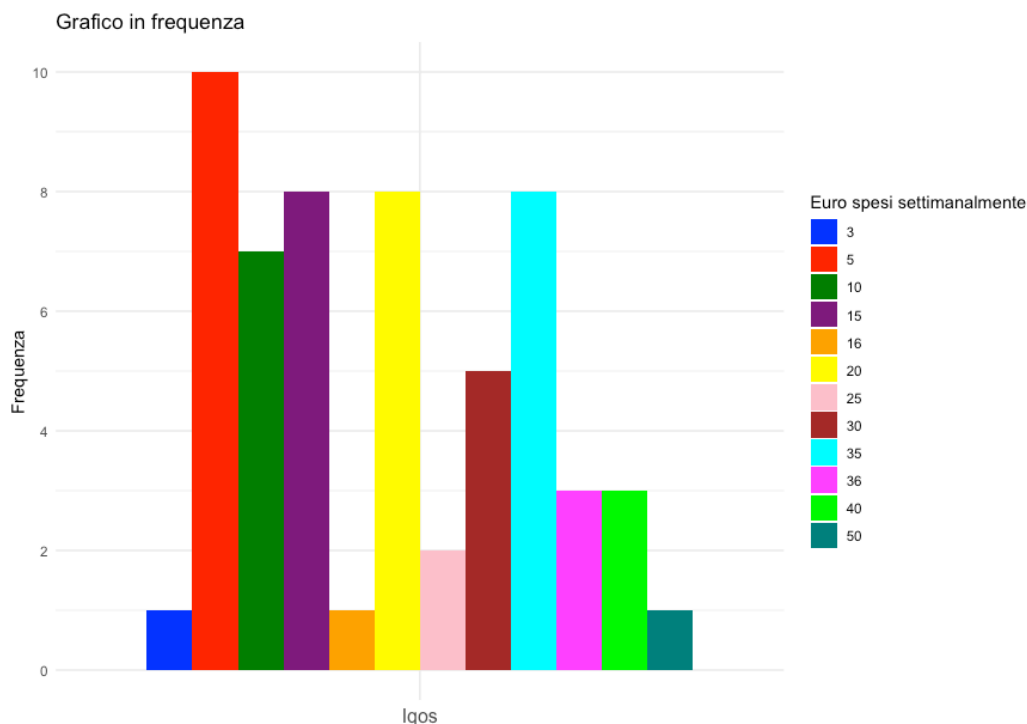


Grafico 4 Distribuzione in frequenza degli euro spesi dal sotto categoria che fumano l'iqos Fonte: Creato dall'autore

L'analisi dei dati ha evidenziato che la maggior parte degli utenti del dispositivo "Iqos" tende a spendere importi variabili, ma la mediana di 20 euro suggerisce che questa sia una cifra rappresentativa della spesa settimanale comune tra gli intervistati. Il grafico a barre mostra la distribuzione delle spese settimanali per i partecipanti che utilizzano "Iqos".

Ogni barra rappresenta una diversa fascia di spesa settimanale in euro (espressa in "Euro spesi settimanalmente"), mentre il colore delle barre indica il numero di partecipanti che rientra in ciascuna fascia di spesa.

L'analisi dei dati evidenzia che la fascia più comune di spesa settimanale tra gli utenti di "Iqos" è di circa 15-20 euro, con 16 partecipanti che rientrano in questa categoria. Ci sono anche altre fasce significative, come 10-15 euro e 20-35 euro, che rappresentano rispettivamente 16 e 16 partecipanti. Tuttavia, la mediana di 20 euro riflette la spesa settimanale centrale, indicando che la maggior parte dei partecipanti ha una spesa prossima a questa cifra.

Questi dati sono importanti perché forniscono informazioni sulla spesa finanziaria associata al consumo di tabacco tra gli utenti di "Iqos". Tale informazione

potrebbe essere utile per comprendere meglio il costo delle alternative al tabacco tradizionale e per sviluppare strategie mirate per la promozione di comportamenti di consumo più salutari.

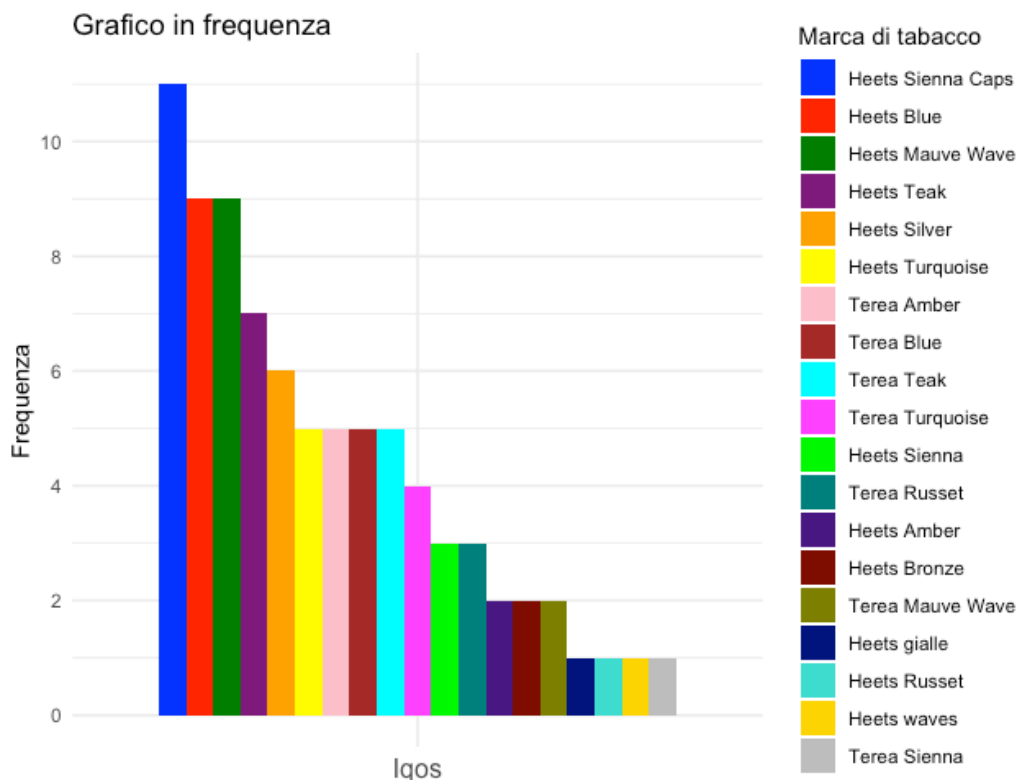


Grafico 5 Distribuzione in frequenza ordinata delle varie tipologie di sigarette fumate Fonte: Creato dall'autore

Tra gli individui intervistati, sono emerse chiare preferenze per determinate tipologie di sigarette senza combustione. Cinque delle varianti più popolari includono Heets Sienna Caps, Heets Blue, Heets Mauve Wave, Heets Teak e Heets Silver. Questi dati riflettono le inclinazioni specifiche dei partecipanti nei confronti delle varie marche o varianti di sigarette senza combustione.

Il grafico a barre offre una rappresentazione visiva della frequenza di utilizzo delle diverse marche o varianti di sigarette senza combustione tra gli utenti di "Iqos". Ogni barra nel grafico rappresenta una specifica marca o variante di sigarette, mentre il colore delle barre indica il numero di partecipanti che utilizzano quella particolare marca.

L'analisi dei dati rivela chiaramente che Heets Sienna Caps è la marca più ampiamente preferita tra gli utenti di "Iqos", con ben 11 partecipanti che scelgono

questa variante. Seguono poi Heets Blue, Heets Mauve Wave, Heets Teak e Heets Silver, con un numero significativo di utenti che le utilizzano regolarmente.

Questi risultati forniscono un'importante panoramica delle preferenze dei consumatori di "Iqos" in termini di marche o varianti di sigarette senza combustione. I dati indicano che queste sottovarianti godono di una buona accettazione tra gli utenti di "Iqos" prima dell'introduzione di "iluma".

L'uscita del nuovo dispositivo Iqos "iluma" ha portato con sé anche una serie di sigarette specificamente progettate per questo dispositivo, le sigarette di tipo Terea che rappresentano una nuova opportunità di crescita per questo tipo di prodotti. Attraverso il grafico possiamo notare che le frequenze relative delle sottovarianti di Terea sono effettivamente più basse rispetto alle corrispondenti sottovarianti di Heets che sono state introdotte precedentemente sul mercato.

Questi dati indicano che al momento dell'analisi le sottovarianti di Terea non sono state ampiamente adottate rispetto alle varianti di Heets già disponibili sul mercato. Questa differenza nelle frequenze suggerisce che le sigarette Terea e il relativo dispositivo "iluma" potrebbero non essere ancora così popolari come le varianti di Heets tra gli utenti di "Iqos".

Se "iluma" e le sigarette Terea vengono ben accettate dal mercato degli utenti di "Iqos," potremmo aspettarci una crescita della frequenza di utilizzo delle sottovarianti Terea nelle future analisi. Questo potrebbe essere indicativo di una tendenza generale di preferenza per le sigarette senza combustione in dispositivi avanzati come "iluma".

Le risposte alle domande relative ai problemi di salute respiratoria mostrano un interessante incrocio di dati, con 22 risposte positive e 86 risposte negative tra gli individui intervistati. Questo dato può essere ulteriormente analizzato considerando il grafico incrociato che evidenzia la relazione tra la presenza di problemi di salute respiratoria ("Sì" per risposte positive e "No" per risposte negative) e il numero di anni di consumo di sigarette senza combustione.

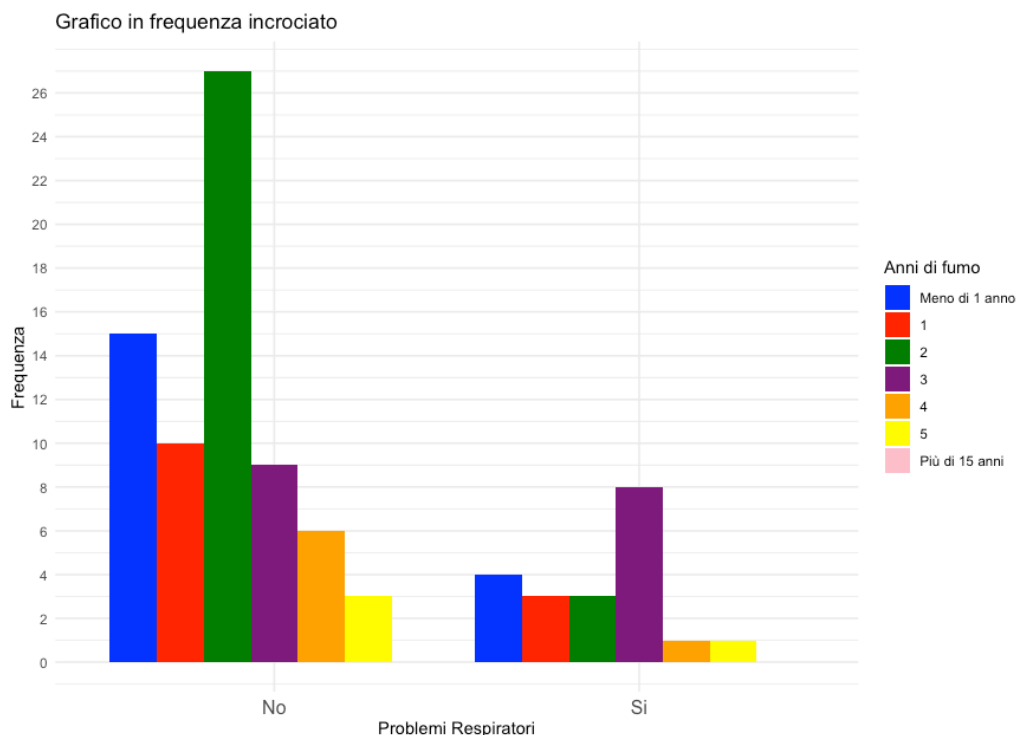


Grafico 6 Grafico a barre multiplo tra le risposte ai problemi respiratori e gli anni di fumo Fonte: Creato dall'autore

Le barre nel grafico rappresentano la frequenza con cui queste combinazioni si verificano nel campione.

L'analisi dei risultati del grafico incrociato rivela alcune tendenze interessanti:

la maggior parte delle risposte positive riguardo ai problemi di salute respiratoria si concentra tra coloro che hanno consumato sigarette senza combustione per un anno o più. Possiamo ipotizzare un'associazione tra l'uso prolungato di questi dispositivi e una maggiore probabilità di riscontrare problemi respiratori.

Tra coloro che hanno risposto "No" alla domanda sui problemi respiratori, la frequenza è significativamente più alta in tutti i gruppi di anni di consumo di sigarette senza combustione, ma è particolarmente elevata tra coloro che non riportano problemi respiratori e che hanno consumato queste sigarette per un anno. È interessante notare che, nonostante la frequenza delle risposte positive sia più bassa rispetto alle risposte negative, esistono alcune risposte positive anche tra coloro che consumano sigarette senza combustione da meno di un anno. Questo potrebbe indicare che i problemi respiratori possono verificarsi anche con un utilizzo relativamente breve.

In generale, l'analisi del grafico incrociato suggerisce una possibile correlazione tra l'uso prolungato di sigarette senza combustione e la presenza di problemi di salute respiratoria. Tuttavia, è importante sottolineare che il grafico non può stabilire una relazione causale, ma solo evidenziare associazioni. Altri fattori, come lo stile di vita e le condizioni di salute individuali, potrebbero influenzare tali risultati. Pertanto, ulteriori studi sono necessari per approfondire questa relazione e confermare le conclusioni.

i. SOTTOGRUPPO DEGLI EX FUMATORI DI SIGARETTE TRADIZIONALI TRA I FUMATORI SENZA COMBUSTIONE

All'interno del gruppo dei fumatori di sigarette senza combustione, troviamo un sottogruppo di 93 individui che avevano precedentemente fumato sigarette tradizionali.

Gli individui ex-fumatori oggetto di studio, che hanno effettuato una transizione verso l'uso di sigarette senza combustione, presentano una mediana di esperienza pari a 2 anni in questo tipo di consumo. Nel loro passato, tali individui fumavano in media 8 sigarette tradizionali al giorno, sostenendo una spesa settimanale mediana di 15 euro.

Per analizzare questa transizione, sono stati condotti due grafici.

Il primo grafico rappresenta un confronto tra la quantità di sigarette al giorno consumate con le sigarette senza combustione rispetto a quelle tradizionali fumate in passato. In questo contesto, ciascun punto nel grafico rappresenta un partecipante, in cui la dimensione del punto è proporzionale al numero attuale di sigarette al giorno senza combustione e il colore del punto è associato al numero di sigarette tradizionali fumate in passato. Questo grafico consente di visualizzare le variazioni nell'uso delle sigarette al giorno in relazione all'uso passato.

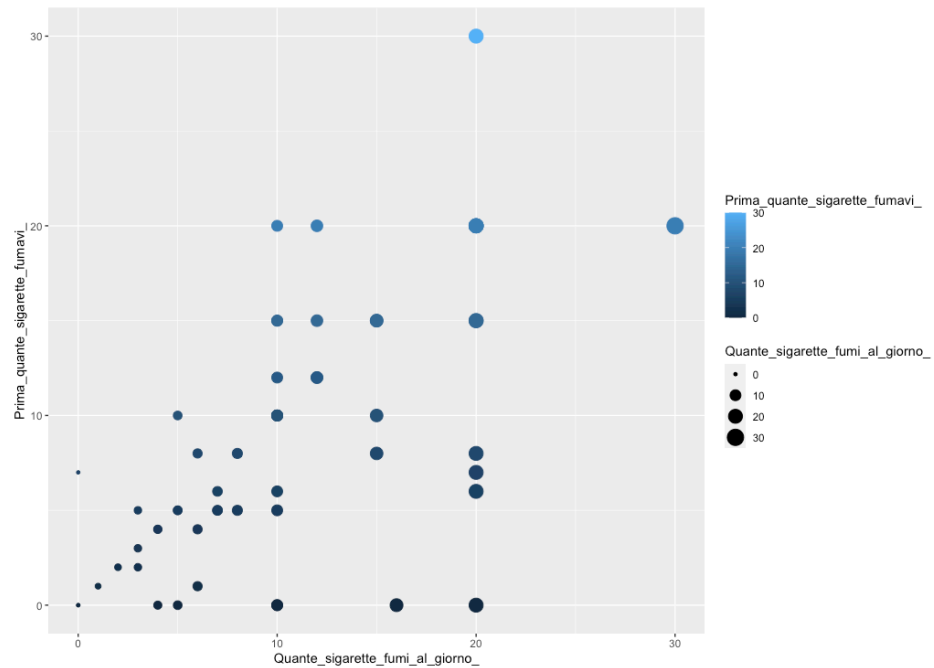


Grafico 7 Grafico a dispersione mettendo a confronto il numero di sigarette tradizionali fumate prima e dopo Fonte: Creato dall'autore

Il secondo grafico focalizza l'analisi sul confronto tra la spesa settimanale per le sigarette senza combustione rispetto a quella sostenuta per le sigarette tradizionali in passato. Anche in questo caso, ogni punto rappresenta un partecipante, in cui la dimensione del punto riflette la spesa settimanale attuale per le sigarette senza combustione e il colore riflette la spesa settimanale passata per le sigarette tradizionali.

Questo grafico fornisce un'analisi approfondita delle tendenze di spesa settimanale in relazione all'uso passato di sigarette tradizionali.

Entrambi i grafici costituiscono strumenti utili per comprendere il cambiamento comportamentale di individui ex-fumatori che abbracciano l'uso di sigarette senza combustione, mettendo in luce le differenze nei consumi e nelle spese tra il passato e il presente.

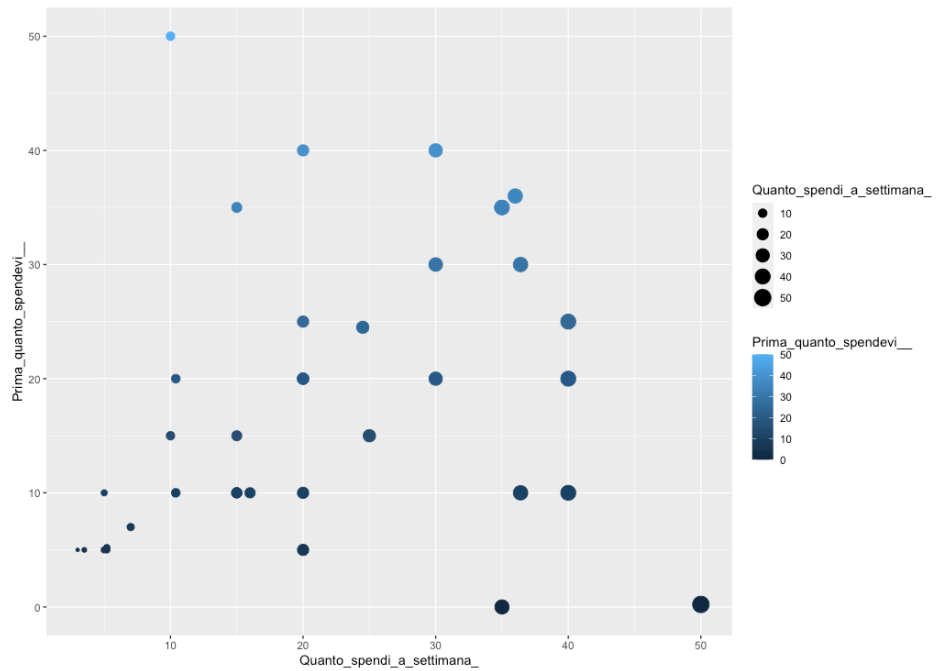


Grafico 8 Grafico a dispersione mettendo a confronto la spesa precedente e attuale per l'acquisto di sigarette Fonte: Creato dall'autore

Osservando il grafico a torta si evince che il 56% degli individui di questo sottogruppo hanno riportato miglioramenti nella loro salute a seguito del passaggio alle sigarette senza combustione riveste notevole importanza all'interno del contesto dell'analisi dei dati. Tali risultati suggeriscono un potenziale impatto positivo del cambiamento dall'uso di sigarette tradizionali a sigarette senza combustione sulla salute degli individui coinvolti.

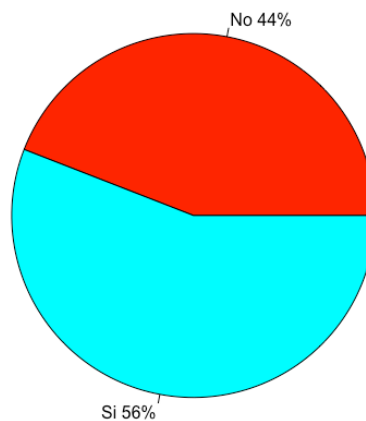


Grafico 9 Grafico a torta con le percentuali delle risposte ad aver avuto miglioramenti dal punto di vista salutare Fonte: Creato dall'autore

2. FUMATORI DI SIGARETTE TRADIZIONALI

Questo campione è composto da un totale di 106 individui, dei quali 56 sono donne e 49 uomini. La mediana dell'età di questo gruppo è di 22 anni.

Il grafico a torta presentato nella figura offre un'affascinante panoramica delle preferenze di consumo di nicotina tra il campione di individui sottoposti all'indagine.

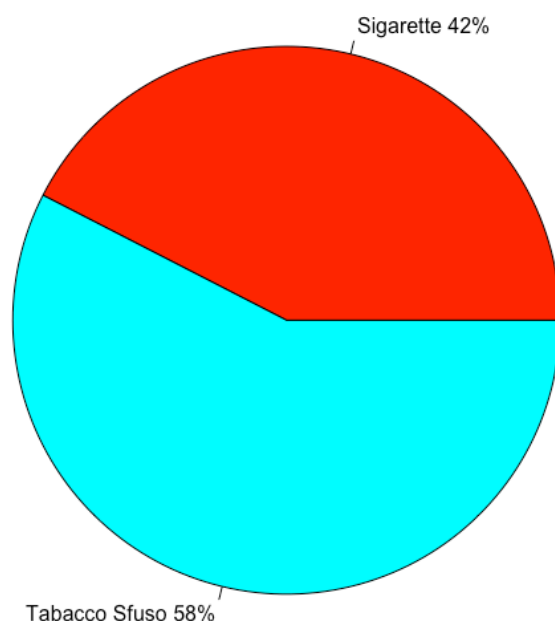


Grafico 10 Tipologia di assunzione di tabacco in percentuale Fonte: Creato dall'autore

Di particolare rilevanza è l'osservazione che il "Tabacco Sfuso" rappresenta il metodo di assunzione più diffuso, con il 58% dei partecipanti che lo ha scelto come preferenza. Questo risultato può essere interpretato in vari contesti, tra cui quello economico. È importante notare che il "Tabacco Sfuso" potrebbe essere preferito per la sua economicità rispetto alle "Sigarette" confezionate. L'accessibilità economica del tabacco sfuso potrebbe rappresentare una decisione ponderata da parte degli individui, specialmente in un'epoca in cui le considerazioni finanziarie svolgono un ruolo significativo nelle scelte quotidiane. L'importanza di considerare non solo le preferenze di consumo ma anche i fattori economici e sociali che possono influenzarle.

In media, gli individui di questo gruppo consumano 8,5 sigarette tradizionali al giorno, osservando i grafici che rappresentano il consumo giornaliero di sigarette da parte di due gruppi distinti possiamo determinare che:

Nel grafico relativo al consumo di tabacco sfuso, è possibile osservare che la maggior parte delle persone consuma da 2 a 10 sigarette al giorno. In particolare, il picco di frequenza si verifica per coloro che fumano 10 sigarette al giorno, con una frequenza di 14. Questo risultato è coerente con l'affermazione di consumo medio di 8,5 sigarette al giorno, in quanto la maggior parte delle persone nel gruppo di consumo di tabacco sfuso si colloca entro questa fascia di consumo.

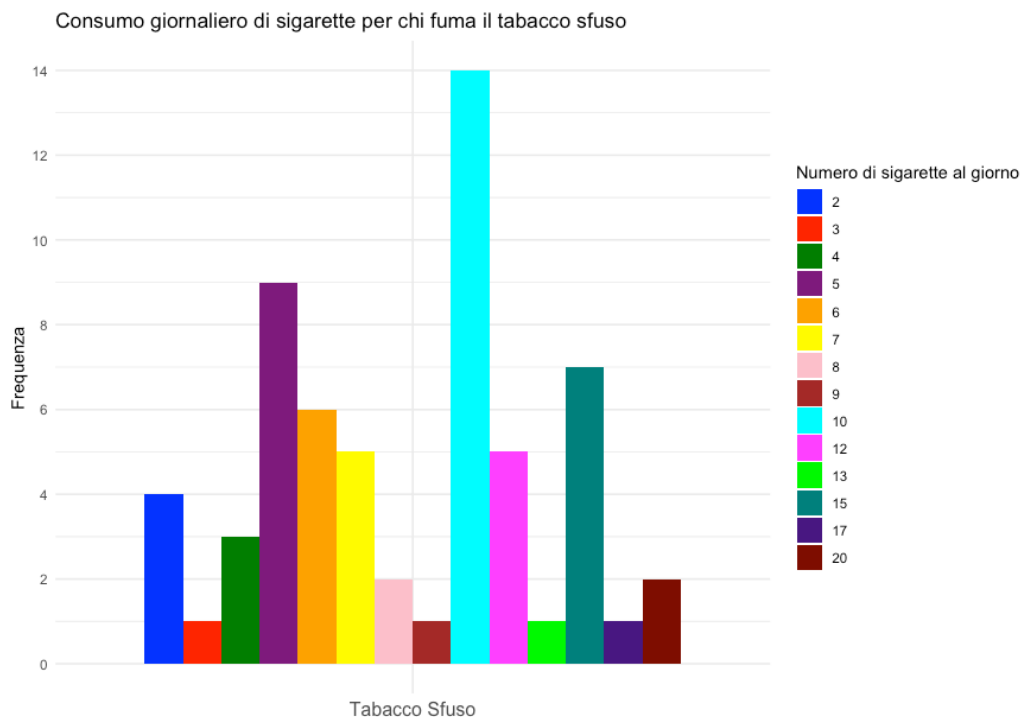


Grafico 11 Distribuzione in frequenza delle sigarette fumate al giorno della sotto categoria che fuma tabacco sfuso Fonte: Creato dall'autore

Il grafico mostra una distribuzione che sembra essere approssimativamente simmetrica e tende a seguire una forma di distribuzione normale o gaussiana. Questo significa che la maggior parte delle osservazioni si concentra intorno al valore medio o centrale e diminuisce gradualmente man mano che ci si sposta verso valori più alti o più bassi.

Non ci sono valori estremi o *outlier* evidenti nel grafico. Tutti i valori del consumo giornaliero di sigarette sembrano raggrupparsi intorno al centro della distribuzione, senza deviazioni significative.

Osserviamo alcuni picchi nei dati intorno a determinati valori specifici, come 5, 10 e 15 sigarette al giorno. Questi picchi indicano che ci potrebbe essere una preferenza o una tendenza significativa tra le persone a consumare un certo numero di sigarette al giorno.

In conclusione, la distribuzione del consumo giornaliero di sigarette per chi fuma il tabacco sfuso sembra seguire una distribuzione approssimativamente normale, con una maggior parte degli individui che consuma un numero moderato di sigarette al giorno e una distribuzione equilibrata senza deviazioni estreme. Questa distribuzione suggerisce che le abitudini di consumo di sigarette tra gli individui inclusi nello studio sono relativamente omogenee e non presentano valori estremi evidenti.

Nel grafico relativo al consumo di sigarette confezionate, il quadro è leggermente diverso. Qui, la maggior parte delle persone consuma da 1 a 10 sigarette al giorno, con un picco di frequenza per coloro che fumano 10 sigarette al giorno, con una frequenza di 8. Sebbene ci siano alcune persone che consumano un numero significativo di sigarette al giorno (ad esempio, 20, 25, 28, 30), la frequenza è relativamente bassa in queste categorie.

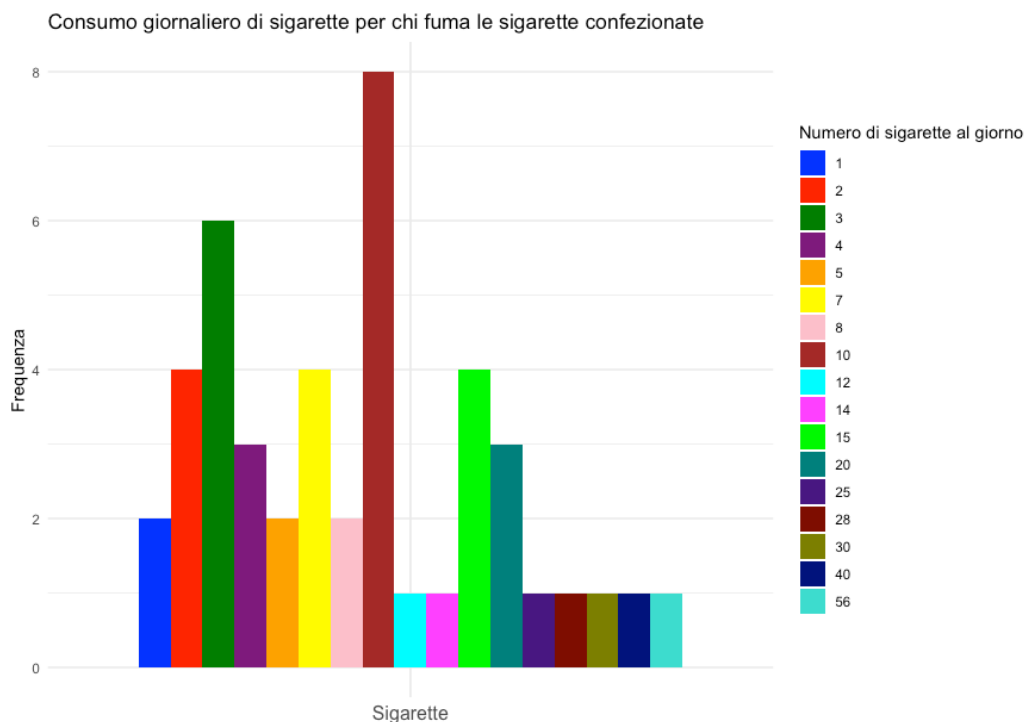


Grafico 12 Distribuzione in frequenza delle sigarette fumate al giorno dalla sotto categoria che fuma sigarette Fonte: Creato dall'autore

La distribuzione dei dati sembra essere asimmetrica e tende a seguire una forma di distribuzione a coda destra, anche nota come distribuzione positivamente asimmetrica o *right-skewed*. Questo significa che la maggior parte delle osservazioni si concentra verso il lato sinistro del grafico, con un numero crescente di individui che consuma un numero sempre maggiore di sigarette al giorno. Ci sono alcuni valori estremi o *outlier* sul lato destro del grafico, indicando che alcune persone consumano un numero significativamente elevato di sigarette al giorno. Questi punti sono visibili come barre più alte e distanti dalla parte centrale del grafico. Osserviamo picchi nei dati intorno a certi valori specifici, come 3, 7, 15, e 20 sigarette al giorno. Questi picchi indicano che ci potrebbe essere una preferenza o una tendenza significativa tra le persone a consumare un certo numero di sigarette al giorno.

In conclusione, la distribuzione del consumo giornaliero di sigarette per chi fuma le sigarette confezionate segue una distribuzione positivamente asimmetrica, con una maggioranza di individui che consuma un numero relativamente basso di sigarette al giorno e un numero limitato di casi estremi con consumi molto elevati. Questa distribuzione suggerisce una variazione significativa nelle abitudini di consumo tra gli individui inclusi nello studio.

In entrambi i grafici, la frequenza del consumo di sigarette tende a diminuire man mano che il numero di sigarette al giorno aumenta. Ciò è coerente con l'affermazione di consumo medio di 8,5 sigarette al giorno, poiché la maggior parte delle persone sembra consumare meno di questa quantità.

In sintesi, i due grafici offrono una rappresentazione visiva dei modelli di consumo di sigarette tra i due gruppi, evidenziando le fasce di consumo più comuni e dimostrando come la frequenza diminuisca con l'aumento del numero di sigarette al giorno. Tale analisi visiva può contribuire a una comprensione più chiara dei comportamenti di consumo di nicotina nei due gruppi distinti.

Facendo riferimento ai dati delle preferenze dei consumatori riguardo alle marche di tabacco, possiamo andare ad analizzare grazie a due grafici a barre la frequenza di consumo delle diverse marche all'interno di ciascuna categoria.

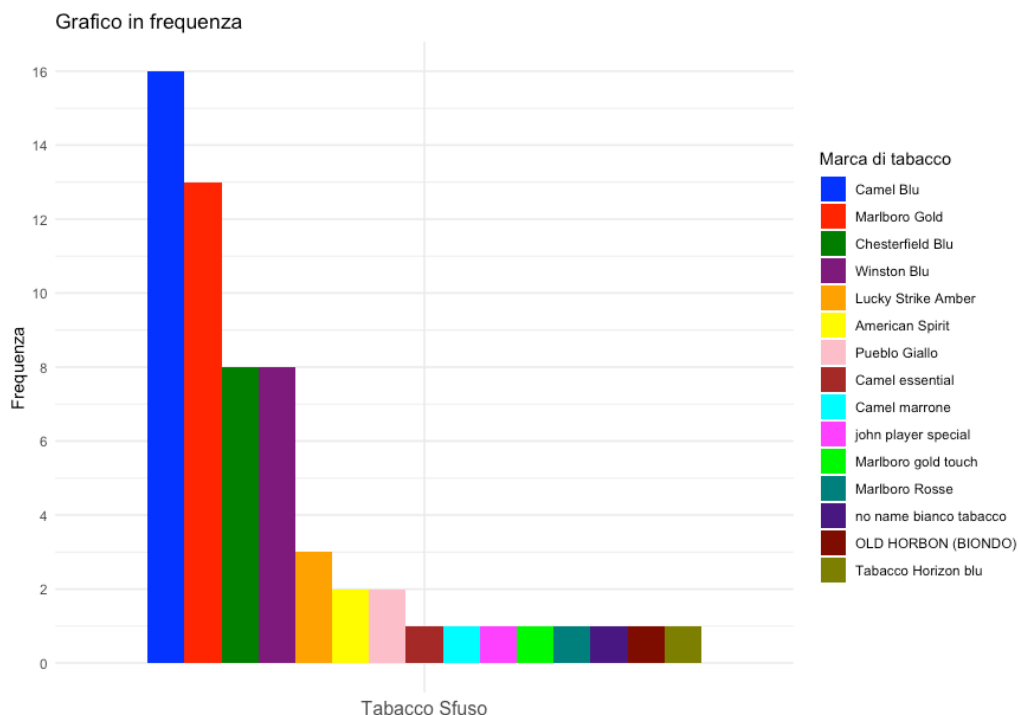


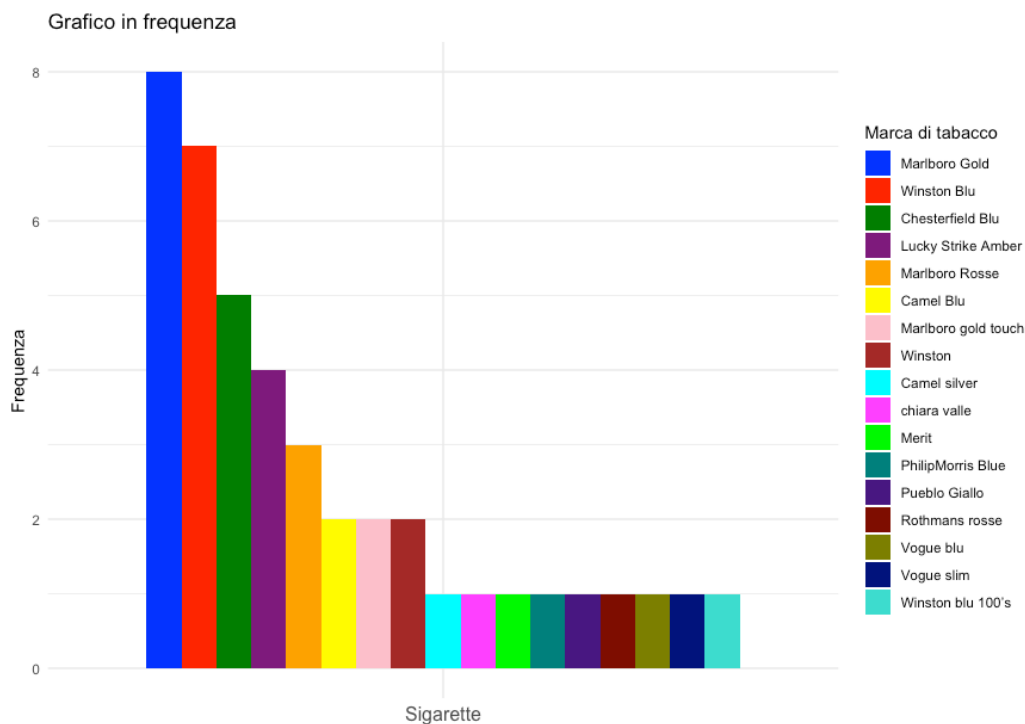
Grafico 13 Distribuzione in frequenza ordinata delle marche fumate dagli individui che fumano tabacco sfuso Fonte: Creato dall'autore

Il primo grafico, riferito al tabacco sfuso, è stato generato utilizzando i dati relativi alle preferenze dei consumatori per le marche di tabacco sfuso.

Questi dati sono stati ottenuti dopo aver escluso le risposte contenenti termini inappropriati o fuori contesto.

Il grafico mostra chiaramente la distribuzione delle preferenze per frequenza delle marche di tabacco sfuso tra i partecipanti all'analisi. Alcune marche, come Camel Blu e Marlboro Gold, sembrano essere più popolari, mentre altre, come American Spirit e Camel essential, sono meno comuni.

Il secondo grafico è relativo alle sigarette e rappresenta la frequenza delle diverse marche di sigarette tra i consumatori. Questo tipo di analisi può essere utile per comprendere quali marche di sigarette siano più popolari o comunemente consumate tra i partecipanti dello studio. Questo tipo di informazione può essere importante per scopi di ricerca di mercato o per valutare eventuali tendenze nel consumo di sigarette. Come nel caso precedente, le risposte inappropriati sono state eliminate per garantire l'integrità dei dati.



*Grafico 14 Distribuzione in frequenza ordinata delle marche fumate dagli individui che fumano sigarette
Fonte: Creato dall'autore*

Da questo grafico emerge che Chesterfield Blu, Marlboro Gold e Winston Blu sono tra le marche più frequentemente consumate tra i partecipanti.

Per interpretare in modo più approfondito queste analisi, è importante considerare alcune osservazioni chiave:

Dai grafici emerge che esistono differenze significative nelle preferenze delle marche tra i consumatori di tabacco sfuso e sigarette. Ad esempio, Marlboro Gold sembra essere una marca popolare in entrambe le categorie, mentre altre marche mostrano differenze notevoli tra i due gruppi di consumatori.

Questi dati possono essere di interesse per l'industria del tabacco e per il settore del marketing. Comprendere quali marche sono più popolari può influenzare le strategie di marketing e pubblicità, nonché le decisioni aziendali sul lancio di nuovi prodotti o sulla promozione di quelli esistenti.

L'analisi delle preferenze delle marche può anche offrire insight sul comportamento dei consumatori. Ad esempio, i consumatori potrebbero basare le loro scelte su fattori come il gusto, la disponibilità o il prezzo delle sigarette o del tabacco sfuso.

In conclusione, l'analisi delle preferenze delle marche di tabacco all'interno di queste due categorie di prodotti fornisce una panoramica delle abitudini di consumo dei partecipanti all'indagine. Tuttavia, è importante ricordare che queste preferenze possono essere influenzate da una serie di fattori e che ulteriori ricerche potrebbero essere necessarie per comprenderle appieno.

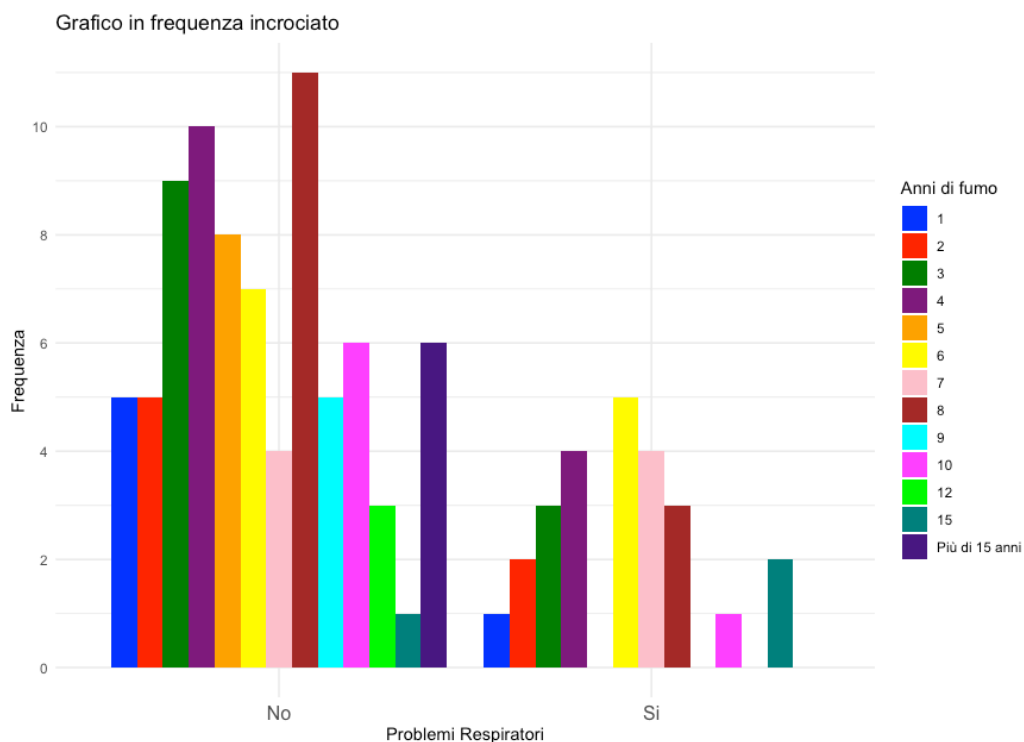


Grafico 15 Grafico a barre multiplo tra le risposte ai problemi respiratori e gli anni di fumo Fonte: Creato dall'autore

Nel contesto dell'analisi dei dati relativi alle risposte fornite in merito a problemi di salute respiratoria, emerge un interessante pattern rappresentato da 25 risposte positive e 81 risposte negative. Questi dati sono stati oggetto di studio tramite un grafico a barre multiplo progettato per esplorare la relazione tra la presenza o l'assenza di problemi respiratori e il numero di anni di fumo, che equivale ad una mediana di 6 anni di esperienza nel fumare sigarette tradizionali all'interno del gruppo di individui che ne fa uso.

Il grafico mostra chiaramente come l'incidenza dei problemi respiratori sia legata al periodo di esposizione al fumo di sigarette tradizionali. Mentre nel caso dei partecipanti che hanno dichiarato di non avere problemi respiratori (categoria "No"), si osserva una distribuzione in cui la maggioranza di essi ha trascorso da 1

a 10 anni fumando, con picchi di frequenza intorno ai 4 e 8 anni di fumo. In contrasto, nei partecipanti che hanno riportato problemi respiratori (categoria "Si"), emerge una distribuzione più ampia, con una maggiore variabilità nell'esperienza di fumo. Tuttavia, si osserva che anche in questa categoria la frequenza tende a concentrarsi maggiormente nei primi 10 anni di fumo, con una diminuzione significativa dopo tale periodo.

Questi risultati suggeriscono che vi sia una relazione tra l'abitudine al fumo di sigarette tradizionali e l'insorgenza di problemi respiratori, con una maggiore probabilità di tali problemi nei primi anni di esposizione al fumo. Tuttavia, è importante sottolineare che questa analisi è di natura descrittiva e non consente di stabilire una relazione causale tra fumo e problemi respiratori. Ulteriori indagini e studi longitudinali sarebbero necessari per confermare tali conclusioni e indagare ulteriormente questa associazione. Inoltre, la distribuzione dei dati nella categoria "Si" sembra seguire una distribuzione unimodale con una leggera tendenza verso una distribuzione asimmetrica a destra (positiva). Questo significa che la maggior parte dei dati "Si" si concentra in un'area specifica, ma potrebbero esserci alcuni valori più elevati che si allontanano da questa concentrazione principale. In altre parole, ci sono individui con problemi respiratori che rientrano in una gamma di valori, ma la maggior parte di essi si trova nella parte inferiore di questa gamma.

Invece, La distribuzione dei dati nella categoria No sembra seguire una distribuzione unimodale con una tendenza più simmetrica o equilibrata. Ciò indica che la maggior parte dei dati No è concentrata intorno a un valore centrale e mostra meno dispersione rispetto alla distribuzione Si. Gli individui senza problemi respiratori sembrano avere risposto in modo più uniforme senza una forte concentrazione in una particolare fascia di valori.

In sintesi, la distribuzione Si è caratterizzata da una maggiore variabilità e una leggera asimmetria positiva, mentre la distribuzione No è più concentrata intorno a un valore centrale ed è meno asimmetrica. Queste differenze potrebbero riflettere variazioni nella gravità dei problemi respiratori tra gli individui inclusi nello studio o altri fattori correlati alle risposte dei partecipanti.

La spesa settimanale mediana per il fumo in questo gruppo è di 10 euro. Questa cifra rappresenta la mediana della spesa settimanale, il che significa che la metà degli individui nel gruppo spende meno di 10 euro a settimana per il fumo, mentre l'altra metà spende più di 10 euro. Per analizzare ulteriormente la spesa settimanale dei partecipanti, sono stati creati due grafici distinti per le due principali tipologie di prodotti del tabacco: tabacco sfuso e sigarette.

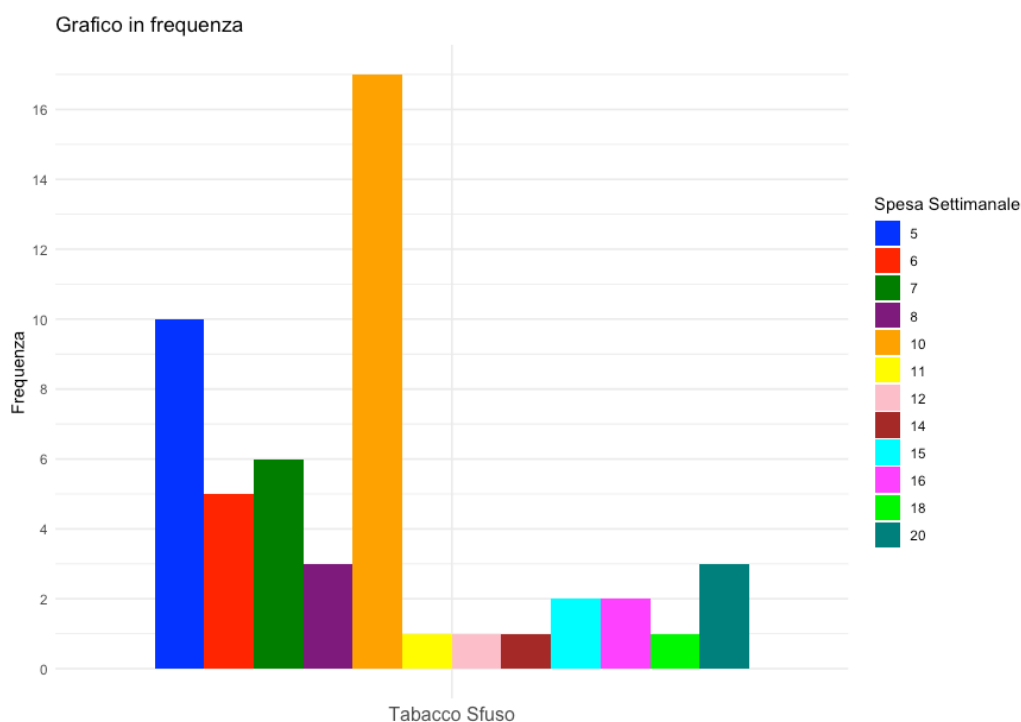


Grafico 16 Distribuzione della spesa settimanale per il tabacco sfuso Fonte: Creato dall'autore

Nel grafico relativo al tabacco sfuso, l'asse delle x rappresenta i livelli di spesa settimanale, mentre l'asse delle y indica la frequenza con cui ciascun livello di spesa settimanale si verifica nel gruppo. Ad esempio, si può notare che la spesa settimanale più comune è di 10 euro, con 17 partecipanti che spendono questa cifra. Questo dato è coerente con la mediana di 10 euro, che suggerisce che molti partecipanti si collocano in questa fascia di spesa. In generale, la distribuzione mostra una varietà di livelli di spesa, con alcune persone che spendono meno e altre di più. La distribuzione della spesa settimanale per il tabacco sfuso sembra essere asimmetrica e potrebbe essere descritta come una distribuzione di tipo positivamente asimmetrica o a destra. Questo significa che la coda della

distribuzione si estende verso destra, indicando che ci sono alcune osservazioni con spese settimanali molto elevate rispetto alla maggioranza degli individui nel gruppo. Nel grafico, si può notare che la maggior parte delle osservazioni si concentra intorno a spese settimanali più basse, come 10 euro, che è anche la mediana. Questo suggerisce che la maggioranza dei partecipanti ha una spesa settimanale relativamente modesta per il tabacco sfuso.

Tuttavia, ci sono alcune osservazioni, sebbene meno numerose, che presentano spese settimanali molto più elevate, come 20 euro o più. Queste osservazioni contribuiscono alla coda destra della distribuzione asimmetrica. Questo tipo di distribuzione suggerisce che, sebbene la spesa settimanale media sia moderata, esistono alcune eccezioni con spese settimanali notevolmente più elevate, il che potrebbe essere interessante da indagare ulteriormente nell'analisi dei dati.

Invece nel grafico relativo alle sigarette, le stesse considerazioni si applicano.

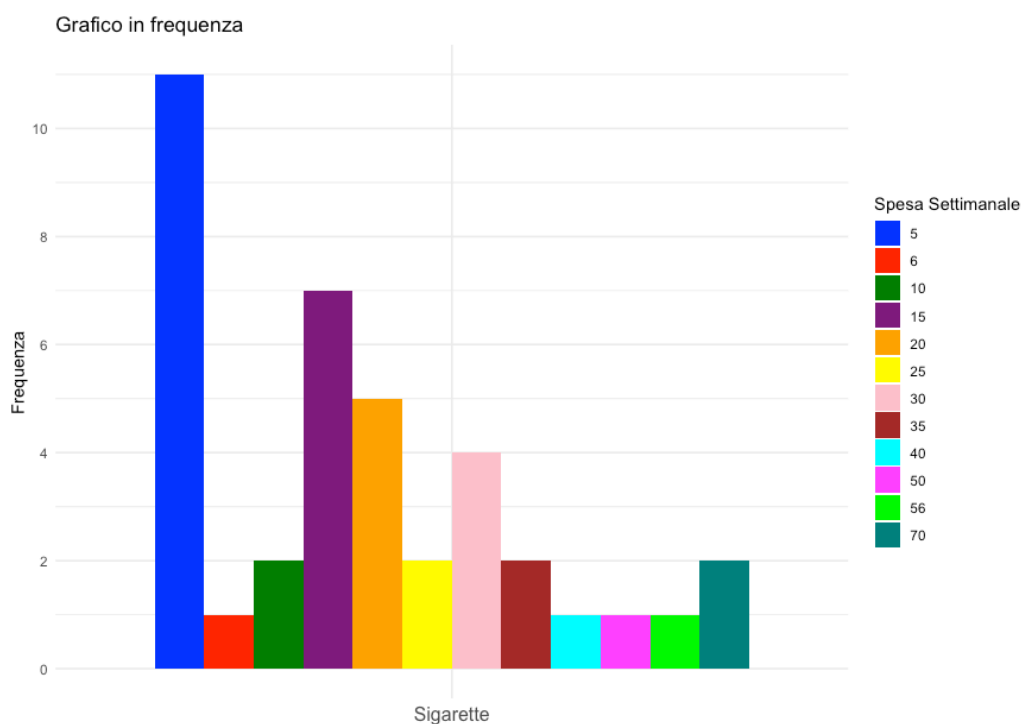


Grafico 17 Distribuzione della spesa settimanale per il tabacco sfuso Fonte: Creato dall'autore

La spesa settimanale più comune è di 5 euro, con 11 partecipanti che spendono questa cifra. Anche qui, la distribuzione evidenzia che la spesa settimanale varia considerevolmente tra i partecipanti.

In entrambi i casi, i grafici mostrano la diversità nella spesa settimanale per il fumo all'interno del gruppo di studio. Mentre la mediana rappresenta una misura centrale, è importante notare che ci sono individui che spendono sia di più che di meno rispetto a questa cifra. La comprensione di queste differenze nella spesa settimanale può essere utile per valutare l'impatto economico del fumo su questo gruppo di partecipanti. La distribuzione rappresentata nel grafico segue una distribuzione asimmetrica positiva, nota anche come distribuzione a coda lunga o skew positivo. Questo tipo di distribuzione è caratterizzato dalla presenza di una lunga coda verso destra, dove si verificano valori estremamente elevati rispetto alla media.

Nel contesto specifico del grafico, la maggior parte delle persone all'interno del gruppo ha spese settimanali moderate per le sigarette, come 5, 10, o 15 euro. Questi valori sono concentrati nella parte sinistra del grafico. Tuttavia, ci sono alcune eccezioni rappresentate da una coda lunga sulla destra, che include valori di spesa molto più elevati, come 50, 56, o 70 euro.

Questa asimmetria positiva indica che, sebbene la maggior parte delle persone all'interno del gruppo abbia spese relativamente basse o moderate per le sigarette, alcune persone hanno una spesa molto elevata, contribuendo a una variazione significativa all'estremo superiore della distribuzione. Questa variazione è tipica di comportamenti di spesa non uniformi all'interno del gruppo e suggerisce che alcune persone potrebbero avere una maggiore dipendenza dal fumo o preferire marche di sigarette più costose.

Inoltre, la distribuzione a coda lunga indica che le spese settimanali molto alte sono rare ma possono avere un impatto significativo sulla media delle spese, portando a una mediana relativamente bassa rispetto al valore massimo delle spese osservate.

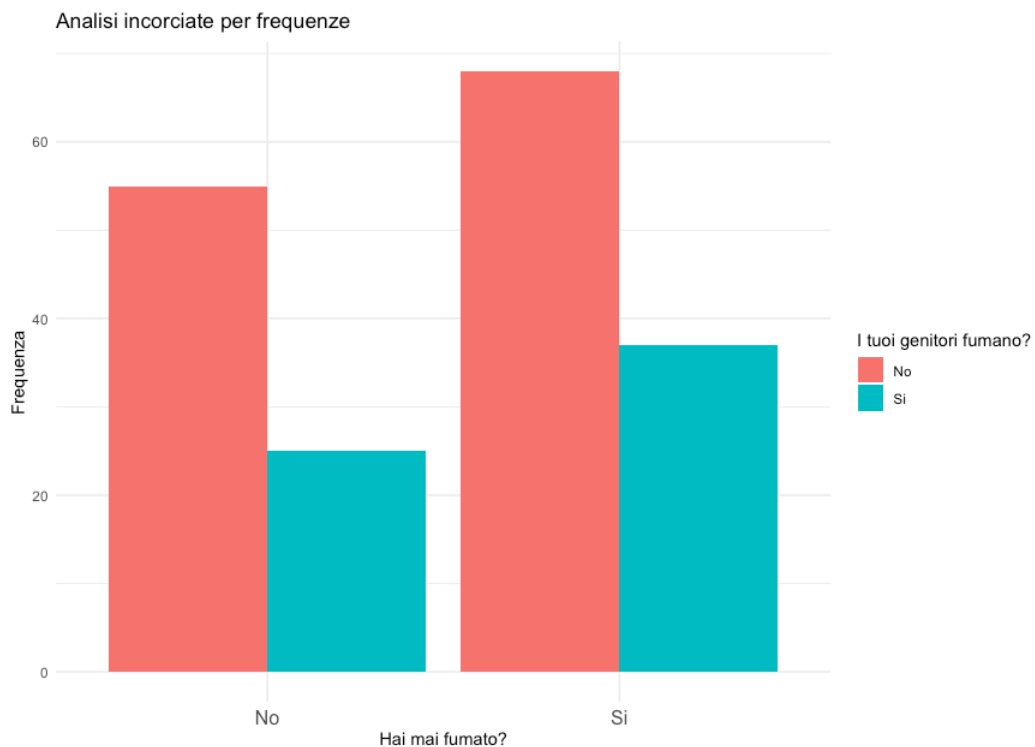
In sintesi, la distribuzione delle spese settimanali per le sigarette in questo gruppo segue una distribuzione asimmetrica positiva, con la maggior parte delle persone che spende moderatamente e alcune persone con spese molto elevate.

Infine, in questo gruppo di individui si evidenzia che 57 individui hanno genitori non fumatori, mentre 49 hanno almeno un genitore fumatore.

3. NON FUMATORI

Il gruppo dei non fumatori è composto da 186 individui, tra cui 132 sono donne e 51 sono uomini, con una mediana di età di 23 anni.

Dall'analisi incrociata rappresentata nel grafico, emergono due categorie principali. La prima categoria comprende coloro che non hanno mai fumato ("No"), con un totale di 80 individui. La seconda categoria include gli ex fumatori ("Si"), per un totale di 105 partecipanti.



*Grafico 18 Grafico a barre incrociando i dati relativi ad aver mai fumato ed avere un genitore che fuma
Fonte: Creato dall'autore*

Questo grafico offre un'analisi incrociata tra due variabili categoriche: "Hai mai fumato?" e "I tuoi genitori fumano?". È possibile osservare come la risposta alla prima variabile influenzi la seconda, evidenziando eventuali associazioni tra il proprio comportamento di fumo e l'abitudine dei propri genitori.

Le variabili "Hai mai fumato?" e "I tuoi genitori fumano?" sono entrambe variabili categoriche binarie, dove i partecipanti sono assegnati a una delle due categorie. Di conseguenza, queste variabili seguono una distribuzione di Bernoulli, dove ogni categoria ha una probabilità di successo (p) e di fallimento ($1 - p$).

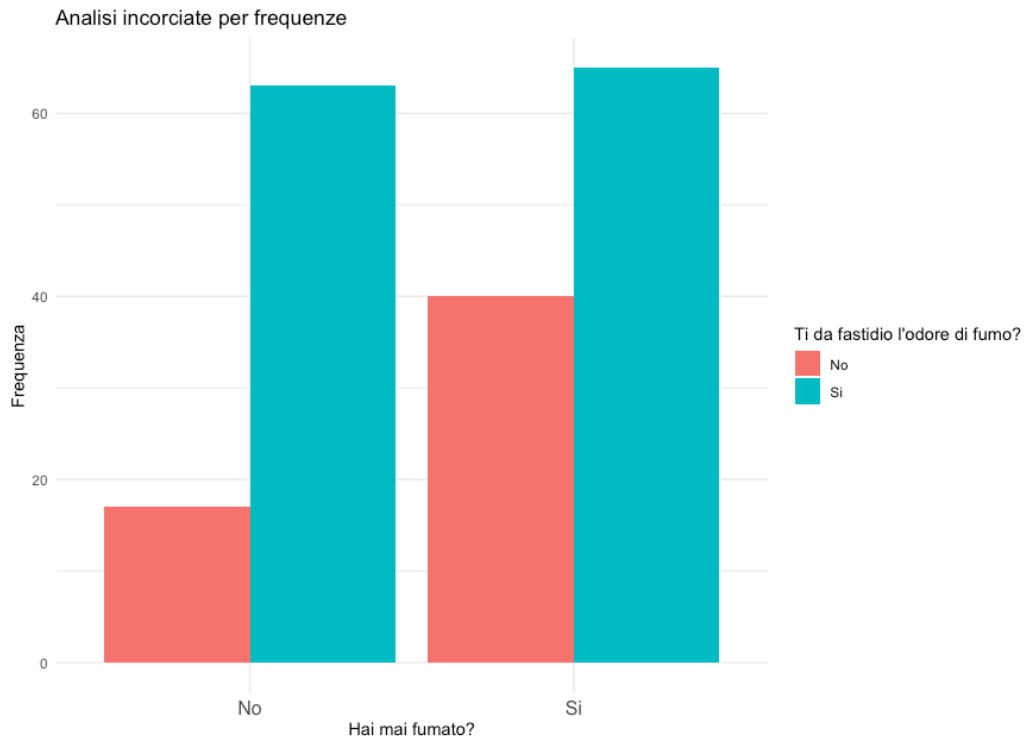


Grafico 19 Grafico a barre incrociando i dati relativi ad aver mai fumato ed essere intolleranti all'odore di fumo Fonte: Creato dall'autore

L'analisi incrociata evidenzia due principali categorie di partecipanti. La prima categoria comprende coloro che non hanno mai fumato ("No"), con 17 individui che non sono infastiditi dall'odore del fumo, contrariamente 63 individui ne sono infastiditi. Mentre la seconda categoria dei non fumatori ("Si"), vi sono 65 individui infastiditi dall'odore del fumo e i rimanenti 40 di questo gruppo non ne sono infastiditi.

Questo grafico rappresenta un'analisi incrociata tra "Hai mai fumato?" e "Ti da fastidio l'odore di fumo?". Attraverso questa visualizzazione, si esplora se i non fumatori siano maggiormente infastiditi dall'odore del fumo rispetto agli ex fumatori.

Le variabili "Hai mai fumato?" e "Ti da fastidio l'odore di fumo?" sono anch'esse variabili binarie, dove ogni partecipante è assegnato a una delle due categorie. E anche in questo caso, seguono una distribuzione di Bernoulli.

Infine, l'ultima analisi incrociata rivela diverse categorie. Alcuni partecipanti non hanno mai fumato ("No") e non hanno alcuna forma di dipendenza (79 individui). Altri hanno dichiarato di avere una dipendenza dall'alcol ("Alcol" - 4 individui), dalla droga ("Droga" - 3 individui) o dalle scommesse ("Scommesse" - 1

individuo). Coloro che hanno fumato ("Si") sono 98 individui e non presentano dipendenze.

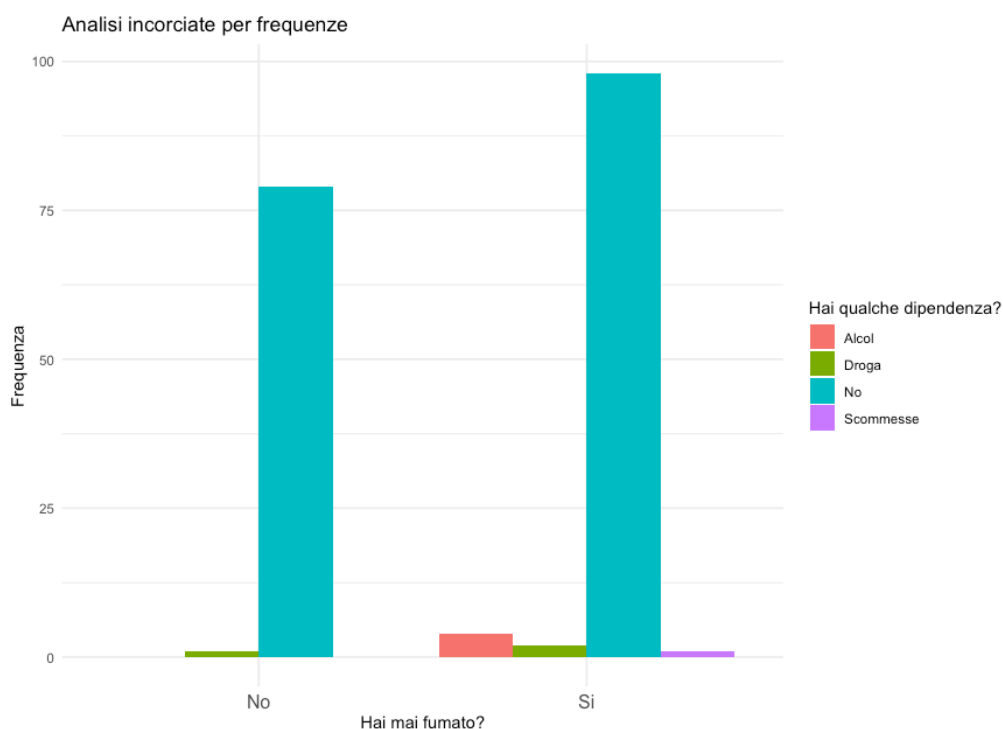


Grafico 20 Grafico a barre multiplo tra l'aver mai fumato ed avere qualche dipendenza Fonte: cerato dall'autore

Questo grafico rappresenta l'analisi incrociata tra "Hai mai fumato?" e "Hai qualche dipendenza?". L'obiettivo è valutare se vi sia una correlazione tra l'esperienza di fumo e l'eventuale presenza di dipendenze da alcol, droga o scommesse.

La variabile "Hai mai fumato?" è una variabile binaria, seguendo una distribuzione di Bernoulli. D'altra parte, la variabile "Hai qualche dipendenza?" è categorica con più di due categorie (alcol, droga, scommesse), quindi segue una distribuzione multinomiale.

Questa analisi descrittiva generale ci offre una comprensione approfondita dei dati raccolti e costituisce la solida base per ulteriori analisi e conclusioni nella nostra ricerca sul tabagismo in Italia.

MCA (ANALISI DELLE CORRISPONDENZE MULTIPLE)

La *Multiple Correspondence Analysis* (MCA), nota anche come Analisi delle Corrispondenze Multiple, rappresenta una tecnica di analisi multivariata avanzata ampiamente utilizzata in ambito statistico e scientifico. Essa si colloca nell'ambito delle analisi delle corrispondenze, le quali sono destinate all'analisi e all'interpretazione di dati categorici o nominali. La MCA, in particolare, è una tecnica che si prefigge di esplorare e rappresentare le relazioni sottostanti tra variabili categoriche, consentendo una visualizzazione efficace e una comprensione approfondita della struttura dei dati multidimensionali. Da un punto di vista statistico, la MCA è una tecnica di riduzione delle dimensioni che si basa su principi di linearità e ortogonalità. Essa opera attraverso la costruzione di una serie di tavole di contingenza tra le variabili categoriche in esame e successivamente sfrutta la decomposizione spettrale di queste tavole per ridurre la complessità dei dati originali. Le nuove dimensioni ottenute, chiamate "dimensioni di corrispondenza", sono interpretate come combinazioni lineari delle variabili originali e permettono di evidenziare le relazioni e le strutture nascoste nei dati.

La MCA trova applicazione in una vasta gamma di discipline, inclusi studi di marketing, sociologia, scienze politiche, ecologia e molti altri campi, in cui è cruciale esplorare e rappresentare le relazioni tra variabili categoriche. Questo capitolo è dedicato a una discussione approfondita della teoria, dei metodi e delle applicazioni della MCA, offrendo una prospettiva critica sull'impiego di questa tecnica nell'ambito della ricerca scientifica e statistica.

DESCRIZIONE DELLA METODOLOGIA MCA E LA SUA APPLICAZIONE NELL'AMBITO DELL'ANALISI DEI DATI CATEGORICI

L'analisi delle corrispondenze multiple (MCA) rappresenta un approccio di rilievo nell'ambito dell'analisi multivariata dei dati, in particolare nei contesti che coinvolgono variabili categoriche o nominali. Questa tecnica è ampiamente utilizzata per esplorare le associazioni tra le categorie di variabili categoriche, ad esempio, indicatori di identità nazionale o altre variabili con molteplici categorie.

Per iniziare, si aggregano tutte le tavole incrociate delle variabili in esame, ottenendo una matrice quadrata di tabelle incrociate, comunemente denominata "matrice di Burt"

(indicata come C nella letteratura CA). Un approccio alternativo implica la costruzione di una "matrice indicatore" basata sui dati originali. Questa matrice indicatore, spesso chiamata "matrice di variabili dummy," è una tabella in cui le righe rappresentano gli intervistati e le colonne rappresentano le categorie di risposta. Gli elementi della matrice sono principalmente zeri, tranne per le posizioni che indicano le categorie di risposta di ciascun intervistato.

La relazione tra la matrice di Burt e la matrice indicatore è definita come:

$$C = Z^T Z$$

Nel contesto dell'analisi delle corrispondenze multiple (MCA), questa tecnica è impiegata per esplorare le relazioni complesse tra più di due variabili categoriche. In MCA, non vi è alcuna distinzione tra variabili descrittive e variabili da descrivere, come si può avere nelle analisi delle corrispondenze ordinarie (CA), che si applicano a singole tabelle o pile di tabelle. In MCA, tutte le variabili condividono lo stesso status.

Sono presentati due approcci distinti per condurre l'analisi delle corrispondenze multiple:

Correlazione Canonica: Questo approccio esamina misure di associazione tra variabili categoriche. Diverse misure di associazione sono utilizzate in base al tipo di dati categorici (nominale o ordinale) e alla natura della relazione tra le variabili. È importante scegliere attentamente la misura di associazione in base alla tipologia dei dati.

Approccio Geometrico: Questo approccio adotta una prospettiva geometrica, utilizzando la teoria della "*Singular Value Decomposition*" (SVD) per stabilire una mappa della tabella di contingenza. Questa decomposizione matriciale permette di ottenere tre componenti principali: la matrice U , la matrice Σ (Sigma) e la matrice V . Queste componenti forniscono informazioni sulle relazioni tra righe e colonne dei dati.

Per il ridimensionamento dei dati, è possibile utilizzare le coordinate principali ottenute dalla SVD. Queste coordinate, ottenute a partire dalle coordinate standard moltiplicate per i valori singolari corrispondenti, consentono una rappresentazione efficace dei dati in una mappa.

Nel contesto della distanza chi-quadrato, la CA è giustificata principalmente dalla sua applicazione a tabelle a due vie, utilizzando la distanza chi-quadrato come misura di dissomiglianza tra profili di riga e profili di colonna. Tuttavia, quando si applica MCA alla matrice del superindicatore Z o alla matrice di Burt, l'uso della distanza chi-quadrato è meno giustificato.

Il biplot, una componente dell'approccio geometrico, è utilizzato per visualizzare congiuntamente le righe e le colonne dei dati in una mappa, offrendo una prospettiva completa sull'associazione tra le variabili. Questo approccio è particolarmente utile per la visualizzazione dei dati e condivide somiglianze con l'analisi delle componenti principali. In sintesi, l'analisi delle corrispondenze multiple (MCA) rappresenta una tecnica sofisticata per esplorare le relazioni complesse tra variabili categoriche, offrendo due approcci principali: correlazione canonica e approccio geometrico. Entrambi forniscono una prospettiva approfondita sull'associazione tra variabili categoriche in un contesto multivariato.

La MCA si sviluppò in Francia negli anni '70 e gradualmente divenne uno standard per l'analisi dei questionari. La "matrice di Burt" venne ampiamente utilizzata in questo contesto. Si sviluppò anche un software specifico per l'analisi MCA.

Le fasi dell'analisi MCA seguono un approccio simile a quello dell'Analisi delle Componenti Principali (PCA):

Costruzione della tabella individui \times variabili:

Questa fase coinvolge l'analisi e la codifica preliminare dei dati, la scelta degli individui attivi e supplementari, delle variabili attive e supplementari, e dei fattori strutturanti.

Interpretazione degli assi:

Qui si determinano gli autovalori, le coordinate principali e le contribuzioni delle categorie agli assi, e si decide quanti assi interpretare. L'interpretazione degli assi implica l'esame delle domande rilevanti e delle categorie significative.

Esplorazione del gruppo di individui:

Questa fase implica l'esplorazione del gruppo di individui alla luce delle domande di interesse e la classificazione euclidea degli individui, con successiva interpretazione.

L'analisi MCA trova ampio utilizzo nella ricerca scientifica, soprattutto nell'analisi di questionari e dati che coinvolgono variabili categoriche. Ad esempio, Bourdieu e la sua scuola hanno utilizzato l'Analisi delle Corrispondenze Multiple per rappresentare il concetto di spazio sociale e le relazioni tra categorie e individui. Offre una rappresentazione di due spazi complementari: lo spazio delle categorie e lo spazio degli individui. Questi spazi sono utilizzati per spiegare le posizioni degli individui nello spazio

sociale in base a fattori strutturanti e per comprendere come le posizioni degli individui influenzino le loro posizioni su questioni politiche o ambientali, ad esempio.

ALGORITMO DI IMPUTAZIONE MULTIPLA MICE

La "*multiple imputation*" rappresenta una tecnica statistica avanzata utilizzata per mitigare gli effetti dei dati mancanti all'interno di analisi quantitative. La sua applicazione può risultare complessa per i neofiti, poiché differisce in modo sostanziale dagli approcci statistici convenzionali. In conseguenza di ciò, i ricercatori alle prime armi possono trovarsi in difficoltà nel navigare tra modelli di imputazione, meccanismi di dati mancanti, molteplici versioni dei dati, processi di combinazione (pooling), e così via.

Nella presente sezione, si presenta un approccio modulare alla "*multiple imputation*", noto come MICE (*Multivariate Imputation by Chained Equations*). La filosofia alla base della metodologia MICE sottolinea l'importanza di suddividere il processo di imputazione multipla in una serie di passaggi più semplici, ognuno dei quali può richiedere una verifica diagnostica.

APPROCCIO MODULARE ALLA MULTIPLE IMPUTATION

La *multiple imputation* è una procedura che comprende tre fasi principali: l'imputazione dei dati mancanti, l'analisi dei dati imputati e la combinazione dei risultati ottenuti. In questa metodologia, i risultati di ciascuna fase sono archiviati in classi specifiche: *mids*, *mira* e *mipo*. A seguire, esamineremo in dettaglio ciascuna di queste fasi.

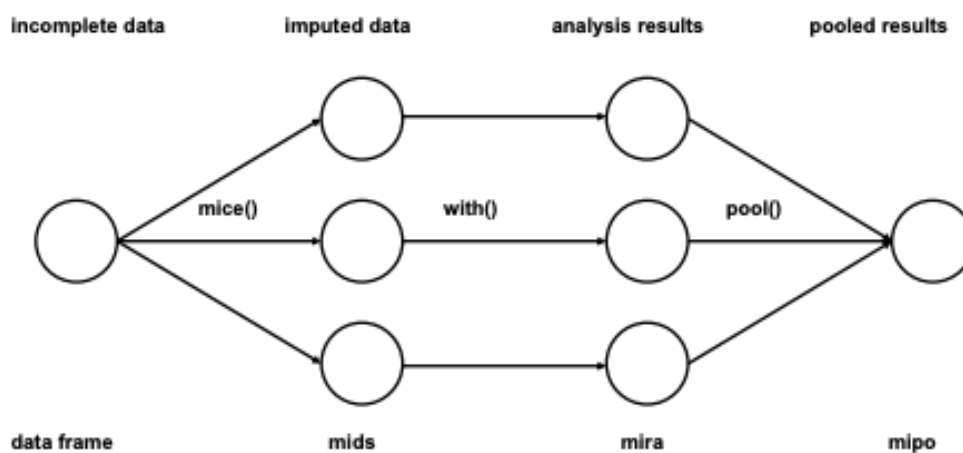


Figura 1 Spiegazione grafica dei processi atti all'imputazione con la libreria *Fonte: Journal of Statistical Software*

Iniziamo con un insieme di dati osservati e incompleti (Y_{obs}), rappresentato nella parte sinistra dell'immagine. Il problema principale consiste nel fatto che non è possibile stimare la quantità di interesse Q dai dati osservati Y_{obs} senza effettuare delle assunzioni irrealistiche sui dati non osservati. La multiple imputation rappresenta un approccio generale che prevede la creazione di diverse versioni imputate dei dati, sostituendo i valori mancanti con valori plausibili. Tali valori plausibili sono estratti da distribuzioni specificamente modellate per ciascuna voce mancante. In MICE, questa operazione viene eseguita attraverso la funzione `mice()`. La figura 1 raffigura l'esistenza di $m = 3$ set di dati imputati $Y(1), \dots, Y(3)$. Questi tre set di dati imputati sono identici per le voci di dati non mancanti, ma differiscono per i valori imputati, il cui scarto riflette l'incertezza riguardo al valore da imputare. È degno di nota che il pacchetto MICE utilizza una classe appositamente progettata per conservare i dati imputati: un insieme di dati imputati multipli della classe `mids`.

La seconda fase prevede la stima della quantità di interesse Q su ciascun set di dati imputati. Questa stima è solitamente ottenuta attraverso l'applicazione del metodo che sarebbe stato utilizzato se i dati fossero stati completi. Questo passaggio è semplificato dal fatto che tutti i dati sono ora completi nei set di dati imputati. Il modello applicato ai dati $Y(1), \dots, Y(m)$ è generalmente lo stesso. La funzione `with.mids()` contenuta in MICE 2.9 esegue questa analisi, rimpiazzando le funzioni `lm.mids()` e `glm.mids()`. Le stime $Q^{(1)}, \dots, Q^{(m)}$ varieranno tra loro a causa delle differenze nei dati di input. È essenziale comprendere che queste differenze derivano dall'incertezza riguardo ai valori da imputare. In MICE, i risultati dell'analisi sono archiviati collettivamente in un oggetto R di classe `mira`, rappresentando un'analisi ripetuta multipla imputata.

L'ultima fase consiste nell'aggregare le m stime $Q^{(1)}, \dots, Q^{(m)}$ in una stima unica Q^- e nello stimare la sua varianza. Per le quantità Q approssimativamente distribuite normalmente, è possibile calcolare la media delle $Q^{(1)}, \dots, Q^{(m)}$ e sommare le varianze intra-imputazione e inter-imputazione seguendo il metodo delineato da Rubin (1987, pp. 76-77). La funzione `pool()` contiene i metodi per

l'aggregazione delle quantità secondo le regole di Rubin. I risultati di questa funzione vengono archiviati come un oggetto di risultati aggregati multipli imputati di classe mipo.

ALGORITMO MICE

L'aspetto chiave dell'algoritmo MICE risiede nella specifica di modelli di imputazione che soddisfino i seguenti principi:

Considerazione del processo che ha causato i dati mancanti.

Preservazione delle relazioni nei dati.

Preservazione dell'incertezza relative a tali relazioni.

L'obiettivo è ottenere imputazioni statisticamente corrette per una vasta gamma di quantità Q , come descritto da Rubin (1987, Capitolo 4). Tuttavia, l'imputazione di dati mancanti può incontrare diverse sfide, tra cui:

- Per una data variabile Y_j , i predittori Y - j utilizzati nel modello di imputazione possono anch'essi essere incompleti.
- Può verificarsi una dipendenza circolare, in cui Y_1 dipende da Y_2 e Y_2 dipende da Y_1 a causa di correlazioni tra le variabili.
- In particolare, con un gran numero di variabili p e un campione limitato n , possono sorgere problemi di collinearità e celle vuote.
- Le righe o le colonne dei dati possono essere ordinate, ad esempio, in dati longitudinali.
- Le variabili possono essere di diversi tipi (ad esempio, binarie, non ordinate, ordinate, continue), rendendo teoricamente inappropriata l'applicazione di modelli convenienti, come la normale multivariata.
- La relazione tra Y_j e Y - j potrebbe essere complessa, ad esempio non lineare o soggetta a processi di censura.
- L'imputazione può generare combinazioni impossibili (ad esempio, padri incinti) o distruggere relazioni deterministiche nei dati (ad esempio, punteggi somma).
- Le imputazioni possono essere insensate (ad esempio, temperatura corporea di un morto).

- I modelli per Q che verranno applicati ai dati imputati potrebbero non essere noti (ancora).

Per affrontare queste sfide, l'approccio MICE adotta una strategia che prevede la specifica separata di modelli di imputazione per ciascuna variabile nei dati. Ciò consente di trattare in modo efficace le complessità specifiche delle variabili e dei dati stessi. L'algoritmo MICE ottiene la distribuzione posteriore di θ campionando iterativamente dalle distribuzioni condizionali di ciascuna variabile.

$$P(Y_1|Y_{-1}, \theta_1) \dots P(Y_p|Y_{-p}, \theta_p)$$

Questo approccio, noto come "equazioni concatenate", è un campionamento di Gibbs che effettua disegni iterativi per le variabili imputate.

$$\begin{aligned}\theta_1^{*(t)} &\sim P(\theta_1|Y_1^{\text{obs}}, Y_2^{(t-1)}, \dots, Y_p^{(t-1)}) \\ Y_1^{*(t)} &\sim P(Y_1|Y_1^{\text{obs}}, Y_2^{(t-1)}, \dots, Y_p^{(t-1)}, \theta_1^{*(t)}) \\ &\vdots \\ \theta_p^{*(t)} &\sim P(\theta_p|Y_p^{\text{obs}}, Y_1^{(t)}, \dots, Y_{p-1}^{(t)}) \\ Y_p^{*(t)} &\sim P(Y_p|Y_p^{\text{obs}}, Y_1^{(t)}, \dots, Y_{p-1}^{(t)}, \theta_p^{*(t)})\end{aligned}$$

Grafico 21 Equazione concatenante utilizzando un campionamento di Gibbs Fonte: *Journal of Statistical Software*

dove $Y^{(t)}_j = (Y^{\text{obs}}_j, Y^{*(t)}_j)$ è la j -esima variabile imputata all'iterazione t . Si osservi che le precedenti imputazioni $Y^{*(t-1)}_j$ entrano in $Y^{*(t)}_j$ solo attraverso la sua relazione con altre variabili, e non direttamente. La convergenza può quindi essere piuttosto rapida, a differenza di molti altri metodi MCMC. È importante monitorare la convergenza, ma nella nostra esperienza il numero di iterazioni può spesso essere piccolo, diciamo 10-20. Il nome equazioni concatenate si riferisce al fatto che l'algoritmo MICE può essere facilmente implementato come concatenazione di procedure univariate per compilare i dati mancanti. La funzione `mice()` esegue m flussi in parallelo, ognuno dei quali genera un set di dati imputato. In conclusione, l'approccio MICE rappresenta una soluzione efficace per affrontare il problema dei dati mancanti, consentendo agli analisti di ottenere

stime affidabili e rappresentative nelle loro analisi statistiche, nonostante le sfide legate alla complessità dei dati mancanti nei contesti reali.

PREPARAZIONE DEI DATASET ALL'ANALISI

Nel contesto di un'analisi delle corrispondenze multiple (MCA), è spesso necessario affrontare il problema dei dati numerici continui, che possono rendere complessa l'interpretazione e l'analisi delle relazioni tra variabili. Per affrontare questa sfida e facilitare l'analisi, è comune eseguire una procedura di categorizzazione delle variabili numeriche. La categorizzazione consiste nell'aggregare i valori continui in intervalli discreti, noti come categorie o classi. Questo processo fornisce una rappresentazione più chiara e interpretabile dei dati, permettendo di esplorare le associazioni tra variabili in modo più agevole. Ad esempio, una delle variabili considerate è stata

"Quante_sigarette_fumi_al_giorno___esprimere_il_valore_numerico".

Questa variabile rappresenta il numero di sigarette fumate al giorno. È stata suddivisa in categorie utilizzando intervalli di valori, come da 0 a 5 sigarette, da 5 a 10 sigarette e così via. Lo stesso approccio è stato applicato ad altre variabili, come la durata del consumo di sigarette, la spesa media settimanale per il fumo e l'età dei partecipanti.

L'obiettivo di questa procedura di categorizzazione è quello di preparare i dati per un'analisi delle corrispondenze multiple, semplificando la comprensione delle relazioni tra le variabili e facilitando l'interpretazione dei risultati dell'analisi.

Dopo aver eseguito la categorizzazione delle variabili numeriche, è possibile che i dataset ottenuti contengano ancora dei valori mancanti. I valori mancanti possono rappresentare un ostacolo significativo nell'analisi dei dati, poiché possono influire sulla validità delle analisi statistiche e sulla completezza delle informazioni. Per affrontare questa sfida e garantire la coerenza dell'analisi, si è fatto ricorso alla tecnica dell'imputazione multipla.

L'imputazione multipla è una strategia avanzata per gestire i dati mancanti, che prevede la creazione di diverse versioni complete del dataset, come visto in precedenza. In altre parole, anziché eliminare le osservazioni con dati mancanti o sostituire i valori mancanti con una singola stima, si generano molteplici versioni dei dati, ognuna delle quali riflette una possibile imputazione dei valori mancanti. Questo processo viene eseguito in modo iterativo e coinvolge una serie di passaggi chiave che verranno spiegate successivamente.

L'obiettivo dell'imputazione multipla è quello di fornire stime accurate e rappresentative, tenendo conto dell'incertezza associata ai valori mancanti. Questa tecnica è particolarmente utile quando i dati mancanti non sono completamente casuali e quando esistono relazioni significative tra le variabili. La sua applicazione consente di preservare la completezza dei dati e di ottenere risultati più affidabili nelle analisi statistiche.

Nel contesto dell'analisi di imputazione multivariata, questa affermazione sottolinea l'importanza di questa tecnica quando ci sono dati mancanti non casuali e relazioni complesse tra le variabili. Ora esaminiamo come determinare il numero di imputazioni e iterazioni nell'ambito di questa tecnica:

- **Dimensione del campione e percentuale di dati mancanti:**

Il numero di imputazioni dovrebbe essere sufficiente per garantire una rappresentazione accurata della variabilità dei dati. In generale, una regola empirica è quella di impostare il numero di imputazioni almeno pari alla percentuale di dati mancanti. Ad esempio, se il 10% dei dati è mancante, potrebbe essere opportuno eseguire almeno 10 imputazioni.

- **Grado di incertezza desiderato:**

Un numero maggiore di imputazioni riduce l'incertezza nelle stime imputate. Se è richiesta una maggiore precisione nelle stime, è consigliabile aumentare il numero di imputazioni. Tuttavia, ciò comporterà un costo computazionale più elevato.

- **Convergenza delle stime:**

Monitorare la convergenza delle stime durante le iterazioni è importante. Se le stime mostrano una stabilità significativa dopo un certo numero di iterazioni, è possibile interrompere il processo di imputazione.

La scelta del numero di iterazioni ottimale dipenderà dalla specifica applicazione e dalla complessità dei dati. Più basse sono le percentuali di dati mancanti, più

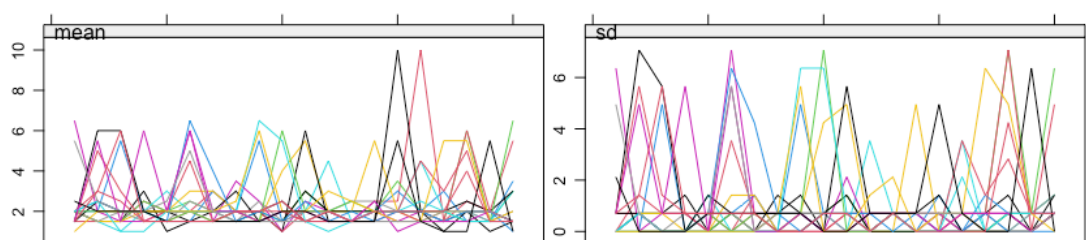


Grafico 22 Andamento delle iterazioni di una variabile imputata Fonte: creato dall'autore

- **Confronto con dati reali:**

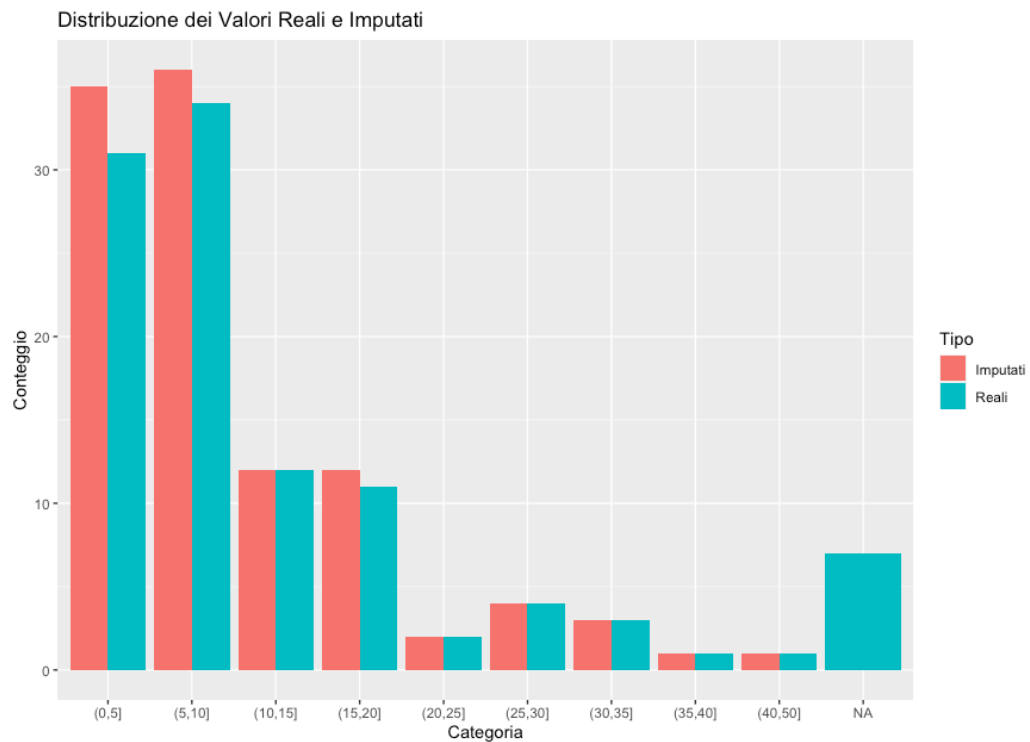


Grafico 23 Grafico a barre multiplo per controllare le categorie dei valori imputati Fonte: creato dall'autore

È possibile utilizzare grafici di confronto tra dati reali e imputazioni per valutare come le stime imputate si avvicinano ai dati osservati. Questo può aiutare a determinare se il numero di imputazioni e iterazioni è sufficiente per ottenere stime accurate.

RISULTATI DELL'MCA UTILIZZANDO L'INTERA POPOLAZIONE

Nel contesto di questo capitolo, esploreremo dettagliatamente i risultati emersi dall'implementazione dell'Analisi delle Componenti Principali Multiple (MCA) sulla totalità della popolazione di studio. In questa sezione, non ci limiteremo semplicemente a presentare i dati numerici derivanti dalla procedura MCA, ma condurremo altresì un'analisi dei grafici prodotti. Tali rappresentazioni grafiche rivestono una rilevanza fondamentale nel contesto della comprensione delle interazioni tra le variabili oggetto di studio.

La nostra analisi MCA è stata eseguita utilizzando l'intero set di dati e adottando il metodo "burt". Inoltre, abbiamo provveduto a rimuovere specifiche variabili di minore interesse, con l'obiettivo di concentrarci in modo più mirato sulle dinamiche sottostanti.

Durante il corso di questo capitolo, ci concentreremo sull'esame approfondito dei risultati ottenuti mediante l'analisi dei grafici generati, con l'intento di acquisire una comprensione completa e dettagliata delle complesse relazioni intercorrenti all'interno della nostra popolazione di studio.

SCREE PLOT (GRAFICO DELLO SCARTO)

Lo scree plot, detto anche grafico dello scarto, costituisce uno strumento di notevole rilevanza per l'interpretazione dei risultati derivanti dall'Analisi delle Componenti Principali (PCA) o, come nel nostro caso, dall'Analisi delle Componenti Principali Multiple (MCA). Tale grafico, di natura baricentrica, offre una rappresentazione visiva della frazione di varianza spiegata da ciascuna componente principale o dimensione dell'analisi.

Nello specifico:

Sull'asse delle ascisse (x), sono riportate le diverse componenti principali o dimensioni, disposte in ordine decrescente in base alla quantità di varianza da esse spiegata. La prima componente principale (PC1) detiene il maggiore contributo nella spiegazione della varianza nei dati, seguita dalla seconda componente principale (PC2), la terza (PC3), e così via.

L'asse delle ordinate (y) rappresenta la varianza spiegata da ciascuna componente principale. Le barre verticali associate a ciascuna componente indicano la proporzione della varianza totale da essa spiegata. Tipicamente, le barre sono disposte in ordine decrescente, riflettendo il contributo decrescente di ciascuna componente alla varianza complessiva.

L'elemento centrale nella lettura dello Scree plot è individuare il "punto di flessione" o "gomito", ossia il punto in cui le barre cessano di diminuire rapidamente e iniziano a presentare un profilo piatto o di diminuzione più lenta.

L'interpretazione cruciale consiste nell'identificare il numero di componenti principali da mantenere nell'analisi, in funzione della percentuale desiderata della varianza totale da conservare. Solitamente, si scelgono le prime componenti principali che spiegano una percentuale significativa della varianza, ad esempio il

70% o l'80%, al fine di semplificare l'analisi preservando al contempo il contenuto informativo sostanziale dei dati.

Pertanto, se osserviamo un punto di flessione ben definito dopo le prime due componenti principali nello scree plot, potremmo optare per l'inclusione esclusiva di PC1 e PC2 nell'analisi, dato che queste componenti esplicano la quota preponderante della varianza e semplificano la comprensione dei dati, parallelo al nostro approccio.

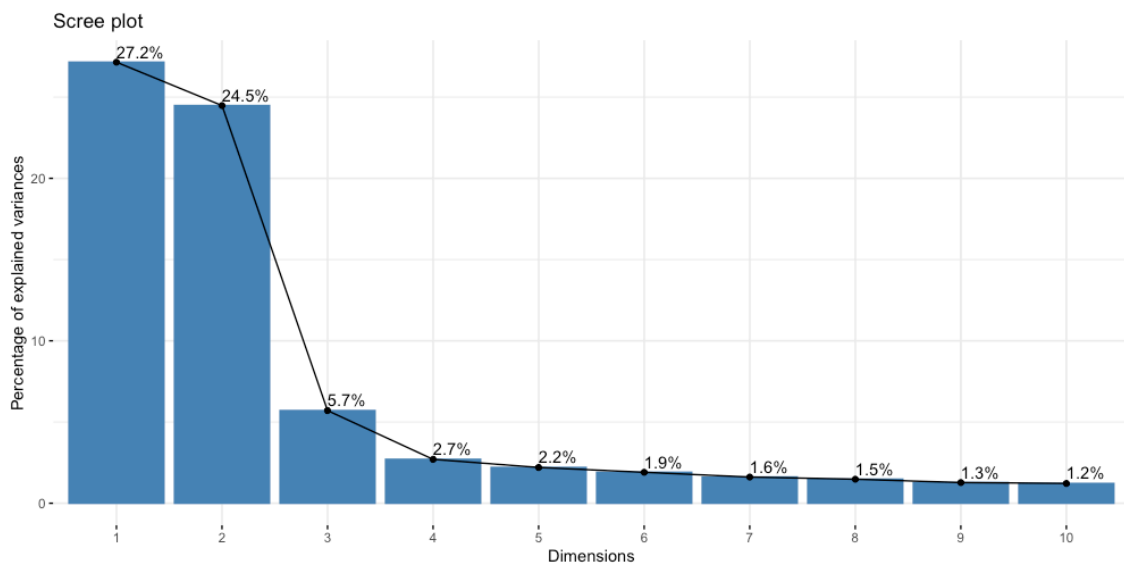


Grafico 24 Varianza dell'intero dataset spiegata da ciascuna componente delle dimensioni dell'analisi Fonte: creato dall'autore

BIPLOT

Il grafico biplot, è una rappresentazione visuale dei risultati dell'analisi applicata a un set di dati contenente variabili categoriche. In particolare, questo grafico è costruito utilizzando le prime due componenti principali (Dim.1 e Dim.2) estratte dall'analisi MCA per spiegare la variazione nei dati.

Interpretazione del Grafico:

- **Punti e Categorie:**

Nel grafico, ogni punto rappresenta una categoria specifica delle variabili.

Le categorie possono includere le diverse tipologie di fumatori, come "Non fumatore", "Fumatore di sigarette tradizionali", "Fumatore di sigarette senza combustione".

- **Posizionamento dei Punti:**

La posizione dei punti sul grafico riflette la relazione tra le categorie di fumatori in base alle componenti principali Dim.1 e Dim.2. Ad esempio, le categorie che sono più vicine tra loro sul grafico sono quelle che hanno una maggiore similarità o associazione in base ai dati.

- **Asse Dim.1 e Dim.2:**

Gli assi Dim.1 e Dim.2 sono le due componenti principali più significative dell'analisi MCA e sono utilizzate per rappresentare la struttura dei dati. Dim.1 spiega la massima variazione nei dati, seguita da Dim.2. Pertanto, la posizione delle categorie su questi assi fornisce informazioni sulla loro importanza nella variazione complessiva dei dati.

- **Colore delle Categorie:**

Le categorie sono colorate in modo diverso per aiutare a distinguere e identificare le diverse tipologie di fumatori. Questo rende più facile associare i punti alle categorie corrispondenti.

- **Interpretazione degli Insights:**

Il grafico biplot consente di identificare modelli, tendenze o associazioni tra le diverse categorie di fumatori in base alle componenti principali Dim.1 e Dim.2. Ad esempio, come possiamo vedere le tre categorie graficate sono posizionate lontana l'una all'altra sul grafico, questo potrebbe indicare che queste tre tipologie di fumatori non condividono alcune caratteristiche simili o non sono strettamente correlate nei dati.

In definitiva, questo grafico offre una prospettiva visuale sulla struttura dei dati relativi alle tipologie di fumatori, consentendo di esplorare le relazioni tra le categorie e le loro influenze sulle componenti principali Dim.1 e Dim.2 dell'analisi MCA.

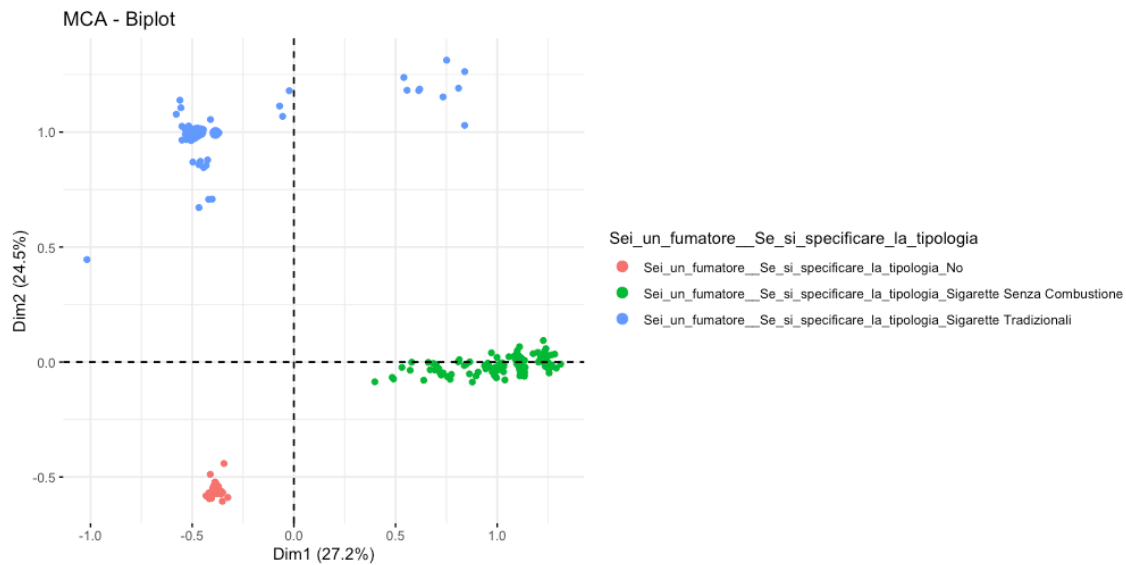


Grafico 25 Rappresentazione della Dimensione 1 e 2 colorando i tre gruppi di individui Fonte: Creato dall'autore

Pertanto, è osservabile che la suddivisione dell'intera popolazione può essere categorizzata in tre distinti gruppi, in linea con quanto precedentemente menzionato. Ciascun gruppo manifesta tratti distinti. Il gruppo dei non fumatori si posiziona prevalentemente nella regione inferiore sinistra del grafico, con scarso disaccoppiamento tra le variabili. In contrasto, i due gruppi restanti mostrano una dispersione più ampia, suggerendo una maggiore variazione tra gli individui all'interno di tali gruppi. È da notare che il gruppo dei fumatori tradizionali presenta una maggiore dispersione dei punti, indicando la possibile presenza di notevoli deviazioni tra gli individui all'interno di questo specifico gruppo. Ulteriori analisi dettagliate per ciascuna tipologia di individuo saranno presentate di seguito.

RISULTATI MCA SIGARETTE TRADIZIONALI

In questo capitolo, esploreremo i risultati l'analisi delle corrispondenze multiple (MCA) focalizzandoci specificamente sulle "Sigarette Tradizionali".

Se si identificheranno dati anomali ("outlier"), effettueremo ulteriori analisi considerando la rimozione dei valori anomali per ottenere una visione più accurata delle dinamiche sottostanti. Questo approccio ci consentirà di esaminare in modo completo e dettagliato

le relazioni e le tendenze legate alle " Sigarette Tradizionali " all'interno del nostro campione di studio, tenendo conto delle sfide legate alla qualità dei dati e ai dati anomali.

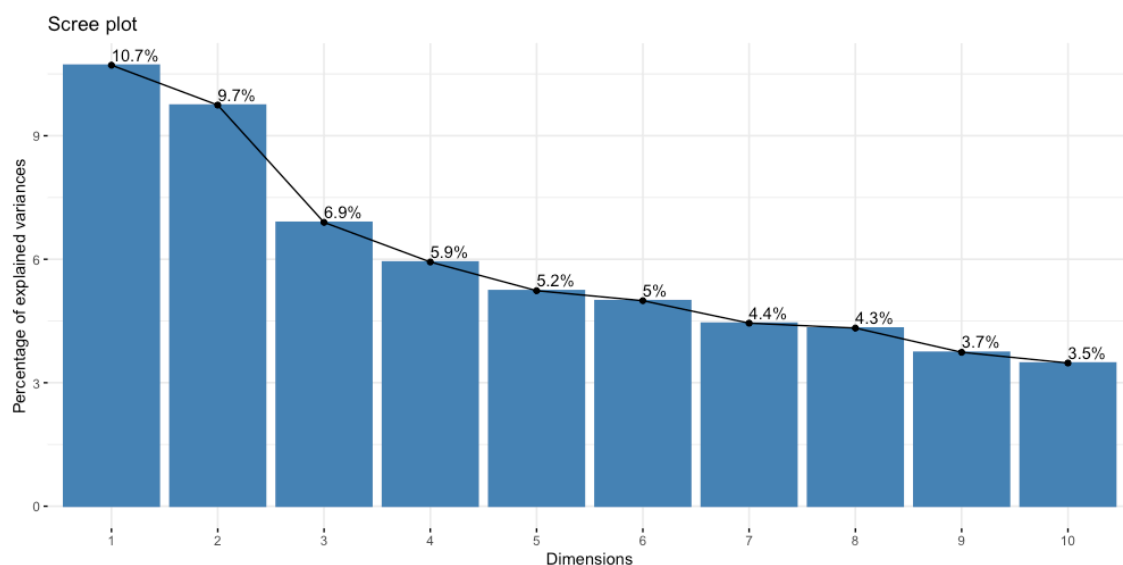


Grafico 26 Varianza del dataset dei fumatori tradizionali spiegata da ciascuna componente delle dimensioni dell'analisi Fonte: creato dall'autore

Dalla valutazione del grafico, emerge chiaramente che le prime due dimensioni (Dim.1 e Dim.2) spiegano in modo significativo la varianza contenuta nel dataset.

Questo suggerisce che il contributo informativo fornito da queste due dimensioni è sostanziale e che esse sono in grado di catturare gran parte delle relazioni e delle variazioni presenti nei dati relativi alle " Sigarette Tradizionali ". Pertanto, l'analisi successiva si concentrerà in modo particolare su queste prime due dimensioni, in quanto offrono una rappresentazione efficace e informativa del dataset preso in esame.

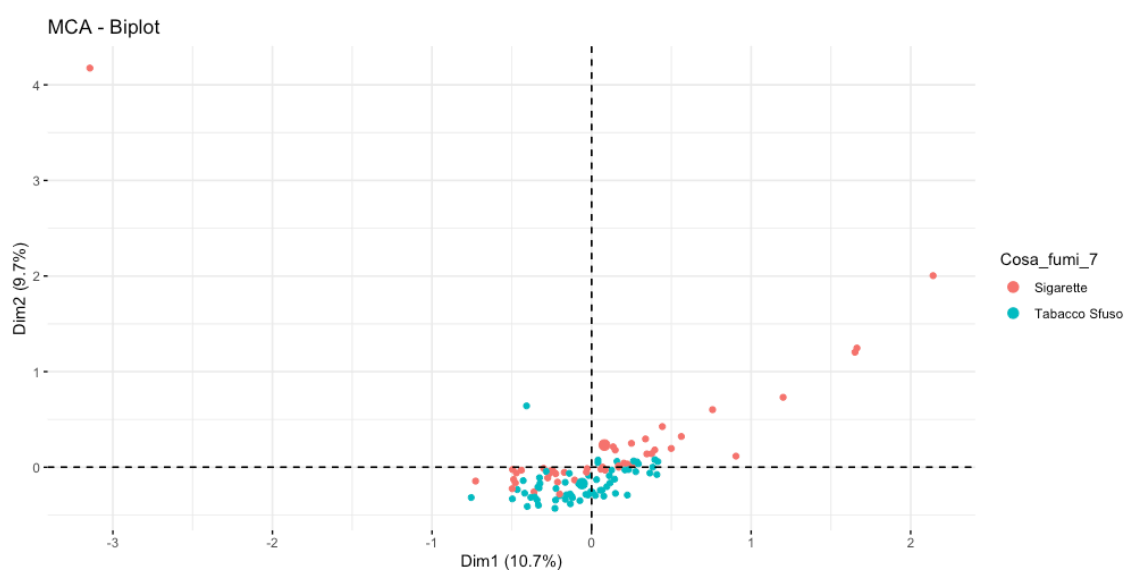


Grafico 27 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie di assunzione di nicotina Fonte: creato dall'autore

Questo biplot visualizza le relazioni tra le variabili categoriche presenti nei dati, insieme alle relazioni tra le categorie di una variabile specifica denominata "Cosa_fumi_7".

La disposizione dei punti nello spazio biplot riflette le associazioni tra le categorie di questa variabile e le dimensioni estratte dall'analisi MCA.

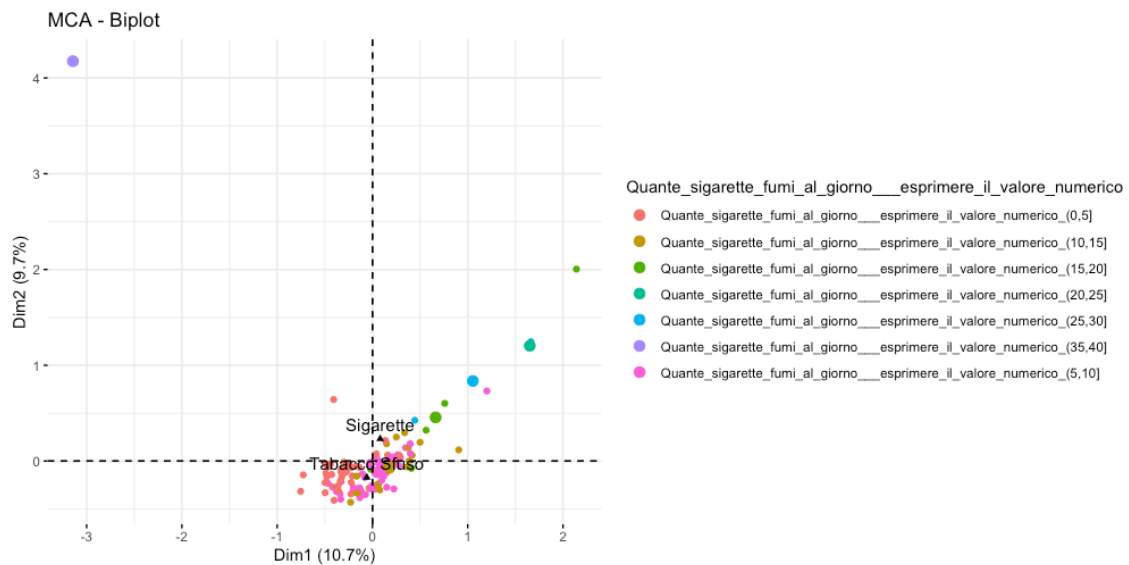


Grafico 28 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie di sigarette fumate al giorno Fonte: creata dall'autore

Osservando il grafico, è possibile notare come la categoria di "Cosa_fumi_7" siano posizionate nello spazio in base alle relazioni tra di loro e rispetto alle dimensioni estratte. Ad esempio, categorie simili o correlate possono trovarsi in posizioni vicine nello spazio biplot, mentre categorie distinte o indipendenti saranno disperse in posizioni più distanti l'una dall'altra. Possiamo inferire che le due categorie di fumatori tradizionali evidenzino una notevole coesione, denotando così una significativa somiglianza reciproca, con l'eccezione di alcune osservazioni che manifestano comportamenti divergenti rispetto alla maggioranza degli individui.

Per esplorare ulteriormente i comportamenti divergenti, è possibile esaminare le altre variabili, come ad esempio il numero di sigarette fumate al giorno. Come evidenziato dal grafico, questa variabile fornisce chiarezza sulle abitudini di alcuni individui. Ad esempio, è possibile osservare un punto in alto a sinistra del grafico che rappresenta un individuo con un alto consumo di sigarette al giorno. Inoltre, c'è un gruppo di individui posizionato a destra del grafico, oltre la coordinata 1 della prima dimensione, che presenta un valore di sigarette fumate al giorno che si discosta significativamente dalla maggior

parte degli individui. Questa divergenza potrebbe contribuire a spiegare la struttura del grafico e le relazioni tra le variabili in esame.

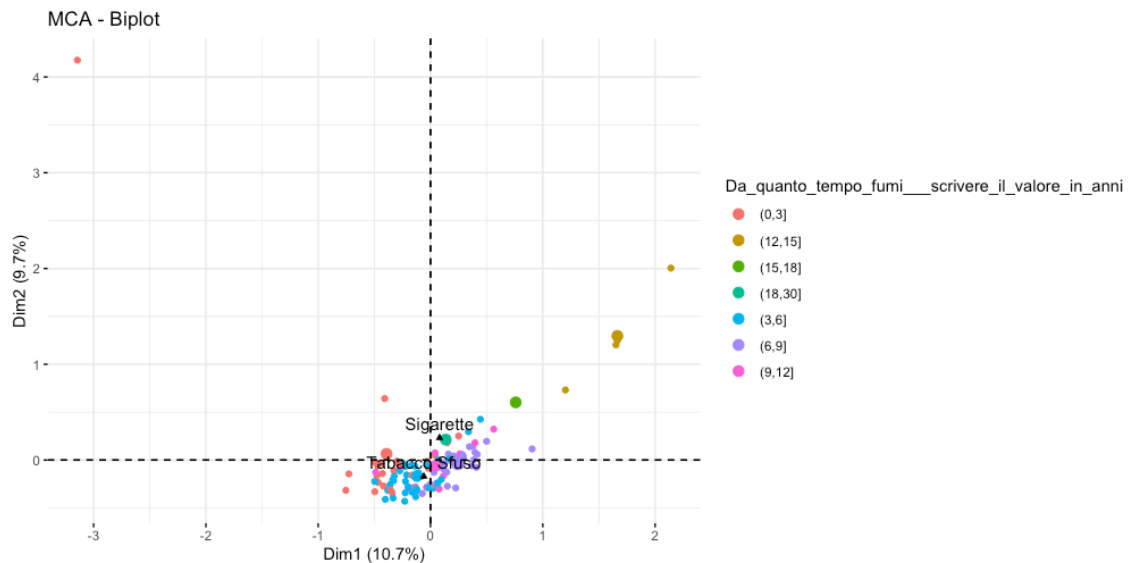


Grafico 29 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti agli anni di fumo Fonte: creata dall'autore

Proseguendo l'analisi delle diverse variabili, è possibile formulare l'ipotesi che il gruppo di individui posizionato a destra del grafico menzionato in precedenza possa essere influenzato dalla variabile relativa agli anni di consumo di tabacco. È plausibile che questo gruppo rientri nella categoria di fumatori con un periodo di consumo di tabacco compreso tra 12 e 15 anni. D'altra parte, il gruppo centrale sembra presentare caratteristiche simili, con un intervallo di anni di consumo di tabacco compreso tra 0 e 9. Queste considerazioni potrebbero contribuire a spiegare la distribuzione spaziale degli individui nel grafico e le relazioni emergenti tra le variabili esaminate.

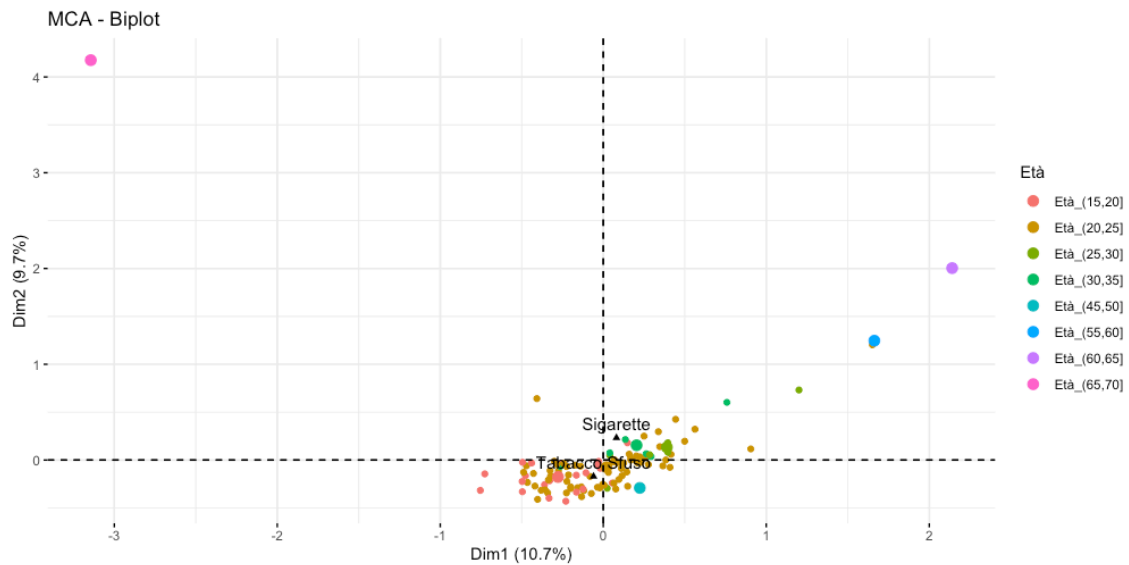


Grafico 30 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie di età degli individui Fonte: creata dall'autore

Un'ulteriore variabile che può contribuire a fornire ulteriori chiavi di comprensione riguardo alle relazioni e alle distanze tra gli individui è l'età. L'analisi del grafico sovrapposto evidenzia che gli individui che si discostano maggiormente dal gruppo centrale presentano un'età compresa tra i 55 e i 70 anni. In contrasto, il gruppo centrale sembra composto principalmente da individui con un'età compresa tra 0 e 35 anni. Questa analisi dell'età contribuisce a delineare ulteriormente la struttura delle relazioni tra gli individui nel contesto dell'analisi.

ANALISI SENZA GLI OUTLINER

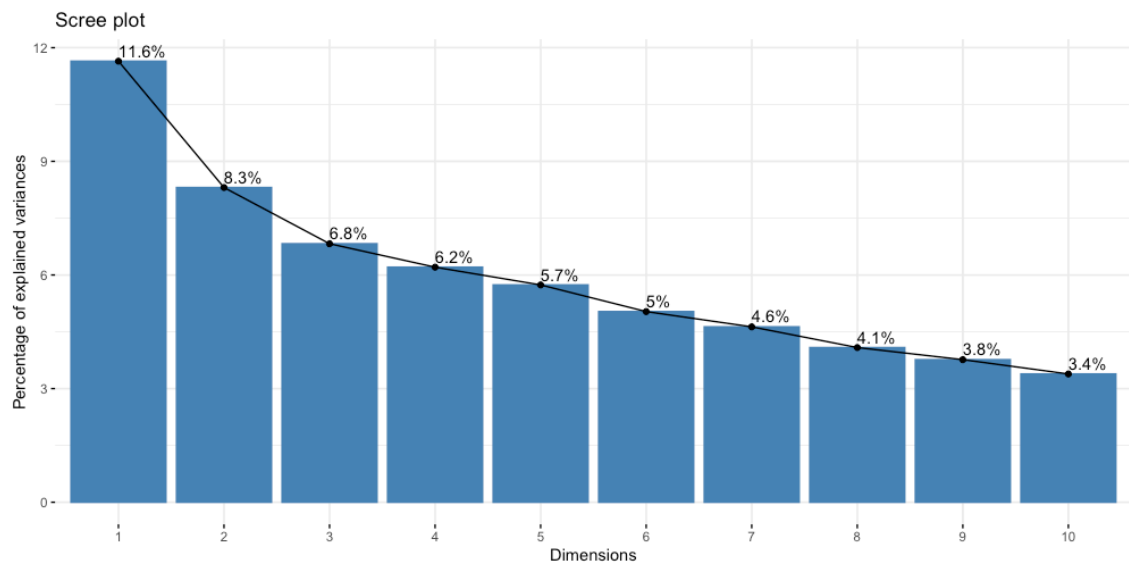


Grafico 31 Varianza del dataset di fumatori tradizionali, senza outlier, spiegata da ciascuna componente delle dimensioni dell'analisi Fonte: creato dall'autore

Attraverso la creazione di un nuovo modello che esclude gli individui precedentemente identificati come *outlier*, si osserva nello Scree plot un aumento significativo della percentuale di varianza spiegata dalla prima dimensione, accompagnato da una diminuzione nelle dimensioni successive, in contrasto con quanto osservato nello scree plot precedente.

L'analisi dei biplot generati da questo nuovo modello rivela che in questa configurazione vi sono meno individui che si discostano significativamente rispetto all'analisi precedente. Ciò implica che la presenza o l'assenza di valori anomali svolga un ruolo cruciale nel determinare i trend e comprendere le correlazioni tra gli individui. In questo nuovo modello, la distinzione tra chi fuma sigarette e chi utilizza tabacco sfuso risulta essere quasi netta, ad eccezione di alcune sovrapposizioni tra le due tipologie di individui.

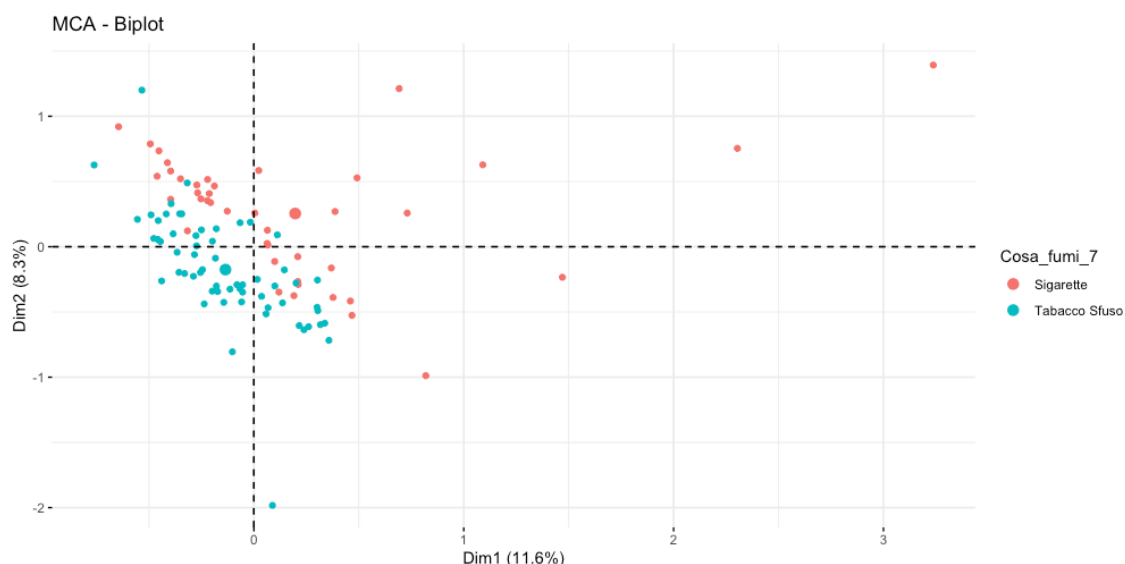


Grafico 32 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti al tipo di assunzione di nicotina
Fonte: creata dall'autore

Per quanto riguarda la variabile relativa al numero di sigarette fumate al giorno, si osserva una notevole somiglianza tra i gruppi nelle loro abitudini di fumo, con la maggior parte degli individui che si colloca nelle stesse aree del grafico. Le eccezioni sono rappresentate da coloro che fumano più di 15 sigarette al giorno, i quali si discostano dagli altri.

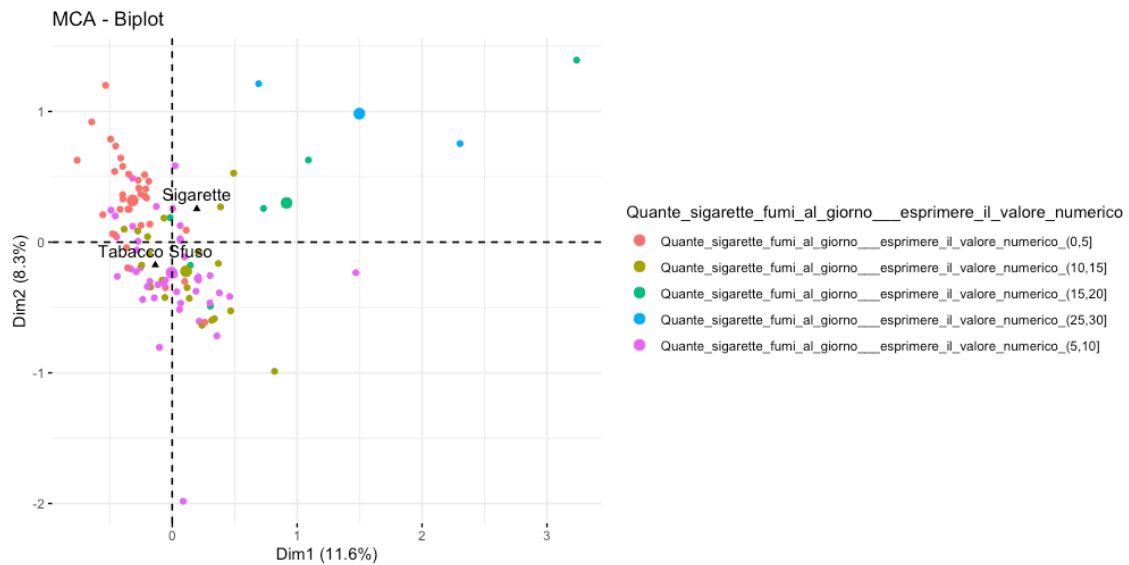


Grafico 33 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti alle sigarette fumate giornalmente Fonte: creata dall'autore

Invece, per quanto riguarda la variabile relativa agli anni di consumo di tabacco, è evidente che il gruppo di individui che fuma da 6 a 9 anni occupa principalmente la parte bassa del grafico, mentre chi fuma da meno, ad esempio nei gruppi 0-3 e 3-6 anni, si trova nella parte opposta del grafico.

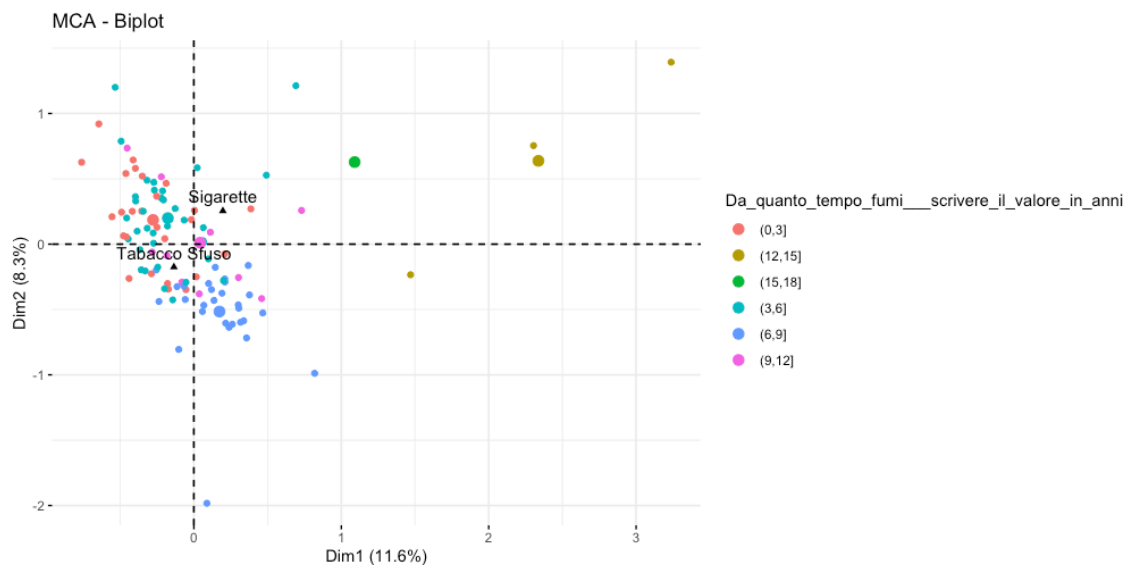


Grafico 34 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti agli anni di fumo Fonte: creata dall'autore

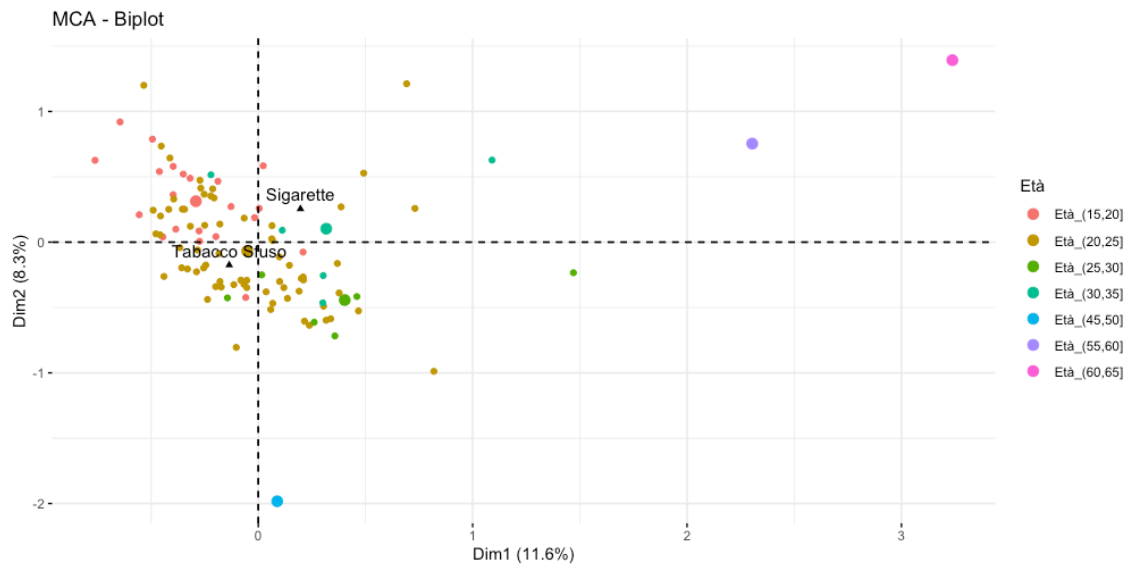


Grafico 35 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti all'età degli individui Fonte: creata dall'autore

Infine, riguardo alla variabile età, si nota chiaramente che il gruppo di individui con un'età compresa tra 15 e 30 anni è situato principalmente al centro del grafico, mentre coloro che hanno un'età più avanzata sono posizionati ai margini del grafico.

In sintesi, il gruppo centrale è costituito da individui che hanno un'età compresa tra 15 e 30 anni, che fumano in media da 0 a 15 sigarette al giorno e che hanno un'esperienza di fumo compresa tra 3 e 9 anni. Questi risultati sottolineano l'importanza di considerare la presenza di valori anomali nelle analisi multivariate e come tali valori possano influenzare significativamente le relazioni tra le variabili.

RISULTATI DELL'MCA SIGARETTE SENZA COMBUSTIONE

In questo capitolo, procediamo con un'analisi dettagliata dei risultati ottenuti mediante l'Applicazione dell'analisi delle corrispondenze multiple (MCA) sul contesto delle "Sigarette Senza Combustione". È fondamentale evidenziare che i dataset utilizzati per questo studio sono stati creati mediante l'impiego di tecniche di imputazione dei dati. In caso di rilevamento di dati anomali, adotteremo un approccio di modellazione alternativa che prevede l'esclusione di tali punti al fine di valutare l'influenza degli outlier sulle nostre analisi. Tale approccio ci consentirà di esplorare in profondità le relazioni e le tendenze intrinseche alle "Sigarette Senza Combustione" all'interno del nostro campione di studio. Inoltre, considereremo con attenzione le sfide legate alla qualità dei dati, alle irregolarità

nelle risposte e alle strategie di modellazione, al fine di fornire un'analisi completa e accurata di questo specifico ambito all'interno del contesto di ricerca.

Iniziamo questa analisi distinguendo i fumatori di "Sigarette Senza Combustione" da un sottogruppo di ex fumatori di "Sigarette Tradizionali". Tale distinzione è necessaria poiché le risposte a domande specifiche riguardanti le "Sigarette Senza Combustione" si applicano solo a un subset di individui all'interno del gruppo considerato. È importante notare che il subset include anche le risposte alle domande relative alle "Sigarette Senza Combustione". Nel seguito di questo capitolo, presenteremo il modello sviluppato specificamente per il gruppo dei fumatori di "Sigarette Senza Combustione" che hanno fumato sigarette tradizionali, illustrando i risultati ottenuti mediante l'analisi. Dall'analisi dello scree plot relativo a questo nuovo modello, emerge che il punto di flessione non è così distintamente definito come nel modello precedente. Nonostante ciò, abbiamo scelto di utilizzare le prime due dimensioni, in quanto contribuiscono in misura maggiore alla spiegazione della varianza nelle nostre variabili.

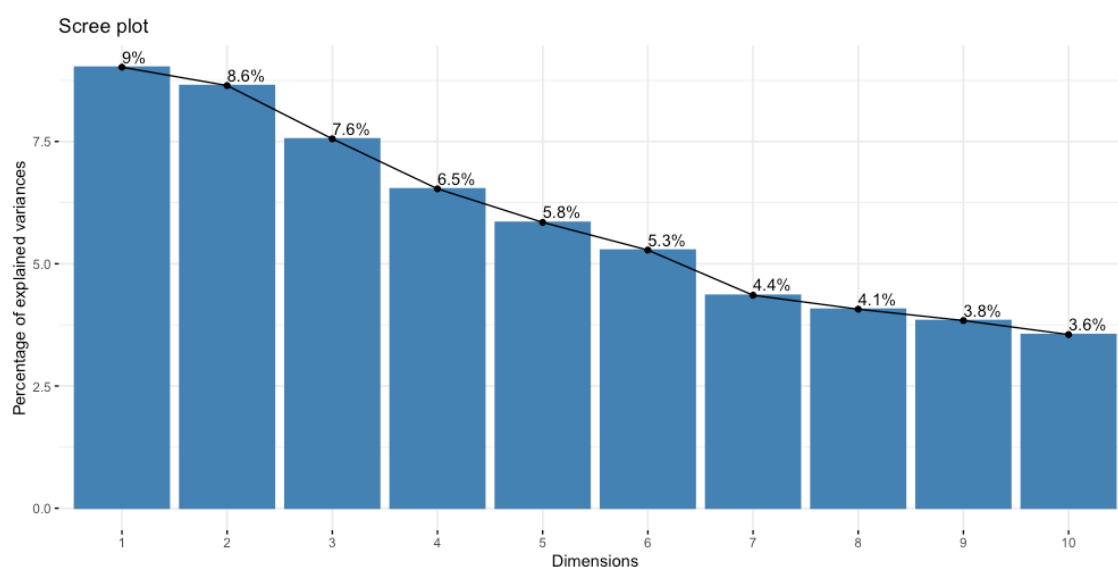


Grafico 36 Varianza del dataset dei fumatori senza combustione spiegata da ciascuna componente delle dimensioni dell'analisi Fonte: creata dall'autore

Il biplot che segue ci consente di visualizzare la disposizione degli individui in base al tipo di dispositivo utilizzato per il consumo di "Sigarette Senza Combustione". È evidente a prima vista che il dispositivo "Iqos" prevale, come già indicato nell'analisi descrittiva. Inoltre, notiamo una vicinanza tra i dispositivi "Iqos" e "Gloo", suggerendo similitudine nelle abitudini dei consumatori tra questi due gruppi. Al contrario, le "Sigarette Elettroniche" e le "Sigarette Usa e Getta" sono distinte dal gruppo principale, possibilmente a causa

delle differenze nei dispositivi stessi e nelle abitudini di consumo. Altro aspetto di rilievo è rappresentato dalla posizione degli individui che fumano "Iqos" all'interno del grafico. La stragrande maggioranza di tali individui è posizionata nella parte negativa della seconda dimensione, contrariamente agli utenti di altri dispositivi, che perlopiù occupano la parte positiva della seconda dimensione. Questo suggerisce una chiara distinzione nell'orientamento delle preferenze dei consumatori di "Iqos" rispetto agli utenti di altri dispositivi, evidenziando una diversa dinamica di consumo tra i gruppi.

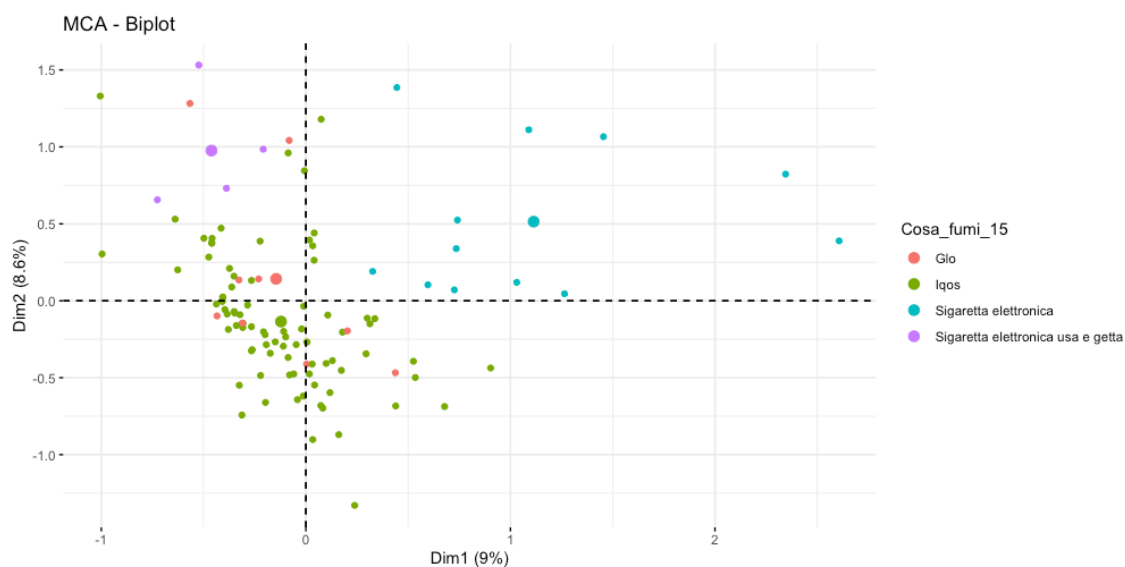


Grafico 37 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti al tipo di dispositivo utilizzato Fonte: creata dall'autore

Altro aspetto di rilievo è rappresentato dal genere dei partecipanti, evidenziando come la popolazione femminile occupi principalmente la parte sinistra del grafico. Questo suggerisce che le donne potrebbero essere più propense a fumare dispositivi come "Iqos" e "Gloo".

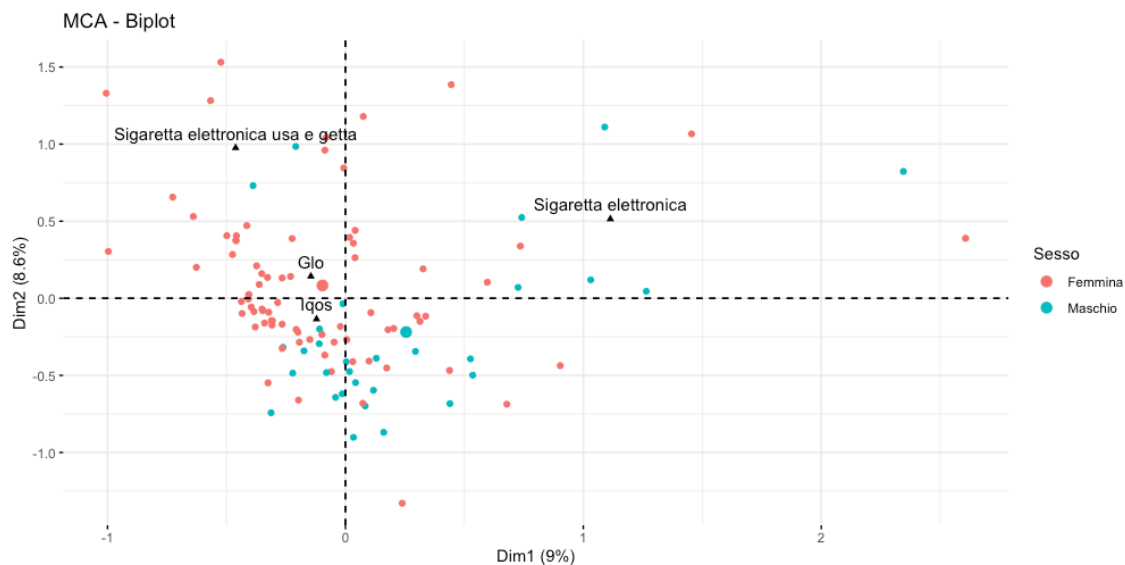


Grafico 38 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti al genere degli individui
Fonte: creata dall'autore

Nel contesto del consumo giornaliero di "Sigarette Senza Combustione", osserviamo che il gruppo precedentemente identificato come femminile, situato sulla sinistra del grafico, presenta un range di consumo giornaliero tra 0 e 10 sigarette. Al contrario, il gruppo di individui di genere maschile, situato sulla destra del grafico, mostra un range di consumo giornaliero compreso tra 10 e 20 sigarette.

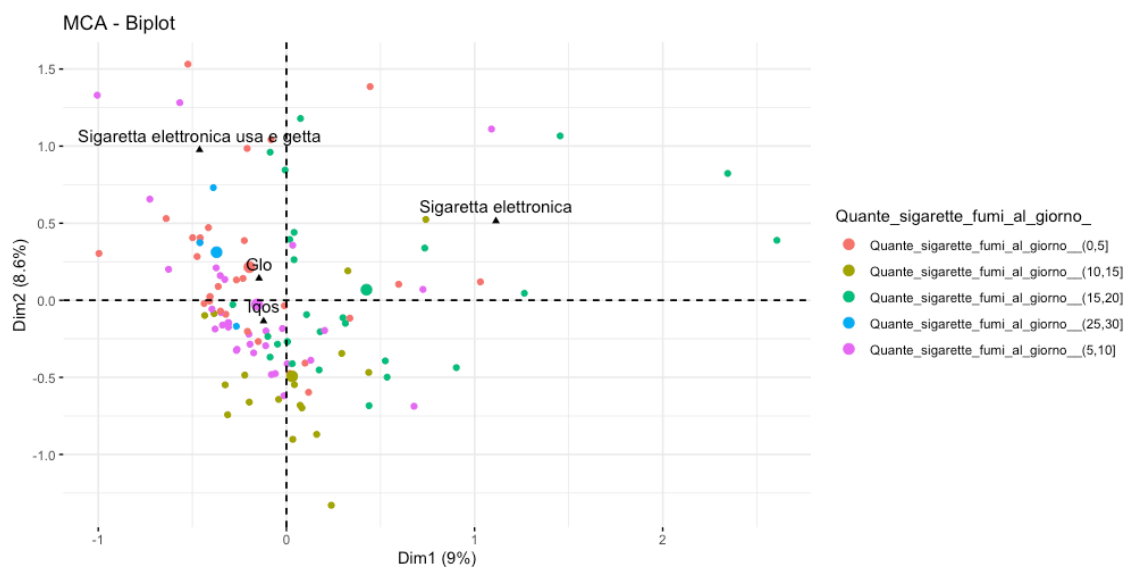


Grafico 39 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti alla quantità di sigarette fumate al giorno
Fonte: creata dall'autore

Per quanto riguarda la variabile relativa alla spesa settimanale, notiamo che non emergono distinti raggruppamenti, ma piuttosto una sovrapposizione di categorie che condividono la stessa area. Ad esempio, il gruppo situato in basso a sinistra del grafico

presenta una spesa settimanale compresa tra 10 e 30 euro, il che potrebbe indicare consumatori di "Gloo", dato il costo relativamente contenuto di queste sigarette rispetto alle "Iqos". Al contrario, il gruppo situato in basso a destra sembra rappresentare i consumatori di "Iqos" con una spesa settimanale tra 30 e 40 euro.

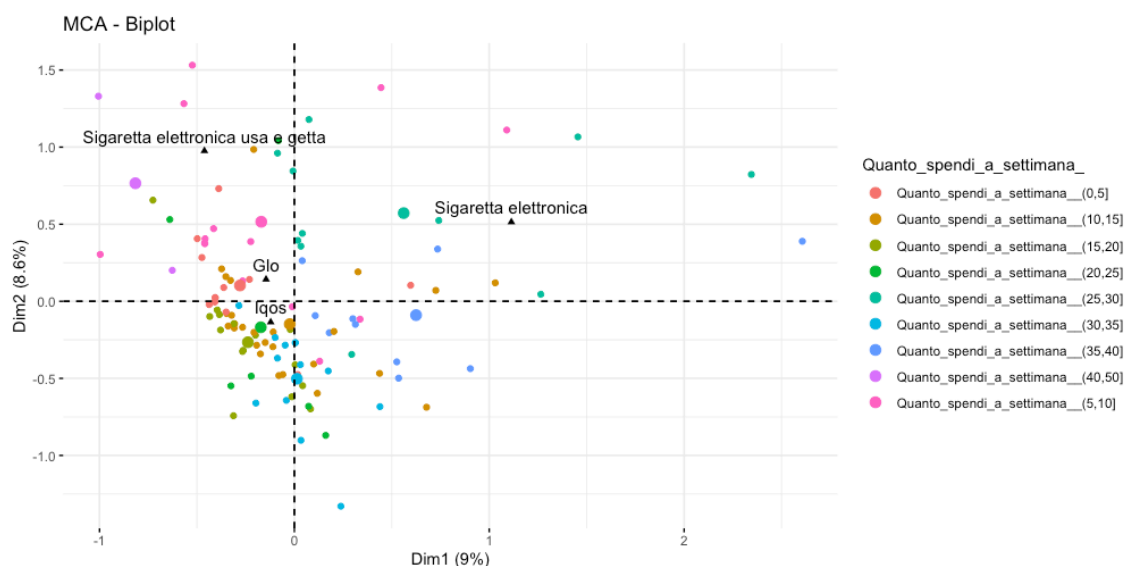


Grafico 40 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti alla spesa settimanale Fonte: creata dall'autore

Infine, considerando la variabile legata all'età, notiamo che i fumatori di "Iqos" sono prevalentemente compresi nella fascia di età tra i 20 e i 25 anni, mentre la parte superiore del grafico mostra una diminuzione dell'età tra i 15 e i 20 anni.

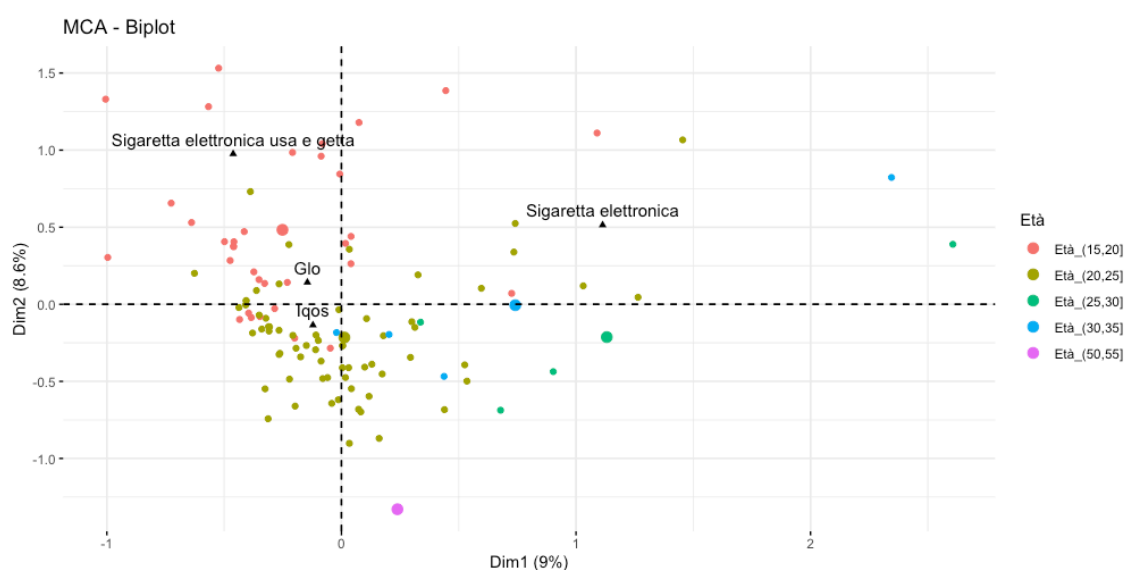


Grafico 41 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti all'età degli individui Fonte: creata dall'autore

In sintesi, l'analisi rivela che la popolazione analizzata tende a consumare "Iqos" principalmente tra i 20 e i 25 anni, con una predominanza di individui di genere femminile. Questi fumatori presentano un consumo giornaliero di 5-15 sigarette e una spesa settimanale di 15-35 euro.

SOTTOGRUPPO DI EX FUMATORI DI SIGARETTE TRADIZIONALI

All'interno di questa analisi specifica, ci concentriamo sul sottogruppo di individui che attualmente fumano sigarette senza combustione ma hanno una storia pregressa di consumo di sigarette tradizionali. Questa suddivisione è stata intrapresa al fine di investigare le dinamiche distintive che caratterizzano questo particolare sottoinsieme di partecipanti. In questo contesto, la finalità principale consisteva nell'individuare le dimensioni più rilevanti per spiegare la varianza osservata nei dati. L'identificazione del punto di flessione all'interno dello Scree plot ha rappresentato una tappa cruciale. Tuttavia, per scopi interpretativi, ci si concentrerà sulle prime due dimensioni, considerate le più informative per le variabili coinvolte nell'analisi.

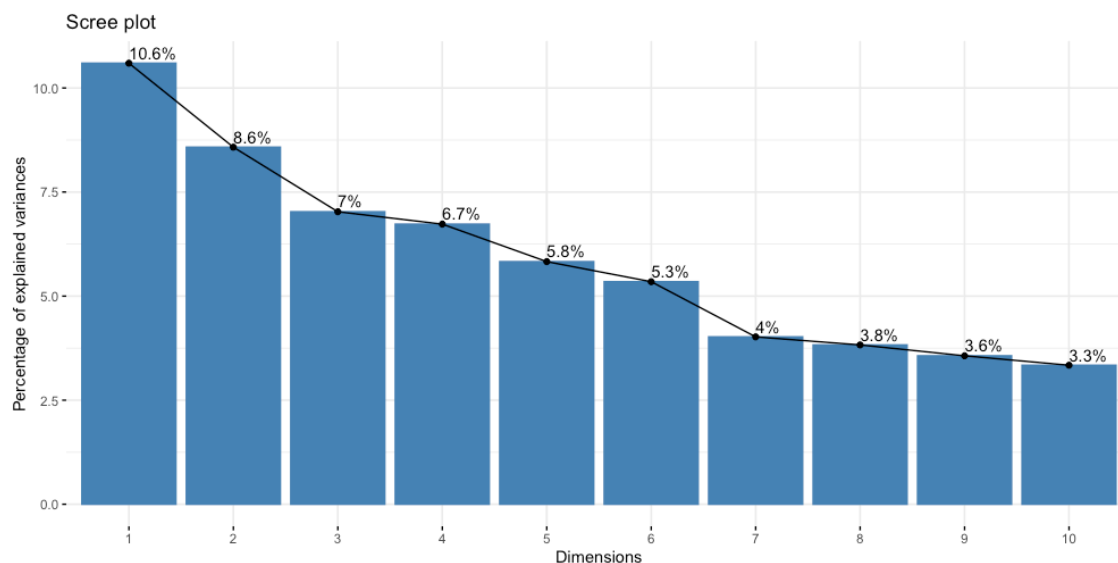
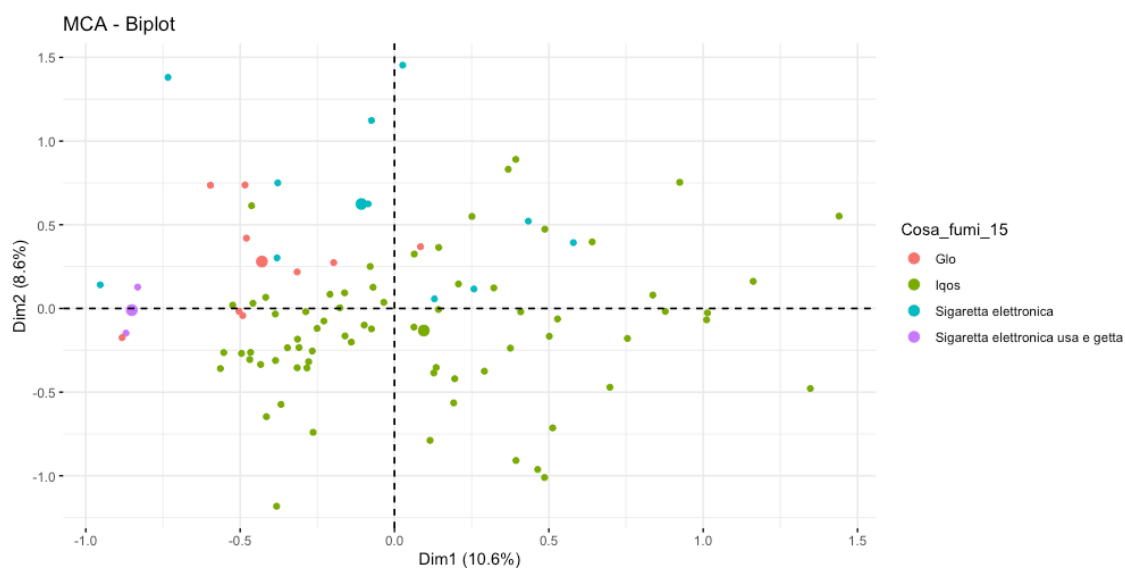


Grafico 42 Varianza del dataset di fumatori senza combustione spiegata da ciascuna componente delle dimensioni dell'analisi Fonte: creato dall'autore

Il biplot che segue fornisce una rappresentazione grafica delle categorie relative al tipo di dispositivo utilizzato da ciascun individuo all'interno di questo sottogruppo. Emergono alcune tendenze interessanti: la maggior parte degli utilizzatori di dispositivi Iqos è collocata nella parte negativa della seconda dimensione, mentre le altre categorie si trovano prevalentemente nella parte superiore. Questa distribuzione potrebbe indicare

delle differenze significative nelle abitudini dei consumatori di Iqos rispetto agli altri dispositivi.



*Grafico 43 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti al tipo di dispositivo utilizzato
Fonte: creata dall'autore*

Successivamente, nell'analisi della durata dell'uso dei dispositivi senza combustione espressa in anni, è possibile notare una predominanza di individui posizionati nella parte negativa della seconda dimensione, soprattutto tra coloro che hanno utilizzato tali dispositivi per un periodo compreso tra 1 e 3 anni. Tuttavia, è fondamentale sottolineare che questa osservazione attualmente assume un carattere ipotetico. Tale cautela deriva dal fatto che, nonostante si osservi una tendenza di aggregazione in questa direzione, non emergono in modo chiaro aggregazioni o cluster distinti che confermino in modo inequivocabile questa specifica dinamica.

In aggiunta, è rilevante evidenziare la presenza di un sottoinsieme di individui che rientrano nella categoria dei fumatori di sigarette senza combustione con un'esperienza di 6 a 9 anni. Questi individui sono posizionati nella parte destra del grafico e appaiono isolati dalle altre variabili. Questa particolare distribuzione suggerisce la presenza di un numero limitato di fumatori che ricadono in questa categoria specifica all'interno del sottogruppo esaminato. Questo dato, sebbene non fornisca evidenza di cluster distinti, suggerisce che esista un numero limitato di individui con un lungo periodo di utilizzo di dispositivi senza combustione, il che potrebbe essere oggetto di ulteriori indagini per una comprensione più approfondita delle loro abitudini e comportamenti.

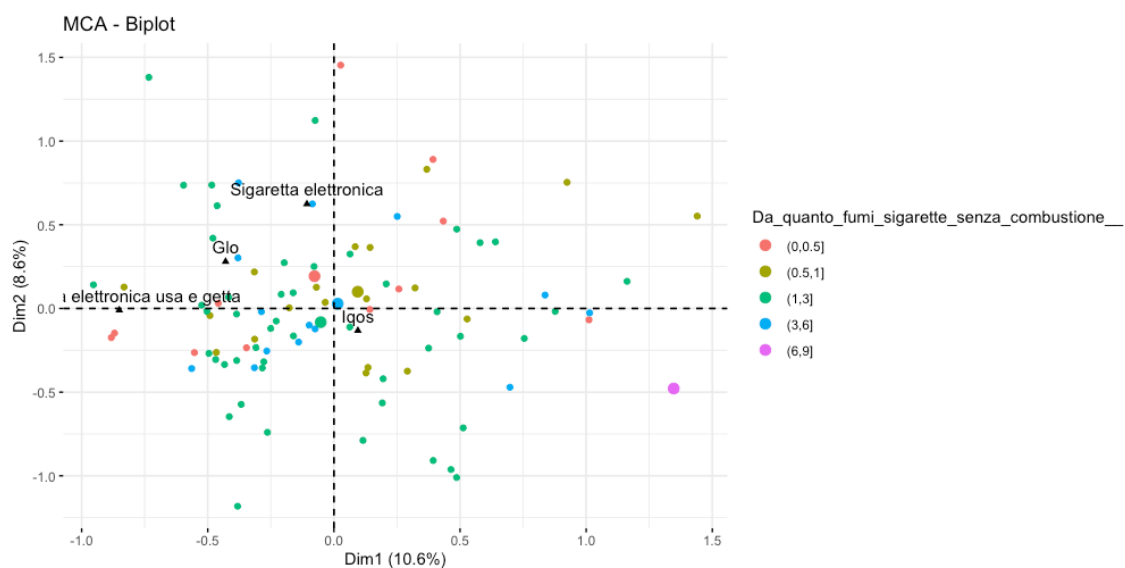


Grafico 44 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti agli anni di utilizzo di dispositivi senza combustione Fonte: creata dall'autore

Proseguiamo con l'analisi dei biplot relativi alla spesa settimanale per le sigarette attuale e passata, così come al consumo giornaliero attuale e passato.

Mettendo a confronto queste visualizzazioni, notiamo che nella parte negativa della prima dimensione non emergono differenze significative, ad eccezione di alcuni individui che hanno ridotto il consumo di sigarette. Dall'altra parte, nella parte positiva della prima dimensione, molti individui hanno aumentato il consumo di sigarette, ricadendo principalmente nella categoria tra le 15 e le 20 sigarette al giorno.

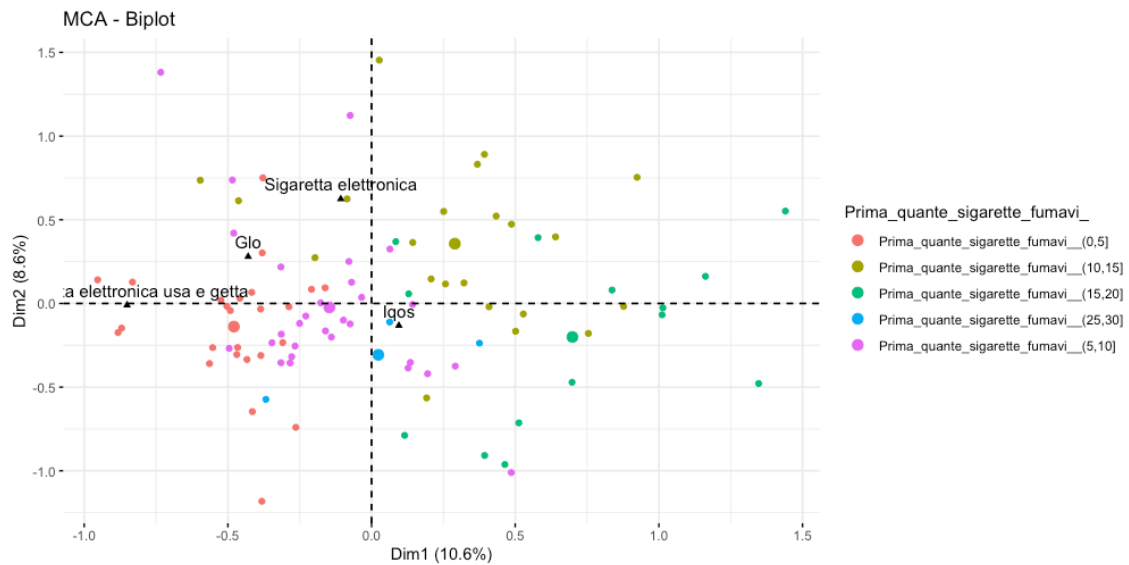


Grafico 45 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti alla quantità di sigarette fumate precedentemente Fonte: creata dall'autore

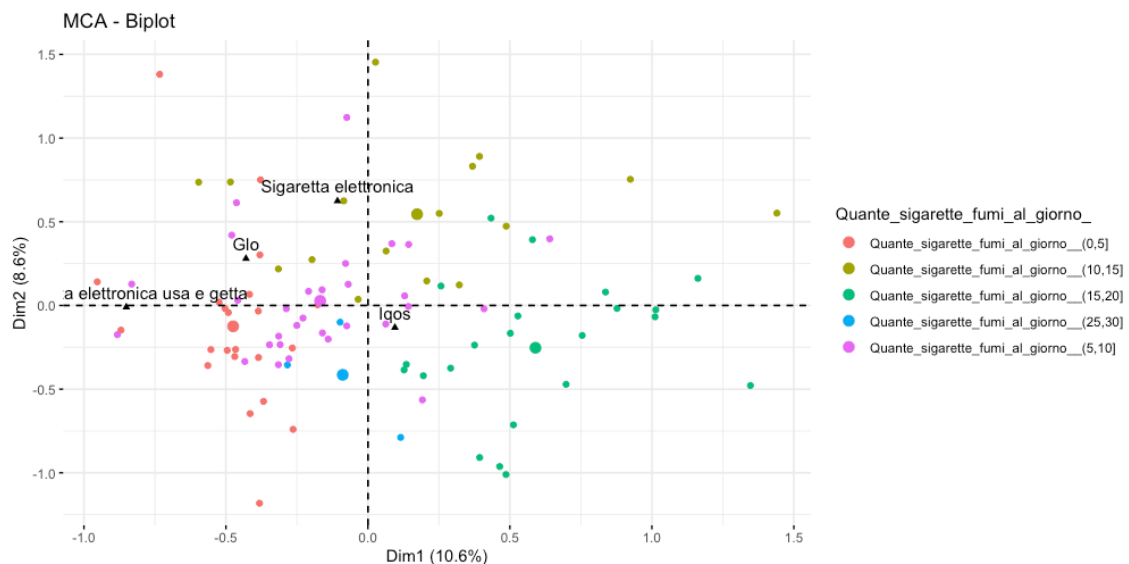


Grafico 46 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti alla quantità di sigarette fumate attualmente Fonte: creata dall'autore

Per quanto riguarda la spesa settimanale, possiamo osservare che nella parte positiva della prima dimensione la spesa media è aumentata, passando da un intervallo tra i 15 e i 30 euro settimanali a un intervallo più elevato, compreso tra i 30 e i 40 euro settimanali. Nella parte negativa della prima dimensione, la spesa è cresciuta da un intervallo tra i 5 e i 10 euro settimanali a un intervallo di spesa tra 10 e 25 euro settimanali. Questa osservazione suggerisce che l'uso di dispositivi senza combustione potrebbe influire sull'incremento del consumo di nicotina e delle relative spese.

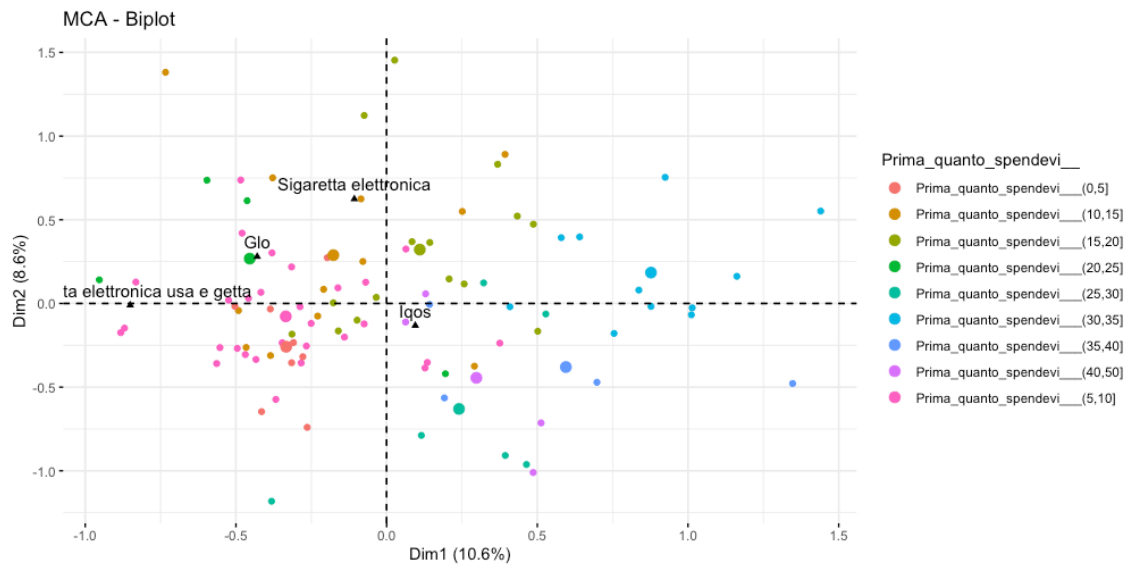


Grafico 47 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti alla spesa settimanale precedente Fonte: creata dall'autore

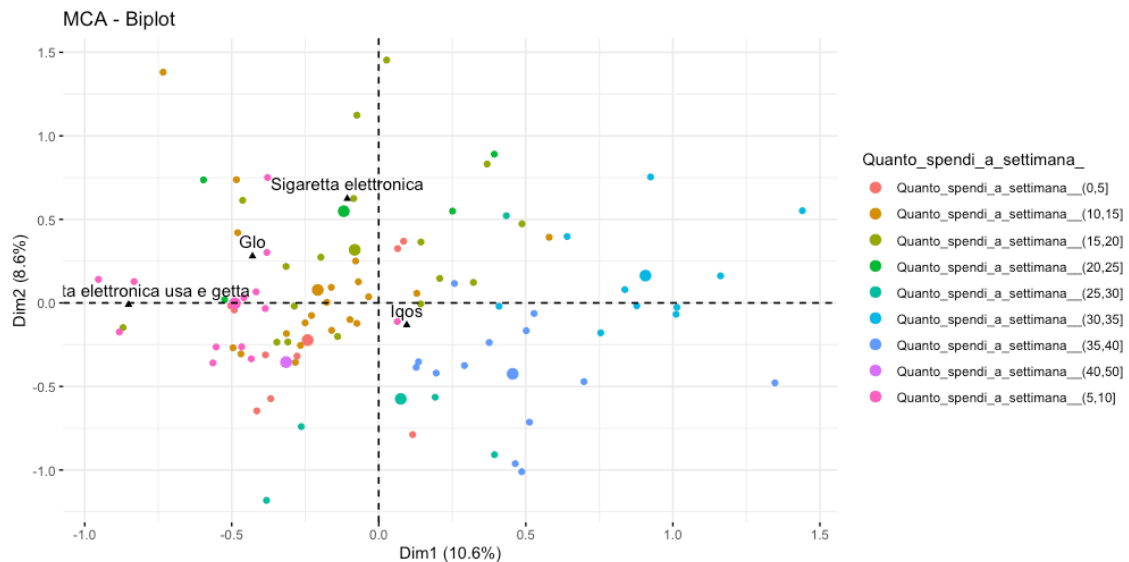


Grafico 48 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie inerenti alla spesa settimanale attuale Fonte: creata dall'autore

Un aspetto significativo da considerare è l'effetto sulla salute degli individui. Dall'analisi emerge che la maggior parte dei fumatori di Iqos, collocati principalmente nella parte negativa della seconda dimensione, non ha riportato benefici significativi in termini di salute.

In contrasto, i fumatori di altri dispositivi, situati principalmente nella parte positiva della seconda dimensione, affermano di aver riscontrato miglioramenti derivanti dall'uso di tali dispositivi.

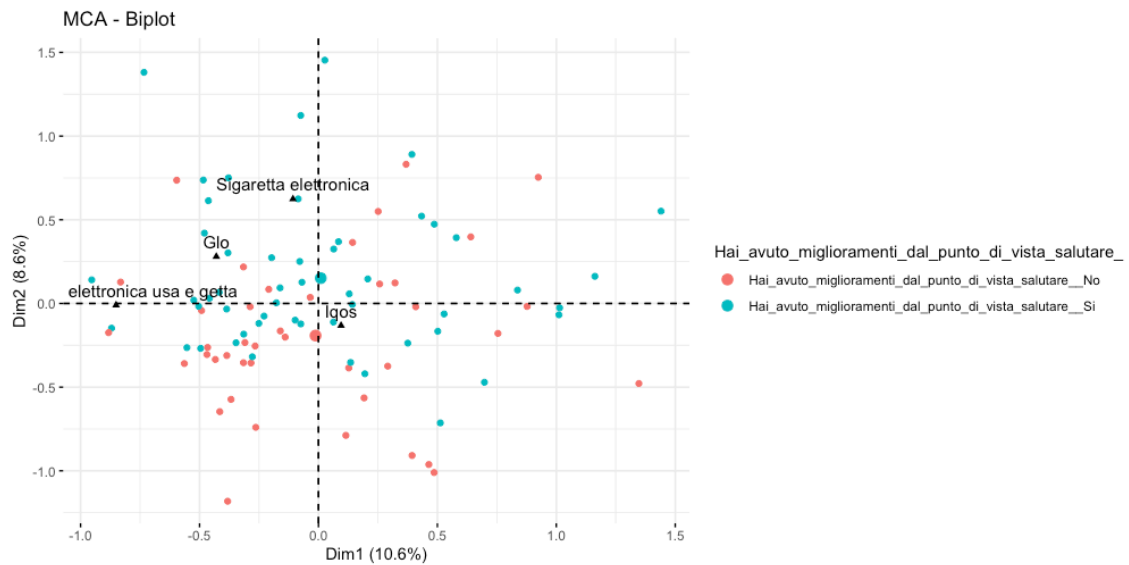


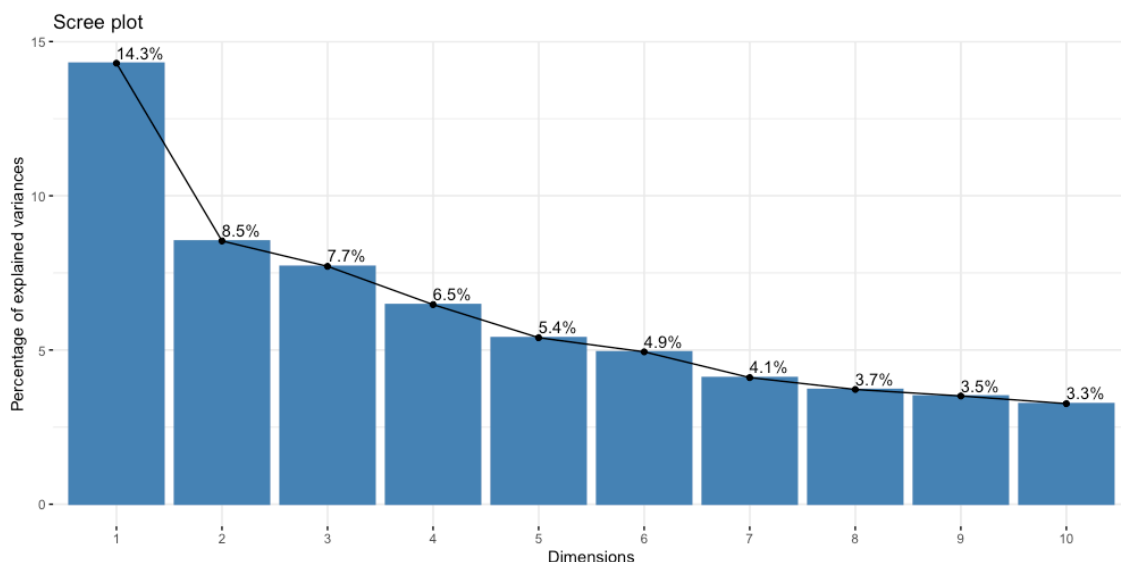
Gráfico 49 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie ai miglioramenti di salute Fonte: creata dall'autore

RISULTATI MCA NON FUMATORI

In questo sottocapitolo, esploreremo i risultati ottenuti dall'Analisi delle Componenti Multipla (MCA) applicata al gruppo di individui che non fumano. Questa analisi mira a fornire una visione dettagliata delle caratteristiche e delle relazioni presenti tra i non fumatori all'interno del nostro campione di studio. Saranno esaminati i dati relativi alle variabili pertinenti ai non fumatori, e le informazioni ottenute contribuiranno a delineare una panoramica esaustiva di questo specifico gruppo all'interno del contesto della nostra ricerca. Inoltre, verranno affrontate le sfide legate alla qualità dei dati, se presenti, al fine di garantire una valutazione accurata e approfondita delle dinamiche che caratterizzano i non fumatori all'interno della nostra indagine.

Nel grafico dello Scree plot è possibile identificare un punto di flessione nella 2^a dimensione. Questo indicatore suggerisce che le prime due dimensioni del modello di Analisi delle Componenti Multipla (MCA) contribuiscono in modo significativo alla spiegazione della varianza presente nei dati, mentre le dimensioni successive hanno un contributo progressivamente minore. In questa analisi, l'attenzione principale è stata rivolta alla 1^a e alla 2^a dimensione per la rappresentazione grafica dei risultati. Questa scelta è stata motivata dalla rilevanza di queste due dimensioni nel catturare le informazioni chiave dei dati, mentre le dimensioni successive apportano contributi relativamente minori nella spiegazione della variazione nei dati. L'utilizzo della 1^a e della 2^a dimensione agevola l'interpretazione e la visualizzazione dei risultati derivanti

dall'analisi MCA, semplificando la comprensione dei modelli e delle relazioni intrinsecamente presenti nei dati oggetto di studio.



*Grafico 50 Varianza del dataset dei non fumatori spiegata da ciascuna componente delle dimensioni dell'analisi
Fonte: creato dall'autore*

Nel corso di questa analisi, abbiamo proceduto con la rappresentazione grafica dei dati, evidenziando la variabile relativa all'esperienza pregressa di fumo. Notiamo che tale rappresentazione grafica riflette una distinzione evidente tra due categorie principali, ossia i partecipanti che hanno dichiarato di aver mai fumato ("si") e coloro che hanno risposto negativamente ("no"). Tuttavia, è importante notare che, nonostante questa distinzione, l'analisi rivela una variazione limitata tra gli individui, con alcune eccezioni rappresentate da pochi individui distanti dal gruppo centrale. Questa variazione anomala può essere attribuita a fattori aggiuntivi non rappresentati nella visualizzazione attuale, che richiederebbero ulteriori indagini.

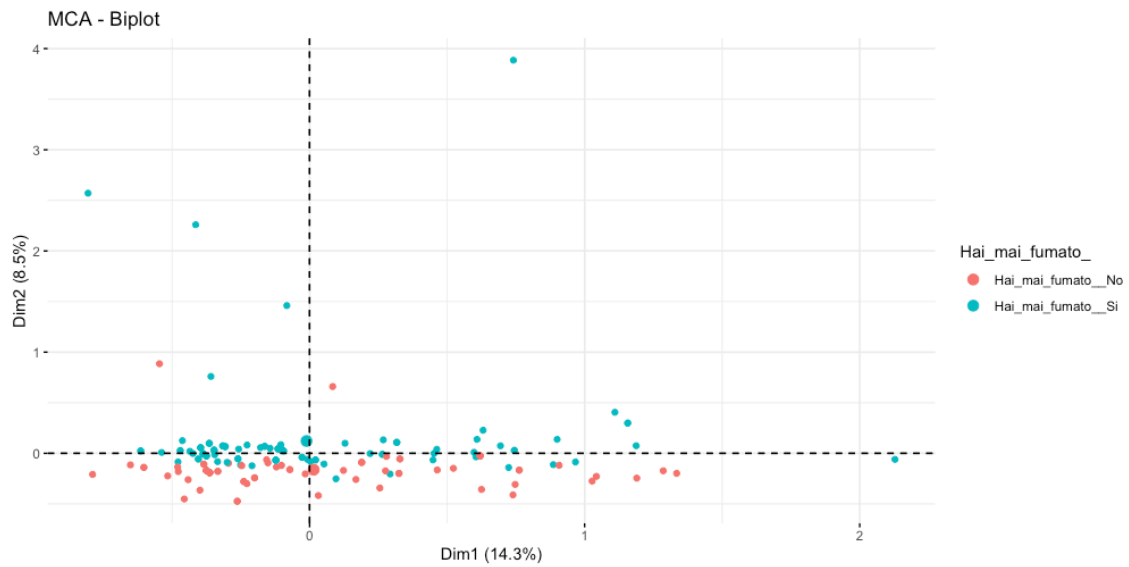


Grafico 51 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie relative all'aver mai fumato
 Fonte: creata dall'autore

In seguito, esaminando la variabile relativa al genere, notiamo una distribuzione più omogenea. Gli individui di genere femminile occupano prevalentemente la parte sinistra del grafico, mentre quelli di genere maschile si concentrano nella parte opposta, tuttavia, la separazione tra i due generi non è altamente distintiva.

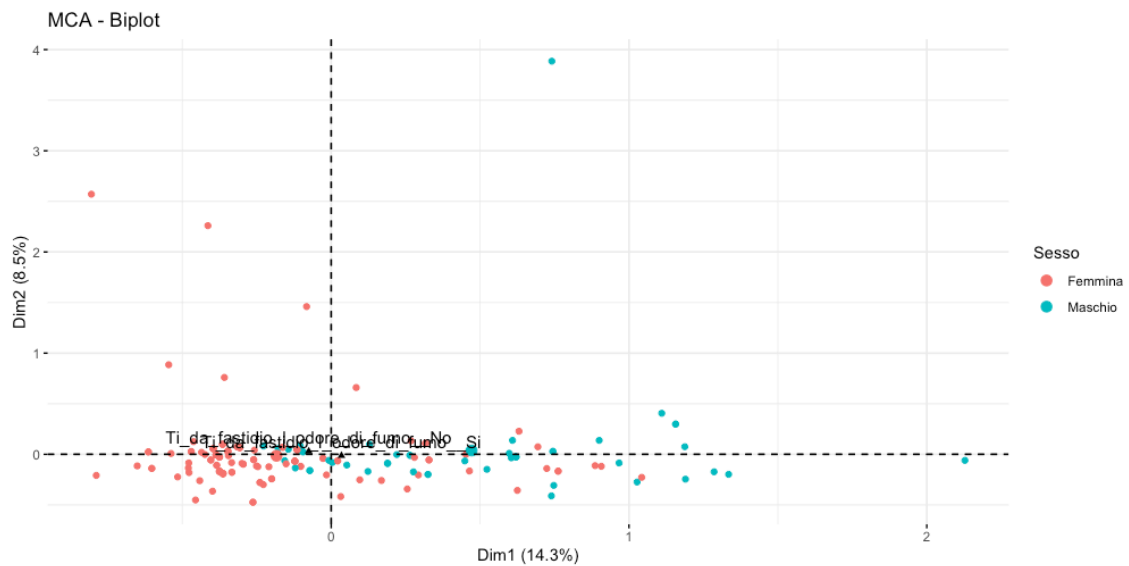


Grafico 52 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie di genere del gruppo analizzato
 Fonte: creata dall'autore

Riguardo alla variabile relativa all'avversione all'odore di fumo, osserviamo un altro scenario. Qui, le due categorie non mostrano una separazione netta, ma piuttosto convivono all'interno dello stesso intervallo. Questa situazione riflette una limitata varianza tra gli individui in termini di fastidio verso l'odore di fumo.

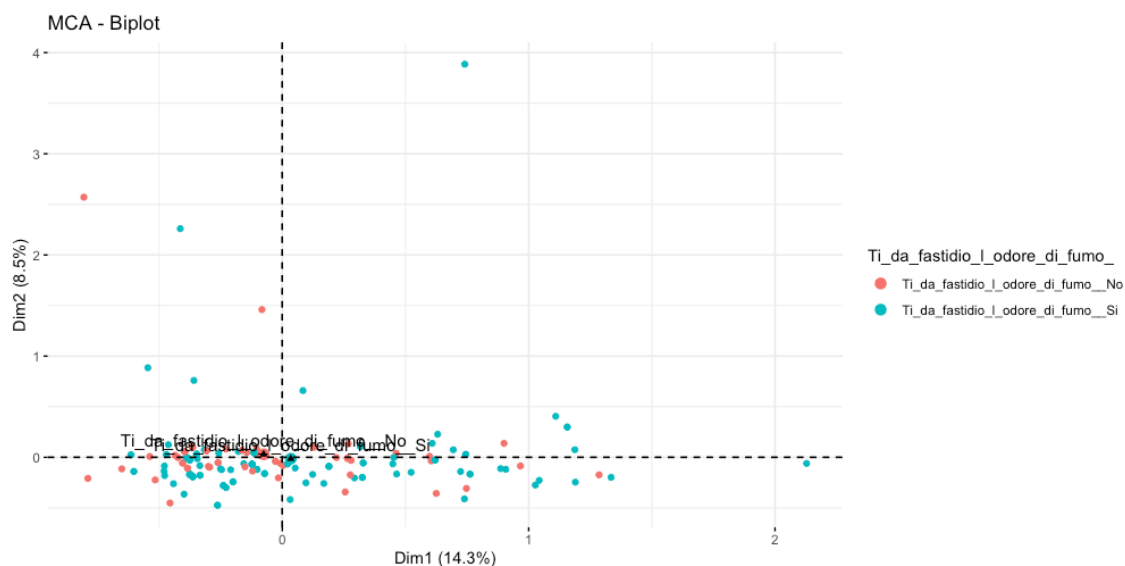


Grafico 53 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie relative al fastidio prodotto dall'odore di fumo Fonte: creata dall'autore

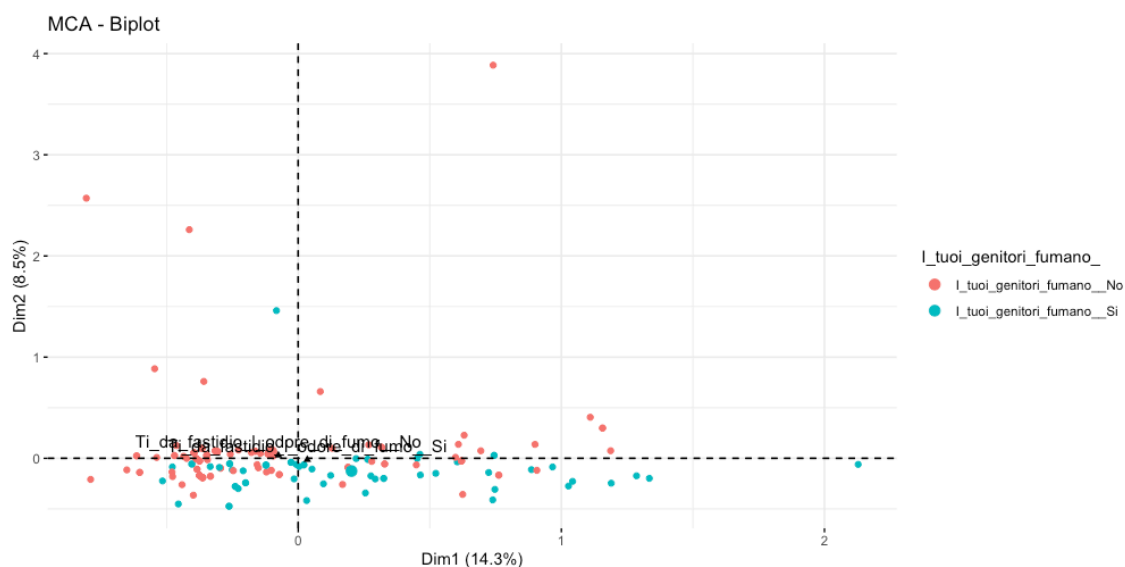


Grafico 54 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie relative ad avere genitori fumatori o meno Fonte: creata dall'autore

Per quanto riguarda la variabile relativa alla presenza o assenza di genitori fumatori, notiamo una sovrapposizione significativa tra le categorie, con l'eccezione di alcuni valori

isolati nella parte superiore del grafico, che corrispondono a individui senza genitori fumatori.

Infine, esaminando la categoria di dipendenze non correlate all'uso di nicotina, notiamo che il gruppo di individui distanti dal gruppo centrale è associato a dipendenze dall'alcol. Inoltre, nella parte destra del grafico, sono presenti anche altre due tipologie di dipendenze, ovvero scommesse e droga.

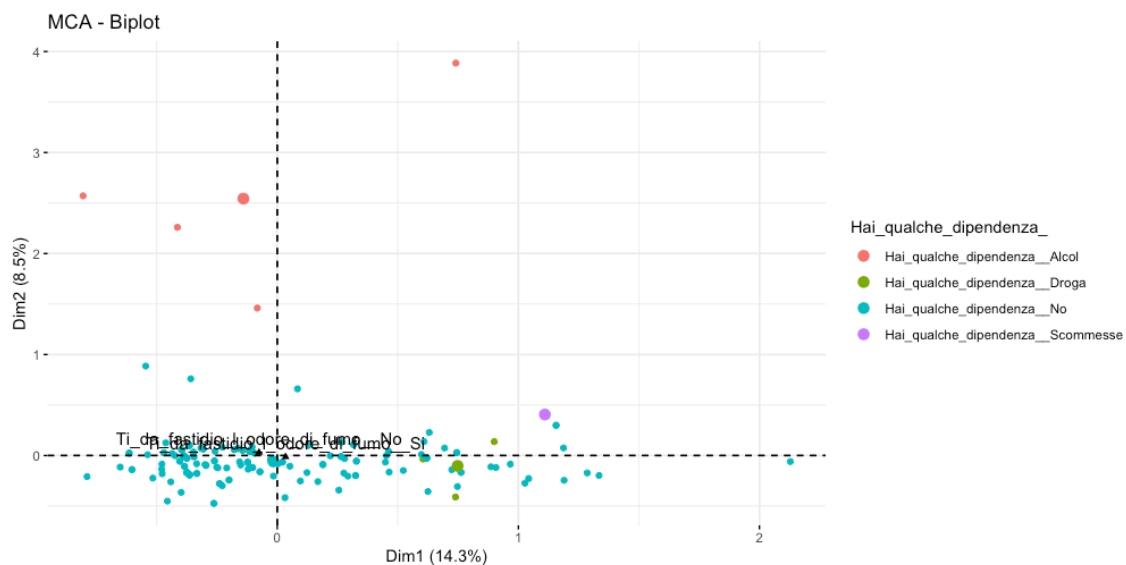


Grafico 55 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie relative all'avere qualche dipendenza
Fonte: creata dall'autore

Per quanto riguarda la variabile età, osserviamo una distinzione notevole tra i gruppi. Nella parte negativa della dimensione 1, sono concentrati gli individui con età compresa tra i 15 e i 25 anni, mentre nella parte positiva della dimensione 1, troviamo gli individui di età compresa tra i 25 e i 60 anni.

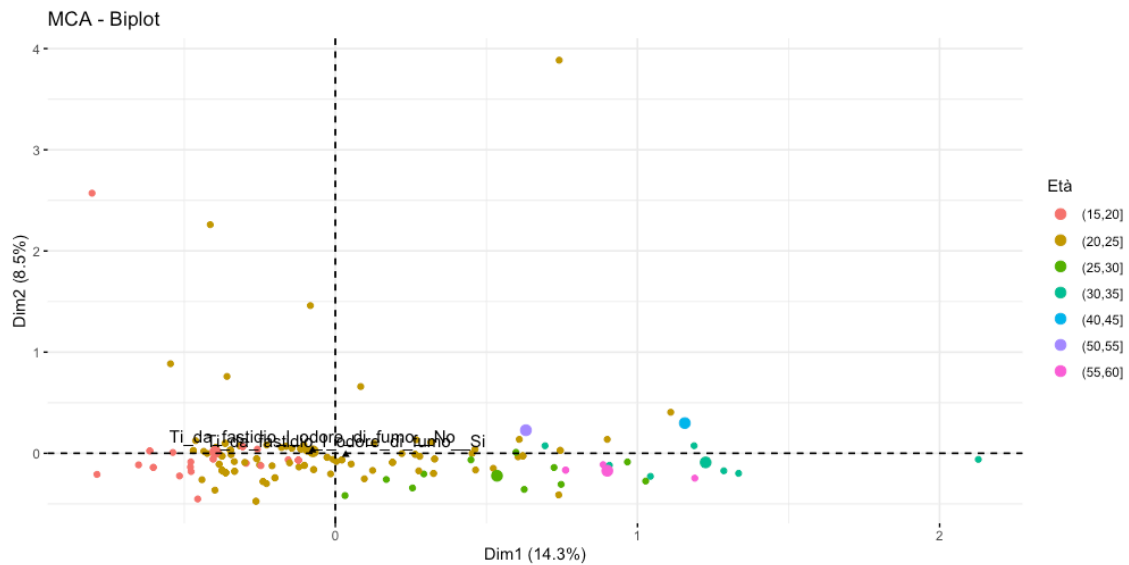


Grafico 56 Rappresentazione della Dimensione 1 e 2 visualizzando le categorie relative all'età Fonte: creata dall'autore

In sintesi, questa analisi rivela una chiara differenziazione tra le categorie di partecipanti solo per la variabile dell'età e delle dipendenze, mentre per la maggior parte delle altre variabili, le distinzioni sono meno evidenti. La variazione osservata in alcune categorie può essere attribuita a fattori non considerati nell'analisi attuale, richiedendo ulteriori approfondimenti.

Conclusioni

RIASSUNTO DEI PRINCIPALI RISULTATI OTTENUTI

Nel primo capitolo, abbiamo delineato il contesto del fenomeno del tabagismo in Italia, evidenziando un aumento significativo del tasso di fumo nel 2022. Abbiamo anche sottolineato l'importanza di questo studio data la gravità dei problemi di salute legati al consumo di tabacco.

Nel secondo capitolo, abbiamo esaminato il processo di acquisizione dei dati attraverso i Moduli Google, evidenziando i vantaggi dell'approccio online e le sfide legate all'accessibilità.

Nella prima parte del terzo capitolo, abbiamo condotto un'analisi descrittiva generale dei dati raccolti, suddividendo i partecipanti in tre categorie principali: non fumatori, utilizzatori di sigarette senza combustione e fumatori tradizionali.

Nella seconda parte del terzo capitolo, ci siamo concentrati sui consumatori di sigarette senza combustione, esaminando le loro preferenze per i dispositivi, il consumo medio, la spesa settimanale e le preferenze per le marche o varianti. Abbiamo anche iniziato a esplorare i problemi di salute respiratoria tra questi consumatori.

Nel sottogruppo degli ex fumatori di sigarette tradizionali tra i fumatori senza combustione, abbiamo analizzato le tendenze di consumo e di spesa e rilevato miglioramenti nella salute dopo il passaggio alle sigarette senza combustione.

Nel quarto capitolo, abbiamo introdotto l'Analisi delle Corrispondenze Multiple (MCA) e l'Algoritmo di Imputazione Multipla MICE come strumenti analitici. Successivamente, abbiamo condotto l'analisi MCA sui dati relativi alle "Sigarette Tradizionali", evidenziando l'importanza della gestione dei dati mancanti e dei valori anomali.

Nel quinto capitolo, abbiamo esaminato i risultati dell'analisi MCA per le "Sigarette Senza Combustione", prendendo in considerazione diversi aspetti, tra cui dispositivi utilizzati, genere dei partecipanti, consumo giornaliero, spesa settimanale, età e effetti sulla salute. Abbiamo anche analizzato i non fumatori nel campione.

DISCUSSIONE DELLE IMPLICAZIONI DEI RISULTATI PER LA COMPrensIONE DEL FENOMENO DEL TABAGISMO IN ITALIA

I risultati ottenuti forniscono una visione approfondita delle abitudini dei consumatori di tabacco in Italia, con particolare attenzione alle sigarette senza combustione. Tali risultati hanno importanti implicazioni per la comprensione del fenomeno del tabagismo in Italia: Aumento del Tasso di Fumo: L'aumento del tasso di fumo nel 2022 è preoccupante e richiede una maggiore attenzione da parte delle autorità sanitarie. È necessario sviluppare strategie di prevenzione mirate per contrastare questo trend.

Preferenza per l'Iqos: Il predominio dell'Iqos tra i consumatori di sigarette senza combustione suggerisce la necessità di esaminare più approfonditamente questo dispositivo e le sue implicazioni per la salute pubblica.

Genere e Età: Le differenze di genere e le fasce di età dei consumatori possono influenzare l'efficacia delle campagne di prevenzione. Le strategie dovrebbero essere adattate in base a queste variabili.

Spesa Settimanale: La spesa settimanale varia notevolmente tra i consumatori di tabacco. Questo può influire sulla loro dipendenza e rappresenta un aspetto importante da considerare nelle politiche di contrasto al tabagismo.

Effetti sulla Salute: La presenza di problemi di salute respiratoria tra i consumatori di sigarette senza combustione richiede ulteriori ricerche per comprendere meglio questi effetti e le potenziali correlazioni.

Non Fumatori: L'analisi dei non fumatori ha rivelato alcune caratteristiche significative, come l'età e le dipendenze, che possono influenzare le dinamiche del tabagismo in Italia.

POSSIBILI DIREZIONI FUTURE PER LA RICERCA

Le implicazioni dei risultati suggeriscono diverse direzioni future per la ricerca sul tabagismo in Italia:

Studio Approfondito sull'Iqos:

Un'analisi approfondita sull'Iqos potrebbe fornire ulteriori informazioni sulla sua diffusione e sugli effetti sulla salute.

Indagini Longitudinali:

Gli studi longitudinali potrebbero seguire i consumatori di tabacco nel tempo per valutare cambiamenti nelle loro abitudini e condizioni di salute.

Comportamenti di Dipendenza:

Approfondire la comprensione dei comportamenti di dipendenza legati al tabagismo e alle sigarette senza combustione potrebbe guidare l'implementazione di programmi di trattamento più efficaci.

Campagne di Sensibilizzazione:

Sviluppare campagne di sensibilizzazione mirate a specifici gruppi demografici e considerare i fattori socio-culturali che influenzano il tabagismo.

Monitoraggio Costante:

Il monitoraggio costante del consumo di tabacco e delle politiche di prevenzione è essenziale per adattare le strategie alle mutevoli dinamiche del fenomeno.

In conclusione, la ricerca offre una panoramica dettagliata del tabagismo in Italia, con importanti implicazioni per la salute pubblica e suggerimenti per future indagini e interventi. Per contrastare il fenomeno del tabagismo la prima regola è non cominciare per nessun motivo.

BIBLIOGRAFIA

1. Greenacre, M., & Blasius, J. (Eds.). (2006). *Multiple Correspondence Analysis and Related Methods*. [Inglese]. Copyright 2006.

2. Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3).
URL: <http://www.jstatsoft.org/>

SITOGRAFIA

1. Epicentro - Istituto Superiore di Sanità. Epidemia di fumo in Italia. URL: <https://www.epicentro.iss.it/fumo/epidemiologia-italia> (Accesso il [25/06/23])
2. Epicentro - Istituto Superiore di Sanità. Costi sociali del fumo in Italia. URL: <https://www.epicentro.iss.it/fumo/pdf/sgambato%20-%20Costi.pdf> (Accesso il [13/07/23])
3. Epicentro - Istituto Superiore di Sanità. Ratifica della Convenzione quadro dell'OMS. URL: <https://www.epicentro.iss.it/fumo/RatificaConvenzioneOms> (Accesso il [20/07/23])
4. Ministero della Salute. Allegato alla Circolare 1614. URL: https://www.salute.gov.it/imgs/C_17_normativa_1614_allegato.pdf (Accesso il [21/07/23])
5. Ministero della Salute. Dipendenza da fumo di tabacco. URL: <https://www.salute.gov.it/portale/fumo/dettaglioContenutiFumo.jsp?lingua=italiano&id=5580&area=fumo&menu=vuoto> (Accesso il [21/07/23])
6. Humanitas. Dipendenza da nicotina. URL: <https://www.humanitas.it/malattie/dipendenza-da-nicotina/> (Accesso il [21/08/23])

RINGRAZIAMENTI

B I M O L B G A F F C L Z H O U S
 X E H U X C N M A R W I A D S W D
 S G C N H P W H U N J L I S W U Y
 M A M D W A X G U Z L T E A R I G
 I D O W K P K Z C U Q D N B E O F
 T H Z M K E V D C L X N C Q L Q M
 T O N A I T S I R C O G D Y E B S
 E N E K S Y N E E N W P I P U B Y
 P R R K Y A Y G N R J M M T N O B
 I J O T Z J M U E F D L A M A K B
 Z A L I W C Z I Q Y U L S O M L U
 D U O W C K Y D V V F T T E E C H
 L N V F T W X I K Y D O R T K O C
 E I X Q U U E L R V S D O T X M L
 T S D B H K K Y A G N I P A J G K
 B O I R E L A V Q P U I A M K W R
 A I G L M S V E C W O T O P M X Y
 I L C A R B O N E W L W L E C L Y
 O D X A X Y Y X B S M X O N Y O T
 C B Z P M A R A B R A B A I Z C I
 C D O C U M U A V X N G Q Z D T E
 O Z P B B D A D B T H D T I T G J
 Q Z H E Z N O M B V I H Q O I J L
 U W B L B R O T H E R S U R L Q Z
 V W U C X H X W Y A N R B Q D U A
 R P K B I R A E Y R A V G D S R P
 U O R A Z Z A L H M Y D M J E B T
 O M E V K K D N O T D M A R V U Z
 Q Y W I B B G L N J W N R J V V I
 C M L X J V L S S H T R K V O H O
 U N B H I E T E N N O H L W J O E M
 Y W S L N M N K N N T M A V G C A
 X Y B O J A I I Z W A F D U J D U
 M W T P I D E I A G U O U L X B R
 J N P K R T O D M G O B Z B E W I
 A F Q D T W B A T P I O M O F N Z
 J Z N A X B I M L P M A X K K R I
 O D R A O D E L O N A I M A D N O
 Z H E H X O R K N S I L G N F O Y